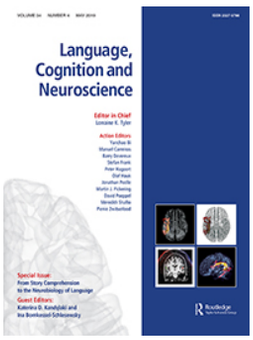

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

M/EEG analysis of naturalistic stories: a review from speech to language processing

Phillip M. Alday

To cite this article: Phillip M. Alday (2019) M/EEG analysis of naturalistic stories: a review from speech to language processing, Language, Cognition and Neuroscience, 34:4, 457-473, DOI: 10.1080/23273798.2018.1546882

To link to this article: https://doi.org/10.1080/23273798.2018.1546882

Published online: 19 Nov 2018.

LANGUAGE, COGNITION AND NEUROSCIENCE
2019, VOL. 34, NO. 4, 457–473
https://doi.org/10.1080/23273798.2018.1546882

Routledge
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

M/EEG analysis of naturalistic stories: a review from speech to language processing

Phillip M. Alday

Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

ABSTRACT

M/EEG research using naturally spoken stories as stimuli has focused largely on speech and not language processing. The temporal resolution of M/EEG is a two-edged sword, allowing for the study of the fine acoustic structure of speech, yet easily overwhelmed by the temporal noise of variation in constituent length. Recent theories on the neural encoding of linguistic structure require the temporal resolution of M/EEG, yet suffer from confounds when studied on traditional, heavily controlled stimuli. Recent methodological advances allow for synthesising naturalistic designs and traditional, controlled designs into effective M/EEG research on naturalistic language. In this review, we highlight common threads throughout the at-times distinct research traditions of speech and language processing. We conclude by examining the tradeoffs and successes of three M/EEG studies on fully naturalistic *language* paradigms and the future directions they suggest.

ARTICLE HISTORY
Received 1 March 2018
Accepted 2 November 2018

KEYWORDS
M/EEG; language; speech; naturalistic stimuli; auditory perception

Naturalistic language processing – language processing with non-trivial context, beyond the single-sentence level, in a modality used in everyday language use – has become an increasingly popular area of research in recent years (cf. Brennan, 2016; Willems, 2015a). Auditory stories are perhaps the most popular stimulus type for naturalistic language experiments using neuroimaging methods such as fMRI.[1] The great challenge in examining naturalistic input is the comparative lack of experimental control in the stimulus. Although modern statistical techniques such as mixed-effects models more easily allow for the inclusion of potential confounding covariates, one aspect of naturalistic input remains impossible to control purely statistically: temporal duration or extent of spoken language. While it is possible to control for temporal extent during stimulus selection and construction, this is of course an inherent tradeoff on the "artificial, experimentally controlled" – "naturalistic, experimentally variable" spectrum (cf. "controlled, simplified stimuli" and "ecological laboratory" traditions in Willems, 2015b). This tradeoff has had profound implications for research into auditory story comprehension in terms of neuroimaging used, leading to a preference for fMRI and a curious omission of M/EEG from most discussions of naturalistic language processing (cf. Andric & Small, 2015; Hasson & Egidi, 2015)

In particular, fMRI's poor temporal resolution is actually advantageous for temporally variable stimuli. In some sense, an impulse stimulus is essentially identical with a stimulus lasting up to about a second. Meanwhile, M/EEG's exceptionally high temporal resolution would appear particularly disadvantageous. Even controlling for frequency and morpho-syntactic or semantic effects, many manipulations suffer from a fundamental length confound. For example, "yellow" and "red" differ in length by at least 100ms, which is already the latency of the earliest ERP components for impulse stimuli. This is part of the reason why auditory ERPs in language studies look so different from visual ERPs – they reflect the temporal spread of the stimulus (cf. Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008 where the same study was conducted in the auditory and visual modalities, see also Dambacher et al., 2012 for the effect of SOA in the visual modality, and Hosemann, Herrmann, Steinbach, Bornkessel-Schlesewsky, & Schlesewsky, 2013; van der Brink & Hagoort, 2004 for the impact of recognition point).

This has generally resulted in a focus on *speech* processing instead of *language* processing for M/EEG research investigating narrative or otherwise more naturalistic auditory stimuli. The fine, acoustic structure of speech has much less of the problematic temporal

CONTACT P. M. Alday phillip.alday@mpi.nl

variability than language. In a particular telling use of terminology, Wöstmann, Fiedler, and Obleser (2016a)'s review of techniques in 'speech and speech comprehension' highlights the bias towards speech and away from language due to methodological challenges in addressing language comprehension. Moreover, Wöstmann and colleagues emphasise that even at the level of speech processing, the temporal variability still poses a large challenge. However, a broad spectrum of researchers have slowly closed the gap between speech and language processing such that we are now able to begin addressing auditory *story comprehension* and more generally language processing under naturalistic stimulation with M/EEG. In the following, we show how the two largely independent research traditions into the electrophysiology of (1) speech and (2) language processing of auditory stories and other naturalistic stimuli provide a foundation for testing recent, integrative theories on the neural encoding of linguistic structure (Giraud & Poeppel, 2012) that depend on predictions too temporally precise to be tested with fMRI.

1. Structure and overview

The present work is not intended as a tutorial, nor a detailed review of all work on naturalistic language processing. Instead, the goal is to highlight previous work with M/EEG spanning the spectrum ranging from "audition" to "language processing" that researchers at one or the other end of the spectrum may not be aware of. To that end, we begin with a somewhat shallow overview of previous work, showing how focus on speech vs. language perception created two largely distinct research literatures, compounded by methodological differences between MEG and EEG or between ERPs and oscillations. Although the speech tradition is somewhat more extensive than the language tradition for naturalistic auditory stimuli, the language tradition is more extensive than generally realised. In the second part, we focus on three publications (recent as of mid-2018) that provide the current state of the art in analysing the electrophysiology of naturalistic language processing. Highlighting the problem of distinct research traditions, none of these three papers cite the related work in the other two, nor of much of the previous literature. The methodological successes and tradeoffs of these publications are therefore discussed and contrasted against each other in more detail than the studies in earlier sections of this paper. We conclude with a discussion of opportunities and open problems in the electrophysiology of naturalistic language stimulation. In particular, we return repeatedly to the inherent difficulty in mapping temporally variable linguistic units to temporally precise measures and discuss potential solutions and promising directions for future research.

1.1. Literature search

As lack of awareness of previous literature was the motivation for the present manuscript, it is important to specify how literature was sought out for the current review. Several attempts were made to find appropriate literature in a (semi-) systematic way so as not to bias the perspective presented here too much by the author's own, previous "chance" exposure to the literature. In particular, we used several searches with Google Scholar using the search terms `naturalistic language`, `natural language`, `auditory story` combined with either `eeg` or `meg`. These yielded some interesting, recent articles but largely produced unrelated texts, often related to the more engineering-oriented fields of brain-computer interfaces (BCI) and natural language processing by computers (NLP). This was followed by a search of PubMed using the advanced search term `((MEG) OR (EEG)) and (((naturalistic) OR (natural) language) OR (auditory story))`. The PubMed search was somewhat more successful, yielding approximately 250 hits; however, only approximately 40 were even remotely relevant.[2] Beyond the aforementioned issues from technical fields, many results were for "natural speech" in the sense of "speech produced by humans", i.e. in contrast to "machine-generated speech" or "rapid serial visual presentation". In many cases, it became clear that the study of language processing – modern psycho- and neurolinguistics as opposed to the older field of aphasiology – has changed dramatically in what it considers "natural". For example, Kutas and Hillyard (1980a) used "natural sentence processing" to describe an RSVP experiment manipulating the typeface of single, sentence-final words (see also Lotze, Tune, Schlesewsky, & Bornkessel-Schlesewsky, 2011). Nonetheless, these searches yielded a number of results from both the speech processing and language processing traditions that provided a starting point for a less systematic "snowball" search through the reference lists of already collected articles.

1.2. Intentional omissions

It is critical to note that many of the results presented here are "replications", or more precisely, phenomena previously observed at smaller time scales or single sentences. This is a multi-edged sword. On the one hand, this serves as a sanity check for the more naturalistic experiments and as weak validation for the typical

"sterile" laboratory environment: the overall pattern of results is similar, so careful control did not distort our picture of language processing too greatly. Nonetheless, the naturalistic paradigms remain crucial for understanding speech and language processing in its full complexity, where interactions between both local linguistic and global contextual features, both linguistic (cf. Alday, Schlesewsky, & Bornkessel-Schlesewsky, 2017) and paralinguistic (Tromp, Peeters, Meyer, & Hagoort, 2017), matter as well as their constant, continuous overlap (cf. Steinberg, Truckenbrodt, & Jacobsen, 2012, who found that stimulus splicing and the resultant misleading formant transitions impact phonological processing). In the following, we present results from our search and not the more fundamental, original results from non-naturalistic studies. As such, this is not a comprehensive review of any of these individual phenomena and associated research programmes, but rather a synthesis of research traditions. This also implies that prominent work for our understanding of speech and language processing under artificial conditions that has not yet been extended to naturalistic stimulation (e.g. Ding, Melloni, Zhang, Tian, & Poeppel, 2016) is not part of this review, although the hope is that this review is useful in developing those naturalistic extensions.

Finally, this review is focused on the electrophysiology of naturalistic language stimulation in the auditory modality and not on other modalities, such as free reading, or methods, such as fMRI. The reading literature is quite extensive and involves a spectrum from visual perception up to language comprehension paralleling the speech vs. language divide in audition. The fMRI literature on naturalistic language processing is much more extensive than the M/EEG literature and more general works on naturalistic language processing tend to focus on fMRI research (cf. Brennan, 2016; Willems, 2015a, many of the other articles in this special issue).

2. Clues from the study of speech perception and first attempts at naturalistic language

The temporal complexity of language embedded in the speech signal arises at many levels: the difference in the length of phonemes, the number of phonemes required to realise a word, the number of words in a phrase, the number of phrases in a sentence, etc. Note that phonemes are not present at any instant in the acoustic signal, so, even as the smallest unit here, phonemes still encode temporally diffuse aspects of the latent linguistic signal. The temporal variability within each level only increases the temporal variability of all the levels above it. Moreover, the largest levels are much larger (seconds) than the time scale of the neural response (milliseconds), so even if the variability were controlled for, it can be difficult to model event-related changes when the event is spread out in time (although this does appear possible in certain situations with a threshold, cf. O'Connell, Dockree, & Kelly, 2012). For example, when does upspeak (rising intonation over the course of a sentence) occur? When is upspeak treated as a regional accent (as stereotyped in the American "valley girl" accent) or as a discourse marker for questions?

2.1. Entrainment in and across electrophysiological frequency bands

The first and most obvious way to avoid the temporal complexity is to compare signals on similar time scales such as the raw acoustic signal and raw electrophysiological response.[3] Especially after suitable filtering and downsampling, the envelope of speech has been repeatedly shown to be coupled to various aspects of the electrophysiological signal (see below). This coupling is often expressed as the electrophysiological signal being *entrained* or phase-locked to the auditory signal. Note that this phase-locking may be lagged and does not imply phase synchrony/identity, but rather that the phases co-vary.

Focusing on speech processing, Gross et al. (2013) showed that the speech envelope in a 7-minute-long continuous story entrains the phase of low-frequency bands (theta, delta) and the amplitude of gamma in MEG, with a larger effect for intelligible, forward speech compared to unintelligible, backward speech. Riecke, Formisano, Sorger, Başkent, and Gaudrain (2018) have recently suggested a causal role for this entrainment via transcranial stimulation with a waveform matching the speech envelope enhancing speech intelligibility and a mismatched waveform damaging intelligibility. In a subsequent EEG study, Kayser, Ince, Gross, and Kayser (2015) manipulated speech rates of 6-minute-long speech samples and found overall similar results. They additionally observed an alpha-delta correspondence, which they interpret as a top-down regulatory mechanism. Park, Ince, Schyns, Thut, and Gross (2015) re-analysed this data using transfer entropy[4] to perform causal connectivity analysis that supported Kayser and colleagues' conclusion of the regulatory role of alpha on delta. Using the same 7-minute story and MEG, Keitel, Ince, Gross, and Kayser (2017) provided additional converging evidence for the role of alpha in central areas in regulating delta in left Heschl's Gyrus and anterior STG, in addition to delta-beta and delta-theta entrainment in other, distinct networks. In brief, entrainment occurs at

multiple scales, both between the brain and the external stimulus and within different signals in the brain.

Entrainment has been suggested to play an important role in attention or more precisely the attending to salient stimuli. Horton, D'Zmura, and Srinivasan (2013) demonstrated entrainment in an EEG study for both attended and unattended stimuli via correlation analyses; however, the direction of the correlation was reversed for unattended stimuli. Whether the weak entrainment of unattended stimuli reflects a failure to completely suppress the unattended stimuli or some measure of non-suppressible automatic processing (or a mixture thereof) is not clear, although this remains an important question both in naturalistic and traditional experiments (Zion Golumbic, Poeppel, & Schroeder, 2012). This study also highlights the necessity of more naturalistic paradigms as the stimuli consisted of independent sentences from an audio corpus that had been concatenated together to exceed 22 seconds in length and thus lacked the overall coherence and shared context of natural language use, which may impact language processing. Regularity in the speech signal was also not achieved by prosodic rhythms but rather by convolution with a carefully controlled modulation function with the stimulus envelope.

Attentional effects have also been demonstrated in truly naturalistic stimulation. Using more explicit statistical modelling of the time course, Ding and Simon (2012) demonstrated that evoked spectral power was sensitive to manipulations of the attended but not the unattended speaker in 1-minute narratives in MEG (see also Ding & Simon, 2013). Using dynamic imaging of coherent sources (DICS, Gross et al., 2001) in MEG, Bourguignon et al. (2012) found that prosodic rhythms (as measured by the envelope of F0) entrain the delta band, even in hummed speech or speech in an unknown language, but theta coherence was only present for speech in a known (here: native) language, which can be thought of as a form of automatic linguistic attention. Moreover, the delta-band modulation was stronger in the pSTS and pSTG for linguistic than non-linguistic stimulation; with the latter showing its peak coherence in the auditory cortex. Wöstmann, Herrmann, Maess, and Obleser (2016b) found the previously reported regulation of delta via alpha in a cocktail party setting; in particular, alpha power and its lateralisation was modulated by delta phase coherence and its lateralisation (cf. Kayser et al., 2015). Taken together, these results start to separate the processing of speech as complex auditory stimulus from the processing of speech as the physical medium of language.

We can also consider language proficiency as related to the ability to attend to speech. Reiterer and colleagues show that the attenuation of the expected spectro-temporal response pattern correlates inversely with L2 proficiency (Reiterer, Hemmelmann, Rappelsberger, & Berger, 2005; Reiterer, Pereda, & Bhattacharya, 2011). The absence of entrainment under appropriate stimulation has also been proposed as a diagnostic and assessment tool for clinical research (Liberto & Lalor, 2017) and the development of appropriate prosthetics in hearing loss (Vanthornhout, Decruy, Wouters, Simon, & Francart, 2018; see also Peelle & Wingfield, 2016; Petersen, Wöstmann, Obleser, & Lunner, 2017, for more general comments on the neural impacts of hearing loss across the lifespan).

More recently, cross-modal studies have shown that visual information aligned with speech enhances entrainment. Prosodically timed emphatic hand gestures in speeches (Biau, Torralba, Fuentemilla, de Diego Balaguer, & Soto-Faraco, 2015) enhance entrainment, as do lip movements (stimuli approximately 8 minutes in both Giordano et al., 2017; Park, Kayser, Thut, & Gross, 2016). Using EEG and short cartoon clips (161 seconds on average), Cohen and Parra (2016) found that memorable scenes (approximately 18 seconds) were better remembered when there was higher neural synchrony during the scene and that this effect was in the multisensory scenes compared to the audio-only condition. Furthermore, Dikker et al. (2017) showed that this type of synchrony is predictive of group dynamics in a classroom setting.

2.1.1. Statistical models of the temporal response

The majority of work thus far has focused on inference or testing of hypotheses about the relationship between the electrophysiological signal and the stimulus. However, it is also possible to take a more generative or predictive approach to data analysis (Yarkoni & Westfall, 2017); the rise in the popularity of so-called "decoder" methods in cognitive neuroscience captures some aspects of this perspective (Holdgraf et al., 2017). In speech processing, Koskinen et al. (2012) used a similar decoding approach with canonical correlation analysis (a functional connectivity measure, cf. Carbonell, Worsley, Trujillo-Barreto, & Sotero, 2009, for M/EEG) to have the model learn the MEG "fingerprints" of short segments of speech (2–3 s) taken from an hour-long news broadcast. They demonstrated that their decoder could generalise from the observed data to have a reliable association between the spectro-temporal form of the auditory stimulus and of the resultant MEG signal. This parallels some work in the fMRI tradition such as Haxby et al. (2011), although full inter-subject alignment via an abstract representation has not yet been demonstrated for the electrophysiology of language.

Hypotheses about the role of the various frequency bands in language processing and more general cognitive function, such as those put forth by Giraud and Poeppel (2012) and more recently tested in an artificial context by Ding and colleagues (Ding, Melloni, Tian, & Poeppel, 2017; Ding et al., 2016) are a first step, but more work is needed to establish stronger hypotheses linking language to its implementation in the brain (cf. Poeppel & Embick, 2005, see also Section 3.4 below).

Beyond decoding models, another way to start establishing such linking hypotheses are generative models. Ding and Simon (2012) and Ding and Simon (2013) formulated their analysis as "(spectro-)temporal response functions" or (S)TRF, which can essentially be thought of as a generalisation of the event-related potential to a complex stimulus (see below). In particular, the "spectro" portion of the name refers not to the electrophysiological response but to the stimulus, with "the" TRF as a whole resulting from the summing of piecewise TRFs to individual portions of the frequency spectrum of the stimulus. In defining TRFs, it is also possible to include additional covariates, such as linguistic information. Di Liberto and colleagues (DiLiberto, O'Sullivan, & Lalor, 2015; Liberto & Lalor, 2016) took a model-based approach to entrainment and found that the inclusion of phonemic labels in addition to the speech envelope improve the model's ability to predict EEG data in all frequency bands. Note phonemes are not present at any instant in the acoustic signal, so phonemic labels encode temporally diffuse aspects of the latent linguistic signal. This suggests that such linguistic labels capture some part of the response not directly captured by the acoustic properties of the stimulus. As such, this model provides evidence for the hypothesis that the abstract entity "phoneme" from linguistic theory captures something that is relevant for the processing of language in the brain.

2.2. The temporal response as an impulse response through embedded probes

In addition to the previously discussed time-frequency based entrainment analyses, Kayser et al. (2015) also examined the evoked potential. Here, they applied the second common trick to avoid the temporal complexity of natural speech: they inserted sharply defined events into the continuous speech stream. Their speech rate manipulation consisted of manipulating the gaps between syllables, which effectively creates an impulse-like aspect to the onset of the next syllables. This again has its parallel in the traditional ERP literature, where the usual exogenous components – N1, etc. – are visible at the start of the auditory stimulation, even if that is rarely a critical position. This type of "gap splicing" can be somewhat problematic (Steinberg et al., 2012) when studying phonological processing because it removes co-articulation; however, here it is less problematic as the speech envelope and not the role of any particular phoneme or higher processing was the object of interest. Kayser and colleagues found no effect of the speech rate manipulation, although the amplitude of the evoked potential did increase with the duration of pause, regardless of overall speech rate.

In such contexts, it becomes clearer that the classical, peaky evoked potential is essentially an impulse response, although the "impulse" may be at a more abstract level. In traditional rapid serial visual presentation, the impulse is simultaneous at several levels. The early exogenous components reflect the impulse response at the level of visual processing, while the N400 reflects impulse responses at more abstract linguistic levels such as words. A useful metaphor is that of a ringing bell. The impulse response reflects the sound, including the continued ringing of the bell, when it is struck once. Note, however, that if we strike the bell again, the second strike also produces an impulse response, albeit convolved with the first impulse response. From this perspective, the sharp shape of components in traditional experiments reflects the relative separation of the various impulses. Introducing a longer pause as Kayser et al. (2015) did, reduces the amount of convolution and provides a clearer perspective of the impulse response. Nonetheless, speech and language rarely occur as a series of sharply defined impulses, but rather as more continuous stimulation.

In continuous, naturalistic stimulation, the evoked potential is viewed slightly differently than in the traditional ERP literature and is often described as the temporal response function or temporal receptive field (both abbreviated TRF), as it shows the time course of the neural sensitivity, when distinct receptive fields are estimated across sensors.[5] Although the TRF perspective was originally developed for low-level perceptual experiments in psychophysics, its potential for application to language was quickly realised (where it is called Auditory Evoked Spread Spectrum Analysis (AESPA) in parallel to the auditory evoked potential (AEP); Lalor & Foxe, 2010). The TRF is often a more abstract perspective, using predictions from more complicated models than "mere" (grand) averaging,[6] such as explicitly convolution-based perspectives (see below) or linear model-based perspectives (e.g. the rERP framework; Smith & Kutas, 2015a, 2015b).

The TRF thus reflects the impulse response at the level of the manipulation, but not necessarily at the level of physical realisation of the stimulus in sound or light, because the impulse may be more abstract than the

physical stimulus at an instant in time. In particular, linguistic impulses such as words are not acoustic impulses, and this is exactly why embedded probes are useful to extract ERPs or more generally TRFs. By embedding targets constrained to be acoustically similar and/or temporally compact, the linguistic impulse becomes more consistent across trials and it becomes easier to extract away its distinct time course, i.e. its impulse response. In these terms, analysing the TRF is an issue of ways to isolate the correct abstract impulse and its response. This illuminates the contrast to fMRI: due to its poor temporal resolution, physical and linguistic impulses are forced onto the same time scale.

Perhaps the most temporally precise application of the evoked TRF to date is isolating the auditory brainstem response (ABR) by Maddox and Lee (2018) who used 64-second clips of an audiobook and EEG sampled at 10kHz. The high sampling rate and minimal filtering is necessary to determine the extremely low latencies characteristic of ABRs and stands in stark contrast to nearly all of the other studies here, which often downsampled their stimulus to around 100 Hz and applied extremely strong bandpass filters, although there is no shortage of literature discussing problems with filtering (see below for further discussion on the role of filtering in naturalistic stimulation; Acunzo, MacKenzie, & van Rossum, 2012; Maess, Schröger, & Widmann, 2016; Tanner, Morgan-Short, & Luck, 2015; Tanner, Norton, Morgan-Short, & Luck, 2016; Van Rullen, 2011).

A number of studies have investigated the response to probes inserted at various levels along the sound-speech-language continuum using classical ERP techniques. In the simplest cases, either linguistic (such as minimal syllables like /da/) or non-linguistic sounds (such as short buzzes) are inserted into the audio stream and serve as time-locking events. While these stimuli are short and uniform in length, their presence dramatically reduces the naturalness of the stimulus. Nonetheless, such stimuli do occur in real life situations, e.g. poor telephone connections or a babbling baby in the background. These stimuli are often used in conjunction with a cocktail-party task that additionally manipulates attention (cf. da Rocha, Foz, & Pereira, 2015; Karns, Isbell, Giuliano, & Neville, 2015; Nager, Dethlefsen, & Münte, 2008; Sanders, Stevens, Coch, & Neville, 2006; Stevens, Sanders, & Neville, 2006; see Table 1 for more details).

As a concrete example of probes in a cocktail party situation, Getzmann and Falkenstein (2011) had participants listen for the stock price of a particular company in a stock-price-ticker cocktail party situation. They found that age did not have a large behavioural impact, but did influence several classical attention and target related components such as the P3a and P3b.

These attention manipulations and associated behavioural tasks, although in many ways somewhat artificial, span parts of the gap from speech perception towards linguistic processing. Categorising and responding to linguistic features of the acoustic-speech signal requires some of level of language processing. However, traditional topics of psycholinguistic research – syntax and semantics in their broadest sense – remain unmanipulated and uncontrolled for in these studies, and the linguistic nature of the stimulus is ignored, instead treating the speech signal as just another acoustic signal which humans have particular expertise on. In other words, the linguistic content of the stimulus largely did not matter and language was just a carrier for a complex perceptual categorisation task. Nonetheless, these studies provided insights into how to deal with the complexity of continuous, naturalistic stimulation. Oscillatory entrainment, full modelling of the time course with covariates (TRFs), and the use of sharply defined probes provide the basic toolkit for studying the brain's response to temporally diffuse stimuli. And so we now shift our focus to language processing, where language exists as a complex system shaping and shaped by its acoustic realisation as speech.

3. From speech to language

Classical ERP research on sentence processing was quickly extended to accommodate contexts, both in the visual and auditory domains (cf. e.g. Kutas & Federmeier, 2011, for a review of the N400, including contextual effects). In both modalities, this was generally done with a deliberately constructed context and a carefully controlled critical word or even entire sentence. This was done for all levels of linguistic processing, e.g. prosodic processing (Dimitrova, Stowe, Redeker, & Hoeks, 2012) and the interaction of syntax and semantics (Nieuwland & Berkum, 2006). More recently, this has been attempted in cross-modal contexts, including virtual reality (e.g. Tromp et al., 2017, for an N400 elicited by mismatch to the virtual context). In all these cases, the context was directly, intentionally part of the manipulation and not part of a truly more naturalistic paradigm with e.g. long contexts, variability in speech, and without tasks and related effects (such as modulation of late positivities, cf. Haupt, Schlesewsky, Roehm, Friederici, & Bornkessel-Schlesewsky, 2008).

3.1. First steps: probes and parsers

The highly artificial nature of language in the laboratory and resultant issues of ecological validity were cause for concern, even before fully naturalistic language in the

laboratory was conceivable. Quite early in the characterisation and study of the electrophysiology of language, O'Halloran, Isenhart, Sandman, and Larkey (1988) used a standard semantic-anomaly N400-elicitation manipulation with less acoustic-phonological control to examine many of the concerns that we are now trying to address with naturalistic stimuli. Two tasks were used (content question following a block and judgement task), and co-articulation effects were avoided without splicing by careful selection of the preceding word to have a terminal voiceless stop. In modern terminology, they found a biphasic N400-P600 pattern (similar to the one in Kutas & Hillyard, 1980b), with an enhanced positivity for the judgement task (cf. Haupt et al., 2008). Moreover, it appears that they allowed for natural volume variation, stating "none of the stimulus words was electronically altered in any way" (p. 245). This was taken as support for the use of similar, but non-identical targets across conditions in the auditory modality. As such, this study provided a critical result for more naturalistic designs, namely that linguistic effects can still be detected in acoustically variable stimuli, despite its use of single sentences as stimuli.

Until relatively recently, restricted contexts with controlled critical sentences and O'Halloran and colleagues' single-sentence study without volume normalisation were the limits of naturalistic language processing as studied in EEG. One notable exception used a technique similar to the linguistic probe methodology in the cocktail-party literature (above). Shafer, Kessler, Schwartz, Morr, and Kurtzberg (2005) used an audio rendition of a children's story in which all the instances of the definite article *the* were replaced by a single exemplar. This was repeated for a nonsense (jabberwocky) condition, and the articles in the two conditions were compared. Using a single, short word (76 ms) with a single acoustic realisation elicited clear ERPs, including the early exogenous components; however, even with an additional task manipulation in a follow-up experiment, it is hard to interpret this experiment beyond something akin to "context impacts the processing of a single function word". This study does however provide a very precisely estimated TRF for a particular linguistic impulse, including how the embedding context changes its appearance. In addition, the clear waveforms here for a short, frequent word suggest an alternative approach to analysing naturalistic data – focus on a few frequent, short tokens such as personal pronouns and many confounds (acoustics, frequency, etc.) will resolve themselves to a level approaching a controlled experiment (see Brilmayer, Werner, Primus, Bornkessel-Schlesewsky, & Schlesewsky, 2018, for more with this approach).

Even with appropriate linking hypotheses for estimating the time scale of the linguistic impulse, the aggregate noise and drift in the time-domain electrophysiological signal over multi-second time scales makes extraction of the associated response impossible with current techniques. Meanwhile, in fMRI, a number of methods and approaches blossomed (see the rest of this issue for some examples and additional literature review). Haxby's hyperalignment approach, based on multivariate pattern analysis, proved particularly fruitful for audiovisual scenes (cf. Guntupalli et al., 2016; Haxby et al., 2011). Huth, de Heer, Griffiths, Theunissen, and Gallant (2016) were able to create a semantic map of the cortex from language input. For auditory stories, Whitney et al. (2009) examined narrative shifts, while Hasson and collaborators used the temporal hierarchical organisation of speech (e.g words within sentences within "paragraphs") to examine the hierarchy of temporal processing within perisylvian cortex (Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Lerner, Honey, Katkov, & Hasson, 2014; Lerner, Honey, Silbert, & Hasson, 2011; Stephens, Honey, & Hasson, 2013). Narrative shifts and multiword organisational units are on a time scale simply incompatible with event-related electrophysiology.

One fMRI technique, however, can be transferred to methods with high temporal resolution: the use of parser metrics as regressors. Brennan et al. (2012) used the current node-count from a context-free parser at a given word in a sentence as a proxy regressor for syntactic structure building. They found that node count correlated with activity in the left anterior temporal lobe, not with activity in inferior frontal areas, i.e. Broca's area, commonly associated with syntax, and attributed this difference to the naturalistic environment without tasks or violations. Willems, Frank, Nijhof, Hagoort, and van den Bosch (2015) expanded this to use a "syntax-free" trigram and entropy-based parser to model data without assuming any traditional theory of syntax. Building on the work of Hale (2001) and Levy (2008), who proposed mappings from natural language processing methods (parsers and Shannon entropy), the use of NLP models as regressors has expanded in fMRI research (cf. Brennan, Stabler, Van Wagenen, Luh, & Hale, 2016) and is slowly making the transition towards M/EEG (see Armeni, Willems, & Frank, 2017; Brennan, 2016 for review, as well as van Schijndel, Murphy, & Schuler, 2015 for an early attempt in the time-frequency domain with only three participants).

The state of the art of this approach as of mid-2018 is the use of probabilistic grammars with beam search. Beam search keeps a ranked collection ("beam") of parser states compatible with the current input that may be pruned when presented with future, incompatible input. Hale, Dyer, Kuncoro, and Brennan (2018) used

recurrent neural network grammars (RNNG) combined with word-synchronous beam search to model not just a single, deterministic partial parse, but also local (syntactic) ambiguity. Both RNNG and beam search are advances compared to previous parsing-based models. RNNGs are generative models of both deep and surface structure, in that they generate both an observable word string and a hidden tree structure. This distinguishes them from traditional parsers, which derive trees from a given word string. Meanwhile, the state of the beam is both a measure of the local syntactic ambiguity and a model of parallelism in human sentence processing.

In their study, Hale and colleagues recorded EEG data at 500 Hz from subjects listening to the first chapter of *Alice's Adventures in Wonderland* and bandpass filtered from 0.5–40 Hz. The authors used both sample-wise (timepoint-wise) regression within subjects and electrodes as well as mixed-effects models within time windows and topographical regions of interests to compare different incremental complexity metrics derived from the RNNG. They found that their complexity models captured activity corresponding to an early anterior component as well as the P600, but not activity corresponding to the N400. Moreover, they found that the early activity was attributable to syntactic composition, while the later P600-like activity was attributable to the overall syntactic effort ("distance" in their language). In brief, using innovations from computational linguistics, Hale and colleagues quantitatively demonstrate the explanatory power of purely syntactic models and discern between aspects of syntactic processing. This stands in contrast to the other approaches that appeared at roughly the same time, which focused more on semantics and word-level phenomena, as shown in the next section.

3.2. Scaling up: using modern statistical and computational approaches

At the end of 2017 and the beginning of 2018, three closely related approaches have emerged that build upon different aspects of ERP and fMRI traditions to auditory story comprehension. Although the surface methods appear quite distinct, all of these methods depend to a greater or lesser extent on general linear model-based convolution, a method which has found success in other areas of EEG research, e.g. removing eye-movement artefacts (cf. Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011).

3.2.1. Estimating full TRFs via filtering

Broderick, Anderson, Liberto, Crosse, and Lalor (2018) linearly regressed EEG data onto (dis)similarity measures for distributional semantics derived from the Word2Vec tool in NLP (Mikolov, Chen, Corrado, & Dean, 2013).[7] EEG data were recorded from subjects listening to 20 3-minute audiobook snippets, downsampled to 128 Hz and bandpass filtered from 1–8 Hz. The authors used ridge regression to calculate the TRF and found a negativity with a latency and distribution corresponding to a traditional N400. Broderick and colleagues hesitated to use the N400 label, however, as they had no cloze probability manipulation; nonetheless, in light of more recent N400 perspectives (Kutas & Federmeier, 2011) beyond discrete components (Alday et al., 2017), the resulting component can safely be classified as an N400. Furthermore, noise and cocktail-party attention manipulations attenuated this response. In brief, Broderick and colleagues found a classic N400 response in a rich context, lacking explicit violations.

Brodbeck, Presacco, and Simon (2018) used linear kernel estimation (a form of convolution) combined with minimum norm source localisation to model the electrophysiological response in an MEG experiment using 1-minute-long segments from an audiobook. The MEG data were downsampled to 100 Hz and bandpassed filtered from 1 to 40 Hz. They clearly place their work in the Lalor temporal response function tradition and used continuous signals of the type often used in Lalor and colleagues' previous TRF work (see Section 2.2 above). Beyond the standard acoustic envelope, they also created continuous signals corresponding to word frequency and semantic composition. For both word frequency and semantic composition, the signal was taken to maintain a constant value over the duration of the relevant critical word. In the case of semantic composition, the variable was simply a binary indicator of semantic composition, following the definition given by Westerlund, Kastner, Kaabi, and Pylkkänen (2015), who examined single sentences and essentially used "composition" to indicate that two adjacent words belong to the same phrase. Although these predictor signals were fit in a linear kernel (convolution) model via boosting, Brodbeck and colleagues note that there is no reason why other fitting techniques such as ridge regression could not be used. This is a clear parallel to the work of Lalor and Foxe (2010) who suggest both standard ordinary least squares and ridge regression. The big innovation here is a model of the electrophysiological response in source and not sensor space. Even beyond the differing predictors, the description in terms of scalp components vs. source time courses makes Brodbeck's and Broderick's (and respective colleagues) approaches as much complementary as overlapping.

Both of these studies used filtering to constrain the electrophysiological response to time scales of interest.

The 1–8 Hz filter used by Broderick et al. (2018) captures events at the scale of language (syllables, words and some phrases), but does not capture events at speech scale. The 1-40 Hz filter used by Brodbeck et al. (2018) includes time scales at the level of speech in its pass-band. These filters remove the signal drift common in electrophysiological signals at the cost of potentially introducing artefacts (Acunzo et al., 2012; Maess et al., 2016; Tanner et al., 2015, 2016; Van Rullen, 2011). However, they are computationally simpler than using temporal offset across the entire recording and not just within epochs when modelling the full time courses within epochs.

3.2.2. Estimating peak TRF via temporal constraints

In contrast to the above, Alday et al. (2017) used a more traditional ERP analysis, focusing on mean EEG voltage in a fixed time window (300–500 ms), time locked to word onset. Although basic results were presented for an rERP-perspective on TRF (no autocorrelation, regression coefficients interpreted as response), the core focus was complex interactions in a longer (23 minute) narrative in German. Instead of an explicit convolution model, linear mixed-effects models were used (as is increasingly the case in the ERP literature on language processing) and as many interactions of linguistic features (e.g. frequency, morphological case, word order, animacy) as possible as well as position within the story were modelled. Mixed-effects models can pool information from every trial from every subject (and item) to simultaneously capture both subject and population level effects. This partial pooling of information between subjects improves inference at both the participant and population level. In contrast to the other two naturalistic M/EEG studies in this section, filtering was comparatively mild (0.3 Hz highpass, with line-noise removal via sine wave fitting), albeit with artefact correction via ICA. Position within the story ("index") was included as a regression covariate, which served to model both the impact of context and overall signal drift. Modelling was performed in a unified hierarchical step accommodating inter- and intra-subject variability without the need for the two-stage approach often used in machine-learning based procedures, such as those used in the M/EEG studies above.

The key finding of this study also highlights the importance of naturalistic stimuli. Experimental control in traditional psycholinguistic experiments not only eliminates confounds, but introduces them, especially in terms of timing and co-occurrence. Despite the fixed offset to onsets of words of vastly different lengths embedded in phrases of different lengths in a morphological rich language – i.e. capturing different amounts of a given word or noun phrase – we found the usual pattern of N400 results. This suggests that the electrophysiological response observed at the usual N400 latency is not tied to words per se, but rather information units at a time scale that corresponds to individual words in more carefully controlled experiments. In other words, the linguistic impulse may not be words per se. Instead, word-level phenomena in traditional, controlled designs may reflect a confound between the time scale of words and the time scale of the linguistic impulse.

3.3. The way forward

Ultimately, these approaches highlight the way forward. All of them are based essentially on linear models. Alday et al. (2017) and Hale et al. (2018) both used a straightforward application of contemporary single-trial ERP analysis with mixed-effects (and two-stage) regression models to accommodate the covariates outside of experimental control (cf. Sassenhagen & Alday, 2016). Broderick et al. (2018) and Brodbeck et al. (2018) both used linear kernel models that are equivalent to linear regression with lags included in the predictors, although they used different fitting methods.[8] These all represent a different take on what is perhaps most widely known in the ERP literature as rERP (Smith & Kutas, 2015a, 2015b). Each of these approaches addresses individually the first challenge set forth in Wöstmann et al. (2016a) ("how to assess event-related responses to temporally varying speech stimuli?") and presents a solution for assessing event-related responses to linguistic information within a continuous speech stream.

As of now, entrainment phenomena at the level of language have not been examined under naturalistic stimulation (see below for speculation as to why). From a methodological perspective, we need similar hierarchical models as extensions to methods currently available to research entrainment phenomena, such as DICS. Driven auto-regressive models (la Tour et al., 2017) are currently formulated in a manner applicable to single subjects, but it is relatively straightforward to reformulate them as a hierarchical model capable of being fit in the mixed-effects framework. The "regression" framework underlies the majority of our modern statistical technique and this could serve a lingua franca for researchers from different areas.

The different perspectives from deeply related techniques are not per se problematic and indeed may even be helpful, just as frequency domain and temporal domain analyses reveal different aspects about fundamentally the same information (at least at the single-trial level; averaging of course loses information). Ultimately, it is useful to be able to combine different

tools to study the full, complex nature of language processing (da Rocha et al., 2015). The important thing is that we do not allow the development of multiple parallel research traditions with little exchange as happened previously for MEG and EEG (cf. Salmelin & Baillet, 2009). With competing measurement techniques (MEG, EEG), modelling techniques (convolution, regression, etc.) and objects of interest (speech vs. language), there are already enough hurdles to overcome in communication assuming fully collaborative research traditions. Neither Broderick, Brodbeck nor Alday cited each other or seemed aware of each other's work before publication. Moreover, Smith's rERP papers never mention any of Lalor's work, although the ideas are extremely closely related, and the only connection to convolution is a brief reference to the hemodynamic response function in fMRI. This suggests a dangerous disconnect amongst subfields.

3.4. Methodological issues are theoretical issues

The lack of coordination between research traditions – whether speech vs. language or MEG vs. EEG – also reflects a lack of a coherent, integrative theory of language processing from perception to comprehension.[9] The speech signal is physically or "surface" observable; its impulse is the rapid changes in air pressure that the human ear is sensitive to. There are agreed-upon methods for measuring acoustics, based on more general work in signal processing and physics, and the temporal properties of sound are well understood. The language signal meanwhile is largely latent. Even basic, relatively well-accepted units such as words are not clearly defined cross-linguistically nor are they apparent from the raw speech signal, as the difficulty of forced alignment shows. Moreover, linguistic features are often strongly correlated and confounded in natural language use, which undermines efforts to find linguistic primitives. In other words, we lack not just linking hypotheses between linguistic computations/representations and neural computations/representations (cf. Embick & Poeppel, 2014; Martin, 2016; Poeppel & Embick, 2005), but even conclusive evidence that the postulated linguistic representations have neurobiological reality. This has a parallel in the history of psycholinguistics. While the wug test was initially taken to demonstrate the psychological reality of linguistic rules (cf. Berko, 1958), the rise of connectionism and the subsequent past-tense debate showed that experimental explanatory power is necessary but not sufficient to demonstrate the psychological existence of a latent construct such as morphological rules (cf. Rumelhart & McClelland, 1986).

We can address this issue by using neurobiology to inform linguistic theory (cf. Duncan, Tune, & Small, 2016) instead of just looking for correlates of linguistic theory in brain activity. As a complementary approach, we can take George Box's aphorism "All models are wrong, but some models are useful" to heart when looking for correlates of linguistic structure in the brain. As an example, both the differences and the similarities in how well the performance of different parsers maps to neurophysiological signals provides insight into which proposed constructs from linguistic theory may be utilised by the brain (cf. Brennan, 2016). This comprehensive methodology – comparing and contrasting neurobiologically informed linguistic models – must also be applied to language in its full complexity, which means that integrative, naturalistic paradigms are an absolute requirement (cf. Small & Nusbaum, 2004).

The difficulty in comparing "red" and "yellow" in auditory EEG lies not in any individual level (acoustic, phonological, lexical, etc.), but in mapping between those levels. We can easily compare these two words at the level of speech and phonemes; we can also easily compare them when speech and its temporal dimension is removed in rapid serial visual presentation. The challenge comes then in the temporal mapping of the series of acoustic impulses to a series of linguistic ones. This in turn requires better linking hypotheses of how linguistic levels map onto time and space in the brain.

In brief, a core challenge for developing sufficient linking hypotheses lies in addressing both temporal regularity and temporal variation. This is ultimately what separates electrophysiology from methods with a lower temporal resolution such as fMRI: we cannot ignore the temporal variation. Combining analysis of methods spanning the range of space and time will show where the granularity of our linking hypotheses fail (cf. Haufe et al., 2018).

4. Conclusion

In this brief review, we have looked at two parallel traditions in studying the electrophysiology of naturalistic auditory linguistic stimulation. The first, "speech processing", has focused largely on oscillatory responses to acoustic aspects of the stimulus. The second, "language comprehension", has largely focused on the evoked, temporal response to carefully controlled manipulations embedded in larger, less controlled contexts. Recent work has however demonstrated the feasibility of examining language comprehension in fully naturalistic environments. This opens up a new frontier for testing combined theories of speech and language comprehension spanning the range from milliseconds to seconds

such as the AST and its descendants (Giraud & Poeppel, 2012; Poeppel, 2003). The full temporal complexity of natural speech eliminates rhythmic confounds in short, isochronous speech (cf. Ding et al., 2016; Frank & Yang, 2018, and the broader debate about the origins of the observed rhythms) and is a challenge we can now take on. The necessary statistical and signal-processing methods have come of age; electronic recordings of longer auditory stimuli are more widespread than ever in the form of audiobooks and podcasts; and NLP advancements (parsers, forced-alignment systems) make annotating them easier than ever. Controlled laboratory experiments remain invaluable, but only by embracing naturalistic designs as part of a comprehensive examination of language will we begin to understand language processing in its full natural complexity.

Notes

1. An obvious exception is co-registration of EEG and/or fMRI with eye movements for examining the dynamics of natural reading, see e.g. Kretzschmar et al. (2013).
2. Only approximate numbers are given as these results are of course subject to change, even on short notice; moreover many entries were duplicates.
3. Throughout the text "electrophysiological response" is used to describe the underlying signal measured by both MEG and EEG, as the magnetic fields measured by MEG are secondary to the electrical fields used in neural computation – the underlying physiology is electrical, even if we sometimes measure the magnetic portion of the resulting change in the electromagnetic field.
4. Very roughly, this is internal unpredictability in one signal appearing at a constant temporal lead to unpredictability in another signal, which by the same logic as Granger causality is assumed to reflect the transfer of (Shannon) entropy between signals and thus causal coupling.
5. While these concepts are slightly different formulations of the same idea, there are some differences in their terminological use. In general, the temporal response function perspective focuses on the impact of the stimulus on the electrophysiological response, while the temporal receptive field focuses on the portions of the electrophysiological response that are sensitive to the stimulus. This is often a distinction without a difference, but it has some impact on other parts of the terminology. In particular, the prefix "spectro" can somewhat confusingly refer either to the stimulus or to the electrophysiological response, although there does seem to be tendency for "spectro-temporal *response function*" to refer to the stimulus spectrum and its mapping to electrophysiology, while "spectro-temporal *receptive field*" emphasises the portions of the electrophysiological response actually impacted. Fundamentally, all of these perspectives refer to the same general phenomena, namely the time course of the electrophysiological modulation driven by the stimulus, and this is the abstraction we use here.
6. Although often not viewed as such, averaging is a model of the "expected value" of a variable. Indeed, "expected value" is a technical term in statistics, which can be conceived of as an abstract generalisation of the average/arithmetic mean.
7. Word2Vec can be thought of as a generalisation of co-occurrence statistics and a successor to techniques such as latent semantic analysis (LSA).
8. A similar model could also be made using an appropriate autoregressive covariance structure instead of explicitly including lags in the predictors.
9. We would like to thank an anonymous reviewer for encouraging us to include such a section in the manuscript and for raising some interesting discussion points.

Acknowledgements

I would like to thank two anonymous reviewers, whose feedback greatly improved the readability of manuscript and encouraged more precise yet more concise formulations of summaries throughout the text. Reviewer 2 in particular suggested the current structuring of the paper. Additionally, FK, IB, GK, and AK all read various drafts of the manuscript and provided feedback. All remaining mistakes are my own.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Phillip M. Alday http://orcid.org/0000-0002-9984-5745

References

Acunzo, D. J., MacKenzie, G., & van Rossum, M. C. W. (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. *Journal of Neuroscience Methods*, *209*(1), 212–218. doi:10.1016/j.jneumeth.2012.06.011

Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2017). Electrophysiology reveals the neural dynamics of naturalistic auditory language processing: Event-related potentials reflect continuous model updates. *eNeuro*, *4*(6). doi:10.1523/ENEURO.0311-16.2017.

Andric, M., & Small, S. L. (2015). fMRI methods for studying the neurobiology of language under naturalistic conditions. In R. M. Willems (Ed.), *Cognitive neuroscience of natural language use* (pp. 8–28). Cambridge: Cambridge University Press.

Armeni, K., Willems, R. M., & Frank, S. L. (2017). Probabilistic language models in cognitive neuroscience: Promises and pitfalls. *Neuroscience & Biobehavioral Reviews*, *83*, 579–588. doi:10.1016/j.neubiorev.2017.09.001

Berko, J. (1958). The child's learning of english morphology. *Word*, *14*, 150–177. doi:10.1080/00437956.1958.11659661

Biau, E., Torralba, M., Fuentemilla, L., de Diego Balaguer, R., & Soto-Faraco, S. (2015). Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. *Cortex*, *68*, 76–85. doi:10.1016/j.cortex.2014.11.018

Boudewyn, M., & Carter, C. (2018). I must have missed that: Alpha-band oscillations track attention to spoken language. *Neuropsychologia*, *117*, 148–155. doi:10.1016/j.neuropsychologia.2018.05.024

Bourguignon, M., Tiège, X. D., de Beeck, M. O., Ligot, N., Paquier, P., Bogaert, P. V., … Jousmäki, V. (2012). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, *34*(2), 314–326. doi:10.1002/hbm.21442

Brennan, J. (2016). Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass*, *10*(7), 299–313. doi:10.1111/lnc3.12198

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*(2), 163–173. doi:10.1016/j.bandl.2010.04.002

Brennan, J., & Pylkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *NeuroImage*, *60*(2), 1139–1148. doi:10.1016/j.neuroimage.2012.01.030

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157–158*, 81–94. doi:10.1016/j.bandl.2016.04.008

Brilmayer, I., Werner, A., Primus, B., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2018). The exceptional nature of the first person in natural story processing and the transfer of egocentricity. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2018.1542501

Brodbeck, C., Presacco, A., & Simon, J. Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage*, *172*, 162–174. doi:10.1016/j.neuroimage.2018.01.042

Broderick, M. P., Anderson, A. J., Liberto, G. M. D., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Current Biology*, *28*(5), 803–809.e3. doi:10.1016/j.cub.2018.01.080

Carbonell, F., Worsley, K. J., Trujillo-Barreto, N. J., & Sotero, R. C. (2009). Random fields-union intersection tests for detecting functional connectivity in EEG/MEG imaging. *Human Brain Mapping*, *30*(8), 2477–2486. doi:10.1002/hbm.20685

Cohen, S. S., & Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *eNeuro*, *3*(6). doi:10.1523/eneuro.0203-16.2016

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia*, *50*(8), 1852–1870. doi:10.1016/j.neuropsychologia.2012.04.011

da Rocha, A. F., Foz, F. B., & Pereira, A. (2015). Combining different tools for EEG analysis to study the distributed character of language processing. *Computational Intelligence and Neuroscience*, *2015*, 865974. doi:10.1155/2015/865974

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., … Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, *27*(9), 1375–1380. doi:10.1016/j.cub.2017.04.002

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*(19), 2457–2465. doi:10.1016/j.cub.2015.08.030

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and eeg in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, *140*(4), 552–572. doi:10.1037/a0023885

Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. J. (2012). Less is not more: Neural responses to missing and superfluous accents in context. *Journal of Cognitive Neuroscience*, *24*(12), 2400–2418. doi:10.1162/jocn_a_00302

Ding, N., Melloni, L., Tian, X., & Poeppel, D. (2017). Rule-based and word-level statistics-based processing of language: Insights from neuroscience. *Language, Cognition and Neuroscience*, *32*(5), 570–575. doi:10.1080/23273798.2016.1215477

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. doi:10.1038/nn.4186

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*(29), 11854–11859. doi:10.1073/pnas.1205381109

Ding, N., & Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *Journal of Neuroscience*, *33*(13), 5728–5735. doi:10.1523/jneurosci.5297-12.2013

Duncan, E. S., Tune, S., & Small, S. L. (2016). The neurobiology of language: Relevance to linguistics. *Yearbook of the Poznan Linguistic Meeting*, *2*(1), 49–66. doi:10.1515/yplm-2016-0003

Embick, D., & Poeppel, D. (2014). Towards a computational(ist) neurobiology of language: Correlational, integrated and explanatory neurolinguistics. *Language, Cognition and Neuroscience*, *30*(4), 357–366. doi:10.1080/23273798.2014.980750

Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PloS One*, *13*(5), e0197304. doi:10.1371/journal.pone.0197304

Getzmann, S., & Falkenstein, M. (2011). Understanding of spoken language under challenging listening conditions in younger and older listeners: A combined behavioral and electrophysiological study. *Brain Research*, *1415*, 8–22. doi:10.1016/j.brainres.2011.08.001

Giordano, B. L., Ince, R. A. A., Gross, J., Schyns, P. G., Panzeri, S., & Kayser, C. (2017). Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLife*, *6*, e24763. doi:10.7554/elife.24763

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. doi:10.1038/nn.3063

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752. doi:10.1371/journal.pbio.1001752

Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, *98*(2), 694–699. doi:10.1073/pnas.98.2.694

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A model of representational spaces in human cortex. *Cerebral Cortex*, *26*(6), 2919–2934. doi:10.1093/cercor/bhw068

Hale, J. (2001). *A probabilistic early parser as a psycholinguistic model*. Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies, NAACL '01, Stroudsburg, PA (pp. 1–8). Association for Computational Linguistics.

Hale, J., Dyer, C., Kuncoro, A., & Brennan, J. (2018). *Finding syntax in human encephalography with beam search*. Proceedings of the 56th annual meeting of the Association for Computational Linguistics (Vol. 1: Long Papers), Melbourne, Australia (pp. 2727–2736). Association for Computational Linguistics.

Hasson, U., & Egidi, G. (2015). What are naturalistic comprehension paradigms teaching us about language? In R. M. Willems (Ed.), *Cognitive neuroscience of natural language use* (pp. 228–255). Cambridge: Cambridge University Press.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *The Journal of Neuroscience*, *28*(10), 2539–2550. doi:10.1523/JNEUROSCI.5487-07.2008

Haufe, S., DeGuzman, P., Henin, S., Arcaro, M., Honey, C. J., Hasson, U., & Parra, L. C. (2018). Elucidating relations between fMRI, ECoG, and EEG through a common natural stimulus. *NeuroImage*, *179*, 79–91. doi:10.1016/j.neuroimage.2018.06.016

Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., & Bornkessel-Schlesewsky, I. (2008). The status of subject–object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, *59*, 54–96. doi:10.1016/j.jml.2008.02.003

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., … Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. doi:10.1016/j.neuron.2011.08.026

Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., & Theunissen, F. E. (2017). Encoding and decoding models in cognitive electrophysiology. *Frontiers in Systems Neuroscience*, *11*, 61. doi:10.3389/fnsys.2017.00061

Horton, C., D'Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology*, *109*(12), 3082–3093. doi:10.1152/jn.01026.2012

Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German sign language. *Neuropsychologia*, *51*, 2224–2237. doi:10.1016/j.neuropsychologia.2013.07.013

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. doi:10.1038/nature17637

Karns, C. M., Isbell, E., Giuliano, R. J., & Neville, H. J. (2015). Auditory attention in childhood and adolescence: An event-related potential study of spatial selective attention to one of two simultaneous stories. *Developmental Cognitive Neuroscience*, *13*, 53–67. doi:10.1016/j.dcn.2015.03.001

Kayser, S. J., Ince, R. A. A., Gross, J., & Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *Journal of Neuroscience*, *35*(44), 14691–14701. doi:10.1523/jneurosci.2243-15.2015

Keitel, A., Ince, R. A. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage*, *147*, 32–42. doi:10.1016/j.neuroimage.2016.11.062

Koskinen, M., & Seppä, M. (2014). Uncovering cortical MEG responses to listened audiobook stories. *NeuroImage*, *100*, 263–270. doi:10.1016/j.neuroimage.2014.06.018

Koskinen, M., Viinikanoja, J., Kurimo, M., Klami, A., Kaski, S., & Hari, R. (2012). Identifying fragments of natural speech from the listener's MEG signals. *Human Brain Mapping*, *34*(6), 1477–1489. doi:10.1002/hbm.22004

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Subjective impressions do not mirror online reading effort: Concurrent EEG-eyetracking evidence from the reading of books and digital media. *PLoS One*, *8*(2), e56178. doi:10.1371/journal.pone.0056178

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. doi:10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980a). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language*, *11*(2), 354–373. doi:10.1016/0093-934x(80)90133-9

Kutas, M., & Hillyard, S. A. (1980b). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. doi:10.1126/science.7350657

Lalor, E. C., & Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*, *31*(1), 189–193. doi:10.1111/j.1460-9568.2009.07055.x

la Tour, T. D., Tallot, L., Grabot, L., Doyère, V., van Wassenhove, V., Grenier, Y., & Gramfort, A. (2017). Non-linear auto-regressive models for cross-frequency coupling in neural time series. *PLoS Computational Biology*, *13*(12), e1005893. doi:10.1371/journal.pcbi.1005893

Lauteslager, T., O'Sullivan, J. A., Reilly, R. B., & Lalor, E. C. (2014). Decoding of attentional selection in a cocktail party environment from single-trial EEG is robust to task. *Proceedings of 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *2014*, 1318–1321. doi:10.1109/embc.2014.6943841

Lerner, Y., Honey, C. J., Katkov, M., & Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of Neurophysiology*, *111*(12), 2433–2444. doi:10.1152/jn.00497.2013

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *The Journal of Neuroscience*, *31*(8), 2906–2915. doi:10.1523/JNEUROSCI.3684-10.2011

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. doi:10.1016/j.cognition.2007.05.006

Liberto, G. M. D., & Lalor, E. C. (2016). Isolating neural indices of continuous speech processing at the phonetic level.

Advances in Experimental Medicine and Biology, *894*, 337–345. doi:10.1007/978-3-319-25474-6_35

Liberto, G. M. D., & Lalor, E. C. (2017). Indexing cortical entrainment to natural speech at the phonemic level: Methodological considerations for applied research. *Hearing Research*, *348*, 70–77. doi:10.1016/j.heares.2017.02.015

Lotze, N., Tune, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Meaningful physical changes mediate lexical-semantic integration: Top-down and form-based bottom-up information sources interact in the n400. *Neuropsychologia*, *49*, 3573–3582. doi:10.1016/j.neuropsychologia.2011.09.009

Maddox, R. K., & Lee, A. K. C. (2018). Auditory brainstem responses to continuous natural speech in human listeners. *eNeuro*, *5*(1). doi:10.1523/ENEURO.0441-17.2018

Maess, B., Schröger, E., & Widmann, A. (2016). High-pass filters and baseline correction in M/EEG analysis. Commentary on: "How inappropriate high-pass filters can produce artefacts and incorrect conclusions in ERP studies of language and cognition". *Journal of Neuroscience Methods*, *266*, 164–165.

Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology*, *7*, 120. doi:10.3389/fpsyg.2016.00120

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv, page 1301.3781v3.

Nager, W., Dethlefsen, C., & Münte, T. F. (2008). Attention to human speakers in a virtual auditory environment: Brain potential evidence. *Brain Research*, *1220*, 164–170. doi:10.1016/j.brainres.2008.02.058

Nieuwland, M. S., & Berkum, J. J. A. V. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111. doi:10.1162/jocn.2006.18.7.1098

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735. doi:10.1162/jocn.2006.18.7.1098

O'Halloran, J. P., Isenhart, R., Sandman, C. A., & Larkey, L. S. (1988). Brain responses to semantic anomaly in natural, continuous speech. *International Journal of Psychophysiology*, *6*(4), 243–254.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., … Lalor, E. C. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. doi:10.1093/cercor/bht355

Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*(12), 1649–1653. doi:10.1016/j.cub.2015.04.049

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife*, *5*, e14521. doi:10.7554/eLife.14521

Peelle, J. E., & Wingfield, A. (2016). The neural consequences of age-related hearing loss. *Trends in Neurosciences*, *39*(7), 486–497. doi:10.1016/j.tins.2016.05.001

Petersen, E. B., Wöstmann, M., Obleser, J., & Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *Journal of Neurophysiology*, *117*(1), 18–27. doi:10.1152/jn.00527.2016

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, *41*(1), 245–255. doi:10.1016/s0167-6393(02)00107-3

Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 103–118). Mahwah, NJ: Lawrence Erlbaum Associates.

Reiterer, S., Hemmelmann, C., Rappelsberger, P., & Berger, M. L. (2005). Characteristic functional networks in high- versus low-proficiency second language speakers detected also during native language processing: An explorative EEG coherence study in 6 frequency bands. *Cognitive Brain Research*, *25*(2), 566–578. doi:10.1016/j.cogbrainres.2005.08.010

Reiterer, S., Pereda, E., & Bhattacharya, J. (2011). On a possible relationship between linguistic expertise and EEG gamma band phase synchrony. *Frontiers in Psychology*, *2*, 334. doi:10.3389/fpsyg.2011.00334

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*, *28*(2), 161–169.e5. doi:10.1016/j.cub.2017.11.033

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of english verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models*. Cambridge, MA: Bradford Books/MIT Press.

Salmelin, R., & Baillet, S. (2009). Electromagnetic brain imaging. *Human Brain Mapping*, *30*(6), 1753–1757. doi:10.1002/hbm.20795

Sanders, L. D., Stevens, C., Coch, D., & Neville, H. J. (2006). Selective auditory attention in 3- to 5-year-old children: An event-related potential study. *Neuropsychologia*, *44*(11), 2126–2138. doi:10.1016/j.neuropsychologia.2005.10.007

Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language*, *162*, 42–45. doi:10.1016/j.bandl.2016.08.001

Shafer, V. L., Kessler, K. L., Schwartz, R. G., Morr, M. L., & Kurtzberg, D. (2005). Electrophysiological indices of brain activity to 'the' in discourse. *Brain and Language*, *93*(3), 277–297. doi:10.1016/j.bandl.2004.10.008

Small, S. L., & Nusbaum, H. C. (2004). On the neurobiological investigation of language understanding in context. *Brain and Language*, *89*, 300–311. doi:10.1016/S0093-934X(03)00344-4

Smith, N. J., & Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*(2), 157–168. doi:10.1111/psyp.12317

Smith, N. J., & Kutas, M. (2015b). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*(2), 169–181. doi:10.1111/psyp.12320

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2012). The role of stimulus cross-splicing in an event-related potentials study. misleading formant transitions hinder automatic phonological processing. *The Journal of the Acoustical Society of America*, *131*(4), 3120–3140. doi:10.1121/1.3688515

Stephens, G. J., Honey, C. J., & Hasson, U. (2013). A place for time: The spatiotemporal structure of neural dynamics during natural audition. *Journal of Neurophysiology*, *110*(9), 2019–2026. doi:10.1152/jn.00268.2013

Stevens, C., Sanders, L., & Neville, H. (2006). Neurophysiological evidence for selective auditory attention deficits in children with specific language impairment. *Brain Research*, *1111*(1), 143–152. doi:10.1016/j.brainres.2006.06.114

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *52*(8), 997–1009. doi:10.1111/psyp.12437

Tanner, D., Norton, J. J. S., Morgan-Short, K., & Luck, S. J. (2016). On high-pass filter artifacts (they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis. *Journal of Neuroscience Methods*, *266*, 166–170. doi:10.1016/j.jneumeth.2016.01.002

Tromp, J., Peeters, D., Meyer, A. S., & Hagoort, P. (2017). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behavior Research Methods*, *50*(2), 862–869. doi:10.3758/s13428-017-0911-9

van der Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience*, *16*(6), 1068–1084. doi:10.1162/0898929041502670

Van Rullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 365. doi:10.3389/fpsyg.2011.00365

van Schijndel, M., Murphy, B., & Schuler, W. (2015). *Evidence of syntactic working memory usage in MEG data*. Proceedings of the 6th workshop on Cognitive Modeling and Computational Linguistics, Denver, CO (pp. 79–88). Association for Computational Linguistics.

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *Journal of the Association for Research in Otolaryngology*, *19*(2), 181–191. doi:10.1007/s10162-018-0654-z

Westerlund, M., Kastner, I., Kaabi, M. A., & Pylkkänen, L. (2015). The LATL as locus of composition: MEG evidence from English and Arabic. *Brain and Language*, *141*, 124–134. doi:10.1016/j.bandl.2014.12.003

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage*, *47*, 360–366. doi:10.1016/j.neuroimage.2009.04.037

Willems, R. M. (Ed.) (2015a). *Cognitive neuroscience of natural language use*. Cambridge: Cambridge University Press.

Willems, R. M. (2015b). Cognitive neuroscience of natural language use: Introduction. In R. M. Willems (Ed.), *Cognitive neuroscience of natural language use* (pp. 1–7). Cambridge: Cambridge University Press.

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2015). Prediction during natural language comprehension. *Cerebral Cortex*, *26*(6), 2506–2516. doi:10.1093/cercor/bhv075

Wolff, S., Schlesewsky, M., Hirotani, M., & Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language*, *107*, 133–157. doi:10.1016/j.bandl.2008.06.003

Wöstmann, M., Fiedler, L., & Obleser, J. (2016a). Tracking the signal, cracking the code: Speech and speech comprehension in non-invasive human electrophysiology. *Language, Cognition and Neuroscience*, *32*(7), 855–869. doi:10.1080/23273798.2016.1262051

Wöstmann, M., Herrmann, B., Maess, B., & Obleser, J. (2016b). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proceedings of the National Academy of Sciences*, *113*(14), 3873–3878. doi:10.1073/pnas.1523357113

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. doi:10.1177/1745691617693393

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language*, *122*(3), 151–161. doi:10.1016/j.bandl.2011.12.010

Appendix. Tabular summary of reviewed literature

Table A1. Summary of relevant electrophysiology literature, as based on the search procedure described in the Section 1.1. Note this is not simply all literature cited or mentioned in this review (for that, see the bibliography), but rather an overview of the literature actually reviewed, including some articles mentioned in passing. The primary criterion speech from language in the table is whether the participants actually had to comprehend the meaning conveyed by speech to achieve the experiment's goals. This stands in contrast to a task related to the categorization of the speech signal as a complex perceptual stimulus, as in probe tasks where subjects press a button in response to a particular word. This distinction is course not completely clear for all studies.

Reference	Speech/language	M/EEG	Method	Stimulus
Alday et al. (2017)	language	EEG	fixed-window regression (ERP)	23-minute story
Biau et al. (2015)	speech	EEG	phase-locking value	17-minute video of politician's speech
Boudewyn and Carter (2018)	speech aiming for language	EEG	alpha power at probe and its relationship to behavioral indices	two stories of about 36 minutes each for a total of 72.8 minutes
Bourguignon et al. (2012)	speech	MEG	dynamic imaging of coherent sources (DICS)	5 minutes of text read by a live speaker
Brennan and Pylkkänen (2012)	language	MEG	sample-wise mixed-effects regression	Sleeping Beauty (82 sentences and 1404 words total), presented RSVP
Brodbeck et al. (2018)	language	MEG	linear kernel estimation of convolution resulting from stimulus (TRF)	Two one- minute samples of an audio book repeated three times (6 minutes total)
Broderick et al. (2018)	language	EEG	ridge-regression based estimation of the evoked potential over time (TRF)	20 trials of 180s long audiobook clips
Cohen and Parra (2016)	language	EEG	spectral power, intersubject correlation	cartoon clips of approximately 161 seconds
Dikker et al. (2017)	language	EEG	inter-subject coherence	classroom setting
Ding and Simon (2012)	speech	MEG	evoked potential as a function of time and stimulus spectrum (STRF)	1 minute audiobook excerpts
Ding and Simon (2013)	speech	MEG	evoked potential as a function of time (TRF)	50 second audiobook excerpts
Getzmann and Falkenstein (2011)	speech	EEG	fixed-window ANOVA (ERP)	Cocktail-party task with a stock ticker
Giordano et al. (2017)	speech	MEG	directed functional connectivity, mutual information	6 minute videos
Gross et al. (2013)	speech	MEG	mutual information	7-minute-long story
Hale et al. (2018)	language	EEG	Sample-wise by-participant regression	first chapter of *Alice's Adventures in Wonderland*
Haufe et al. (2018)	as analysed, speech	EEG		Subset dependent. 325s long movie clip, multiple 7 minute audio clips for EEG. Study aggregates across multiple fMRI, EcoG and EEG studies
Horton et al. (2013)	speech	EEG	cross correlation	cocktail party with sentence-length trials
Karns et al. (2015)	speech	EEG	ANOVA of peak latency to probes (single syllables or buzzes)	dichotic cocktail party with two stories, length unclear
Kayser et al. (2015)	speech	EEG	mutual information, ITC	6 minute speeches derived from TED talks
Keitel et al. (2017)	speech	MEG	mutual information	7 minute "real life" story
Koskinen et al. (2012)	speech	MEG	canonical correlation analysis	collection of short news articles totalling 58 minutes
Koskinen and Seppä (2014)	speech	MEG	canonical correlation analysis	repetitions of a 1-minute-long audiobook passage
Lalor and Foxe (2010)	speech	MEG	auditory evoked epread spectrum analysis (AESPA / TRF)	181s clips from an audio book, each two of the three test subjects listened to over 46 or 47 segments, while the final subject only listened to 16
Lauteslager, O'Sullivan, Reilly, and Lalor (2014)	speech	EEG	decoder model discriminating between the response for attended and unattended speech	60 second audiobook clips, presented dichotically in a cocktail-party paradigm
Di Liberto et al. (2015)	speech aiming towards language	EEG	TRF models augmented by phonemic labels	28 trials of 155s samples from an audiobook
Liberto and Lalor (2016)	speech aiming towards language	EEG	TRF models to the speech envelope, spectrogram, phonetics, and phonemes	short speech segments, each presented in a vocoded-original-vocode sequence
Liberto and Lalor (2017)	speech	EEG	evoked potential as a function of time (TRF)	2.5 minute audiobook clips
Maddox and Lee (2018)	speech	EEG	auditory brainstem response (ABR) to complex stimuli (TRF)	64-second clips of an audiobook
Nager et al. (2008)	speech	EEG	fixed-window ANOVA (ERP) to an overlaid probe syllable	a complex cocktail party task (3 simultaneous audiobooks presented in 8 3-minute chunks)
O'Halloran et al. (1988)	language	EEG	fixed-window regression (ERP)	7-8 word sentences without any volume normalization or other acoustic control
O'Sullivan et al. (2014)	speech	EEG	decoder model discriminating between the response for attended and unattended speech	60 seconds of an audiobook for each trial
Park et al. (2015)	speech	MEG	transfer entropy	7 minute "real life" story

(*Continued*)

Table A1. Continued.

Reference	Speech/language	M/EEG	Method	Stimulus
Park et al. (2016)	speech	MEG	dynamic imaging of coherent sources (DICS)	7–9 minute video clips of a professional male speaker
Reiterer et al. (2005)	speech? (ability to attend to speech in a foreign language)	EEG	coherence	2–3 minute video clips of television news
Reiterer et al. (2011)	speech? (ability to attend to speech in a foreign language)	EEG	coarse-graining of Markov chains, coherence, phase-lag index	2–3 minute video clips of television news
Sanders et al. (2006)	speech	EEG	fixed-window ANOVA (ERP) to a probe	children's stories of about 2.5 to 3.5 minutes in length
van Schijndel et al. (2015)	language	MEG	coherence between one anterior and one posterior sensor, focused on alpha band	80 minute audiobook of *Heart of Darkness*
Shafer et al. (2005)	Speech-to-language	EEG	fixed-window ANOVA (ERP) to a probe word, current source density (CSD)	children's story and a jabberwocky / nonsense syllable control, presented with pauses between sentences
Stevens et al. (2006)	speech	EEG	fixed-window ANOVA (ERP) to a probe word, current source density (CSD)	
Tromp et al. (2017)	language	EEG	fixed-window ANOVA (ERP)	single sentences embedded in virtual reality context
Vanthornhout et al. (2018)	speech	EEG	evoked potential as a function of time (TRF)	15 minute story, list of single sentences
Wöstmann et al. (2016b)	speech	MEG	intertrial phase coherence	cocktail party task of digit sequences