



Interdependence of “What” and “When” in the Brain

Sanne Ten Oever^{1,2,3}  and Andrea E. Martin^{1,2}

Abstract

■ From a brain’s-eye-view, when a stimulus occurs and what it is are interrelated aspects of interpreting the perceptual world. Yet in practice, the putative perceptual inferences about sensory content and timing are often dichotomized and not investigated as an integrated process. We here argue that neural temporal dynamics can influence what is perceived, and in turn, stimulus content can influence the time at which perception is achieved. This computational principle results from the highly interdependent relationship of *what* and *when* in the environment. Both brain processes and perceptual events display strong temporal

variability that is not always modeled; we argue that understanding—and, minimally, modeling—this temporal variability is key for theories of how the brain generates unified and consistent neural representations and that we ignore temporal variability in our analysis practice at the peril of both data interpretation and theory-building. Here, we review *what* and *when* interactions in the brain, demonstrate via simulations how temporal variability can result in misguided interpretations and conclusions, and outline how to integrate and synthesize *what* and *when* in theories and models of brain computation. ■

INTRODUCTION

“What” and “when” seem like simple concepts, but when we think about them pertaining to events, and the occurrence of those events in time and space, things quickly become complicated. Modern physics tells us that the relationship between time and space is much more intimate than previously thought, namely, that time appears to be a function of position in space in relation to mass (Rovelli, 2019). This example shows that intuitively independent concepts (time and space) can actually share a close interdependent relationship. Here, we illustrate that a similar interdependent relationship between *what* and *when* exists in the neural operations pertaining to the representations and computations of events in the perceptual world. Naturally, every stimulus input that enters the brain is processed with its own unique temporal dynamics (Köhler, 1967), and concomitantly, different stimulus features can be processed at different timescales (Moutoussis & Zeki, 1997b). Perceptual operations are highly influenced by top-down processes that are context dependent and have their own timescale (Frith & Dolan, 1997). What is more, temporal features can also determine the content of a stimulus (e.g., when determines what; Ten Oever & Sack, 2015; Rosen, 1992). One example is voice-onset-time (VOT), in which the identification of a syllable fully depends on the time of the onset of vocal vibrations (Lisker & Abramson,

1967). This means that besides the interdependence of *what* and *when* in neural computation, there is also an interdependence of *what* and *when* in the stimulus input that arrives in the brain. Nonetheless, to interpret the changing environment, a core function of our perceptual system must be to extract what information is happening when. Understanding how the brain does this has driven the fields of perceptual and systems neuroscience since their inception; it also has profound implications for cognition (Dennett, 1993; Köhler, 1967).

A leitmotif of the problem in miniature can be seen in the dichotomy between two methodological backbones in neuroscience: EEG and fMRI. Most classical fMRI studies investigate where specific content is processed in the brain, whereas most classical EEG studies have focused on when information is processed. Over the last 50 years, more complex paradigms and data analysis techniques, combined with theoretical advances, have pushed the field forward, but the separate treatment of *what* and *when* information seems to persist. For example, an important theoretical principle that has been around since the time of Heraclitus, Bayes, and Helmholtz has more recently evolved the field of perceptual neuroscience: It poses that perception not only depends on passive bottom-up interpretation of perceptual input but also on the current state of the brain in terms of “belief” about what information is going to happen (De Lange, Heilbron, & Kok, 2018; Noppeney & Lee, 2018; Bastos et al., 2012; Trommershauser, Kording, & Landy, 2011; Ernst & Bühlhoff, 2004; Rao & Ballard, 1999; Frith & Dolan, 1997; Helmholtz, 1878) at what time (Nobre & van Ede, 2018; Schroeder & Lakatos, 2009). Although *what* and *when* information

¹Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, ²Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands, ³Maastricht University, The Netherlands

are strongly represented in these top-down accounts of perception, most theoretical and empirical accounts still make a stark, implicit division between *what* and *when*, an orthogonalization: The brain either predicts what is going to happen or it predicts when something is going to happen, but few integrative accounts are provided (Ballard, 2015), never mind synthetic ones, where *what* and *when* interrelate both in neural representation and processing—where the intertwining of *what* and *when* becomes part and parcel of neural representation.

Top-down effects have a direct consequence for the perceptual inference process; as such, it has been robustly shown that knowledge about the content and timing of upcoming stimuli can bias what is perceived (Ernst & Bühlhoff, 2004), as well as the time at which something is perceived (Eggleman, 2008). The McGurk illusion is an example of a content bias (McGurk & MacDonald, 1976). In this illusion, the presentation of mouth movements corresponding to a /ga/ can bias the percept of an auditory /pa/ toward a merged audiovisual percept resembling a /da/. An example of a top-down temporal bias was shown in Stetson, Cui, Montague, and Eggleman (2006). Here, motor feedback was systematically delayed causing events that occur simultaneous with a motor response to be perceived as occurring before the response. The circumstances under which content or time perception is altered are still not fully understood. However, it is unlikely that top-down modulations influence the *what* and *when* perceptual inferences process independently, as predictive top-down processes in the brain are known to bias neural processing of specific neural representations as well as decreasing neural latencies (Trommershauser et al., 2011; Frith & Dolan, 1997). It is still unclear whether, and how, predictive processing speeds up or biases neural processing. Instead, neuronal responses that result in neural bias or decreased latencies are often interpreted as evidence for top-down predictive models of perception, rather than being explicitly included in the models either as principles, mechanisms, or representational states.

We argue that it is vital to our theories and models to stop separating *what* and *when*, and to provide an integrated, synthetic account of *what* and *when* in perception, accounting both for interdependencies in neural computation and for interdependencies in environmental input. *What* and *when* are not independent in the brain, and neither are the statistical regularities pertaining to *what* and *when* in the environment. What is more, assuming their independence in our analyses strategies can result in misinterpretation of neural data and, thereby, inaccurately serve as evidence in favor of a given theory or mechanism in the brain. As such, modeling the indivisible nature of *what* and *when* would strongly improve our theories of perception and cognition. Here, we review how *what* and *when* interact, we highlight how one can easily inappropriately interpret data when *what/when* interactions are not explicitly taken into account, and what steps

can be taken to move forward with the indivisible nature of *what* and *when*.

TEMPORAL VARIABILITY IN THE BRAIN IS OFTEN OVERLOOKED

It is indisputable that the brain takes time to process information and that different brain areas have different latencies at which input can arrive (Keitel & Gross, 2016; Purushothaman, Patel, Bedell, & Ogmen, 1998). This is a consequence not only of the synaptic delays between consecutive processing areas (Bi & Poo, 1999) but also of the fact that every processing stage likely requires time to finish its own computation (viz., which might include integration, lateral inhibition). As sequential processing stages relate to gradually increasing cognitive complexity, the field has classically interpreted earlier ERP components as reflecting early sensory processing stages, whereas later components are interpreted as being modulated by higher-order cognitive factors, such as attention, decision-making, and “cognitive processing” (Luck, 2014; Luck & Hillyard, 1990). The temporal dynamics of these components also follow a systematic pattern: The later the processing stage, the lower the frequency of the averaged ERP components. The earliest components such as brain stem responses can be up to 50 Hz (Starr & Achor, 1975; Jewett, Romano, & Williston, 1970), whereas later attention-modulated components can go down to 1 Hz (Spencer, Dien, & Donchin, 2001), or even lower, in the case of the contingent-negative variation component, whose temporal profile depends on the time interval between a warning and a go signal (Walter, Cooper, Aldridge, McCallum, & Winter, 1964).

This pattern of interpretation seems to suggest that later stages of brain computation take more time, or have a longer duration, during the processing of a stimulus. However, this is not the whole story. An ERP frequency reduction does not necessarily have to be a consequence of needing a longer duration to perform a cognitive process. Such a delay in response could also be caused by the summation of temporal variability that increases as sequential processing steps increase (Figure 1A). If the latter, the brain processes a stimulus at different latencies depending on the trial (viz., the brain time does not run in parallel with stimulus time). Knowing which of the two situations are at play is crucial to understand the underlying temporal brain dynamics. Averaging cannot tell us the difference between the two states of affairs. One option to disentangle these situations is to look at ERP components that are identifiable on a single trial, such as the P300, which is apparent based on its large amplitude. Looking at the frequency response of the single trials within a single participant, frequencies between 2.7 and 4.1 Hz (at 1 *SD* from the mean) at an average of 3.5 Hz can be observed for the P300 component. This average drops down to 2.9 Hz when investigating the average of the trials (Figure 1B and C). The difference in frequency

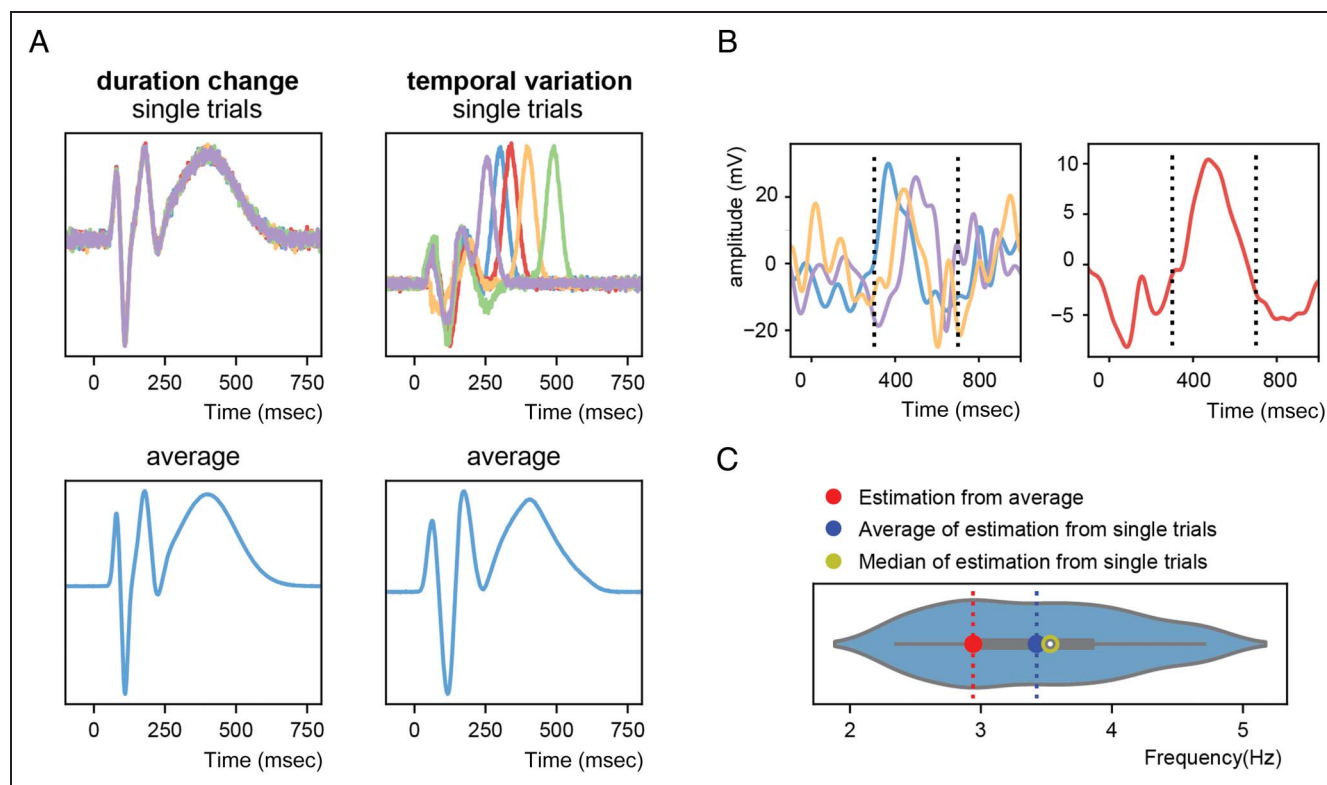


Figure 1. Slower components as a consequence of duration change or temporal variation. (A) ERPs can look the same when effects either are a consequence of duration changes or temporal variation. (B) Filtered (1–10 Hz) examples of single trial P300 (left) and the average P300 (right). Dotted lines are for comparisons between single trials and average. (C) Peak frequency estimation of single trials and of the average (in red) for one participant.

content is consistent across participants and shows that the exact time at which the P300 component varies. Thus, this example illustrates how temporal variability can cause a reduction in the peak frequency of the signal, which, without the acknowledgment of the variability, might lead to an unsound conclusion, for example, stating that neural processing is occurring at this lower frequency. One might even say that averaging in this circumstance reduces the effective temporal resolution of EEG. Thus, it seems that classical ERP methods cannot tell us whether an observed frequency is the product of a neural process with a longer duration or of temporal variability in the execution of a process (where that variability itself could arise from multiple possible sources). Nor can they tell us whether the brain computes at a slower rate in lockstep with a long duration or a low-frequency stimulus, or if what we are observing is simply a consequence of brain computation not having a linear relationship with stimulus time (see Luck, 2014).

How can we quantify how much temporal variability is present in the neural signal? One classical way to avoid this problem is simply not to average, but instead extract the frequency content of the individual trials and perform the analyses on the spectral content of the individual trials (Tallon-Baudry & Bertrand, 1999; Makeig, 1993). Although this ensures that no new frequency content is introduced through averaging, temporal variability can still contaminate the signal in different ways. First, in continuous data

where we cannot clearly separate trials, repeated high-frequency evoked responses may still appear in lower frequencies (see the simulations in Variability in Neural Dynamics and Analytical Interpretation section). Second, temporal variability on single-trial spectral estimates can still lead to temporal smearing of the frequency content across the estimated window. For example, Jones (2016) has shown that an activity in the beta frequency band lasting around 1 sec in the average power spectra does not necessarily reflect beta oscillations that are active for 1 sec. Instead, strong discrete beta events occur at different moments in different trials, which leads to a beta-band effect. Averaging across the power spectra of single trials smears this temporal variance across time, suggesting that an event previously described as a sustained rhythmic event can better be described as discrete events that have strong temporal trial-by-trial variability (reviewed in Tal, Neymotin, Bickel, Lakatos, & Schroeder, 2020). The same could hold for other reported low-frequency effects attributed to true oscillators, or to stimulus tracking such as the phrase-tracking by delta oscillations (Coopmans, De Hoop, Hagoort, & Martin, 2022; Rimmele, Poeppel, & Ghitza, 2021; Ten Oever & Martin, 2021; Kaufeld, Bosker, et al., 2020; Meyer, 2018; Ding, Melloni, Zhang, Tian, & Poeppel, 2016). Again, using single trial spectral methods cannot diagnose whether a brain event is occurring at a specific moment relative to a stimulus event. As such, one needs to exclude alternative explanations such as

trial-by-trial temporal variability changes in brain responses; these changes could be random, or systematic, and, in either case, would induce low-frequency or temporally smeared effects as described here. Alternatively, condition manipulations might systematically change temporal variability at particular moments in time (see the simulations in Variability in Neural Dynamics and Analytical Interpretation section). This situation could also lead to low-frequency effects being misattributed to other factors.

Temporal variability in brain responses can thus lead to a lowering of apparent frequency content, as well as temporal smearing of a short-duration spectral event. Importantly however, we do not have to regard temporal variability as unfortunate neural noise, or the byproduct of neural computation, as not all temporal variability is random. For example, neural response latencies to stimuli reduce as a function of stimulus intensity, such that the higher the intensity, the faster or earlier the response latency (Brisson, Robitaille, & Jolicœur, 2007). A rich literature of RT studies (both in the *what* and *when* domain) have also shown that expected input is processed faster and more efficiently (Parasuraman, 1980) and that, overall, decisions are made faster when input is expected (Sternberg, 1966). This indicates that somewhere along the processing pipeline, processing was sped up or occurred earlier as a function of expectation (Duncan-Johnson & Donchin, 1980; Parasuraman, 1980; Kutas, McCarthy, & Donchin, 1977); in other words, *when* something was processed interacted with *what* was processed and perceived. We know that these top-down effects are already present at early sensory states (Kok, Jehee, & de Lange, 2012), although latency shifts at early sensory components because of attention or expectancy are not typically found in ERP research. It is, however, likely that although the initial bottom-up sensory activation does not result in a latency shift, top-down processing likely affects the efficiency of the computation and, thereby, its effective completion time (Bastos et al., 2012). Indeed, decoding of expected phonemes is possible at an earlier latency than the processing of unexpected phonemes (Gwilliams, King, Marantz, & Poeppel, 2020). This pattern fits with predictive coding, top-down, and Bayesian frameworks, which would also predict optimized, and thereby faster, processing for stimuli that are expected.

Although we know that neural latencies can depend on the exact content that is being processed, the temporal variability of brain processes is rarely included as a factor in models of brain processing at the population level. For example, although we theorize that processing is dynamically optimized as a function of predictability, we typically do not model that processing is finished earlier or occurred at a faster speed (i.e., two different states of affairs). We suggest that deeper investigation into how *what/when* interactions can influence neural temporal variability is warranted, both analytically and theoretically. These interactions do not only pertain to latency shifts dependent on

the content of the input as described above (such that “brain time” thus depends on environmental content), but also relate to perceptual inferences that change as a function of the presentation time (where “brain content” thus depends on environmental time). In addition, such *what/when* interactions are also relevant when perceptual inference relies on learned associations, for example, between the exact moments in time that an event occurs, and the particular content of the event. In this case, the brain learns a mapping or a systematic statistical relationship between environmental content and timing. Acknowledging that perception in the brain is often based on all these types of *what/when* interactions, which in turn result in temporal variability in brain computation, will also dramatically impact our developing theories of cognition.

WHAT/WHEN INTERDEPENDENCIES IN THE ENVIRONMENT AND THE BRAIN

What/When Interdependencies in the Environment

Classically, perceptual paradigms rely on stochasticity in stimulus presentation time and concomitantly extracting brain dynamics after stimulus onset. Although well-controlled, it is far from how we are naturally presented with stimuli in the environment. In fact, in the environment, there is a strong temporal dynamic in which different input sequentially arises at effectors, receptors, and sensory organs at semipredictable times. As such, most would agree with a characterization of perception as the monitoring of an ever-changing environment (White, 2018). Consequently, current brain states influence ongoing and upcoming perceptual processes because they are manifestations of carrying over information from previous moments of perception into the next timestep. Indeed, brain state influences perceptual processing—for example, it has been shown that the brain state at the time of stimulus onset influences the detection (Ten Oever, Van Atteveldt, & Sack, 2015; Henry & Obleser, 2012; Mathewson, Gratton, Fabiani, Beck, & Ro, 2009) and categorization (Ten Oever, Meierdierks, Duecker, De Graaf, & Sack, 2020; Thézé, Giraud, & Mégevand, 2020; Kayser, McNair, & Kayser, 2016; Ten Oever & Sack, 2015) of presented stimuli, suggesting that the brain integrates the current presented information with the processing occurring at that moment. In a natural setting where information is a continuous stream, the computations of the brain relating to a previously presented stimulus influence the processing of a subsequent stimulus (or in a true continuous sense as a dynamical temporal interaction; Summerfield & de Lange, 2014; Friston, 2011; Rao & Ballard, 1999; Frith & Dolan, 1997). Likely, initial information alters brain processing to optimize the brain’s computation streamline of a subsequently presented stimulus (Summerfield & de Lange, 2014; Markman & Otto, 2011; Schroeder & Lakatos, 2009;

Niemi & Näätänen, 1981). The brain can do this as the natural environment contains structural regularities (statistical or otherwise) that can be used to predict what information is coming next.

Whenever a statistical regularity occurs in the environment, the brain could in principle exploit it. For example, if you drop a ball from some height, it will fall on the floor at a predictable time and make a predictable sound. When one knows from previous experience what the most likely future time and content of an event are, predictions about sensory events can be made. It is well-established that the brain creates these predictions to prepare for new perceptual input (Summerfield & de Lange, 2014; Frith & Dolan, 1997). This is a well-adapted mechanism that optimizes the processing of the right information at the right time. Studies investigating predictions in the *what* and *when* domain typically vary the temporal predictability as well as the content predictability and investigate their relative neural instantiation (Figure 2A; Auksztulewicz et al., 2018; Morillon, Schroeder, Wyart, & Arnal, 2016). As such, the field has gained much understanding in the descriptions of neural activity during predictive processing (Bastos et al., 2012; Friston, 2005). However, what is rarely done, either in presentation, analysis, modeling, or theorizing, is to make the content prediction contingent upon the event time (Los et al., 2021; Ten Oever & Sack, 2019; van de Ven, Kochs, Smulders, & De Weerd, 2017). By ignoring this relation, one implicitly assumes that *what* and *when* are independent in the brain: One can predict the content, one can predict the time, but there is no integrated content–time prediction, or reliance of one prediction upon the other (Figure 2A). We are convinced this is inaccurate because content and time predictions are often integrated based on experience with the environment.

For example, imagine watching the finish of a professional cycling race on a crowded street. You have a good view of the finish line, but many people are occluding your view of the main street. Suddenly, you get a glimpse of two cyclists sprinting to the finish, but your view of the last several meters is blocked by other spectators. This does not

bother you because you can just redirect your view to the finish line, where the cyclists will be in a matter of seconds. Five seconds pass, but still no cyclists appear in your view of the finish line. Now, your expectations have been violated. You return your gaze to the street and just realize that both cyclists had turned a corner to another street, taking an alternate route to the finish line that took them around a corner, hidden from view (Figure 2B).

This example demonstrates that based on an earlier perception, we can form strong expectations about when the cyclists will arrive at the finish line; these predictions can be formed in a matter of hundreds of milliseconds. Once this time has passed, your expectation shifts to another visual location to see if they are borne out. In other words, the location of your attention depended on the time passed (i.e., when influences what you expect). In this particular example, the expectation was so strong that some might call it a violation of an expectation (Squires, Squires, & Hillyard, 1975) rather than different types of expectations at different moments in time. However, it clearly demonstrates that time can be an indicator for the expected content in the environment; we will further discuss instances where expectation of different perceptual contents across time can be equal, leading to interesting behavioral patterns.

Speech acoustics strongly demonstrate how our perceptual system is sensitive to statistics about *what/when* relationships. Many phonetic classes are fully determined by the temporal course of the spectral information in the acoustics (Kaufeld, Ravenschlag, Meyer, Martin, & Bosker, 2020; Martin, Monahan, & Samuel, 2017; Ghitza, 2013; Salomon, Espy-Wilson, & Deshmukh, 2004). One example is VOT (Lisker & Abramson, 1967). The exact time at which the vocal cords start vibrating can determine the difference between perceived consonants (e.g., [p] is unvoiced and [b] is voiced). Simply by manipulating the time at which voicing starts, one can influence the phones and syllable that you hear—so *when* literally becomes *what*. Similarly, audiovisual temporal statistics in speech can also determine and influence phoneme, syllable,

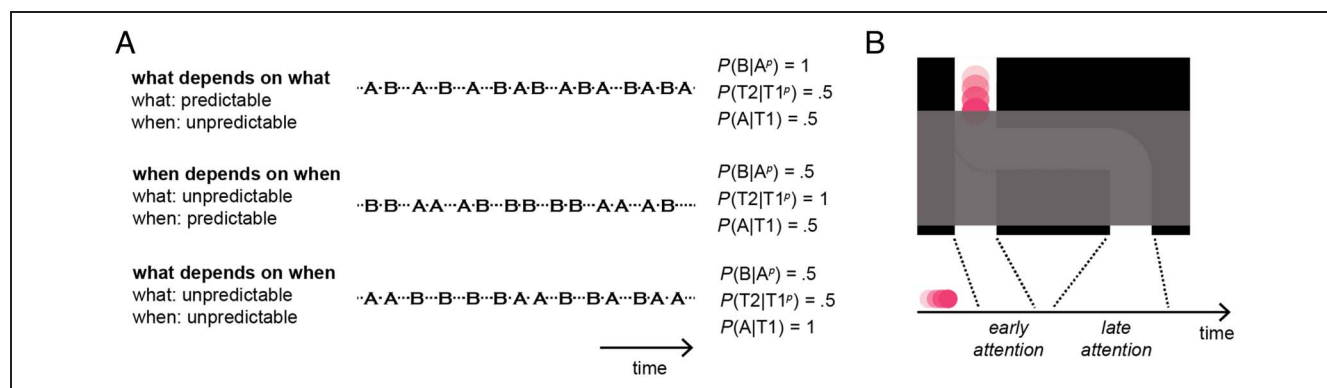


Figure 2. What and when are likely dependent on each other. (A) Most studies independently vary the what and when predictions, but rarely are what and when predictions made dependent on each other. (B) In natural situations, the time of input has a consequence on your content expectations. When we drop a ball with an occluded path, early expectations will be guided to the left potential path, but later expectations will shift to the right potential path.

and word identification (Ten Oever, Sack, Wheat, Bien, & van Atteveldt, 2013; Van Alphen & McQueen, 2006). Finally, speeding up speech can also influence the perception of subsequent words (Kaufeld, Naumann, Meyer, Bosker, & Martin, 2020; Bosker, 2017a, 2017b). Understanding speech is dependent on the knowledge of how content–time patterns link to abstract phonetic, phonological, prosodic, syllabic, morphemic, and lexical representations; in fact, the *when/what* synergy is so fundamental to speech perception and language comprehension that we are often not aware of it at all. Similarly, in music, strong *what/when* relationships can also be seen. The brain can predict which note is coming when, and it has been shown that in many circumstances, there is an integrative *what/when* representation in music, which can be modeled accordingly (Pearce, 2018; Boltz, 1999).

In the visual domain, various memory studies have shown that when you systematically present a specific item at a specific time, participants will associate that time with that specific item and perform better when the item is presented in the correct temporal context (Cravo, Rohenkohl, Santos, & Nobre, 2017; van de Ven et al., 2017). This indicates that the temporal context (not only the sequence of events, but the precise time) is included in memory representations of events (Ten Oever et al., 2016). These types of effects have been corroborated in a recent association study where different temporal probability distributions were associated with different perceptual cues (Los et al., 2021). Although most of these experiments have studied suprasecond intervals, we have recently shown that also with subsecond intervals these associations can be made (Ten Oever & Sack, 2019). Abstract memory representations of events appear to contain integrated content and temporal information; thus, content and time cannot be thought of independently from the brain's perspective.

The previous sections illustrated that temporal information is not merely an expected or unexpected variable, but rather can serve as a cue to content, which in turn demonstrates how abstract neural representations can be determined by temporal dynamics. For example, VOT cannot be classified as an expected or unexpected timing as it determines—and is therefore part of the representation of—phones and other linguistic categories. Of course, in many other cases, timing alone does not fully determine the percept (as with VOT), but, rather, timing is a probabilistic cue for an item's content (Martin, 2016; Ernst & Banks, 2002). As with any probabilistic cue, the brain will show some robustness to deviations from the expected temporal dynamics (Pefkou, Arnal, Fontolan, & Giraud, 2017; Ghitza, 2014). For example, in memory paradigms, participants can still perform a memory task in the face of unexpected or deviant temporal dynamics, but perform worse if an item is not presented within the correct temporal context. Whether deterministic or probabilistic, it is likely that temporal context is part of the neural representation of an external event, and thus, more research is

needed to understand how perceptual, memory, and cognitive systems deal with these associations.

What/When Interdependencies in the Brain

Because stimulus and event time can determine the interpretation of that stimulus or event, it follows that the particular time at which the brain is computing becomes relevant in theories and models of perception and cognition. The brain does not passively wait until input arrives; as such, the timing of neural activation is not merely the time that the brain detects that an event is happening. Instead, the temporal dynamics of brain activity is important for the interpretation of external events. There is a large literature on temporal properties that modulate the detection of an event—for example, inattentive blindness shows temporal windows during which the brain poorly detects visual stimuli (Simons, 2000). However, here, we also highlight situations in which time influences the qualitative interpretation of the perceptual event (i.e., interpreting an event as Event A or Event B). Observation of neural dynamics indicates that brain time likely has an influence on the interpretation of an event (Mehta, Lee, & Wilson, 2002; Panzeri, Petersen, Schultz, Lebedev, & Diamond, 2001). We here define brain time as the time at which a particular brain area processes an external event, either to be forwarded and processed by another area or to be used for perceptual inference.

When stimulus input arrives, the initial sweep of information might be very coarse, and only neuronal populations that are receptive to the input will activate (Petras, Ten Oever, Jacobs, & Goffaux, 2019; Kay & Yeatman, 2017). It has been proposed that the first active populations might contain the most relevant information (VanRullen, Guyonneau, & Thorpe, 2005; Mehta et al., 2002). In other words, the time at which populations are active is relevant to stimulus interpretation, such that spike timing has been proposed as a way to code stimulus-identity or stimulus-related feature information (Downer, Bigelow, Runfeldt, & Malone, 2021; VanRullen et al., 2005; Panzeri et al., 2001). Even without external events, this temporal code might serve to relay different types of stimulus information in parallel (Lisman & Jensen, 2013). The order of spike times is a consequence of activity patterns aligning to ongoing oscillations that modulate overall excitability of a population (Lakatos et al., 2005; Buzsáki & Draguhn, 2004). Neurons with higher base excitability will be active earlier (at a less excitable phase) relative to other neurons, which creates a phase map of excitability states (Ten Oever & Martin, 2021; Jensen, Gips, Bergmann, & Bonnefond, 2014; Mehta et al., 2002). Saliency dependence (and thereby base excitability variation) in the exact phase of optimal processing has recently been shown in a behavioral and EEG study (Jia, Fan, & Luo, 2022). In addition, different connectivity between neuronal ensembles might engage in spike-time-dependent plasticity, such that connections are solely strengthened when postsynaptic

activity follows at a specific temporal delay from presynaptic activity (Feldman, 2000; Markram, Lübke, Frotscher, & Sakmann, 1997). These types of time- or phase-coded spiking operations could also provide a means for hierarchical structure building (Martin & Dumas, 2017, 2020) and seem vital for efficient communication between distant brain areas (Singer, 2009; Fries, 2005; Von der Malsburg, 1995). These putative neural encoding schemes show that brain time, or the temporal dynamics of population activity, could be an important factor for both theories of information encoding in neural systems and for extracting information from neural readouts during data analysis.

How can the brain leverage temporal dynamics across regions to optimize its cognitive and behavioral outcomes, while concurrently being so biased by timing in its interpretation of its surroundings? Such a computational situation seems to require a control structure for temporal dynamics, such as a stand-alone timer, clock, or counter, which in turn seems highly implausible, both from a formal and neurobiological standpoint. As such, it is difficult to imagine how temporal variation could change and align computation to be sensitive to time–content associations in the brain without some form of dynamic self-organization (see next section).

Mapping *what/when* Dynamics Inside and Outside the Brain

The putative goal of the perceptual system is to determine when what is happening in the environment in the service of cognition and behavior. If, as argued above, the time-of-neural-response is dependent on the content itself, then the brain faces yet another challenging puzzle: to relate its own temporal dynamics of internal computation to ongoing stimulus time. This transduction of stimulus time to brain time is not infallible as evidenced by various temporal illusions. For example, if a moving dot pattern changes both in direction and in color of the dots at the same time, participants will first perceive the color change and only later perceive the movement change (the difference can be as large as 80 msec; Johnston, Arnold, & Nishida, 2006; Moutoussis & Zeki, 1997a). Color and movements are processed in different brain regions and are processed with different temporal dynamics. In the illusion, the speed and latency of color and movement processing differ, which lead to an incorrect perceptual inference of temporal occurrence (Moutoussis & Zeki, 1997b). In the auditory domain, it has been shown that when participants are asked to mark the time at which a noise, such as a click, occurs in ongoing speech, they systematically misalign it to an earlier point in the sentence (Ladefoged & Broadbent, 1960) and typically have a bias toward placing the noise at syntactic boundaries (Fodor & Bever, 1965). In the memory system, temporal order effects also make it more difficult to judge temporal order across encode events compared with within them (Ezzyat & Davachi, 2014; Alvarez, 2011). These and other temporal

illusions demonstrate that the brain's temporal inferences do not reflect the veridical stimulus time of perceptual input (for a review, see Eagleman, 2008).

The brain's inferences about timing also rely on calibration of its estimations in the temporal domain (Vroomen & Keetels, 2010). As the brain copes with different sources of temporal uncertainty, it must also adapt its estimation of event timing based on statistical evidence. If one systematically flashes audiovisual stimuli at an offset, participants will perceive the apparent simultaneity as being shifted toward the offset (Fujisaki, Shimojo, Kashino, & Nishida, 2004). In a similar audiovisual flash paradigm, it has been shown that the neural response to the rate of the flashed audiovisual events shifts its phase together with the perceptual shift in perceived simultaneity, suggesting that neural timing shifts together with perceived timing (Kösem, Gramfort, & van Wassenhove, 2014). Temporal calibration effects have also been reported in the tactile domain (Keetels & Vroomen, 2008) and have been shown to follow Bayesian inference rules (Miyazaki, Yamamoto, Uchida, & Kitazawa, 2006).

Here, we have focused on temporal estimation in perception, but ultimately, timing in perception is crucial to perform coordinated action at the right moment (De Kock, Gladhill, Ali, Joiner, & Wiener, 2021; Georgopoulos, 2002), in cognition, and behavior more broadly. In speech production, coordinated temporally accurate articulation directly determines the content of what is uttered. It is thus no wonder that a widespread motor network has been implicated as the brain basis of temporal perception (Kotz, 2011; Ivry & Spencer, 2004; Harrington, Haaland, & Hermanowitz, 1998). However, it is also evident that not all temporal illusions arise from the motor system (e.g., as the dynamics of temporal estimation of color and motion differ in the early perceptual system). Thus, it is likely that the brain must rely on multiple scales and hierarchies across systems to compute temporal information (Mauk & Buonomano, 2004; Mangels, Ivry, & Shimizu, 1998). Coordinated timed action likely requires learning, which allows for adjustments in temporal estimation, which would be based on the temporal variability that was observed or imposed during perception.

As we have posited thus far, temporal variability in perception likely increases as more computational stages have to be passed (*viz.*, each stage adds some variability). Typically, higher-order regions also must cope with the integration of information either across different locations or different temporal timescales to create abstract representations. It is likely that brain regions and networks have developed temporal invariance or robustness to cope with wider or more variable temporal scales of input. Especially if the time of presentation is part of an abstract (memory) representation, it is likely that this representation can be separated from the particular time of computation during an instance. One way to separate the time of neural activity from the temporal dynamics in the environment is to keep perceptual representations active for a longer duration by

using sustained responses (Courtney, Ungerleider, Keil, & Haxby, 1997). Another way is to use lower frequency oscillations. As these oscillations extend activation for a longer period, and provide longer windows of excitability, the exact moment at which information arrives can be more variable (Herbst & Obleser, 2019; Ten Oever et al., 2015). In addition, low-frequency oscillations could serve to represent information that is integrated across a wider time-scale (Meyer, Sun, & Martin, 2020). This has been proposed in the domain of language processing where objects like phrases and sentences must be represented by (a network of) neural assemblies. Such abstract mental representations are likely to be the product of perceptual inference, as they are not directly perceivable from sensory input alone, and required previous experience (viz., knowledge of a given language) to be perceived (Martin, 2016, 2020). As such, neural dynamics relating to phrase and sentence representation can be separated from the exact time of the computation (Martin & Doumas, 2017). However, as temporal variability likely also increases, the more information is abstracted away from sensation during perception and cognition; it is unclear if low-frequency oscillations occur because of integration (Martin & Doumas, 2017; Poeppel, 2003), because of a mixture of computations (Martin, 2020), or because of temporal variability alone (see Ten Oever, Kaushik, & Martin, 2022, and the simulations in the Variability in Neural Dynamics and Analytical Interpretation section).

Even if the brain indeed predicts what is happening when and sensitizes processing at specific moments for specific content, it remains a puzzle how these expectations are operationalized considering that the brain has to overcome the temporal variability in neural processing as well as deal with known interdependence of *what* and *when* in the environment. Sensitization to predictable stimuli in time and content has been shown to occur close to stimulus onset (e.g., 40 msec before the onset; Kok, Mostert, & De Lange, 2017). Furthermore, sensitization can occur much earlier for stimuli that are predictable in content, but unpredictable in time (Mohanta et al., 2021). However, it is still unclear how the brain creates this temporal prediction, what its content or format is, and how it is deployed. Moreover, it is also unclear whether the brain provides feedback to different neural representations at different moments in time and how it would do so as a system, especially in cases where the content of the stimulus depends on time that has passed or on duration.

Oscillations Make *what/when* Interdependencies More Complex

The brain has strong intrinsic oscillatory dynamics as a function of, and which modulate, the ongoing excitability of neuronal populations (Buzsáki, 2004). The temporal dynamics of oscillations modulate and interact with the impulse responses generated as a consequence of sensation, perception, and cognitive computations (Lakatos,

Karmos, Mehta, Ulbert, & Schroeder, 2008). Oscillations have been argued to fulfill important computational roles in the brain such as the parsing of stimulus input (Giraud & Poeppel, 2012; VanRullen & Koch, 2003), the optimization of communication within (Schroeder & Lakatos, 2009) and across distant brain areas (Fries, 2005), and also the organization of seemingly chaotic responses of individual neurons (Lisman & Jensen, 2013). Adding oscillatory features to a neural system has been argued to increase the representational space to store and differentiate better neural patterns both empirically and in computational models (Martin, 2020; Schaefer, Angelo, Spors, & Margrie, 2006). Whereas on the one hand, oscillations streamline neural processing, they also add another level of complexity to investigate neural time courses.

Just as top-down predictions about what is happening next can sensitize specific neural representations (Kok et al., 2012), oscillations can modulate the excitability of whole regions (Buzsáki & Draguhn, 2004). Therefore, overall activation will occur earlier if oscillations happen to be at an excitable phase; oscillations thus act as a temporal filter (Lakatos et al., 2013). Presenting brief stimuli at low excitability phases of an oscillatory cycle can reduce the chance of detecting the stimulus (Mathewson et al., 2009). On the other hand, when both top-down predictions and oscillations interact, activity can occur during low-excitability phases because the gain from the top-down prediction sensitizes processing (Mehta et al., 2002), leading to a phase gradient of excitability (Ten Oever & Martin, 2021; Jensen, Bonnefond, & VanRullen, 2012; Mehta et al., 2002).

Inputs that arrive during high excitability phases have the highest chance of reaching threshold; it has been proposed that the brain as a system proactively aligns high excitability phases to stimulus input when it is able to predict when a stimulus will occur (Rimmele, Morillon, Poeppel, & Arnal, 2018; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008). However, if the content of subsequently presented stimuli varies, and especially the predictability of the content, it is not evident how the brain could accommodate such an alignment. This situation occurs, for example, in speech, where words following each other have varying content predictability levels. Another example is a dance performance where often particular movements are more predictable than others. In both these examples, the input pertains temporal predictability and we can illustrate how alignment to this temporal input structure becomes difficult when content predictability varies. In Figure 3A, a rhythmic input stream is presented. However, different neural regions might be sensitive to different content in the input stream because of top-down predictive processes (or alternatively they are intrinsically more sensitive). In this example, content that is predicted by a specific region is followed by content that is not predicted. If ongoing neural oscillations align with the expected time of activation based on the predicted content, the phase of the oscillation will not be at an optimal point for the input of

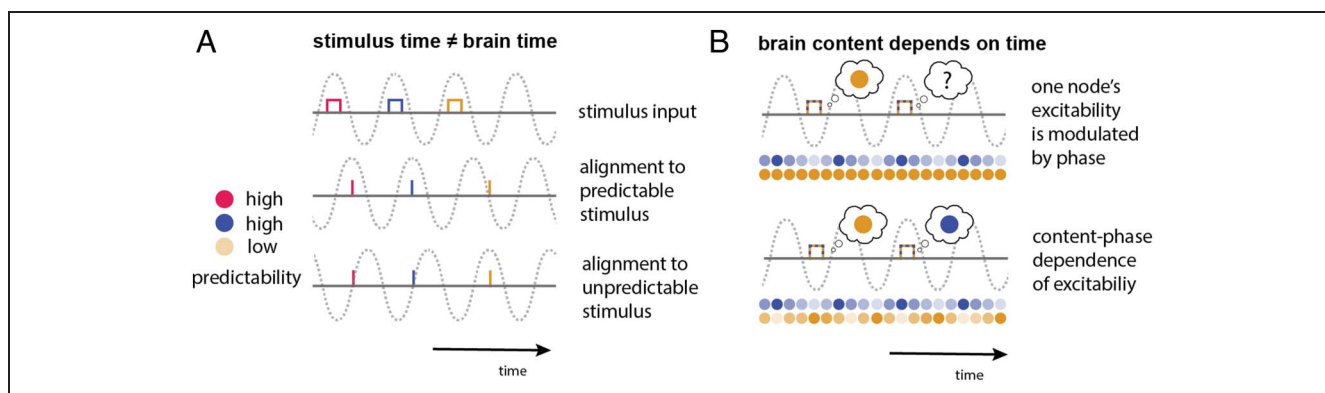


Figure 3. Oscillatory interactions with *when* and *what* predictions. (A) Dependent on the excitability of a neural representation, the neural latency of activation will vary. It is unclear whether neural oscillations will align to predictable input with a high excitable neural representation that activates representations early (top alignment), will align to unpredictable input with a low excitable neural representation that activates representations later (bottom alignment), or will not align at all (not visualized). (B) Activation of a representation might be dependent on the time of activation if excitability is not evenly modulated among neural representations. Here, we only represent the first time a cortical representation is activated.

the unpredicted content, because input to which the brain is already less sensitized will arrive during a lower excitability period, essentially resulting in a net higher threshold for excitability for already lower-sensitized input (top alignment in Figure 3A). This state of affairs does not seem efficient because it would polarize or exaggerate sensitivity to predicted and unpredicted content, possibly to the degree that predicted content is perceived even when it does not occur (viz., resulting in extreme and highly frequent perceptual and cognitive illusions) and unpredicted content is summarily missed or undetectable. This situation echoes neither human behavior nor optimality in neural computation. A putatively better strategy might be to align oscillations to the input to which the brain is less sensitized (bottom alignment in Figure 3A; Ten Oever & Martin, 2021). Alternatively, it is an option to not rely on low-frequency oscillations and revert to a high-frequency domain in which there are many frequently occurring high-excitable phases making it unlikely that important information is missed (Schroeder & Lakatos, 2009; Fries, Nikolic, & Singer, 2007). Note, however, that low-frequency oscillations have often been observed during speech processing, a situation in which both temporal and content predictability variations occur (Giraud & Poeppel, 2012). To sum up, although we know that the brain proactively changes its low-frequency oscillatory phase to optimize processing (Samaha, Bauer, Cimaroli, & Postle, 2015; Lakatos et al., 2008, 2013), it remains unclear how this mechanism is interactively influenced by both the temporal and content predictability of the input.

As outlined in previous sections, the time at which stimulus input arrives can influence what is perceived; this principle is also true for phase, and thus, the phase at which a stimulus arrives now also must be considered. For example, when an ambiguous or noisy input is presented, if a neural representation has a higher base excitability (i.e., the brain is sensitized to a stimulus input), an ambiguous stimulus input at a lower excitability point might only excite that single neural representation. However, during a higher excitability phase, more neural

representation might be available to reach activation, and differentiation between stimulus types would deteriorate (Schaefer et al., 2006). It is even conceivable that different neural representations have different excitabilities depending on oscillatory phase (Figure 3B). Because of the excitability level of an underlying neural representation at a specific phase, one representation or another representation might be activated (see e.g., Ten Oever & Sack, 2015; Lee, Simpson, Logothetis, & Rainer, 2005). In this way, phase of activation is related to the content of the stimulus (Panzeri, Macke, Gross, & Kayser, 2015). These examples demonstrate that time–content interdependence is important to consider when theorizing about the functional role of oscillations in perception, cognition, and behavior.

VARIABILITY IN NEURAL DYNAMICS AND ANALYTICAL INTERPRETATION

Problems for Current Methodologies

In the previous section, we discussed *what/when* interdependence in the environment and in neural computation. When environmental *what/when* interdependence is not explicitly modeled, it leads to the unsound conclusion that timing and content information are processed independently in the brain. Similarly, neglecting *what/when* interdependence in theories of neural computation leads to questionable analytical interpretation of neural readouts, because temporal variability in neural responses is not explicitly taken into account during data analysis. This is an analytical problem for many studies in perception and cognition because brain time is almost exclusively assumed to be relative to the onset of a perceptual event, and most statistical models do not explicitly model temporal shifts. If temporal shifts are absolute, they might be expressed as condition differences, because latency differences can be explicitly modeled by extracting differences in peak RT. If there is not an absolute shift, but rather

increased temporal variation for one condition, latency effects might show up as amplitude or frequency shifts (also see Figure 1).

In recent years, there has been an increase in the application of complex statistical models to extract relevant brain dynamics in electrophysiology. These approaches have strongly increased our ability to dissociate different brain states. It is, for example, now possible to analyze ongoing signals (e.g., ongoing speech) by taking the auto-correlated structure of the electrophysiological data into account using, for example, ridge regression (viz., in the EEG and magnetoencephalography field, this encoding or regression model is commonly known as the *temporal response function* [TRF]; Crosse, Di Liberto, Bednar, & Lalor, 2016; Zion Golumbic et al., 2013; Ringach & Shapley, 2004). In addition, the usage of multivariate approaches has made it possible to identify differences in distributed patterns in the brain rather than being restricted to observing only absolute changes as identified with univariate approaches (Formisano, De Martino, Bonte, & Goebel, 2008; Haxby et al., 2001). However, not many of these approaches have focused on explicitly modeling the fact that temporal shifts, rather than true differences in neural representation of the features, could account for any observed differences by encoding feature.

Insensitivity to Amplitude and Latency Effects versus Temporal Shifts in the Modeling and Decoding of Neural Responses

TRFs are widely used for analyzing ongoing EEG or magnetoencephalography, data especially in the auditory domain (Weissbart, Kandylaki, & Reichenbach, 2020; Brodbeck, Hong, & Simon, 2018; Zion Golumbic et al., 2013). This is because they can handle autocorrelated data well and can thereby decompose temporal overlapping responses typically found in ongoing speech (Crosse et al., 2016). Moreover, they allow incorporation of continuous factors such as the envelope of the data (Hausfeld, Riecke, Valente, & Formisano, 2018). In addition, factors of interest beyond the acoustics that are continuous can be modeled such as the predictability or entropy of words in context (Weissbart et al., 2020; Brodbeck et al., 2018).

The TRF reflects the time course of regression coefficients of the model at different delays; performance of the full regression model can be estimated through cross-validation (Crosse et al., 2016). Of course, from the model's perspective, it does not matter whether amplitude differences are generated by variation in amplitude or by shifts in latency (Figure 4). Any latency shift reflects a local amplitude shift such that when one condition has a later response, there will also be an amplitude difference between conditions at different time points. Therefore, the TRF will look almost identical. TRFs also look similar to amplitude shifts when simply changing condition-specific temporal variation in the response. Similar insensitivity to temporal variability occurs for methods in which

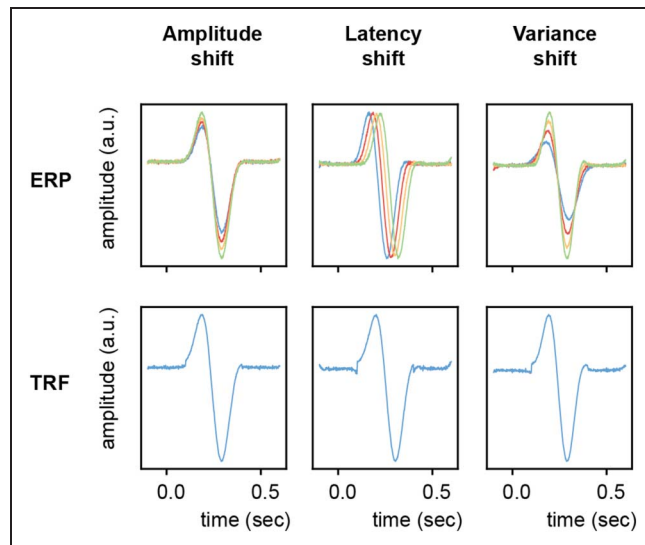


Figure 4. Temporal shifts could underlie found TRF effect. TRFs can be identical for underlying effects caused by amplitude shifts, latency shifts, or variance shifts.

data are collapsed over time. For example, decoding methods often include a range of time points over which decoding occurs, making it impossible to dissociate whether the outcome relates to temporal variation or amplitude changes. Of course, it is possible to decode using another means of variance, such as the spatial or spectral distribution, and then the analysis would be more sensitive to temporal differences and variation.

Temporal Variability Can Induce Low-frequency Effects

Temporal variability in neural response also changes the spectral content of the analyzed signal. Various studies show that although low-frequency spectral content differences do not exist in the acoustics, neural responses at this rate can be still found (Kaufeld, Bosker, et al., 2020; Keitel, Gross, & Kayser, 2018; Ding et al., 2016). One interpretation is that lower frequency oscillations integrate responses across a wider timescale (Henin et al., 2021; Meyer et al., 2020). Although an interesting hypothesis, the mere presence of low-frequency peaks in the spectra (i.e., when the peaks do not exist in the acoustics) does not necessitate the presence or involvement of intrinsic low-frequency oscillations in the brain. It is possible that the latency or variance of neural activation is modulated through, for example, a linguistic event that occurs at a low-frequency rate (Figure 5A). These effects can be seen as low-frequency peaks in either the power spectra or inter-trial coherence (Figure 5B). Of course, the measured signal does contain this low-frequency pattern as the signal is modulated at this rate. However, it is an open discussion whether we should call this an oscillation arising from the brain as a computational means for integration or rather a stimulus-induced response, which happens to occur at a later or

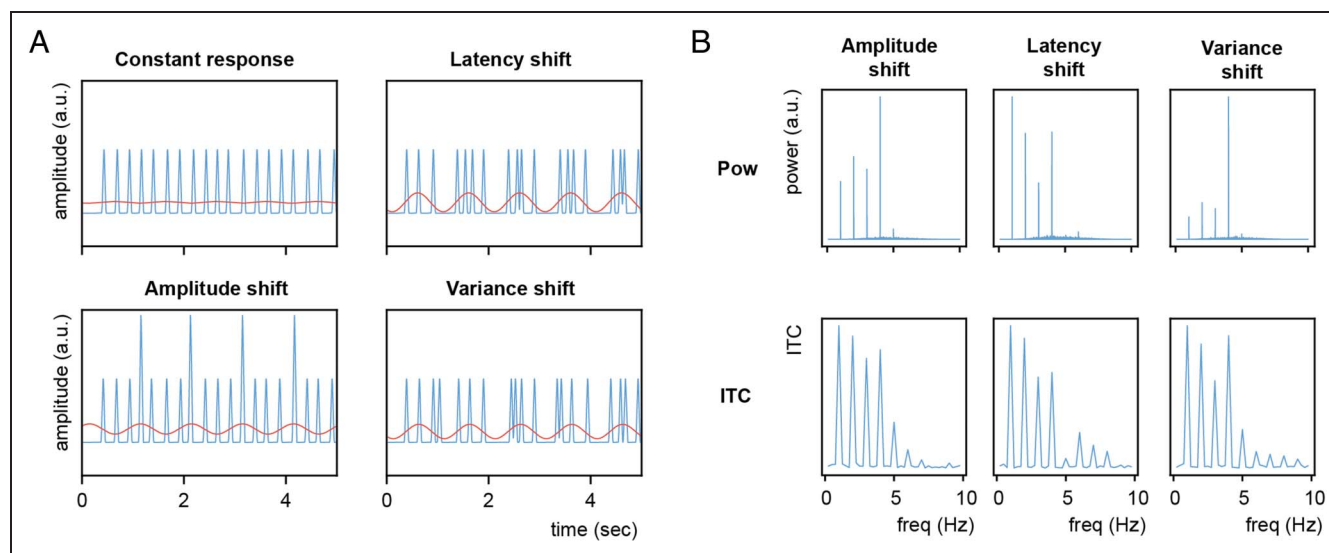


Figure 5. Temporal shifts could underlie found low-frequency effects. (A) One hertz fit (red) on responses that changes amplitude, latency, or variance. (B) Power and ITC effects can be highly similar for underlying effects caused by amplitude shifts, latency shifts, or variance shifts.

with a more/less variable time (Zoefel, Ten Oever, & Sack, 2018). If the latter, one would predict no peak when we correct for the variance in the timing of brain responses (the metatheoretical intricacies of interpreting these types of results is further analyzed in Ten Oever et al., 2022, and Guest & Martin, 2023).

Temporal Variability Can Induce Low-frequency Effects, Even on Filtered Data

For single trials, filtering can solve temporal variability problems as the frequency content of that signal is fully dependent on a single event (Figure 1). When studying continuous data, one filters signals across multiple stimuli. If a single event has a different temporal variability, or contains a latency shift, the filtered data will contain the low-frequency signal (Figure 5). This situation cannot be avoided as a longer window is needed to estimate low frequencies, which, for continuous data, will contain responses to nearby stimuli. When using this type of filtered data in other analyses, one has to be aware that the presence of a low-power signal, or increased variance in the power signal, can lead to potentially spurious increased tracking at that frequency. We here simulate data at an average stimulus presentation rate of 4 Hz similar to the previous section. Every fourth signal, we either modulate the amplitude, induce a temporal latency shift, or vary the temporal variability (Figure 6A). Then, mutual information (MI) (Ince et al., 2017) is calculated between filtered stimulus data (aligned with onset of the stimuli convoluted with a Gaussian) and the filtered response data (responses are broadband 1- to 20-Hz signal of a 0.4 duration plus an additional 1/f noise). The outcome shows that decreasing temporal variability can increase mutual information at delta frequencies even when the stimulus does not contain a 1-Hz signal and the delta itself is not

modulated in the responses (Figure 6B). Instead, merely reducing the temporal variability can cause this effect. This might seem counterintuitive, but having a more systematic neural pattern can have an influence on the mutual information with an external stimulus, for example, because the phase estimation is more consistent.

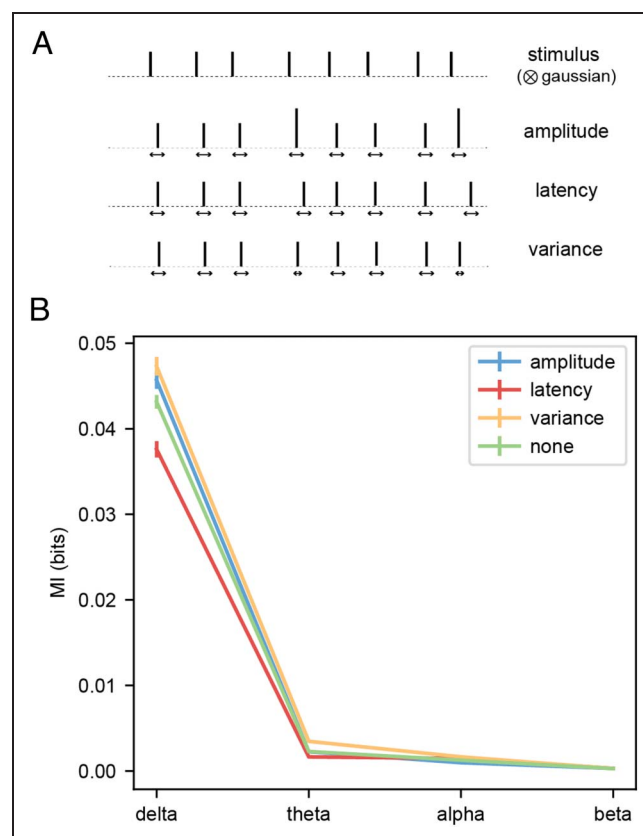


Figure 6. Temporal variability changes can influence tracking results. (A) Conditions for the simulation. (B) Results for the MI analysis.

Note that, in our simulation, this effect extended for the theta frequency band (but not the alpha and beta band). Although we show that temporal variability can explain some of previously reported effects, we do not claim that temporal variability is by definition the mechanism behind these low-frequency effects. Instead, we reiterate that one must be careful when interpreting data and attentive and open to alternative explanations.

Inferences about Brain Computation and Our Analytical Choices

If one is solely interested whether brain responses change as a function of condition or other factors (either included in a regression or part of experimental design), any observed difference in condition would indicate success. However, we hope that, as scientists, we aim for more explanatory satisfaction and, thus, that we try to understand the underlying causal structure that an observed difference reflects. To do so, it is critical to understand what can cause a potential difference in an analysis or model, as it is by no means always obvious. We here show that temporal variability, potentially caused by *what/when* interdependence, is an important factor to take into consideration, especially in the interpretation of magneto- and electrophysiological data. A TRF cannot directly be inferred to reflect amplitude or latency effects. A peak in the spectral domain could be a consequence of amplitude shift, latency shift, or an actual neural oscillation at the specified rate. Excluding alternative explanations for our results in our scientific practice is critically important to better understand the mechanisms by which our effects arise, and thereby build better theories of brain computation (Guest & Martin, 2023).

To investigate whether temporal variability is part of the underlying mechanism driving our findings, below, we propose a few practices and future developments needed to further investigate if effects are a consequence of temporal variability:

- In a TRF analysis, redoing the TRF analysis on a median split of a significant linear factor could provide information about what the underlying causes of the significant TRF effects are. If there are clear latency differences in the two TRFs, then one could assume that the underlying effect is likely because of latency shifts that systematically relate to the factor of interest (Figure 4B). If there are amplitude differences, the cause is more likely an amplitude variation or a change in temporal variability across the linear factor (Figure 4A and 4C). It is more difficult to dissociate whether TRFs are caused by changes in amplitude or changes in temporal variability.
- One approach for the low-frequency effects in a steady-state analysis described in the Temporal Variability Can Induce Low-frequency Effects section is to investigate the instantaneous phase response at

the stimulus repetition rate. For amplitude effects, one would expect that the phase response at the stimulus repetition rate does not change. However, the phase response because of latency shifts or changes in temporal variability should lead to systematically delayed phase responses or more variable phase responses, respectively. Note, however, that effects found via these phase responses could be small and dependent on a waveform shape of the underlying oscillation (Jones, 2016). Therefore, it is difficult to exclude that the absence of a latency/temporal variation increases in the absence of these changes. As with the TRF analysis, it seems easier to find effects because of latency shifts compared with changes in temporal variability.

- Generally, we advise performing analyses that do not collapse across many time points, which is sometimes done in decoding approaches.
- The same goes for collapsing across trials either in the time or frequency domain. More approaches are being developed to extract single trial oscillatory events from electrophysiological data and quantify them (Quinn, Lopes-dos-Santos, Dupret, Nobre, & Woolrich, 2021; Donoghue et al., 2020; Jones, 2016). For example, Shin, Law, Tsutsui, Moore, and Jones (2017) developed a way to extract individual beta events. Something that might also be possible for other frequency bands is as follows: Separating different oscillatory states that occur in parallel in a data driven manner, for example, done in an empirical mode decomposition (Huang et al., 1998), could also aid a single trial analysis.
- New approaches are being developed that perform temporal alignment to optimize the temporal correlation across trials (van Bree et al., 2022; Haxby, Guntupalli, Nastase, & Feilong, 2020; Zoumpoulaki, Alsufyani, Filetti, Brammer, & Bowman, 2015). Some of these methods could be further developed to investigate whether alignment parameters systematically relate to physiological factors of interest.
- Ultimately, the tendency to go for more naturalistic designs makes it more difficult to unequivocally determine the mechanistic underpinnings of *what* and *when* interactions. We therefore believe that to supplement the naturalistic designs, which have clear and unequivocal value, we will need well-designed, experimentally controlled studies to further disentangle these interactions.

WHAT ABOUT WHEN?

The temporal dynamics of brain responses are—definitionally—different, every time a stimulus is presented. This variability depends not only on the features of the perceptual input but also on differences in brain and bodily state at the time of presentation (Schroeder,

Wilson, Radman, Scharfman, & Lakatos, 2010). We here have argued that temporal variability should not be merely seen as a source of noise that can lower the signal-to-noise ratio of our analyses (also see Panzeri et al., 2015). Instead, the time of information in the external world (stimulus time) as well as the timing of perceptual and cognitive computations as realized in the brain (brain time) are important for stimulus information processing and interpretation (Panzeri et al., 2015; Ten Oever & Sack, 2015). Indeed, the environment, content, and timecourse are dependent on each other, and the brain likely has an integrated time–content representation of external stimuli and events in the world. Although this representation does not have to be operationalized at the same timescales as the external world (Dainton, 2010), temporal dynamics of the external world do need to be integrated with the temporal dynamics of the brain. As such, what (content) and when (stimulus time vs. brain time) are highly intertwined if not indivisible in the brain. The time of the presentation cues the content of a stimulus (Reinisch & Sjerps, 2013; Ten Oever et al., 2013; Figure 3A), and the content of a stimulus can influence the timing of processing (Los et al., 2021; Figure 3B).

Similarly, in most data analyses approaches, we ignore temporal variability as a potential explanation of our results at our own peril. The most common data analyses methods do not take in to account inherent temporal variability in neural responses; we have shown that ignoring temporal variability can lead to spurious effects that can be misinterpreted. Our hypotheses should be guided by what we know about temporal dynamics in the brain, and we should therefore not only include variations in content (e.g., pitch, lexical information, word onset, predictability) as factors in our models, but also their interaction with and across time. The relationship between neural computation and time needs deeper attention in our analytical methods, models, and theories, and especially in methods that can deal with expected (but not exact) time shifts (van Bree et al., 2022; Zoumpoulaki et al., 2015). Such an approach will enable us to differentiate temporal variability from other interpretations of our data and to formulate more comprehensive theories that incorporate temporal variability as a feature, not a bug.

To get a better understanding of how *what* and *when* depend upon each other, we no doubt need more experimental and modeling work. In the most basic form, this can be done by experimentally changing the time of stimulus presentation and hypothesizing about behavioral changes. There are many studies that show that the stimulus time of presentation can impact behavioral detection or accuracy (Fiebelkorn, Saalman, & Kastner, 2013; Jones, Johnston, & Puente, 2006), but fewer investigate how temporal dynamics create an overall bias to what participants perceive (Ten Oever & Sack, 2019; van de Ven et al., 2017; Reinisch & Sjerps, 2013). Therefore, more studies are needed to get a better understanding of the *what/when* relationship. Besides these behavioral

studies, we need to go the extra mile to link *what/when* dependence to the ongoing temporal dynamics in the brain (Köseme et al., 2018; Ten Oever & Sack, 2015). This link can be formed by building models (theoretical and computational; Guest & Martin, 2021). Ultimately, from the brain’s-eye-view, *what* and *when* are indivisible aspects of interpreting the perceptual world, cognizing, and behaving in it; this principle needs to be reflected in our theories of brain computation—in fact, it is crucial for any account of how the brain computes and represents the perceptual world, and in turn acts on it via cognition and behavior.

METHODS

ERP Simulations

We simulated how modulations in duration versus temporal variability can lead to similar-looking ERPs. To do so, we simulated five ERP components (P1, N1, P2, N2, and P3) at an average peak time of 80, 100, 180, 220, 400 msec, respectively. Each component was modeled as a Gaussian shape at varying amplitudes. For the duration modulation, the width and amplitude of the Gaussian were 10, 10, 20, 20, and 100 msec and 2, –3, 2.5, –1, and 2.5, respectively. For the duration modulation, the width and amplitude of the Gaussian were 10, 10, 20, 20, and 100 msec and 2, –3, 2.5, –1, and 2.5, respectively. For the temporal variability modulation, the width of the Gaussian was always 25 msec and the amplitude was 2, –3, 2.5, –1, and 6. Temporal variability was extracted from a normal distribution centered around the peak time with a width equal to the duration length for the duration modulation (10, 10, 20, 20, 100 msec). We added noise from a normal distribution (mean: 0, std: 1) multiplied with 0.05. One thousand permutations were run, and the average ERP and the individual trials were extracted.

P300 Results

We used data collected at Maastricht University from one participant who took part in a study investigating P300 (of a total of 14 participants collected for this study). The study was approved by the local ethics board at Maastricht University, and all participants gave informed consent. Data were collected at a sampling rate of 200 Hz with an online filter set at 35 Hz. In this study, participants had to respond to digits 3, 6, 8, and 9. In half of the trials, participants had to respond with the left hand to 3 and 6, and participants had to respond with the right hand to 8 and 9. In other blocks, the mapping was reversed. Digits were presented at a size of $0.87^\circ \times 1.37^\circ$. On half the trials, digits were visually degraded by overlaying a checkerboard pattern. Only trials with a degraded digit were used for the present purpose. ISI was jittered between 3950 and 4450 msec. Preprocessing involved epoching from –0.2 to 1.2 sec around stimulus onset, and removal of eye movement artifacts with independent component analysis.

After that, trials with EEG exceeding 75 μV were detected as artifacts and excluded. For all other trials, we performed a fast Fourier transform (FFT) analysis using Hanning tapers (and no further padding) as implemented in Fieldtrip on single trials or on the trial average and extracted the log of the power. The peak frequency within a range of 2–5 Hz was extracted per trial and for the average (within the known band of the P300 and ensuring no influence of the alpha range, which commonly has high power in single trials). To visualize the ERP and single trial examples, we bandpass-filtered the data between 1 and 10 Hz.

TRF Simulations

We simulated how TRF can look similar when differences are either based on amplitude shifts, latency shifts, or difference in temporal variability. We simulated four conditions with the following base values: amplitude: 1; latency: 0.08 sec; latency variability: 0.03 sec. For the amplitude modulation, we varied the amplitude across conditions (0.85, 0.95, 1.05, 1.15). For the latency modulation, we varied the latency across conditions (0.08, 0.1, 0.12, and 0.14 sec) and set the variability to 0.03. For the variability modulation, we varied the variability across conditions (0.04, 0.03, 0.02, and 0.01 sec). Stimuli were modeled as a 4-Hz sinusoidal with a Hanning taper of a 0.3-sec duration. Stimulus were spaced every 500 msec with 4000 stimuli. Normal noise (mean: 0, std: 1) at an amplitude of 0.2 was added to the stimuli. We calculated the TRF using defining the factor based on the onset of the words at an amplitude ranging from 1 to 4. TRF were estimated at -0.1 to 0.6 sec using the MNE Python *TimeDelayingRidge* implementation.

Low-frequency Simulations

Stimuli were the same stimuli as for the TRF simulations but spaced at a 4-Hz rate. We simulated four conditions with the following base values: amplitude: 1; latency: 0.11 sec; latency variability: 0.05 sec. Then, we varied how every fourth stimulus was modulated (so at a 1-Hz rate). For the amplitude modulation, we varied the amplitude of the fourth stimulus to 5. For the latency modulation, we increased the latency of the fourth stimulus by 0.39–0.5 sec. For the variability modulation, we varied the temporal variability to 0.4 sec. Then, we performed an FFT analysis and extracted the power over the complete time course. We also extracted the intertrial coherence by epoching the data in nonoverlapping 4-sec epochs and calculating inferior temporal cortex from the phase estimated from the FFT estimation. For the visualization of the examples (Figure 5A) of a 1-Hz fit, we kept the same parameters except that the base variability was set to 0 (to improve visualization of the variability) and the amplitude modulation to 2. Fitting was performed using a sinus fit using a least square implementation.

Mutual Information Simulations

The MI simulations were performed in MATLAB using the *gcmi* toolbox (Ince et al., 2017). Responses consistent of evoked responses for Hanning tapered 10- to 20-Hz sinusoids (step size of 1 Hz) at an average phase of 0 (random added noise of 0.4π at a poisson distribution) of a duration of 0.2. $1/f$ noise was added at an amplitude of 0.5. Stimuli consisted of an average 4-Hz repetition of stimuli (stimulus interval were drawn from a random normal distribution of a mean of 0.25 sec and a standard deviation of 0.1 sec. The stimulus with only the onset at 1 were convoluted with a normal distribution with a 0.2-sec deviation (and 0 mean). Default delay, variability, and amplitude were set at 0.2 sec, 0.05 sec, and 1 sec, respectively. As in the low-frequency simulation, we modulated every fourth stimulus. For the amplitude modulation, we varied the amplitude to 1.5. For the latency modulation, we changed the latency to 0.25 sec. For the variability modulation, we changed the fourth variability to 0.03 sec. The MI was calculated on the phase estimations of the Hilbert transformed stimulus and data time courses (filtered with a third order Butterworth filter at 0.5–3 Hz, 3–8 Hz, 8–12 Hz, 12–30 Hz) across eight different delays (0:0.05:0.35 sec) and averaged across delays (Kaufeld, Bosker, et al., 2020; Keitel et al., 2018).

Acknowledgments

A. E. M. was supported by an Independent Max Planck Research Group and a Lise Meitner Research Group “Language and Computation in Neural Systems,” by NWO Vidi grant 016.Vidi.188.029 to A. E. M., and by Big Question 5 (to Prof. Dr. Roshan Cools and Dr. Andrea E. Martin) of the Language in Interaction Consortium funded by NWO Gravitation Grant 024.001.006 to Prof. Dr. Peter Hagoort.

Corresponding author: Sanne Ten Oever, Oxfordlaan 55, 6229EV Maastricht, The Netherlands, or via e-mail: sanne.tenoever@maastrichtuniversity.nl.

Author Contributions

Sanne Ten Oever: Conceptualization; Formal analysis; Methodology; Validation; Visualization; Writing—Original draft; Writing—Review & editing. Andrea E. Martin: Conceptualization; Funding acquisition; Supervision; Validation; Writing—Original draft; Writing—Review & editing.

Funding Information

Andrea E. Martin, Max Planck Institute, Lise Meitner Research Group; NWO, grant numbers: 024.001.006, 016.Vidi.188.029.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were $M(\text{an})/M = .407$, $W(\text{oman})/M = .32$, $M/W = .115$, and $W/W = .159$, the comparable proportions for the articles that these authorship teams cited were $M/M = .549$, $W/M = .257$, $M/W = .109$, and $W/W = .085$ (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15, 122–131. <https://doi.org/10.1016/j.tics.2011.01.003>, PubMed: 21292539
- Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., et al. (2018). Not all predictions are equal: “What” and “when” predictions modulate activity in auditory cortex through different mechanisms. *Journal of Neuroscience*, 38, 8680–8693. <https://doi.org/10.1523/JNEUROSCI.0369-18.2018>, PubMed: 30143578
- Ballard, D. H. (2015). *Brain computation as hierarchical abstraction*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/9780262028615.001.0001>
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76, 695–711. <https://doi.org/10.1016/j.neuron.2012.10.038>, PubMed: 23177956
- Bi, G., & Poo, M. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature*, 401, 792–796. <https://doi.org/10.1038/44573>, PubMed: 10548104
- Boltz, M. G. (1999). The processing of melodic and temporal information: Independent or unified dimensions? *Journal of New Music Research*, 28, 67–79. <https://doi.org/10.1076/jnmr.28.1.67.3121>
- Bosker, H. R. (2017a). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics*, 79, 333–343. <https://doi.org/10.3758/s13414-016-1206-4>, PubMed: 27630058
- Bosker, H. R. (2017b). How our own speech rate influences our perception of others. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 1225–1238. <https://doi.org/10.1037/xlm0000381>, PubMed: 28095012
- Brisson, B., Robitaille, N., & Jolicoeur, P. (2007). Stimulus intensity affects the latency but not the amplitude of the N2pc. *NeuroReport*, 18, 1627–1630. <https://doi.org/10.1097/WNR.0b013e3282f0b559>, PubMed: 17885614
- Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology*, 28, 3976–3983. <https://doi.org/10.1016/j.cub.2018.10.042>, PubMed: 30503620
- Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7, 446–451. <https://doi.org/10.1038/nn1233>, PubMed: 15114356
- Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304, 1926–1929. <https://doi.org/10.1126/science.1099745>, PubMed: 15218136
- Coopmans, C. W., De Hoop, H., Hagoort, P., & Martin, A. E. (2022). Effects of structure and meaning on cortical tracking of linguistic units in naturalistic speech. *Neurobiology of Language*, 3, 386–412. https://doi.org/10.1162/nol_a_00070, PubMed: 37216060
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, 386, 608–611. <https://doi.org/10.1038/386608a0>, PubMed: 9121584
- Cravo, A. M., Rohenkohl, G., Santos, K. M., & Nobre, A. C. (2017). Temporal anticipation based on memory. *Journal of Cognitive Neuroscience*, 29, 2081–2089. https://doi.org/10.1162/jocn_a_01172, PubMed: 28777060
- Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, 10, 604. <https://doi.org/10.3389/fnhum.2016.00604>, PubMed: 27965557
- Dainton, B. (2010). Temporal consciousness. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. Stanford University.
- De Kock, R., Gladhill, K. A., Ali, M. N., Joiner, W. M., & Wiener, M. (2021). How movements shape the perception of time. *Trends in Cognitive Sciences*, 25, 950–963. <https://doi.org/10.1016/j.tics.2021.08.002>, PubMed: 34531138
- De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, 22, 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>, PubMed: 30122170
- Dennett, D. C. (1993). *Consciousness explained*. Penguin UK. <https://doi.org/10.2307/2108259>
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19, 158–164. <https://doi.org/10.1038/nn.4186>, PubMed: 26642090
- Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., et al. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*, 23, 1655–1665. <https://doi.org/10.1038/s41593-020-00744-x>, PubMed: 33230329
- Downer, J. D., Bigelow, J., Runfeldt, M., & Malone, B. J. (2021). Temporally precise population coding of dynamic sounds by auditory cortex. *Journal of Neurophysiology*, 126, 148–169. <https://doi.org/10.1152/jn.00709.2020>, PubMed: 34077273
- Duncan-Johnson, C. C., & Donchin, E. (1980). The relation of P300 latency to reaction time as a function of expectancy. In H.H. Kornhubek & L. Deecke (Eds.), *Progress in brain research* (Vol. 54, pp. 717–722). Elsevier. [https://doi.org/10.1016/S0079-6123\(08\)61693-3](https://doi.org/10.1016/S0079-6123(08)61693-3)
- Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, 18, 131–136. <https://doi.org/10.1016/j.conb.2008.06.002>, PubMed: 18639634
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433. <https://doi.org/10.1038/415429a>, PubMed: 11807554
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8, 162–169. <https://doi.org/10.1016/j.tics.2004.02.002>, PubMed: 15050512
- Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, 81, 1179–1189. <https://doi.org/10.1016/j.neuron.2014.01.042>, PubMed: 24607235

- Feldman, D. E. (2000). Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron*, 27, 45–56. [https://doi.org/10.1016/S0896-6273\(00\)00008-8](https://doi.org/10.1016/S0896-6273(00)00008-8), PubMed: 10939330
- Fiebelkorn, I. C., Saalman, Y. B., & Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. *Current Biology*, 23, 2553–2558. <https://doi.org/10.1016/j.cub.2013.10.063>, PubMed: 24316204
- Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, 4, 414–420. [https://doi.org/10.1016/S0022-5371\(65\)80081-0](https://doi.org/10.1016/S0022-5371(65)80081-0)
- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science*, 322, 970–973. <https://doi.org/10.1126/science.1164318>, PubMed: 18988858
- Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9, 474–480. <https://doi.org/10.1016/j.tics.2005.08.011>, PubMed: 16150631
- Fries, P., Nikolic, D., & Singer, W. (2007). The gamma cycle. *Trends in Neurosciences*, 30, 309–316. <https://doi.org/10.1016/j.tins.2007.05.005>, PubMed: 17555828
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 360, 815–836. <https://doi.org/10.1098/rstb.2005.1622>, PubMed: 15937014
- Friston, K. (2011). Prediction, perception and agency. *International Journal of Psychophysiology*, 83, 248–252. <https://doi.org/10.1016/j.ijpsycho.2011.11.014>, PubMed: 22178504
- Frith, C., & Dolan, R. J. (1997). Brain mechanisms associated with top-down processes in perception. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 352, 1221–1230. <https://doi.org/10.1098/rstb.1997.0104>, PubMed: 9304688
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, 7, 773–778. <https://doi.org/10.1038/nn1268>, PubMed: 15195098
- Georgopoulos, A. P. (2002). Cognitive motor control: Spatial and temporal aspects. *Current Opinion in Neurobiology*, 12, 678–683. [https://doi.org/10.1016/S0959-4388\(02\)00382-3](https://doi.org/10.1016/S0959-4388(02)00382-3), PubMed: 12490258
- Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, 4, 138. <https://doi.org/10.3389/fpsyg.2013.00138>, PubMed: 23519170
- Ghitza, O. (2014). Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology*, 5, 652. <https://doi.org/10.3389/fpsyg.2014.00652>, PubMed: 25071631
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15, 511–517. <https://doi.org/10.1038/nn.3063>, PubMed: 22426255
- Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16, 789–802. <https://doi.org/10.1177/1745691620970585>, PubMed: 33482070
- Guest, O., & Martin, A. E. (2023). On logical inference over brains, behaviour, and artificial neural networks. *Computational Brain & Behavior*, 6, 213–227. <https://doi.org/10.1007/s42113-022-00166-x>
- Gwilliams, L., King, J.-R., Marantz, A., & Poeppel, D. (2020). Neural dynamics of phoneme sequencing in real speech jointly encode order and invariant content. *bioRxiv*.
- Harrington, D. L., Haaland, K. Y., & Hermanowitz, N. (1998). Temporal processing in the basal ganglia. *Neuropsychology*, 12, 3–12. <https://doi.org/10.1037/0894-4105.12.1.3>, PubMed: 9460730
- Hausfeld, L., Riecke, L., Valente, G., & Formisano, E. (2018). Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *Neuroimage*, 181, 617–626. <https://doi.org/10.1016/j.neuroimage.2018.07.052>, PubMed: 30048749
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430. <https://doi.org/10.1126/science.1063736>, PubMed: 11577229
- Haxby, J. V., Guntupalli, J. S., Nastase, S. A., & Feilong, M. (2020). Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *eLife*, 9, e56601. <https://doi.org/10.7554/eLife.56601>, PubMed: 32484439
- Helmholtz, H. (1878). *The facts in perception*. Wesleyan University Press.
- Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., et al. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, 7, eabc4530. <https://doi.org/10.1126/sciadv.abc4530>, PubMed: 33608265
- Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences, U.S.A.*, 109, 20095–20100. <https://doi.org/10.1073/pnas.1213390109>, PubMed: 23151506
- Herbst, S. K., & Obleser, J. (2019). Implicit temporal predictability enhances pitch discrimination sensitivity and biases the phase of delta oscillations in auditory cortex. *Neuroimage*, 203, 116198. <https://doi.org/10.1016/j.neuroimage.2019.116198>, PubMed: 31539590
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences*, 454, 903–995. <https://doi.org/10.1098/rspa.1998.0193>
- Ince, R. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, 38, 1541–1573. <https://doi.org/10.1002/hbm.23471>, PubMed: 27860095
- Ivry, R. B., & Spencer, R. M. (2004). The neural representation of time. *Current Opinion in Neurobiology*, 14, 225–232. <https://doi.org/10.1016/j.conb.2004.03.013>, PubMed: 15082329
- Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, 16, 200–206. <https://doi.org/10.1016/j.tics.2012.03.002>, PubMed: 22436764
- Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neurosciences*, 37, 357–369. <https://doi.org/10.1016/j.tins.2014.04.001>, PubMed: 24836381
- Jewett, D. L., Romano, M. N., & Williston, J. S. (1970). Human auditory evoked potentials: Possible brain stem components detected on the scalp. *Science*, 167, 1517–1518. <https://doi.org/10.1126/science.167.3924.1517>, PubMed: 5415287
- Jia, J., Fan, Y., & Luo, H. (2022). Alpha-band phase modulates bottom-up feature processing. *Cerebral Cortex*, 32,

- 1260–1268. <https://doi.org/10.1093/cercor/bhab291>, PubMed: 34411242
- Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology*, *16*, 472–479. <https://doi.org/10.1016/j.cub.2006.01.032>, PubMed: 16527741
- Jones, S. R. (2016). When brain rhythms aren't 'rhythmic': Implication for their mechanisms and meaning. *Current Opinion in Neurobiology*, *40*, 72–80. <https://doi.org/10.1016/j.conb.2016.06.010>, PubMed: 27400290
- Jones, M. R., Johnston, H. M., & Puenta, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive Psychology*, *53*, 59–96. <https://doi.org/10.1016/j.cogpsych.2006.01.003>, PubMed: 16563367
- Kaufeld, G., Bosker, H. R., Ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*, *40*, 9467–9475. <https://doi.org/10.1523/JNEUROSCI.0302-20.2020>, PubMed: 33097640
- Kaufeld, G., Naumann, W., Meyer, A. S., Bosker, H. R., & Martin, A. E. (2020). Contextual speech rate influences morphosyntactic prediction and integration. *Language, Cognition and Neuroscience*, *35*, 933–948. <https://doi.org/10.1080/23273798.2019.1701691>
- Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2020). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 549–562. <https://doi.org/10.1037/xlm0000744>, PubMed: 31343252
- Kay, K. N., & Yeatman, J. D. (2017). Bottom-up and top-down computations in word-and face-selective cortex. *eLife*, *6*, e22341. <https://doi.org/10.7554/eLife.22341>, PubMed: 28226243
- Kayser, S. J., McNair, S. W., & Kayser, C. (2016). Prestimulus influences on auditory perception from sensory representations and decision processes. *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, 4842–4847. <https://doi.org/10.1073/pnas.1524087113>, PubMed: 27071110
- Keetels, M., & Vroomen, J. (2008). Temporal recalibration to tactile-visual asynchronous stimuli. *Neuroscience Letters*, *430*, 130–134. <https://doi.org/10.1016/j.neulet.2007.10.044>, PubMed: 18055112
- Keitel, A., & Gross, J. (2016). Individual human brain areas can be identified from their characteristic spectral activation fingerprints. *PLoS Biology*, *14*, e1002498. <https://doi.org/10.1371/journal.pbio.1002498>, PubMed: 27355236
- Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, *16*, e2004473. <https://doi.org/10.1371/journal.pbio.2004473>, PubMed: 29529019
- Köhler, W. (1967). Gestalt psychology. *Psychologische Forschung*, *31*, XVIII–XXX. <https://doi.org/10.1007/BF00422382>, PubMed: 5606808
- Kok, P., Jehee, J. F., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, *75*, 265–270. <https://doi.org/10.1016/j.neuron.2012.04.034>, PubMed: 22841311
- Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences, U.S.A.*, *114*, 10473–10478. <https://doi.org/10.1073/pnas.1705652114>, PubMed: 28900010
- Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, *28*, 2867–2875. <https://doi.org/10.1016/j.cub.2018.07.023>, PubMed: 30197083
- Kösem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage*, *92*, 274–284. <https://doi.org/10.1016/j.neuroimage.2014.02.010>, PubMed: 24531044
- Kotz, S. A. (2011). Differential input of the supplementary motor area to a dedicated temporal processing network: Functional and clinical implications. *Frontiers in Integrative Neuroscience*, *5*, 86. <https://doi.org/10.3389/fnint.2011.00086>, PubMed: 22363269
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*, 792–795. <https://doi.org/10.1126/science.887923>, PubMed: 887923
- Ladefoged, P., & Broadbent, D. E. (1960). Perception of sequence in auditory events. *Quarterly Journal of Experimental Psychology*, *12*, 162–170. <https://doi.org/10.1080/17470216008416720>
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*, 110–113. <https://doi.org/10.1126/science.1154735>, PubMed: 18388295
- Lakatos, P., Musacchia, G., O'Connell, M., Falchier, A., Javitt, D., & Schroeder, C. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*, 750–761. <https://doi.org/10.1016/j.neuron.2012.11.034>, PubMed: 23439126
- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, *94*, 1904–1911. <https://doi.org/10.1152/jn.00263.2005>, PubMed: 15901760
- Lee, H., Simpson, G. V., Logothetis, N. K., & Rainer, G. (2005). Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron*, *45*, 147–156. <https://doi.org/10.1016/j.neuron.2004.12.025>, PubMed: 15629709
- Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*, 1–28. <https://doi.org/10.1177/002383096701000101>, PubMed: 6044530
- Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, *77*, 1002–1016. <https://doi.org/10.1016/j.neuron.2013.03.007>, PubMed: 23522038
- Los, S. A., Nieuwenstein, J., Bouharab, A., Stephens, D. J., Meeter, M., & Kruijine, W. (2021). The warning stimulus as retrieval cue: The role of associative memory in temporal preparation. *Cognitive Psychology*, *125*, 101378. <https://doi.org/10.1016/j.cogpsych.2021.101378>, PubMed: 33524889
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Luck, S. J., & Hillyard, S. A. (1990). Electrophysiological evidence for parallel and serial processing during visual search. *Perception & Psychophysics*, *48*, 603–617. <https://doi.org/10.3758/BF03211606>, PubMed: 2270192
- Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, *86*, 283–293. [https://doi.org/10.1016/0013-4694\(93\)90110-H](https://doi.org/10.1016/0013-4694(93)90110-H), PubMed: 7682932
- Mangels, J. A., Ivry, R. B., & Shimizu, N. (1998). Dissociable contributions of the prefrontal and neocerebellar cortex to time perception. *Cognitive Brain Research*, *7*, 15–39. [https://doi.org/10.1016/S0926-6410\(98\)00005-6](https://doi.org/10.1016/S0926-6410(98)00005-6), PubMed: 9714713
- Markman, A. B., & Otto, A. R. (2011). Cognitive systems optimize energy rather than information. *Behavioral and Brain Sciences*, *34*, 207. <https://doi.org/10.1017/S0140525X11000355>
- Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, *275*, 213–215. <https://doi.org/10.1126/science.275.5297.213>, PubMed: 8985014

- Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology*, 7, 120. <https://doi.org/10.3389/fpsyg.2016.00120>, PubMed: 26909051
- Martin, A. E. (2020). A compositional neural architecture for language. *Journal of Cognitive Neuroscience*, 32, 1407–1427. https://doi.org/10.1162/jocn_a_01552, PubMed: 32108553
- Martin, A. E., & Doumas, L. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biology*, 15, e2000663. <https://doi.org/10.1371/journal.pbio.2000663>, PubMed: 28253256
- Martin, A. E., & Doumas, L. A. (2020). Tensors and compositionality in neural systems. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, 375, 20190306. <https://doi.org/10.1098/rstb.2019.0306>, PubMed: 31840579
- Martin, A. E., Monahan, P. J., & Samuel, A. G. (2017). Prediction of agreement and phonetic overlap shape sublexical identification. *Language and Speech*, 60, 356–376. <https://doi.org/10.1177/0023830916650714>, PubMed: 28915783
- Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: Prestimulus alpha phase predicts visual awareness. *Journal of Neuroscience*, 29, 2725–2732. <https://doi.org/10.1523/JNEUROSCI.3963-08.2009>, PubMed: 19261866
- Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience*, 27, 307–340. <https://doi.org/10.1146/annurev.neuro.27.070203.144247>, PubMed: 15217335
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748. <https://doi.org/10.1038/264746a0>, PubMed: 1012311
- Mehta, M. R., Lee, A. K., & Wilson, M. A. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, 417, 741–746. <https://doi.org/10.1038/nature00807>, PubMed: 12066185
- Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, 48, 2609–2621. <https://doi.org/10.1111/ejn.13748>, PubMed: 29055058
- Meyer, L., Sun, Y., & Martin, A. E. (2020). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 35, 1089–1099. <https://doi.org/10.1080/23273798.2019.1693050>
- Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience*, 9, 875–877. <https://doi.org/10.1038/nn1712>, PubMed: 16732276
- Mohanta, S., Afrasiabi, M., Casey, C. P., Tanabe, S., Redinbaugh, M. J., Kambi, N. A., et al. (2021). Predictive feedback, early sensory representations, and fast responses to predicted stimuli depend on NMDA receptors. *Journal of Neuroscience*, 41, 10130–10147. <https://doi.org/10.1523/JNEUROSCI.1311-21.2021>, PubMed: 34732525
- Morillon, B., Schroeder, C. E., Wyart, V., & Arnal, L. H. (2016). Temporal prediction in lieu of periodic stimulation. *Journal of Neuroscience*, 36, 2342–2347. <https://doi.org/10.1523/JNEUROSCI.0836-15.2016>, PubMed: 26911682
- Moutoussis, K., & Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London, Series B*, 264, 393–399. <https://doi.org/10.1098/rspb.1997.0056>, PubMed: 9107055
- Moutoussis, K., & Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual perceptual systems. *Proceedings of the Royal Society of London, Series B*, 264, 1407–1414. <https://doi.org/10.1098/rspb.1997.0196>, PubMed: 9364780
- Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89, 133. <https://doi.org/10.1037/0033-2909.89.1.133>
- Nobre, A., & van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, 19, 34–58. <https://doi.org/10.1038/nrn.2017.141>, PubMed: 29213134
- Noppeney, U., & Lee, H. L. (2018). Causal inference and temporal predictions in audiovisual perception of speech and music. *Annals of the New York Academy of Sciences*, 1423, 102–116. <https://doi.org/10.1111/nyas.13615>, PubMed: 29604082
- Panzeri, S., Macke, J. H., Gross, J., & Kayser, C. (2015). Neural population coding: Combining insights from microscopic and mass signals. *Trends in Cognitive Sciences*, 19, 162–172. <https://doi.org/10.1016/j.tics.2015.01.002>, PubMed: 25670005
- Panzeri, S., Petersen, R. S., Schultz, S. R., Lebedev, M., & Diamond, M. E. (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29, 769–777. [https://doi.org/10.1016/S0896-6273\(01\)00251-3](https://doi.org/10.1016/S0896-6273(01)00251-3), PubMed: 11301035
- Parasuraman, R. (1980). Effects of information processing demands on slow negative shift latencies and N100 amplitude in selective and divided attention. *Biological Psychology*, 11, 217–233. [https://doi.org/10.1016/0301-0511\(80\)90057-5](https://doi.org/10.1016/0301-0511(80)90057-5), PubMed: 7272394
- Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, 1423, 378–395. <https://doi.org/10.1111/nyas.13654>, PubMed: 29749625
- Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A.-L. (2017). θ -Band and β -band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience*, 37, 7930–7938. <https://doi.org/10.1523/JNEUROSCI.2882-16.2017>, PubMed: 28729443
- Petras, K., Ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *Neuroimage*, 186, 103–112. <https://doi.org/10.1016/j.neuroimage.2018.10.086>, PubMed: 30403971
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41, 245–255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
- Purushothaman, G., Patel, S. S., Bedell, H. E., & Ogmen, H. (1998). Moving ahead through differential visual latency. *Nature*, 396, 424. <https://doi.org/10.1038/24766>, PubMed: 9853748
- Quinn, A. J., Lopes-dos-Santos, V., Dupret, D., Nobre, A. C., & Woolrich, M. W. (2021). EMD: Empirical mode decomposition and Hilbert-Huang spectral analyses in python. *Journal of open source software*, 6, 2977. <https://doi.org/10.21105/joss.02977>, PubMed: 33855259
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2, 79–87. <https://doi.org/10.1038/4580>, PubMed: 10195184
- Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, 41, 101–116. <https://doi.org/10.1016/j.wocn.2013.01.002>
- Rimmele, J. M., Morillon, B., Poeppl, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, 22, 870–882. <https://doi.org/10.1016/j.tics.2018.08.003>, PubMed: 30266147

- Rimmele, J. M., Poeppel, D., & Ghitza, O. (2021). Acoustically driven cortical δ oscillations underpin prosodic chunking. *eNeuro*, 8, ENEURO.0562-20.2021. <https://doi.org/10.1523/ENEURO.0562-20.2021>, PubMed: 34083380
- Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, 28, 147–166. https://doi.org/10.1207/s15516709cog2802_2
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 336, 367–373. <https://doi.org/10.1098/rstb.1992.0070>, PubMed: 1354376
- Rovelli, C. (2019). *The order of time*. Riverhead Books.
- Salomon, A., Espy-Wilson, C. Y., & Deshmukh, O. (2004). Detection of speech landmarks: Use of temporal information. *Journal of the Acoustical Society of America*, 115, 1296–1305. <https://doi.org/10.1121/1.1646400>, PubMed: 15058352
- Samaha, J., Bauer, P., Cimaroli, S., & Postle, B. R. (2015). Top-down control of the phase of alpha-band oscillations as a mechanism for temporal prediction. *Proceedings of the National Academy of Sciences, U.S.A.*, 112, 8439–8444. <https://doi.org/10.1073/pnas.1503686112>, PubMed: 26100913
- Schaefer, A. T., Angelo, K., Spors, H., & Margrie, T. W. (2006). Neuronal oscillations enhance stimulus discrimination by ensuring action potential precision. *PLoS Biology*, 4, e163. <https://doi.org/10.1371/journal.pbio.0040163>, PubMed: 16689623
- Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32, 9–18. <https://doi.org/10.1016/j.tins.2008.09.012>, PubMed: 19012975
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, 12, 106–113. <https://doi.org/10.1016/j.tics.2008.01.002>, PubMed: 18280772
- Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, 20, 172–176. <https://doi.org/10.1016/j.conb.2010.02.010>, PubMed: 20307966
- Shin, H., Law, R., Tsutsui, S., Moore, C. I., & Jones, S. R. (2017). The rate of transient beta frequency events predicts behavior across tasks and species. *eLife*, 6, e29086. <https://doi.org/10.7554/eLife.29086>, PubMed: 29106374
- Simons, D. J. (2000). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, 4, 147–155. [https://doi.org/10.1016/S1364-6613\(00\)01455-8](https://doi.org/10.1016/S1364-6613(00)01455-8), PubMed: 10740279
- Singer, W. (2009). Distributed processing and temporal codes in neuronal networks. *Cognitive Neurodynamics*, 3, 189–196. <https://doi.org/10.1007/s11571-009-9087-z>, PubMed: 19562517
- Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, 38, 343–358. <https://doi.org/10.1111/1469-8986.3820343>, PubMed: 11347879
- Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38, 387–401. [https://doi.org/10.1016/0013-4694\(75\)90263-1](https://doi.org/10.1016/0013-4694(75)90263-1), PubMed: 46819
- Starr, A., & Achor, L. J. (1975). Auditory brain stem responses in neurological disease. *Archives of Neurology*, 32, 761–768. <https://doi.org/10.1001/archneur.1975.00490530083009>, PubMed: 1180745
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153, 652–654. <https://doi.org/10.1126/science.153.3736.652>, PubMed: 5939936
- Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, 51, 651–659. <https://doi.org/10.1016/j.neuron.2006.08.006>, PubMed: 16950162
- Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, 15, 745–756. <https://doi.org/10.1038/nrn3838>, PubMed: 25315388
- Tal, I., Neymotin, S., Bickel, S., Lakatos, P., & Schroeder, C. E. (2020). Oscillatory bursting as a mechanism for temporal coupling and information coding. *Frontiers in Computational Neuroscience*, 14, 82. <https://doi.org/10.3389/fncom.2020.00082>, PubMed: 33071765
- Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3, 151–162. [https://doi.org/10.1016/S1364-6613\(99\)01299-1](https://doi.org/10.1016/S1364-6613(99)01299-1), PubMed: 10322469
- Ten Oever, S., Hausfeld, L., Correia, J., Van Atteveldt, N., Formisano, E., & Sack, A. (2016). A 7T fMRI study investigating the influence of oscillatory phase on syllable representations. *Neuroimage*, 141, 1–9. <https://doi.org/10.1016/j.neuroimage.2016.07.011>, PubMed: 27395392
- Ten Oever, S., Kaushik, K., & Martin, A. E. (2022). Inferring the nature of linguistic computations in the brain. *PLoS Computational Biology*, 18, e1010269. <https://doi.org/10.1371/journal.pcbi.1010269>, PubMed: 35900974
- Ten Oever, S., & Martin, A. E. (2021). An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions. *eLife*, 10, e68066. <https://doi.org/10.7554/eLife.68066>, PubMed: 34338196
- Ten Oever, S., Meierdierks, T., Duecker, F., De Graaf, T. A., & Sack, A. T. (2020). Phase-coded oscillatory ordering promotes the separation of closely matched representations to optimize perceptual discrimination. *iScience*, 23, 101282. <https://doi.org/10.1016/j.isci.2020.101282>, PubMed: 32604063
- Ten Oever, S., & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences, U.S.A.*, 112, 15833–15837. <https://doi.org/10.1073/pnas.1517519112>, PubMed: 26668393
- Ten Oever, S., & Sack, A. T. (2019). Interactions between rhythmic temporal and feature predictions for parallel time-content associations. *Frontiers in Neuroscience*, 13, 791. <https://doi.org/10.3389/fnins.2019.00791>, PubMed: 31427917
- Ten Oever, S., Sack, A. T., Wheat, K. L., Bien, N., & van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology*, 4, 331. <https://doi.org/10.3389/fpsyg.2013.00331>, PubMed: 23805110
- Ten Oever, S., Van Atteveldt, N., & Sack, A. T. (2015). Increased stimulus expectancy triggers low-frequency phase reset during restricted vigilance. *Journal of Cognitive Neuroscience*, 27, 1811–1822. https://doi.org/10.1162/jocn_a_00820, PubMed: 25941871
- Th  z  , R., Giraud, A.-L., & M  gevand, P. (2020). The phase of cortical oscillations determines the perceptual fate of visual cues in naturalistic audiovisual speech. *Science Advances*, 6, eabc6348. <https://doi.org/10.1126/sciadv.abc6348>, PubMed: 33148648
- Trommershauser, J., Kording, K., & Landy, M. S. (2011). *Sensory cue integration*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195387247.001.0001>
- Van Alphen, P. M., & McQueen, J. M. (2006). The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 178–196. <https://doi.org/10.1037/0096-1523.32.1.178>, PubMed: 16478335

- van Bree, S., Melcón, M., Kolibius, L., Kerrén, C., Wimber, M., & Hanslmayr, S. (2022). The brain time toolbox, a software library to retune electrophysiology data to brain dynamics. *Nature Human Behavior*, *6*, 1430–1439. <https://doi.org/10.1038/s41562-022-01386-8>, PubMed: 35726055
- van de Ven, V., Kochs, S., Smulders, F., & De Weerd, P. (2017). Learned interval time facilitates associate memory retrieval. *Learning & Memory*, *24*, 158–161. <https://doi.org/10.1101/lm.044404>, PubMed: 28298554
- VanRullen, R., Guyonneau, R., & Thorpe, S. J. (2005). Spike times make sense. *Trends in Neurosciences*, *28*, 1–4. <https://doi.org/10.1016/j.tins.2004.10.010>, PubMed: 15626490
- VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*, 207–213. [https://doi.org/10.1016/S1364-6613\(03\)00095-0](https://doi.org/10.1016/S1364-6613(03)00095-0), PubMed: 12757822
- Von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, *5*, 520–526. [https://doi.org/10.1016/0959-4388\(95\)80014-X](https://doi.org/10.1016/0959-4388(95)80014-X), PubMed: 7488855
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*, *72*, 871–884. <https://doi.org/10.3758/APP.72.4.871>, PubMed: 20436185
- Walter, W. G., Cooper, R., Aldridge, V., McCallum, W., & Winter, A. (1964). Contingent negative variation: An electric sign of sensori-motor association and expectancy in the human brain. *Nature*, *203*, 380–384. <https://doi.org/10.1038/203380a0>, PubMed: 14197376
- Weissbart, H., Kandylaki, K. D., & Reichenbach, T. (2020). Cortical tracking of surprisal during continuous speech comprehension. *Journal of Cognitive Neuroscience*, *32*, 155–166. https://doi.org/10.1162/jocn_a_01467, PubMed: 31479349
- White, P. A. (2018). Is conscious perception a series of discrete temporal frames? *Consciousness and Cognition*, *60*, 98–126. <https://doi.org/10.1016/j.concog.2018.02.012>, PubMed: 29549714
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron*, *77*, 980–991. <https://doi.org/10.1016/j.neuron.2012.12.037>, PubMed: 23473326
- Zoefel, B., Ten Oever, S., & Sack, A. T. (2018). The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses. *Frontiers in Neuroscience*, *12*, 95. <https://doi.org/10.3389/fnins.2018.00095>, PubMed: 29563860
- Zoumpoulaki, A., Alsufyani, A., Filetti, M., Brammer, M., & Bowman, H. (2015). Latency as a region contrast: Measuring ERP latency differences with dynamic time warping. *Psychophysiology*, *52*, 1559–1576. <https://doi.org/10.1111/psyp.12521>, PubMed: 26372033