How do we get the brain to tell us about language computations and representations? Designing and implementing experiments

Miika Leminen & Alina Leminen

Abstract

Planning and conduction of electrophysiological experiments involves considering numerous factors, such as meticulous matching linguistic stimuli by their psycholinguistic features, finding optimal data recording settings, and choosing data-analysis criteria. All these factors may considerably alter the results of your experiment. This chapter is intended for those, who are new to the field and wonder where to start. This chapter is suitable also for those readers, who wish to update their knowledge on the topic of EEG data recording and analysis. While many laboratories around the world already have EEG equipment, this chapter offers recommendations also for those, who wish to update their EEG equipment or lab settings. We focus particularly on the settings that are intended for planning and conducting experiments with linguistic stimuli.

Keywords

experiment, laboratory, EEG, MEG, guidelines, psycholinguistics, cognitive neuroscience, language

1. Introduction: how to get reliable data to test your hypotheses

This chapter is intended particularly for young investigators, who are at the beginning of their path to designing and conducting neuroscientific studies. Building experiments in the field of cognitive neuroscience of language means not only presenting linguistic stimuli, but in a wider scope, the ability to control for different features of language, which enables opening a window to various cognitive aspects of language processing. However, due to some limitations in behavioral and

psychophysiological measurements, some neurocognitive effects are still very difficult, if not impossible, to illuminate. The main bottlenecks one can face during conducting a study are the manipulation of critical stimulus features and the measurement errors. These topics are partly covered in this chapter's Subheadings 2 and 3, respectively. Once the studied phenomena are revealed by a suitable experimental manipulation, the experiment can still be jeopardized by data recording, data analysis, statistical approaches, or result interpretation. Hence, it is essential to understand what can and cannot be measured with present methodology, understand the limitations and use them in a competent manner. This is the reason why both experimental design and data recording issues are linked in our chapter – they go hand in hand and cannot be planned independently. Thus, in our view, the window to linguistic phenomena includes both the limited capability to isolate or categorize abstract features of a language and a limited window to neurocognitive processes.

Clear and simple recommendations would be easy to apply. Unfortunately, such recommendations would be oversimplified and could not cover all the different needs in different labs, experiments, and situations. However, we will try to give some hints that we hope will be useful to the readership. For every topic covered in this chapter, we will first explain the typical setups and try to elaborate different aspects that affect the practical decision making. In the end of each topic, we will give simplified guidelines. The recommendations are based on our experience in using, developing, maintaining, and operating psychophysiology laboratories.

2. Building and running experiments

2.1 Designing experiments

It is important to build general knowledge about different experimental paradigms used in your research field. It can be complicated and time-consuming, particularly because you should be able to collect your "library" or toolkit of competent experiments not only from your own methodological area (e.g., EEG/event-related potentials (ERPs)), but from other areas as well (e.g., functional magnetic resonance imaging (fMRI), behavioral experiments, eye-tracking).

A good way to start is to take an experiment that is already been repeatedly and successfully used in the field (see Box 1 for some examples) and make some modifications to it according to your needs and research questions. This will make sure that interpretations and conclusions are less dependent on the outcome of your experimental design, while your scientific argumentation can be partly corroborated by existing literature. Making some modifications to an existing experimental paradigm allows you to ensure the coverage your phenomena of interest while minimizing the risk of getting null or laboriously explainable results. Developing own experimental design from scratch is undoubtedly possible as well, however, piloting and developing an experiment typically takes much longer, on a timescale of weeks or months.

Box 1. Examples of some typical paradigms used in cognitive neuroscience of language

Isolated phonemes. Used, for instance, in behavioral phoneme discrimination or identification paradigms, or in Mismatch Negativity paradigms (cf. Chapter 6). Phonemes can be recorded naturally or synthetized. There can either be a single item representing the whole class or feature (e.g., /e/), or naturally varying tens of different items belonging to the same class but are different examples of it (e.g., different voices, sexes, intonations, base frequencies etc.). Natural variation can lead to more variance in brain responses.

Isolated syllables. Similar natural variation as in isolated phonemes ***(1, 2)***. In this case a syllable is composed of two or more phonemes. A syllable can exist in a language or it can be novel. This aspect can also lead to large differences in brain responses.

Violation paradigms and linguistic judgment paradigms typically involve modifications violations of one or several linguistic rules (e.g., "The boy is tuning the guitar/*the sock before the concert ***(3, 4, 5)***). The participants typically perform a judgment whether the sentence is acceptable or not or if there is anything strange in the sentence. While violation paradigms can provide interesting insights about language processing in the brain, the processing of violated structures may not be equal to natural language processing.

Listening or reading of lists of isolated words ***(6)***. Using single words stimuli is more straightforward to analyze as compared to clauses or sentences, which can also be its weakness, as a list of words is a mere simplification of a natural language. For example, in a natural language context, the context guides extracting word meanings, and single words are rarely presented alone without any context in naturally unfolding language.

Lexical decision tasks. Typically, word list paradigms that require a lexicality judgment (i.e., whether a linguistic item is a real word or not) ***(7)***.

Natural listening or reading. Presenting a participant with natural texts and passages without strict control of stimulus characteristics ***(8)***.

Priming paradigms. In priming paradigms, linguistic relationships between different words are examined by presenting a prime, such as 'cat' and a target word, such as 'dog'. With this approach, it is possible to investigate the extent a prior presentation of a word facilitates the recognition of another word ***(9)***.

It is also important to note that the neurocognitive processing of some specific task may differ depending on the exact formulation of the participant instruction. For instance, if the participants

are asked to judge whether a word is real or not, some participants might think that they should have heard someone using that word, while others might think that the meaning should be understandable according to linguistic rules irrespective of its usage. Hence, it is crucial to be as explicit as possible and use examples while formulating instructions for the participants. In addition, it is always advisable to use practice trials to ensure that every participant has understood the task instructions properly.

A potential difficulty in many experimental paradigms is related to *behavioral responses* of an experimental task, that is, sometimes patterns of behavioral and brain responses are incompatible with each other. For instance, when comparing two experimental conditions (e.g., words and pseudowords), brain responses can differ significantly, while no significant differences are observed in the behavioral data and vice versa. Such an outcome may certainly create challenges for interpretation of the results and comparing them with previous behavioral findings. On the other hand, if we do not collect behavioral responses during EEG/MEG measurement, we cannot be certain that participants indeed paid attention to the stimuli and performed the task correctly. A typical way to verify this is to use a control condition, in which motor preparation and performance do not differ from the condition of interest. As a rule of thumb—make sure that "you know what your participants are thinking". If you do not collect behavioral evidence, spend some extra time in planning the paradigm.

Other challenges that need to be considered during designing experiments are, for instance, *overlapping brain responses* and *time-locking* of the ERP/ERF responses. Brain responses are often difficult to separate from each other and overlapping brain responses occur both spatially (in space) and temporally (in time). Thus, data analysis approaches are often chosen to validate an assumption that a specific brain response is indeed related to a specific cognitive function and/or activity of

some specific neural network (cf. Chap. 6). A well-designed experiment goes conjointly with the analysis methodology and attempts to minimize the complexity of brain responses as well as allows you to isolate different neurocognitive functions.

Moreover, most electrophysiological analysis techniques require using some point in time to correlate some stimulus features with corresponding neural data. Such time-locking can be done in several ways. The most typical method is using the onset of a stimulus item as a reference time point, such as the onset of an auditorily presented word or the onset of a visually presented word. In natural language, words have very different durations, and if a brain response of interest (e.g., ERP/ERF component) is small and focal, you may lose your effects in averaging ***(10)***. This may happen, for instance, if a response of interest occurs in the end of the word than in the beginning, or at some other point in time when important information becomes available. In this case you could time-lock the responses to another point, such as the suffix onset ***(6, 11)***, a disambiguation point ***(12, 13)*** or a button press. However, these unconventional time-locking methods are often not straightforward to operationalize. For example, strong responses related to word onset can cause unwanted variance in response baseline and can disturb analyses.

Furthermore, processing of linguistic stimuli and electrophysiological responses associated with them do not end after the stimuli are presented to the participants. For instance, when using sentence-level and discourse-level stimuli, integration of individual words to a sentential context may continue several seconds after the stimulus presentation has ended. If a next sentence is presented too soon after the previous one, it is possible that the brain responses of interest will overlap with other responses and will be more difficult to quantify and separate. The same challenge is faced when presenting the target stimulus too soon after the prime stimulus in priming experiments, especially if the prime and target are presented in different stimulus modalities (e.g.,

auditory and visual), and especially if stimulus lengths and durations are not well controlled for. Hence, in such paradigms, we recommend including a sufficiently long inter-stimulus and inter-trial intervals. Box 2 summarizes recommendations presented above.

Box 2. Summary of the guidelines for designing experiments

It is easier and less risky to begin neuroscience experiments using a modification of an existing paradigm, using well-defined and known ERP/ERF components. This also facilitates the formulation of a priori hypotheses.

An entirely novel paradigm requires careful testing and piloting, and it is important to invest time and resources into this preparative work. In addition, novel paradigms may complicate hypothesis setting, as it may be challenging to predict the exact brain responses that will be elicited by the novel paradigm. This may lead to exploratory analyses, which has a weaker explanatory scientific power.

Carefully plan your experiments to isolate ERP/ERF components of interest from other potentially overlapping components; carefully explore previous literature.

Plan to which point in time you time-lock ERP/ERF responses of interest; this may affect the results greatly.

2.2. Experiment building tools

What is a good strategy for obtaining an experiment execution software for a research lab or unit? There is no solution that is suitable for everyone's needs; however, we recommend that the decision is based on strategic planning. When planning and evaluating different experiment software tools,

you should also consider their costs, technical accuracy, usability, and difficulty of implementation. You can then select the software according to your budget, laboratory settings, and/or programming skills. For example, you can program any experiment from scratch by, for instance, a lower-level programming language, such as C or C++. In this case, any solution can be programmed, since development of an experiment is not limited by features missing from an existing software or a toolbox. However, experiment development and programming are time-consuming tasks that require careful testing and piloting to exclude any errors in the code. Moreover, if there is a wide researcher community in the lab already skilled in, for instance, Python or Matlab programming languages, it would be easier to use tools in which this existing expertise can be used. If you are not a programmer and cannot invest months or years to learn a new programming language, a more cost-efficient and time-efficient way is to use a readily available tools to program experiments. These can be code libraries, toolboxes (such as Psychophysics toolbox for the Matlab environment), or software packages (such as PsychoPy, E-prime, Presentation by Neurobehavioral Systems, and Experiment Builder by SR Research). Irrespective of the experiment building software, if you are new to research, designing experiments and/or programming them, you might need support to get you started or when you experience problems or errors in the code. In some labs, every researcher in lab is using different experimental tools, making it undoubtedly more challenging to get support and advice from colleagues. Fortunately, nowadays international online communities do offer plenty of support. It is worth considering, however, whether you would like to develop a library of paradigms that can be used and modified in your lab, ultimately saving time from paradigm software development and testing in future.

Some experiment building software are open source freely available resources, and some are commercial. If a tool is commercial, then it is important to estimate how many separate experiments you would need to run in parallel for the next few years. Licensing options for different software

products can also be quite different. For instance, if each license purchase requires 2000 euros, it might be challenging to obtain 30 licenses for the whole classroom test lab. On the other hand, some software manufacturers offer other licensing options, such as online test licenses with an annual fee, and the costs can be more easily adapted for varying needs. Check the manufacturer's licensing options and discuss it in your lab.

Crucially, different experiment building tools differ from each other in their technical (timing) accuracy. For instance, some tools offer a wide range of algorithms to manipulate visual stimuli but can be rather inaccurate in presenting auditory stimuli. Hence, it is advisable to choose the software or platform depending on the needs of your experiment. Furthermore, different tools require and support different computer hardware. For instance, whereas one software works optimally only with certain manufacturer's audio cards, another may require a different card. Another example could be that a certain video card may be optimized for the high refresh rate for some software, whereas another tool misses some frames every now and then and would run optimally in some other video card. Make sure to verify your lab computer settings before you start programming your experiment. It is also recommended to use free or low-cost demo versions to get some first-hand experience about usability and technical accuracy.

As a final note, it is worth mentioning that implementation of any experimental tool in lab environment requires engineering work and testing. Thus, it can be quite costly to implement more than one experimental environment in the lab, since all of them require resources for development and maintenance. For instance, testing of timing accuracy and other important features (such as presence of skipped frames with visual stimuli) should be done systematically on a regular basis. Hence, the more different tools and systems a lab keeps running in parallel, the more working hours will be needed. As mentioned above, not all hardware combinations are supported by all tools,

which makes it more challenging, if not impossible, to implement simultaneously all possible tools in the same laboratory or experiment unit.

The abovementioned recommendations are simplified in the guidelines presented in Box 3.

Box 3. Guidelines for choosing an experiment building software

Find out which experiment building tools work optimally with linguistic stimuli you frequently use (e.g., auditory, visual, audio-visual).

Plan how many simultaneous experiment units you want to have in your lab and what licensing options are optimal for it.

Consider whether you need online paradigms. Some of the tools support both running the same experiment in the lab and online via an Internet browser (e.g., PsychoPy).

It always takes time to learn to use a new tool. Invest time in it and be patient. Typically, learning a new experiment building tool means learning of a new programming syntax as well as finding optimal technical settings. Moreover, experiment-designing philosophy can be quite different in different software (i.e., how to optimally build hierarchical experiment structures with trials and blocks of trials, and how to control their randomization), and it can be rather time-consuming to transform a ready-made experiment from one platform to another.

2.3. Stimulus preparation and technical setup

Electrophysiological responses are typically very sensitive to physical differences in stimuli. Hence, you must avoid unintentional and undesirable stimulus differences, as they may potentially lead to

differences in brain responses overall, while your effects of interest may remain hidden. Hence, it is essential to invest time in learning to prepare and edit your stimuli meticulously. With visual stimuli you often need to control for at least the *length* of your visual stimuli, such as words, the *size* of an item or text on the display, which is usually reported as visual field angles, taking into account also participants' distance from the display. In addition to size, you need to consider the *resolution* of stimuli or a computer display (i.e., the number of pixels in x and y dimensions), *colors*, brightness (note that the text font thickness affects perceived brightness), loci on the visual field (you should know where a participant's gaze is fixated when your stimuli appear), visual frequencies (density of the lines or stripes--is your image "busy with lines" or not?), and luminance in different spatial locations on the computer display. Be very careful if you aim to present different stimuli in asymmetric locations on the display (e.g., by comparing brain responses to stimuli located in the middle of the display vs. in the top); the differences in response sizes can be easily affected by unwanted difference in stimulus brightness. Note that especially in low quality displays the luminance and other properties vary greatly at different locations of the display. Unfortunately, technical specifications reported by display manufacturers do not often help, but lab setup should be verified by calibration test measurements in a lab.

When preparing auditory stimuli, you need to carefully match at least the *duration* of an auditory item, such as a syllable or a word. Please keep in mind that duration of an audio file can differ from duration of an auditory item, if you include silence in the beginning of your audio file (avoid doing that) and intensity (raw audio intensity, how well the dynamic range is used and how loudly it is presented). Note that loudness depends not only on the raw file but also on the technical setup of stimulus presentation, including volume adjustment on a computer and also a potentially separate audio/headphones amplifier. As mentioned above, check that you do not have silence randomly in the beginning of the audio file. This will lead to jitter in triggering, because your experiment

building software assumes that your stimulus starts when an audio file starts. Simply go through all of your audio files and delete possible unnecessary silence periods. You also need to control for how similar is the perceived intensity between the different stimuli, which can occasionally be quite challenging, as our perception of loudness is different for consonants and vowels, to name a few. Moreover, the algorithm used for measuring intensity always has certain parameters, such as length of a time window and the frequency weight distribution (check these settings with physical loudness meters as well as in audio file loudness normalization algorithms provided by your audio editing software). Hence, it is a good idea to pay attention to these parameters. Keep in mind that if your stimulus is very short in duration, also the time-window for the intensity normalization algorithm should be short.

Furthermore, rapid transitions in audio signal can sometimes cause some additional sounds in headphones or loudspeakers and even generate brain responses of their own. To avoid these unwanted responses, use, for instance, some fade-in and fade-out ramping functions (e.g., linear or logarithmic) to make different stimuli comparable. These fading ramps ensure that your audio signal's amplitude rises/falls from/to zero (silence) in some specific controlled time window. Typically, researchers use, for instance, 5 ms rise/fall ramp times. However, be careful especially with consonants while ramping their onsets. You also need to control for prosody (intonation), sex (male or female speaker), speed and pace of speech as well as other psycholinguistic features of the stimuli (e.g., lexical frequency, phonological neighborhood size etc.). During stimulus recording, make sure that factors such as background noise, microphone quality, and echo are taken care of. You do not necessarily need a professional studio for recordings, but you need at least a high-quality microphone and an audio card. Some higher-class USB microphones are also relevant, as they have an inbuilt AD converter and do not require a separate audio card. Do make sure that the background noise is low in your recording venue. Echoes can be dampened by, for instance, simple

echo attenuators around the microphone. Those are sold, for instance, in music instrument stores. If you go to one, obtain a pop filter as well (a light shield between the microphone and the speaker).

Next, we mention a few words on stimulus triggers and logging your experiment. In our opinion, it is better to store as much information as possible, since it helps solving possible issues that might come up later on. As a rule of thumb, careful planning before data collection always saves time in the analysis phase. It is, thus, crucial to ensure that your triggered and logged timepoints are precisely the ones you need. If there is a systematic delay in stimulus triggering (a gap between the trigger sent from a stimulus computer to an EEG/MEG recording equipment and actual stimulus onset), then you should at least be aware of it and take it into account during data analysis. We also recommend that a lab engineer or a technician should also routinely check that the delay remains constant). In addition to routine checks, measurements of trigger jitter (variance of the delay) should also always be performed after any changes in the lab settings. Trigger jitter can easily cause unwanted disturbance to brain responses and the jitter issue is particularly harmful if you have not measured and minimized it.

Even if your stimulus software is able to send triggers accurately synchronized with a timepoint that is assumed to be the stimulus onset by your stimulus software, it may not necessarily match with the actual stimulus onset. With visual stimuli, first it takes some time for a computer to prepare the stimuli in a video card. Thereafter, it may wait for the next frame and some processing may also take place in the display. Finally, your stimuli are presented on the display. However, even then the onset slope varies depending on the display model and technology of the panel it is using, meaning that the visual stimulus is not visible in full brightness immediately after the onset but fades in over approximately 5 or 30 ms. This onset slope is always present and often you cannot do more than test several monitors to find the optimal onset slope. Similarly, there is an offset slope, that is, a

stimulus does not return to dark immediately but follows a fade out curve. The easiest way to measure the trigger jitter as well as onset and offset slopes is to build a measurement setup with a sufficiently quick photo-sensitive sensor. These onset and offset responses are sometimes the reason why expensive research-specific displays may be worth of an investment. Box 4 summarizes our recommendations presented above.

Box 4. Recommendations for stimulus preparation

Be meticulous in your stimulus preparation--make sure that your stimuli differ from each other only by the desired stimulus characteristics and variables

Accuracy of stimulus triggering is critical especially in electrophysiological recordings, due to their high temporal resolution

In non-optimized stimulus computers jitters in both interstimulus intervals and stimulus-trigger asynchrony can easily be in the same time range with the neural responses of interest, that is, even tens of milliseconds and can thus ruin your responses

2.4. Running experiments

When you are ready to run your experiments with actual participants, it is important to keep recording notes; that will help with possible issues that might need to be solved afterward. It is also essential to minimize the possibility to connect personal identification information with recorded data. Hence, we recommend creating pseudo-identification numbers for all the participants, using this pseudo-ID in all the data, log files, and test documents. Such pseudo-ID numbers can be created in advance, prior to actual measurements. A participant's name should be used in an informed

consent documents only, and these should be stored in a separate place. To ensure privacy of the participants, do not insert even pseudo-ID in the documents that do not require IDs. You can carefully and securely store one mapping table (a paper or an MS Excel sheet) for mapping participant names and pseudo-IDs, so you can easily match the pseudo-IDs with any participant if needed (e.g., you discover that you have to exclude one of the participants during the data analysis phase). Thus, you can also anonymize your data by deleting this mapping file (assuming that no other information in your data allows identification of your participants. Pay extra attention to data security in case you have full head anatomical MR images). Describe this procedure in your ethical application and follow your local institute's instructions.

Remember to provide sufficient breaks for your participants. Make sure to keep the length of the recording session reasonable (maximum 1.5 hr per session) and even much shorter with child participants. Keeping your participants alert will ensure their attention on the task and will minimize unwanted disturbance of your data (e.g., alpha waves caused by fatigue; their signal amplitudes can easily be much larger than the ERP responses of interest). Fatigue also affects cognitive performance of the participants. Hence, it is important to randomize the order of the blocks for each participant to avoid systematic fatigue or movement artifacts for some experimental conditions, which may disturb your neural and behavioral data. Invest in a good armchair so that your participants can sit comfortably throughout the experiment. Offer refreshments when needed and offer several breaks between the experimental blocks.

3. Setting up a psychophysiology laboratory

3.1 Data recording infrastructure

Recently, lower prices of high-quality laboratory equipment have enabled many smaller labs to purchase high quality research facilities. However, there are also equipment on the market that do

not meet standards of a high-quality psychophysiology lab, particularly the one focusing on language research. As with strategic planning of your stimulus setup, it is important to make a strategic plan for developing, running, and maintaining a recording instrumentation. This will help you to decide which features are essential and which are less important to include in the lab.

There are less manufacturers for MEG equipment than for EEG and, hence, less options are available. However, the features built around the basic MEG infrastructure allow for more customization (such as different behavioral response instrumentation, auditory and visual stimulation setups, and simultaneous EEG and eye tracking data acquisition). In our opinion, MEG-compatible (or inbuilt) EEG equipment is highly useful to purchase together with MEG equipment. It will allow you to measure EEG and MEG simultaneously, offering a possibility to use of more advanced neural source modelling techniques, since EEG's lead fields are different from both (MEG) gradiometers' and magnetometers', and, thus, can offer complementary information.

When planning EEG facilities, we recommend to consider at least the number of channels, mobility, and electromagnetic noise shielding features. In many studies on the neurocognition of language, 16, 32, or 64 channels are sufficient, not every study needs to have 128 or 256 EEG channels. In other words, high-quality science is possible to do with less than 64 channels and not every study using over hundred channels is automatically better. What matters is how (and if) you use the advantage of having better spatial information. High-resolution EEG with up to 256 channels is advantageous in neural source modeling and in analysis techniques such as independent component analysis (ICA). The latter, the so-called blind source separation technique can also be used in data cleaning. The disadvantages of multi-channel EEG are larger lab expenses, larger equipment, and slower (and less comfortable) preparation of EEG recordings. In many cases it is optimal to have equipment, which allows one using a different number of channels for different kinds of

experiments. With respect to mobility, consider if you plan to record EEG outside the lab in the so-called naturalistic settings. The highest-quality lab equipment with a large number of channels is often bigger, heavier, and includes many modules with a lot of wires. This makes transportation of equipment to a new recording venue for each recording session rather inconvenient. The recommended options are to have a separate lab and mobile equipment with different qualities or to purchase semi-portable equipment. Furthermore, different equipment has different noise shielding features, such as active shielding (the wires between an electrode and an amplifier are actively shielded), active electrodes (instrumentation buffer integrated in each electrode), length of electrode wires (the longer the wires, the more noise they can pick up), as well as the distance between participant and AD transformation unit. Regarding the latter, in some equipment this distance is the same as the length of electrode wires, but in some equipment digital conversion does not take place in a so-called "headbox" close to participant, but in an amplifier situated farther away. In such a case, the signal between the headbox and the amplifier can be still rather vulnerable. If your lab has low electrical noise environment, such as an electrically shielded room, these features have less power to improve the signal-to-noise ratio, as shielded room is already handling most of the electrical noise and thus shielding your measurement. However, in varying and naturalistic environments in particular, these features are more crucial.

Another important issue to consider while building an EEG lab is triggering possibilities. Some small ambulatory devices or clinical neurophysiology devices have poor triggering interfaces and may even make ERP recordings impossible. A sufficient input port for EEG recordings aiming at ERP analyses are for an 8-bit or 16-bit TTL signal (i.e., the port where a stimulus computer's trigger output is connected). In addition to EEG electrodes, you might need to record data from additional sensors. You might consider how many bipolar electromyogram (EMG) inputs are needed in addition to common referenced EEG inputs. For instance, EMG inputs are typically used

for two bipolar electro-oculograms (EOG), that is, vertical and horizontal EOGs to optimally record eye movements and blinks. Other potential needs for EMG inputs are electrocardiograms (ECG), facial EMG for autonomous responses (cf. Chaps11 and 19), as well as muscle tonus for sleep recordings to distinguish random eye movement (REM) from other sleep stages and the waking period. Four bipolar inputs (8 electrodes altogether) are typically sufficient to cover most of the needs. If your needs change, some amplifiers allow you to upgrade your system with more channels or inputs for different additional sensors.

Next, we will say a few words about amplitude resolution. For the highest-level scientific purposes, EEG equipment should have a 22–24-bit AD converter. With such a converter, the smallest recordable changes in an EEG signal are within a range of few tens of nanovolts, yet the dynamic range is large enough for even larger-scale artefacts and signal changes, while the amplifier will remain in the functional (dynamic) range and not saturate. In cheapest or "consumer" brain-computer interface (BCI) EEG equipment, the resolution is sometimes too poor for scientific ERP studies. Another problem with cheapest EEG equipment is that the dynamic range of the converter is so small that heavy filtering (signal dampening or smoothing) is performed before the AD conversion. Quite often this filter harms your ERP responses, particularly slow language-related ERP components such as N400 and P600.

To run the recordings smoothly, you should make participant preparation as fast and participant-friendly as possible. Children, elderly, and clinical group participants benefit from gentle electrode preparation, leading to possibly less tension or movement of the participants during the recordings. The so-called traditional passive electrodes with electrode gel/paste are most difficult in this sense. Active electrodes with electrode gel are faster to prepare and (gentle) scratching of the participant's skull is often unnecessary (see below for more explanation). Nevertheless, electrode gel application

is still somewhat time-consuming. Electrodes with saline solution pads are fastest and easiest to apply. However, saline pad connected electrodes have a higher risk of varying signal quality due to movement, and during long recordings in particular, the electrodes may dry, leading to poorer electrode connection, and, hence, to poorer signal-to-noise ratio (SNR). In this sense, active electrodes are better, but they are pricier. In addition to electrode preparation, you should pay close attention to how comparable your signal quality is in the beginning and in the end of the recording session. For instance, you might need to compare, brain responses in the beginning and the end of the session in a language learning experiment. In such studies, it is crucial for the SNR to remain the same throughout the recording session. As discussed above, saline pad electrodes in particular do not always meet this criterion.

If the electrodes are not attached permanently to the EEG caps, as is often the case with active electrodes, it is good to purchase EEG caps of several different sizes and only a few sets of (rather expensive) electrodes. This will prolong the life cycle of EEG caps, as you can always get a new cap of an appropriate size. In this case, you will not have to try to use too small caps, as stretching will eventually damage the cap, leading to the loss of its shape. In too large caps (or in ones that have lost their original shape), some electrodes will be loosely connected to the scalp and cause bad or varying signal quality. Generally, any lab should have a variety of EEG caps with different sizes and more copies of those that are used most frequently. Otherwise, you may have to use a wet EEG cap if you have several participants in a row. See Fig. 1 for schematic view of a modular EEG cap/electrode system.

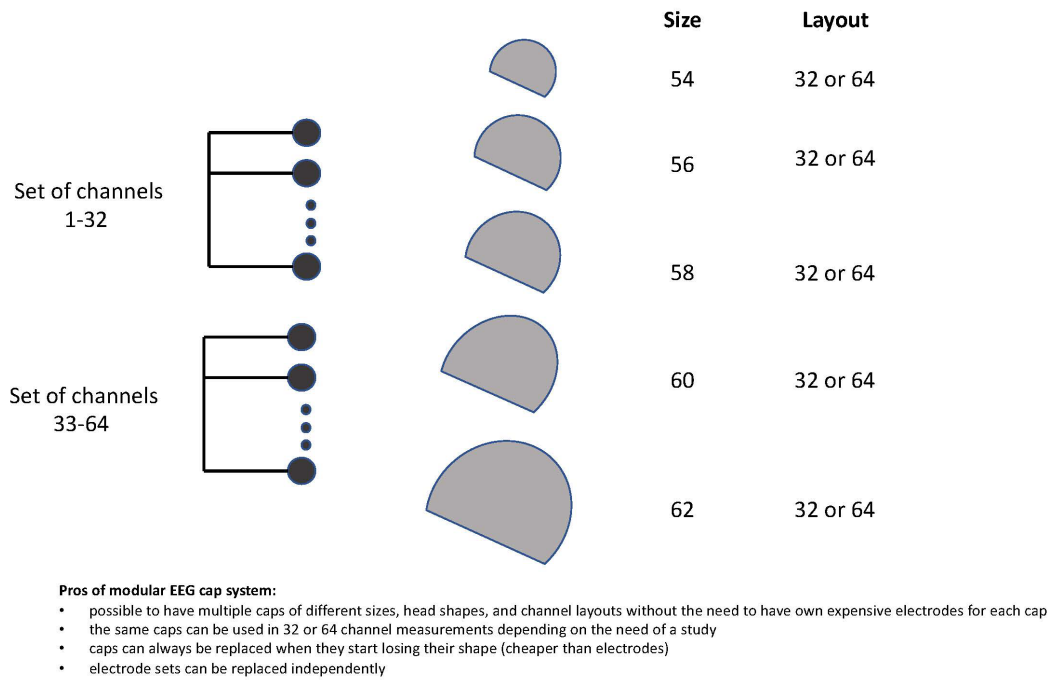

Fig. 1 Schematic overview of a modular EEG/cap system

Furthermore, if an EEG system is actively used, single electrodes and wires are typically broken quite often. It is therefore important to know if single electrodes can be replaced in your lab or if the whole set of electrodes needs to be sent to a manufacturer. Even if replacement of a single electrode is inexpensive, it always means that one of the electrode sets cannot be used for several weeks. Hence, we recommend having a sufficient number of spare electrode sets to avoid interruption in the EEG recordings. You should also do some simple lifecycle cost estimation for your EEG equipment, for instance, take into account the prices of replacing electrodes and caps. After such an analysis, you may find that expensive electrodes turn out to be cheaper in the long run, since in that case you can replace and renew electrode sets independently from caps.

In order to obtain a proper EEG signal, most EEG systems require using electrode gels and pastes. The choice of an appropriate gel or paste depends both on your EEG system and on your research needs. There is a plethora of available options and hence, it not straightforward to choose an

optimal electrode gel. Typically, an EEG system manufacturer can recommend certain gels or pastes, but it is beneficial to learn about other possibilities and test different products to find the ones that are optimal for your own research purposes. When choosing an electrode gel or paste, you need to consider, for example, the properties of the gel, the ease of gel or paste removal, and the stability of the contact impedance during your recordings. A smoothly running gel with high viscosity is usually faster to apply than hard and sticky one, and the difference between them is more significant if you have many electrodes to prepare (i.e., longer preparation times). However, a smoothly running gel tends to leak and does not keep a good electrode contact, unless the electrode holder in the cap or the opening in a circular electrode is tight enough to keep the gel in. On the other hand, removal of thick pastes is time-consuming and requires force during washing of electrodes and caps. This may cause physical damages and shorten the life cycle of caps and electrodes. Thick pastes are also inconvenient for the participants, who need to wash their hair after the experiment.

When the gel or paste is properly attached and the EEG measurement has started, the next task is to verify the stability of the contact impedance during the recording. The change in the contact impedance may be caused by, for instance, drying of a gel or a paste and due to the movement of the electrodes, caused by movement of a participant. If the recording session is short (< 45 min.), this is less likely to cause issues. However, the longer is the recording, the more attention the stability of the contact impedance will require. This is the reason why, for instance, sleep researchers use specialized electrode pastes, which maintain thickness and electrode contact during overnight recordings.

In addition to *keeping* the contact stable, *obtaining* a good electrode contact is a crucial step, which also helps to improve your signal quality during an experiment. In other words, the better the

contact impedance, the better is the signal-to-noise ratio of your EEG signal. In order to get the signals of all EEG electrodes comparable with each other (e.g., if you would like to statistically compare an EEG response amplitude in, say, F3 and F4 electrode sites), all the impedances in different electrodes should be in a comparable range (i.e., quite similar). Typical standards for ERP studies require <10 kΩ contact impedances. It can sometimes be challenging to achieve; yet it is one of the most critical steps in the whole study. In active electrodes, the first instrumentation amplifier buffer electronics already exists in each electrode, making measurements somewhat less sensitive to contact impedance differences between the electrodes, and also buffer the signal against the environmental electrical noise (see Fig. 2). Hence, it is reasonable to have slightly worse contact impedances, while keeping comparability and signal-to-noise ratio in sufficient levels. For passive electrodes, on the other hand, abrasive electrode gels are particularly good in cases when it is very challenging to get sufficiently low contact impedances (< 10 kΩ). Abrasive gels (containing some small particles, usually pumice) help to attain a good electrode contact, but for it to work properly, one needs to scratch a participant's skin with some tool (e.g., with a wooden stick or head of a plastic syringe).

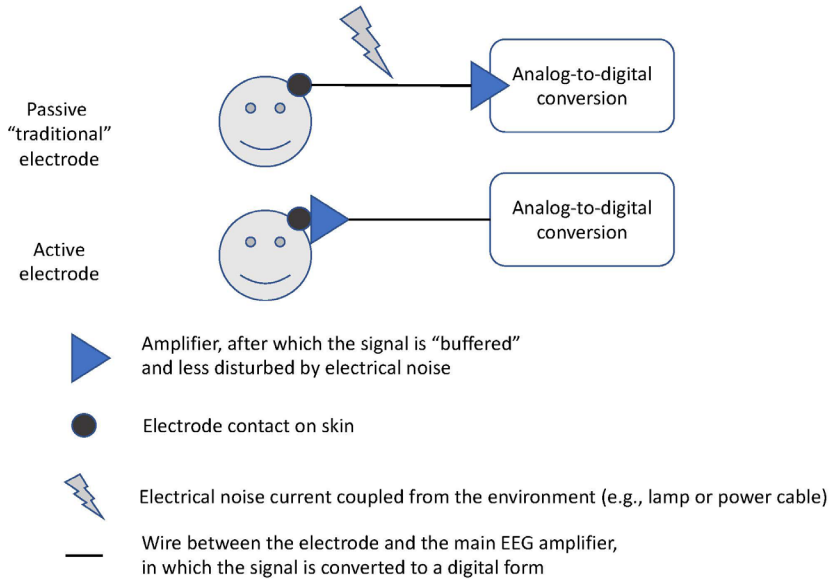

Fig. 2 Schematic overview of active and passive electrodes

3.2. Behavioral recordings

Many EEG studies of language include behavioral responses and we would like to say a few words about them as well. Most of the consumer response equipment such as computer keyboards do not need to be accurate on a millisecond scale, and a timescale of tens or hundreds of milliseconds is usually enough for standard office work. However, in reaction time tasks significant and important differences between different experimental conditions or participant groups can be on a scale of only a few tens of milliseconds, meaning that an effect can be easily lost by an inaccurate measurement. A constant and known delay in all behavioral responses does allow for comparisons between conditions or groups, however, if this constant delay is longer, it is often accompanied by a larger variance, which directly causes more measurement error. However, since these delays and inaccuracies are quite challenging to measure and verify, it is better to invest in proper response devices designed for psychophysiological studies.

In addition to a possible delay, the “feel” and sound of a response button may also matter. To test this, press different buttons that you see around you. Do you get sensory feedback from a button press? How quickly does it return to the original position? How long does it take for a button to return to its place after the press? How strong is the spring force working against your pressing? All these different features may have an impact on reaction time measurements. This is also one of the reasons why it may be challenging to compare exact (absolute) reaction times between different studies, when all or some of these features vary from study to study. An optimal reaction time button should ideally be silent, have a short trajectory, provide clear sensory feedback (“feeling”) about the button press, and should quickly return to its original position. Other important features of a response system include timing accuracy (optimally on a microsecond scale or even smaller) and connection of a response pad to both stimulus hardware and software, such that this time-resolved

functioning can be logged with an optimally short delay and small jitter. Box 5 summarizes our recommendations for building a psychophysiology laboratory.

Box 5. Summary of recommendations for building a psychophysiology lab

Plan carefully if you need a mobile (small, usually less channels) or stable (typically more channels and better qualities) EEG equipment. An ideal is to have separate equipment for both of these needs.

The "consumer-EEG" equipment, costing a few hundreds of USD does not meet the standards for high-quality scientific ERP research, despite the seemingly normal-looking raw EEG verified by visual inspection. However, despite its insufficient quality for a research EEG lab, such equipment may still be suitable for BCI purposes, for instance, measuring alpha range signal power in some electrode sites, which is much stronger signal than many language-related ERP responses.

Using active electrodes is advisable, since they tolerate higher contact impedances without compromising the signal quality. In addition, participant preparation is faster and more comfortable for the participants. Moreover, they buffer the signal against environmental electric noise and are thus, particularly suitable for measurements outside the shielded laboratory, such as offices or classrooms. Obtaining active noise canceling feature improves the signal quality even more.

For reaction time measurements, we recommend to use proper response devices instead of PC keyboards and spend some time to correctly connect them to an experiment building software.

Otherwise, there is a danger of obtaining measurement error that is larger than the effect of interest, even in “simple” reaction time tasks.

3.3. Laboratory practicalities

As mentioned above, keep your participants content and relaxed (but not tired) during participant preparation and throughout the experiment. This is particularly important for children. Provide a pleasant atmosphere both in the measurement room and in the lab; keep the lab tidy and avoid a “wire mess”. This will give a professional yet casual impression of the lab. Some participants can be intimidated and stressed by the hospital-like environment; hence, it is also advisable to avoid wearing white coats.

Since most EEG systems require using gels and pastes, the lab often requires cleaning after the experiment has ended. Therefore, cleaning facilities are needed to be taken into account while designing and building an EEG lab. Ideally, there should be a cleaning spot with a tap and a sink. Note, however, that a metal sink is not a good option because metal ions can attach to the electrodes and thus, worsen the quality of the signal. In order to avoid wire mess, the sink should ideally be located in proximity to a place where one keeps EEG electrodes and EEG caps. You should also provide hair washing facilities for participants. The standard solution is shower, but you can also consider a hairdresser-type of a sink. In our experience, if a research assistant can spend a few minutes to wash a participant’s hair, the participants are pleasantly surprised to receive such a service. It is much more comfortable for participants than taking a shower only to wash away electrode gel. The dressing room is also very important, especially if you have shower facilities. You can also consider purchasing small lockers for participants to leave their personal belongings.

4. Improving data quality

We cannot emphasize enough the importance of data quality. The earlier you can improve the quality of your data over the course of the study, the better. Data quality improvement procedures include avoiding artefacts, instrumentation (i.e., sensors and other instruments that convert a physiological signal of interest to an electric signal, see "Laboratory practicalities"), transfer path (between the electrode and AD conversion, usually a wire between an electrode and amplifier), online data processing (such as online filtering), as well as offline data processing. While many correcting steps can nowadays be performed during an offline data analysis, such correction methods are rarely perfect. Hence, we strongly recommend to first optimize everything possible prior to the analysis phase. Complex data cleaning techniques have their important role in your toolbox, but they also cause non-transparency to your EEG analysis process, while it is possible to collect clean data if you pay attention to it. Below, we discuss these points in more detail.

During MEG or EEG measurement, we aim to record brain response signals that originate from actual neural activity. Here, we define all other signals coming from the same experimental participant as artefacts. These include, for instance, eye movements. An eye is a strong dipole with electrical charge (cf. Chapter 6), and its movement is seen as an electric potential shift in the electrodes. Other artifacts include eye blinks, muscle artefacts (an electrical signal from muscle activity or tension), and movement artefacts (change in the electrode-gel-skin geometry, for example, by pulling an electrode wire too tightly, resulting in turning of that electrode). A common feature of all these artefacts is that they are large in comparison to brain responses and thus, affect the quality of the measurement. There are several signal processing methods to handle these artefacts, but it would be best to avoid having artifacts altogether. The most critical is minimizing eye movements (by participant instructions and using, for instance, fixation crosses for participants

to focus their gaze during an experiment) and movement artefacts (keeping the participant comfortable, avoiding too tight electrode wires).

An EEG signal is very vulnerable in a wire connecting an electrode and amplifier. In this wire, the signal is transformed to a digital form, after which it cannot be easily disturbed by environmental electrical noise. One solution to avoid disturbance is to use active shielding, which actually shields the signal in the wires. Another powerful feature is active common mode signal feedback. The most expensive solution to shield an EEG signal is to shield the whole room (i.e., shielded room). A magnetically shielded room is crucial for a MEG system, while a Faraday's cage type of a measurement room is useful for EEG. Such a room takes care of weak electrical currents coming from the environment, for example, from powerline or lightning, which are too small to be observed, yet large in comparison with the EEG signal.

During recordings, an amplifier performs certain signal processing that cannot be changed afterwards, such as filtering before AD conversion (so called anti-aliasing filters). Some EEG measurement systems allow you to modify processing settings and these processing steps cannot be undone in subsequent offline analysis. In contrast, offline processing steps can be iterated multiple times. It is thus important to perform only very essential processing steps at this stage. If you, for example, decide to perform online filtering with 1 Hz high-pass filter, you cannot return to your data later to extract, for instance, P600 responses (cf. Chapter 6), which are likely to be attenuated using this filter. On the other hand, sometimes certain online data processing steps are required to improve the signal quality. For instance, with amplifiers with a low dynamic range, a signal saturates easily due to signal drifts. Hence, you should filter out these slow drifts and keep the signal in the optimal signal range of the amplifier.

Before quantifying your effect of interest from a raw EEG or MEG signal and performing statistical analyses, offline data processing is needed to improve the signal quality and reduce measurement error. During offline processing, it is possible to try to reduce effects, which were impossible to prevent during earlier steps. Offline methods include, for instance, filtering and artefact rejection or removal techniques ***(7)***.

Box 6. Recommendations for improving data quality

Minimize the need for eye movements and participant's head movements.

A good electrode contact is the key to good EEG data quality, saving you from a lot of trouble in the analysis phase.

Small things in the lab do matter--no loops (or semi-loops/curves) in the electrode wires, no power lines or cables close to EEG wires or the participant.

If your EEG system has a so-called head box, which is NOT an AD converter, the signal continues to be "weak and vulnerable" also in the cable between the headbox and amplifier. Solution: do not put power cables next to it.

5. Combining methodological modalities

Ideally, scientific papers should include all the relevant studies using a wide variety of methods in the introduction, including several neuroimaging methods and other research methods, such as behavioral methods and eye-tracking. In order to widen your theoretical and experimental perspective, learn more than one neuroimaging method; this will also help you to understand

literature better. When you learn several methods, you can use them as a toolbox and pick whatever is the most suitable for your specific research question. During result interpretation, integrate your findings with results obtained with other methods – while this is challenging and requires deep understanding of many different methods, ideally, it will lead to better science. Furthermore, build the "big picture" by combining several theoretical accounts, which are based on different methodologies, and connect new findings with theoretical accounts, based on evidence obtained from different methodologies. Focusing on only one method may limits scientific progress and nowadays many manufacturers offer measurement and analysis software that enable combining methods. Make the most of your experiments!

6. Summary and conclusions

There are no simple solutions for conducting high-quality electrophysiological research but as methodology progresses, so do common practices and guidelines. In this chapter, we attempted to offer a few practices and means for planning, building, and conducting experiments of electrophysiology of language. Our recommendations are by no means exhaustive but are based on almost two decades of experience in electrophysiological research and laboratory build-up and maintenance.

References

1. Näätänen, R. (2001) The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiol 38, 1-21.
2. Sorokin, A., Alku, P., Kujala (2010) Change and novelty detection in speech and non-speech sound streams. Brain Res 1327, 77-90.

3. Hut, S.C.A., Leminen, A. (2017) Shaving bridges and tuning kitaraa: The effect of language switching on semantic processing. Front. Psychol. 8:1438.
4. Leinonen, A. Brattico, P., Järvenpää, M. et al. (2008) Event-related potential (ERP) responses to violations of inflectional and derivational rules of Finnish. Brain Res 1218, 181-193.
5. Leminen, A. Jakonen, S., Leminen, M., et al. (2016) Neural mechanisms underlying word- and phrase-level morphological parsing. J Neuroling 38, 26–41.
6. Leminen, A., Leminen, M., Lehtonen, M. et al. (2011) Spatiotemporal dynamics of the processing of spoken inflected and derived words: A combined EEG and MEG study. Front Hum Neurosci 5, 66.
7. Leinonen, A., Grönholm, P., Järvenpää, M. et al. (2009) Neurocognitive processing of auditorily and visually presented inflected words and pseudowords: Evidence from a morphologically rich language. Brain Res 1275, 54-66.
8. Leminen, A. Verwoert, M., Moisala, M. et al. (2020) Modulation of brain activity by selective attention to audiovisual dialogues. Front Neurosci 781344.
9. Bosch, S., Leminen, A. (2018) ERP priming studies of bilingual language processing. Bilingualism: Lang Cogn 21, 462-470.
10. Kimppa, L., Kujala, T., Leminen, A. et al. (2015) Rapid and automatic speech-specific learning mechanism in human neocortex. NIMG 118, 282-291.
11. Leminen, A., Leminen, M. M., Krause, C. M. (2010) Time course of the neural processing of spoken derived words: an event-related potential study. Neuroreport 21, 948-952.
12. MacGregor, L., Pulvermüller, F., van Casteren, M., Shtyrov, Y. (2012) Ultra-rapid access to words in the brain. Nat Comm 3:711.

13. Pulvermüller, F., Shtyrov, Y., Hauk, O. (2009) Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. Brain Lang 110, 81-94.