Waves of Binding: Neural Oscillations of Visual, Auditory, and Lexical Integration

Hannah Morrow
hannah.morrow@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/gs_theses

Recommended Citation
https://opencommons.uconn.edu/gs_theses/1488

This work is brought to you for free and open access by the University of Connecticut Graduate School at OpenCommons@UConn. It has been accepted for inclusion in Master's Theses by an authorized administrator of OpenCommons@UConn. For more information, please contact opencommons@uconn.edu.
Waves of Binding: Neural Oscillations of Visual, Auditory, and Lexical Integration

Hannah Michiko Morrow

B.S., The University of Kansas, 2016

A Thesis
Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science
At the
University of Connecticut
2020
Waves of Binding: Neural Oscillations of Visual, Auditory, and Lexical Integration

Presented by

Hannah Michiko Morrow, B.S.

Major Advisor________________________________________________
Eiling Yee

Associate Advisor_____________________________________________
Emily Myers

Associate Advisor_____________________________________________
Jon Sprouse

Associate Advisor_____________________________________________
Roeland Hancock

University of Connecticut

2020
Acknowledgements

I would like to express my gratitude to those who made this research possible. First, I would like to thank my advisor, Dr. Eiling Yee, for providing constant insight, support, and encouragement, and teaching me to think critically about all aspects of an experiment. I appreciate her positivity when I can’t find my own, and her commitment to helping me realize my full potential as a researcher. I would also like to thank my committee members Drs. Emily Myers, Jon Sprouse, and Roeland Hancock for their help and feedback throughout the development of this project. Additional, heartfelt thanks to Yanina Prystauka, Dr. Roeland Hancock, and Dr. Jon Sprouse for providing hours upon hours of expertise and advice on EEG methodology and time frequency analysis. A sincere thank you to the Yee and Altmann labs for their generous feedback and assistance, and many exciting conversations. Finally, I thank my family and friends for their love and support; I’m truly blessed to have such kind, honest, and most of all, hilarious people around me.

The research presented in this thesis was funded by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1247393. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work was supported in part by National Science Foundation Research Traineeship (NRT) grant IGE1747486. This work was supported in part by a UConn Brain Imaging Research Center (BIRC) seed grant.
# Table of Contents

Introduction..........................................................................................................................................1  
Semantic Memory Organization..............................................................................................................1  
Neural Synchrony................................................................................................................................3  
The Role of Neuronal Synchrony in Sensory Integration........................................................................7  
Current Study.......................................................................................................................................9  
Methods...............................................................................................................................................10  
Participants.........................................................................................................................................10  
Stimuli..................................................................................................................................................10  
Procedure..........................................................................................................................................12  
EEG Preprocessing...............................................................................................................................14  
Results................................................................................................................................................15  
Behavioral...........................................................................................................................................15  
Event-related Potentials.......................................................................................................................17  
Time Frequency Analysis....................................................................................................................19  
Gamma..................................................................................................................................................19  
Theta....................................................................................................................................................20  
Good primes only...............................................................................................................................21  
Discussion..........................................................................................................................................24  
Future Directions..............................................................................................................................28  
Appendix............................................................................................................................................32  
References..........................................................................................................................................35
Abstract

If our conceptual knowledge about concrete objects is represented (in part) across the brain regions that are active when those objects are perceived, how is information from those regions integrated, or “bound” into a coherent whole? One potential mechanism for binding is synchronized neuronal firing. In particular, high-frequencies (e.g., gamma) may support local interactions, while lower frequencies (e.g., theta) may support longer-range interactions. Previous work has implicated the role of gamma and theta in binding, but not testing both on the same paradigm with a condition intended to create more cortical distance. We recorded participants’ EEG while they were presented with pairs of visual and auditory stimuli that were either congruent or incongruent (e.g., a picture of a lion followed by either a roar, or a bang sound, respectively), predicting that as observed in prior work (Schneider et al., 2008), integration and thus gamma power should be greater for congruent pairs. We also manipulated whether the sounds following the pictures were non-lexical (e.g., a roar sound) or lexical (e.g., the word “roar”), predicting that because interactions involved in integrating lexical information are presumably longer-range (i.e., not just interactions between visual and auditory processing brain areas, but between these regions and lexical processing areas as well) these interactions would produce more sustained theta activity. We found no differences between conditions in gamma. However, contrary to our predictions there was a significant increase in late theta for incongruent compared to congruent lexical trials, which we speculate may be due to working memory load. Our findings underscore the need for work examining the relationship between synchronous neural firing and integration during conceptual processing, as better understanding may lead to improvements in interventions for clinical populations, as well as better models of semantic memory organization and learning.
**Introduction**

You’re walking into your kitchen, and your cat follows you, meowing for food. The meowing sound seems to so naturally be part of the cat that you may not even think of it as a separate characteristic from the cat’s shape – they seem to be parts of an integrated whole. This perception stems from the brain’s ability to integrate information across different sensory modalities. The neural basis of integration, or how exactly our brain is able to “bind” information into a coherent whole (the binding problem, Treisman, 1996), has been a question for cognitive science for decades. In order to examine this question, one must first consider how knowledge about the sensory properties of objects is stored, then move to determining how it may become bound together.

*Semantic Memory Organization:*

Current views of semantic memory (our general knowledge about the world) assume that information about concrete objects is partially stored across the brain regions that are active when those objects are perceived and/or interacted with (e.g., Allport, 1985). Damasio’s (1989) theory of semantic memory organization is an important precursor of most current theories of “embodied” or “grounded” cognition. Based on evidence from lesion patients, Damasio theorized that semantic memory is supported by a series of “convergence zones” or sets of cell assemblies at different levels that feed forward and backward to communicate with one another (Damasio, 1989). The first level of convergence zones consists of assemblies of neurons located in primary sensory and motor regions. These contain patterns of activity that represent features of objects. The next level, downstream from first, holds representations of features that occur perceptually at the same space and time. The final level, still further downstream, holds representations of combinations of those combinations of features.
Other theories have expanded on this idea in multiple ways, but still kept the core feature of modality-specific information being encoded in modality-specific areas, and then becoming integrated (for review, see Simmons & Barsalou, 2003; see also Patterson, Nestor, & Rogers, 2007; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017). The location(s) of where this integration takes place varies among theories, but for the purposes of this paper, the location of a hub or hubs is not a central focus. That is, the focus is on “how” binding might happen, not yet at “where.”

In order to illustrate the problem of why information must be bound together, it is necessary to show that knowledge about different sensory properties of objects is stored in different regions across the brain. Ample evidence from areas outside of the lesion studies that first led to the idea of convergence zones, supports the idea that conceptual knowledge about different sensory properties of objects is stored at or near regions of the brain that participate in the perception of sensory information. For instance, when subjects are asked to generate color or action words related to an object (e.g., yellow or write, respectively, for a picture of a pencil), there are increases in activity in brain regions just anterior to those responsible for color and motion perception, respectively (Martin et al., 1995). Likewise, there is activation of the left ventral premotor cortex when subjects are asked to view or name tools, but not other objects (Chao & Martin, 2000). These findings suggest that knowledge about sensory properties of objects is stored close to the brain regions responsible for perception of that particular sense. Behavioral studies also support the idea of sensory knowledge being stored close to where perception takes place. For example, in one study, when presented objects which were frequently experienced via touch, subjects were less accurate at naming them if they had to perform a concurrent “patty-cake”-like task that was incompatible with how subjects would interact any of
them (Yee, Chrysikou, Hoffman, & Thompson-Schill, 2013). This interference effect did not occur when the concurrent task was not motor-based (i.e., a mental rotation task), indicating that loading motor regions of the brain specifically interferes with the ability to think about concepts highly related to motor information (Yee et al., 2013).

Studies like these provide evidence that sensory knowledge about objects is stored in different regions across the brain. But one missing element from most theories of grounded or embodied cognition is how information from these modality-specific areas is combined. Specifically, what are the mechanisms involved in integrating (or “binding”) information about the shape of a cat and the sound of its meow?

*Neural Synchrony:*

One mechanism that has been proposed to support binding is the synchronized firing of different cell assemblies (Singer & Gray, 1995). The idea of synchronous firing comes from work by Gray and colleagues (1989) in which they recorded from neurons in the visual cortex of cats, and observed that two spatially separated groups of neurons can fire synchronously based on the features of presented stimuli. Two groups of neurons were chosen based on their response to a particular orientation of a bar of light. When two bars of light (shown in separate receptive fields) in that orientation were moved in opposite directions, the two neuronal groups did not fire in synchrony. However, when the two bars of light moved in the same direction, the neuronal groups demonstrated weak synchrony, which increased when instead of two bars, one bar of moving light was shown. Because this synchronization was affected by global features of the stimuli, such as coherent motion and continuity, Gray and colleagues proposed that this synchrony represents global and coherent features of a pattern (Gray, König, Engel, & Singer, 1989).
Relatedly, it has been proposed that if different groups of neuronal cells fire at the same time in response to processing the color vs. shape vs. motion of the object, this synchronous activity occurring at the same time allows for the perception of color, shape, and motion all belonging to the same entity (Cosmelli et al., 2007). As Triesman (1996) argues, single cells cannot achieve binding alone; our perceptual systems must be able to handle an infinite combination of properties, and the number of neurons (though extremely large) is finite. Therefore, it follows that binding would occur not on the level of single neurons, but on the firing patterns of groups of neurons working together, with the idea that this pattern would cause synchrony across the cell assemblies. When cells fire in-phase synchronously, we see increases in power within frequency bands (Buzsáki, Anastassiou, & Koch, 2012). It is worth noting that while in this paper we refer to “synchrony” as the in-phase firing of cells that results in increases in power in, e.g., the frequency bands described above, the term “synchrony” is used to refer to several different things in the brain sciences (see Appendix). Though these increases in power can often be seen in entrainment studies, which look at stability between two or more oscillators (Buzsáki, 2006), we are not looking at oscillators (i.e., sources), but power increases in general. The majority of work we refer to uses increases in power as the measure of synchrony, with some exceptions in the Discussion section.

Recent work has tested this binding-by-synchrony hypothesis by using electroencephalogram (EEG) to detect the electrical activity associated with neuronal firing during presentation of multimodal stimuli. EEG measures electrical activity at the scalp, and can capture power increases in different frequency bands. In 1929, Hans Berger (as described in Buzsáki, 2006) was among the first to demonstrate different rhythms of the brain as being functionally connected to behaviors. Though others had made similar discoveries, Berger is
mentioned here as his experiments stood out for having numerous controls. Berger found that when a subject closed their eyes, 10 Hz waves could be measured from the scalp over their occipital lobe. When the subject’s eyes were open, slightly faster waves were produced. He called the slower waves “alpha” and the faster waves “beta” (Berger, in Buzsáki, 2006). Since then, discoveries of the relation between frequency bands and behavior has blossomed.

In cognition, five major frequency bands have been delineated. However, given that brains are not limited to five cognitive processes, each band has been implicated in more than one, if not several, cognitive processes, and many cognitive processes have been found to be supported by more than one band. A comprehensive review of the roles of different frequency bands/rhythms is well beyond the scope of this thesis, so below I limit discussion to a brief summary of each of the five frequency bands and one or two of the primary cognitive processes that they have been proposed to play a role in, starting from slower waves to faster waves. (Though this thesis does not focus on the neural generators (or sources) of these oscillations, because prior work often refers to sources, we include some source information in this summary.)

**Delta oscillations (0-3.5 Hz)** in frontal and cingulate cortex have been associated with inhibition and attention, particularly in Go/No-Go tasks (Başar-Eroğlu, Başar, Demiralp, & Schürmann, 1992; Harmony, 2013; see Herrmann, Struber, Helfrich, and Engel, 2016 for review).

**Theta oscillations (4-7.5 Hz)** have been implicated in memory, and—based on *in vitro* work in rabbits, cats, monkeys, and rats demonstrating theta generators in the hippocampus (Green & Arduini, 1954; Vanderwolf, 1969; Goutagny, Jackson, & Williams, 2009)—have been assumed to arise from hippocampal communications (Klimesch, 1999; Mitchell,
McNaughton, Flanagan, & Kirk, 2008). Theta also appears to play a role in inhibition, as it is produced in frontal cortex both during a mental task (continuous addition) and during light drowsiness, leading researchers to think theta is linked to inhibition as you have to inhibit information during a task, but may also inhibit information as you are falling asleep (Takahashi, Shinomiya, Morl, & Tachibana, 1997; see Mitchell et al., 2008 for review). Theta increases have also been observed when the demands of a task switch compared to when the demands remain the same (Sauseng et al., 2006). Finally, theta has been tied to integrating information across sensory modalities (van Ackeren & Rueschemeyer, 2014). We elaborate on this last hypothesized function of theta in the next section.

*Alpha oscillation (8-12 Hz)* suppression has been associated with sensory processing (Schürmann & Başar, 2001), as well as memory (i.e., recalling items from a set) or changes in alertness or attention (Klimesch, 1996). Increases in alpha have been associated with functional inhibition (Jensen & Mazaheri, 2010).

*Beta oscillations (12.5 – 30 Hz)* have been observed following motor interactions (Neuper & Pfurtscheller, 2001), and have also been hypothesized to play a role in predictability; if the same task or motor movement is expected to repeat, there is an increase in beta activity, compared to when change is expected (Engel & Fries, 2010).

*Gamma oscillations (30 Hz and above, though often functionally observed at 40 Hz)* have been attributed to cortical activation in general, and thus may be present during any cognitive task (Merker, 2013). However, gamma has been tied to more specific functions when location of oscillatory activity is taken into account. For example, when shown objects that the subject has a long-term memory representation for, there is significantly more gamma over the occipital cortex than for objects the subject had never seen before (Herrmann, Munk, & Engel, 2004). Gamma
has also been implicated in integrating information across different sensory modalities (Schneider et al., 2008), the focus of the current study. We elaborate on this below.

The Role of Neuronal Synchrony in Sensory Integration:

A few studies have looked at the role of frequency bands in integrating sensorimotor stimuli. For instance, when presented with a visual object (e.g., a picture of a sheep) followed by a congruent auditory sound (the sound of a sheep vocalizing what would be described as a baa sound), subjects showed increases in gamma-band activity in left middle temporal gyrus relative to when the visual and auditory stimuli were incongruent (Schneider et al., 2008). This finding suggests that an increase in gamma in left middle temporal gyrus may reflect integration of information across visual and auditory modalities. Previous fMRI work supports this idea: posterior temporal sulcus and middle temporal gyrus show increased responses when visual and auditory features of objects are shown together rather than individually (Beauchamp, Lee, Argall, & Martin, 2004).

In related work, when participants judged whether two words referring to “different modality” features were consistent with a concept (e.g., judging whether the visual and auditory features silver and loud are consistent with whistle) there was a greater increase in theta power in anterior temporal lobe (ATL) compared to when integration demands were arguably lower because the two features were from the same modality (e.g., silver and shiny, which are both visual; van Ackeren & Rueschemeyer, 2014).

Note that Schneider et al. (2008) did not examine theta-band activity, and van Ackeren & Rueschemeyer (2014) did not examine gamma-band activity. If they had, it is possible that both studies may have found similar results. This is because some theories propose that local interactions between neuronal assemblies produce gamma oscillations, and long-range cortical
interactions produce lower-frequency oscillations, including theta (Donner & Siegel, 2011; for support, see Mellem et al., 2013). If true, it seems reasonable that both Schneider et al. (2008) and van Ackeren & Rueschemeyer (2014) would have seen increases in both gamma and theta activity, had they each looked at both frequencies.

In fact, a study by van Ackeren, Schneider, Müsch, & Rueschemeyer (2014) provides further evidence for the local-gamma distal-theta theory: using the same feature word pair paradigm that van Ackeren & Rueschemeyer (2014) employed (where two features from the same modality or two features from two modalities were presented in accordance with a target word), van Ackeren et al., found more gamma power for same modality pairs, and more theta for crossmodality pairs. This finding is consistent with the local-gamma, distal-theta theory in that both types of feature word pairings are related to a target word and thus require integration, but the pairs from the same modality would theoretically activate groups of neurons that are closer together (thus increasing gamma activity), whereas the crossmodality pairs would activate groups of neurons further apart (thus increasing theta activity).

A related possibility for why van Ackeren & Rueschemeyer (2014) may have observed more theta activity and Schneider et al. observed more gamma activity is that van Ackeren & Rueschemeyer showed words compared to Schneider et al.’s images and sounds, thus adding a lexical component. Incorporating frontal-temporal areas supporting lexical processing into the integration process increases cortical distance, so van Ackeren & Rueschemeyer (2014) may have seen either larger or more sustained increases in theta-activity because their stimuli were lexical, compared to Schneider et al.’s (2008) experiment. Another, compatible possibility is that lexical stimuli draw more heavily on working memory, which has also been associated with increases in theta-band activity (Bastiaansen et al., 2002).
To summarize, although some current theories of semantic knowledge organization do theorize about binding, there is little empirical data in the semantic memory literature directly testing the neural mechanisms that may support binding of sensory information. Communication between assemblies of neurons seems to be a plausible mechanism for binding, and recent work on neural synchrony for auditory and visual integration shows promising results.

*Current Study:*

In the current study, we will both attempt to replicate the gamma-band increase observed by Schneider et al. (2008), and also test whether, in a minimally different experimental paradigm, incorporating lexical information into the integration process increases theta-band activity. In the *image plus non-lexical sound* conditions, following Schneider et al. (2008), we will present a visual object (e.g., a picture of a sheep) followed by a congruent or incongruent auditory sound (e.g., a recording of a sheep or a cow sound, respectively). In the *image plus lexical sound* conditions, we will present a visual object (e.g., a picture of a sheep) followed by a congruent or incongruent word that is minimally different from the aforementioned non-lexical sounds (e.g., the word “baa” or the word “moo”). We predict that in both the non-lexical and lexical conditions there will be a “congruency effect” in gamma whereby gamma-band activity will be higher for congruent than incongruent pairs, reflecting integration via neural synchrony. We further predict that in the lexical, but not the non-lexical conditions, we will observe a larger congruency effect in theta-band activity, reflecting either that integrating lexical information involves longer distance cortical interactions or increased working memory demands. Note that in this study, we use the term “integration” to refer to the process of combining different sensory information into a single concept (i.e., the process that allows you to perceive the shape of a cat and the sound of its meow as parts of a coherent whole).
Our primary predictions (described above) are about oscillatory activity. However, in order to situate our study in the context of more traditional EEG analyses, which analyze event-related potentials (ERPs) we analyzed ERPs, predicting that, as observed by Schneider et al., (2008), and following the well-established literature on the N400 (e.g., Kutas & Hillyard, 1980), we would see an N400 component for incongruent compared to congruent stimuli pairs.

Methods

Participants: 29 subjects (12 female, mean age = 18.9, range 18-22) were collected from the University of Connecticut introductory psychology classes and compensated with course credit. Subjects were all right-handed, native English speakers with normal hearing, normal or corrected-to-normal vision, and had no history of neurological or psychiatric illness. This study was approved by the Institutional Review Board at the University of Connecticut, and all subjects gave informed consent prior to data collection. One subject was removed due to misunderstanding the task, and 9 subjects were removed based on noise/artifacts in EEG data, resulting in 19 subjects (7 female, mean age = 18.8, range 18-21) included in the analyses.

Stimuli: One hundred visual and one hundred auditory stimuli were obtained from a pool of color pictures and sounds previously rated in a norming study (Schneider, Engel, & Debener, 2008, who tested German participants). Fifteen pictures from this pool were replaced (with pictures obtained from online image databases) because the original images were judged (by H.M.) to be unfamiliar to local participants (e.g., an image of a German-style blue van for “police car” was replaced with a more USA-typical white and black car). Lexical auditory stimuli were recorded by the first author, with sound intensities equalized at a root mean square power across all sounds and intensity was set at 60 dB. All sound files were 400ms long, and to avoid clipping at onset and offset, sound files were windowed with a 10ms rise and fall time.
Norming data for this stimulus set was collected from 21 subjects (15 female, mean age = 18.8, range 18-20) from the University of Connecticut psychology department participant pool who were compensated with course credit. Subjects were all right-handed, native English speakers with normal hearing, normal or corrected-to-normal vision. This study was approved by the Institutional Review Board at the University of Connecticut, and all subjects gave informed consent prior to data collection. Norming was conducted to determine whether participants judged picture-sound pairs that were intended to be congruent as congruent, and whether picture-sound pairs that were intended to be incongruent were judged as incongruent. To this end, during norming, congruent and incongruent stimulus pairs were presented to each subject, and they were asked “Does the picture match the sound?” For lexical pairs that were intended to be congruent, the mean “match” judgment (i.e., the pair was intended to be congruent and the subject responded that the pair matched) was 64.1% ($SD = 48\%$), and for non-lexical pairs that were intended to be congruent, the mean “match” judgment was 76.4% ($SD = 42.5\%$). For lexical pairs that were intended to be incongruent, the mean “match” judgment (i.e., the pair was intended to be incongruent and the subject responded that the pair actually matched) was 6.5% ($SD = 24.7\%$), and for non-lexical pairs that were intended to be incongruent, the mean match” judgement was 6.2% ($SD = 24.2\%$).\(^1\)

Thus, norming revealed that only about two-thirds (for lexical pairs) to three-quarters (for non-lexical pairs) of the pairs that were intended to be congruent were judged by norming participants to be congruent. Despite this, because our study was designed as a partial replication

\(^1\) Because picture stimuli were presented 4 times to each subject (in congruent lexical, congruent non-lexical, incongruent lexical, incongruent non-lexical trials) and all auditory stimuli appeared 2 times (either in a congruent or incongruent pairing), after the experiment was over, each subject was asked if repeated exposure to the same sound/picture aided in predicting if the trial would be a matching or non-matching pair. Only 1 subject reported attempting to make predictions. Removing this participant from the data analysis produced minimal change in the results.
of Schneider et al., (2008) and most of the stimuli pairs in our non-lexical condition were drawn from the 85 pairs used in that study (with the exception of 15 pairs, which, as described above, were not culturally appropriate), we opted to retain all 100 pairs in each condition. Our reasoning was that doing so would most closely replicate Schneider et al., while still providing the flexibility, in our data analysis, to examine the subset of pairs for which there was behavioral evidence of the intended congruency in agreement with participants’ perception.

Procedure: As in Schneider et al. (2008), each trial began with a 500ms fixation, followed by a visual stimulus followed by either a or non-lexical auditory stimulus or a lexical auditory stimulus (the latter condition was not included in Schneider et al., 2008) using Experiment Builder software (SR Research, Ontario, Canada). We blocked and counterbalanced across subjects whether the visual stimulus was followed by a lexical or non-lexical stimulus. Both auditory stimuli were presented for 400ms, with an inter-stimulus interval of 1000ms. Response intervals began at the offset of the lexical/non-lexical sound (see Fig 2.). If the subject did not respond within 2000ms, a message appeared reminding them to respond faster. There were 400 trials, with 200 in the lexical sound condition and 200 in the non-lexical sound condition. Each condition contained 100 semantically congruent (e.g., a picture of a lion and a voice saying “roar”/the sound of a lion roaring) and 100 semantically incongruent (e.g., a picture of a lion and a voice saying “chirp”/the sound of a bird chirping) pairs.

Each visual stimulus was paired with a corresponding congruent lexical stimulus and non-lexical stimulus (e.g., a picture of a lion would appear with both a “roar” and a roaring sound), and each auditory and visual stimulus was used in both incongruent and congruent conditions (i.e., the same pictures and sounds were used in the incongruent trials by pairing them with incongruent sounds/pictures, e.g., a picture of a lion could be followed by a “chirp” or a
chirping sound, and the “roar” or the lion roaring could appear following an image of a turtle), resulting in each visual stimulus repeating 4 times, and each lexical/non-lexical stimulus repeating twice.

Trials were presented in 4 blocks of 100 trials. Lexicality of the auditory stimuli was blocked (i.e., each block contained either lexical or non-lexical sounds). Half of the trials within each block were congruent and half were incongruent. Block order was counterbalanced, separated into blocks of 100 visual stimulus/lexical stimulus congruent and incongruent trials, and 100 visual stimulus/non-lexical stimulus congruent and incongruent trials. Each block lasted about 15 minutes. As in Schneider et al. (2008), trials were pseudo-randomized to avoid stimuli from the same object-category being presented sequentially (e.g., a trial with either a picture or sound of a lion was not played immediately after a trial with either a picture or sound of a dog, as both of these objects are in the “animal” category).

![Stimuli presentation schematic](image)

Figure 2. Stimuli presentation schematic

Subjects were asked to answer “Does the object that makes the sound fit into a shoebox?” (as in Schneider et al., 2008) by pressing one of two buttons with either their left or right thumb (the side of the “yes” button was counterbalanced across subjects). Whether or not the size of the auditory target would allow it to fit into a shoebox was coded by H. M. Although the
participant’s task was only to respond to the sound, we were concerned that if, e.g., the object in the image did not fit in a shoebox, but the object that made the sound did (or vice-versa), this might produce some degree of implicit conflict, making incongruent trials, on average, harder than congruent trials (in which, by definition, both items would have the same response) for reasons other than the congruency we were interested in. This concern was salient because in Schneider et al., 48% of incongruent trials had a visual stimulus that was response-incompatible with the auditory stimulus (i.e., the visual stimulus did fit in a shoebox when the auditory stimulus did not, or vice versa), and within incongruent trials, response-incompatible trials had a much higher error rate (mean = 60.3%) than response-compatible trials (mean = 30.2%). This difference made us suspect that some of the difference between the congruent and incongruent conditions in Schneider et al. may have been to the more frequent response-incompatibility in the incongruent condition rather than to congruency differences.

To avoid this potential confound, stimuli were designed such that regardless of whether the visual stimulus and lexical/non-lexical stimulus were congruent or incongruent, both stimuli would produce the same task response. In theory, this feature of the design could have allowed participants to respond correctly, by only responding to the image, and ignoring the sound. However, none of our subjects reported noticing that the picture always depicted an object that would elicit the same task response as the sound, nor did any participant report responding to only the image as a strategy. Further, our behavioral results show highly significant priming effects ($p < .001$) which would not be expected if participants were only responding to the picture, as in that case, the congruency of the sound would not be relevant.

**EEG Preprocessing:** Continuous EEG data was collected from Phillips EGI 256 channel cap. Data was collected with a sampling rate of 500Hz, impedances were kept below 50 Ohms.
Analysis was completed in Matlab using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Continuous data was filtered between 0.1 Hz and 100 Hz, with a notch filter at 59-61 Hz, and was also demeaned and detrended. Bad channel identification and repair along with independent component analysis was run with Matlab functions based on the Fully Automated Statistical Thresholding for EEG artifact Rejection (FASTER) method (Nolan, Whelan, & Reilly, 2010), along with visual inspection. Subjects with for whom more than 25% (64 out of 256) of channels were identified as bad were removed from the dataset. Trials that contained artifacts above 100 µV or below -100 µV were removed. This resulted in a final dataset of 19 subjects (7 female, mean age = 18.8, range 18-21), where on average, 18% of channels were removed and repaired with spine interpolation (range: 9% - 23%), and 12% of trials were removed due to artifacts (range: 2% - 31%). This resulted in an average of 87 trials in the congruent word condition (range: 55-98), 87 in the incongruent word condition (range 52-98), 88 in the congruent non-lexical condition (range 71-98), and 88 in the incongruent non-lexical condition (range 73-98).

**Results**

*Behavioral:*

RTs from all 19 subjects were winsorized to replace the top and bottom 5%. RTs for all subjects and all conditions ranged from 215 ms to 1391 ms (mean: 662.7 ms, median: 602 ms; see Figure 2). Dependent-samples t-test were run in R (R Core Team, 2019) using the *t.test* function in the “stats” package. Comparing the RTs between congruent ($M = 701, SD = 175$) and incongruent ($M = 756, SD = 175$) lexical trials revealed that congruent lexical trials were responded to faster than incongruent lexical trials, a difference that was significant both by subjects, $t(18) = -5.05, p < .001$, and items, $t(99) = -3.37, p = .001$. Congruent non-lexical trials
\[ M = 615, SD = 116 \] were also responded to faster than incongruent \[ M = 679, SD = 140 \] non-lexical trials. This difference was also significant both by subjects \[ t(18) = -5.15, p < .001 \], and items \[ t(99) = -3.58, p < .001 \].

![Figure 3. Response times by condition, with 95% confidence intervals as the error bars.](image)

We also compared accuracy between conditions. Congruent lexical trials \( M = 0.71, SD = 0.08 \) were responded to more accurately than incongruent lexical trials \( M = 0.66, SD = 0.07 \), a difference that was significant by subjects \( t(18) = 2.19, p = .04 \), but not items \( t(99) = 1.36, p = .18 \). Similarly, congruent non-lexical trials \( M = 0.81, SD = 0.05 \) were responded to more accurately than incongruent non-lexical trials \( M = 0.77, SD = 0.04 \), and this difference was also significant across subjects, \( t(18) = 5.66, p < .001 \), but not items \( t(99) = 1.55, p = .13 \). We suspect the lack of significant differences in accuracy across items is due to the fact that for some of the audio stimuli, the size of the object that produced them was quite obvious, meaning that there was not much benefit of congruency; for example, when hearing *ribbit*, regardless of whether it was presented after a congruent or incongruent visual stimulus, it may be more clear to subjects
that the audio target pertains to a frog which would fit into a shoebox. And indeed, in both congruent and incongruent conditions, there were a number of items where the average accuracy score was 100%.

Figure 4. Percent accuracy by condition on the task “does the object that makes the sound fit into a shoebox?” 95% confidence intervals used for error bars.

**Event-related Potentials:**

For ERP analysis, we constructed epochs beginning 200ms before the auditory stimulus onset and ending 1000ms post-onset. We used the period of -200 to onset of the auditory stimulus as the baseline for each trial. We then averaged all trials across subjects for each condition.

Cluster-based permutations were performed to conduct pairwise t-tests to examine differences between congruent and incongruent non-lexical trials, and differences between congruent and incongruent lexical trials, from a period between 300-600ms post-onset of the auditory stimulus. Results indicated a significant difference between congruent and incongruent non-lexical trials ($p < .001$), and congruent and incongruent lexical trials ($p < .001$). In order to
visualize the differences, waveforms were created using the average signal from electrodes included in the significant cluster (see Figures 5 & 6). Our waveforms show that the difference between incongruent and congruent trials extends beyond the N400 period for both lexical and non-lexical conditions. This pattern is very similar to Schneider et al.’s (2008) waveforms.

![Figure 5. Average of electrodes found to be significant in cluster-based permutation comparing congruent and incongruent lexical trials. Blue line indicates incongruent, red is congruent, with microvolts on the y-axis ascending negative to positive.](image)
Figure 6. Average of electrodes found to be significant in cluster-based permutation comparing congruent and incongruent non-lexical trials. Blue line indicates incongruent, red is congruent, with microvolts on the y-axis ascending negative to positive.

Time Frequency Analysis:

For time frequency analysis, total power was computed separately for gamma and theta bands.

Gamma

Spectral power increases in gamma (30-100 Hz) were analyzed using multitapers rather than the Morlet wavelets used by Schneider et al. (2008). We hoped this would improve upon Schneider et al.’s analyses, because although Morlet wavelets have higher temporal precision, which is useful for lower frequencies (i.e., <30 Hz), multitapers provide a boost in signal-to-noise ratio, which is useful for higher frequencies (i.e., >30Hz) as they are more affected by noise.
Multitapers were generated using sliding time windows of 200ms multiplied with three orthogonal Slepian tapers. These time windows started at -1000ms before onset of the auditory stimulus, and moved in steps of 10ms until 1150ms after auditory stimulus onset. Multitapers were applied to frequencies starting at 30 Hz and ending at 100 Hz in steps of 2. Resulting power spectra for each frequency were averaged over tapers, resulting in a frequency smoothing of ±5 Hz. All trials for each condition were then averaged for each subject. A period of -300 to -100ms before onset of the auditory stimulus was used as a baseline correction to calculate the relative signal change. Power differences between conditions were calculated using cluster-based randomization, which controls for multiple comparisons by making clusters of neighboring samples in time, frequency, and space domains (Maris & Oostenveld, 2007).

For statistical evaluation of gamma power differences between conditions, we used cluster-based permutations to run pairwise t-tests comparing congruent and incongruent non-lexical trials, and to compare congruent to incongruent lexical trials, at latency 100-300ms, and from 30-50 Hz. Results indicated no significant differences between the congruent and incongruent conditions, for either non-lexical trials (positive cluster, lowest p-value = .4; negative cluster, lowest p-value = .6) or lexical trials (positive cluster, lowest p-value = .6; negative cluster, lowest p-value = .3) trials.

**Theta**

Unlike gamma, theta (4-8 Hz) power was analyzed using Morlet wavelets, because while multitapers improve signal-to-noise ratio, they also result in more temporal and frequency smoothing than Morlet wavelets, thus making it difficult to isolate temporal events for lower frequencies (Cohen, 2014, p. 207). Morlet wavelets were constructed with a fixed width of 3 cycles (e.g., a wavelet for 4 Hz would need to be 750ms long to include 3 cycles at that
frequency, 8 Hz would be 375ms) for frequencies 4 to 8 Hz in steps of 2. These wavelets were applied in sliding time windows starting at -1000ms before onset of auditory stimulus, and moving in steps of 10ms to end at 1150ms after onset. All trials for each condition were then averaged for each subject. A period of -750 to -250 before onset of auditory stimulus was used as a baseline correction to calculate the relative signal change. Like gamma, power differences between conditions were calculated using cluster-based randomization (Maris & Oostenveld, 2007).

For statistical evaluation of theta power differences between conditions, cluster-based permutations were computed for pairwise t-tests comparing congruent to incongruent non-lexical trials, and congruent to incongruent lexical trials, at latency 500-800ms, from 4-8 Hz. No significant difference was found between congruent and incongruent conditions for either non-lexical trials (positive cluster, lowest p-value= .7, negative cluster, lowest p-value = .4) or lexical trials (positive cluster, lowest p-value = .7, negative cluster, lowest p-value = .3).

**Good primes only:**

As described in the methods, the stimuli pairs from this study were normed in a match/mismatch task. As also described above, norming revealed that only about two-thirds (for lexical pairs) to three-quarters (for non-lexical pairs) of the pairs that were intended to be congruent were judged by norming participants to be congruent. This was not surprising, as a 400ms clip of the sound of a tennis racket hitting a tennis ball does not provide much auditory information and therefore might not be perceived as much more congruent with an image of a tennis racket than a 400 ms clip of a basketball bouncing Likewise, hearing the word “smack” might not be perceived as much more congruent with a tennis racket than the word “bounce.”

However, because the goal of this study was to examine the role of different frequency
bands in integration, the most appropriate test items are those for which “congruent” sounds are more likely to be integrated with their paired images than the incongruent sounds for the participants we tested. We therefore inspected the by-item behavioral priming effects with the idea a behavioral priming effect could be interpreted as an index of integration. Following winsorization (by item, at 10%), paired t-tests were performed on the 19 RTs (from the 19 participants) obtained for each item pair, comparing congruent versus incongruent conditions. Using a t-value of 1 ($p < 0.35$) as our (admittedly rather arbitrary) cutoff, meant that we retained 53 visual stimuli/non-lexical item pairs and 47 visual stimuli/lexical item pairs\(^2\).

Time frequency analysis on this subset of items followed the same procedure as before. Cluster-based permutation pairwise t-tests compared congruent and incongruent lexical trials, and congruent and incongruent non-lexical trials, from 500-800ms, at frequencies 4-8 Hz. Results revealed a significant difference between congruent and incongruent lexical trials (negative cluster, $p = 0.008$), indicating more theta power in *incongruent* compared to the congruent lexical trials (Fig. 7). A similar, albeit non-significant, pattern was found for the comparison of congruent and incongruent non-lexical trials (negative cluster, $p = 0.10$).

\(^2\) Because of the exclusion of some individual participants’ trials due to EEG artifacts and noise, this meant that on average, across participants, we retained: 42 trials (range: 32-47) in the congruent lexical condition, 41 trials (range: 31-51) in the incongruent lexical condition, 45 trials (range: 31-51) in the congruent non-lexical condition, and 47 trials (range: 40-52) in the incongruent non-lexical condition. For reference, Schneider et al. (2008), who did not limit analyses to items with a behavioral priming effect, had on average 64 trials (range 27-82) in the congruent condition and 63 trials (range 27-82) in the incongruent condition.
Figure 7. Topographical maps of congruent – incongruent lexical trials. Colorbar represents total oscillatory activity expressed as % change relative to baseline. Asterisks highlight channels within the significant cluster.
Discussion

Our behavioral results indicated that congruent pairs are responded to faster and more accurately than incongruent pairs. Our ERP results also showed an N400 effect for incongruent compared to congruent trials in both lexical and non-lexical trials. Both our behavioral and ERP findings are similar to results from Schneider et al. (2008), leading us to believe that we did succeed in creating a stimulus set and experimental paradigm that was sensitive to congruency effects.

Contrary to our hypothesis that gamma plays a role in integration, we found no increase for congruent pairs compared to incongruent pairs. Also contrary to our predictions, there were no increases in theta power for congruent compared to incongruent word trials. However, in analyses that included only item pairs for which priming was obtained in behavioral data (suggesting that more integration occurred in congruent than incongruent conditions), we observed increased theta for incongruent compared to congruent lexical trials. There was also a trend for increased theta in incongruent compared to congruent non-lexical trials. This theta increase for incongruent trials was unexpected. However, phase and amplitude resetting (which creates synchrony/power increases) in theta activity in response to seeing new words vs. repeated words (a contrast that is analogous to congruent vs. incongruent) has been suggested to contribute to an N400 effect recorded from the hippocampus (Mormann et al., 2005), and semantic violations have been shown to produce increases in theta (in addition to producing the classic N400 effect) compared to semantically correct sentences (Hald, Bastiaansen, & Hagoort, 2006). Taken together, these results suggest that one reason why we saw increases in incongruent compared to congruent conditions could be because increases in theta power may be tied to the N400 effect.
However, this still leaves the question as to why Schneider et al. (2008) found a significant increase in gamma for congruent compared to incongruent non-lexical pairs, while we did not. One reason could be that differences in analysis methods made Schneider and colleagues’ analyses more sensitive to differences in gamma. For instance, they used morlet wavelets to estimate power in gamma band, whereas we used multitapers. As described above, multitapers do have the advantage of boosting signal-to-noise. However, it could be that because multitapers have less temporal precision than Morlet wavelets, we were unable to detect a difference in gamma between congruent and incongruent pairs; it is possible that smoothing over the temporal domain may have hidden very quick periods of change between conditions. Additionally, Schneider et al. (2008) used \textit{a-priori} defined ROIs whereas we used a neighborhood triangulation method to find clusters of electrodes.

To examine whether these differences in analysis were the reason for our differing results, we re-analyzed our data using methods that matched more closely with those described in Schneider et al. (2008), i.e., using Morlet wavelets and \textit{a-priori} defined ROIs similar to theirs. We still observed no differences in gamma in the congruent vs. incongruent conditions.

What else then might be the reason for our null results in gamma? Some researchers suggest that gamma is simply a by-product of neural activity in general, and will thus show increases as difficulty increases in any cognitive task (Merker, 2013). Although our task was designed to be (implicitly) about congruency, the explicit task asked subjects to determine if the auditory stimulus can fit in a shoebox. It is therefore possible that determining the size of the object that makes a given sound may be of equal difficulty in both conditions. However, one would assume that there should still be a slight advantage in congruent conditions, as the visual stimulus should help identify the auditory stimulus, and the large behavioral congruency effects
we observed suggest that the task was indeed easier in the congruent condition. Our ERP finding of a significantly larger N400 in incongruent conditions compared to congruent conditions also supports the idea that task was easier in the congruent condition. Furthermore, our task was identical (withstanding a few stimuli) to that of Schneider et al., (2008) who did observe a difference between congruity conditions.

However, as described above, unlike Schneider et al., we controlled for response-compatibility of our congruent and incongruent pairs (that is, in all of our pairs if the visual stimulus fit in in the shoebox, the auditory stimulus did too, or vice versa), whereas almost half of Schneider et al.’s incongruent pairs were response-incompatible. Thus, what Schneider et al intended to be a congruency manipulation was confounded with response-compatibility, and it could therefore be that differences in response compatibility were what elicited the differences in gamma in that study.

Concerning our unexpected results in theta, because there are 5 types of frequency bands and many more than 5 cognitive processes, it is entirely possible that theta does play a role in binding, but its contribution is masked here by the incongruency detection processes that produce the N400 (Mormann et al., 2005). One could see this as analogous to a busy highway en route to Boston: many of the cars heading in towards Boston are commuters going to work, with only a handful going to Boston to shop/other activities. However, when one is driving to Boston, one tends to believe rush hour traffic is due to people commuting to work, not drivers with other plans. In the same way, as N400 is a fairly robust effect when it comes to mismatch/incongruency detection, it may be that the theta signal from that detection process is much stronger/more prevalent in comparison to the theta signal produced during binding (which one could also argue takes less effort for concepts that are already known, as the accuracy and
RTs suggest), and therefore theta appears weaker in binding contexts compared to mismatch detection contexts. Therefore, it may be useful for future analyses to subtract theta power related to the N400, and analyze the remaining theta power in order to assess the role theta may play in binding separate from incongruency detection; in other words, look at the evoked and induced theta activity separately rather than total oscillatory theta activity (David et al., 2005).

However, theta has also been implicated in inhibitory processes (Takahashi, Shinomiya, Morl, & Tachibana, 1997; see Mitchell et al., 2008 for review), so it is also possible that we would still see increases in theta for incongruent conditions compared to congruent conditions, because one might want to inhibit the visual cue in the incongruent condition to make the size judgment about the sound. Yet if theta does support inhibition and/or the N400, then why did van Ackeren and Rueschemeyer (2014) observe increases in theta linked to integration? Although their task did include incongruent pairs, they were not looking at the difference between congruent pairs and incongruent pairs. Instead, they compared visually-presented word pairs either in a single modality (i.e., silver and shiny for target whistle) or across modalities (i.e., silver and loud for target whistle), and assumed that more integration is required for cross-modality pairs. Thus, they did not compare responses to congruent and incongruent pairs as an index of integration, which may have made their design less susceptible to interference from theta produced by a secondary source, such as the N400 or inhibitory processes.

Thus, our findings do not provide evidence that gamma and theta are involved in “binding,” when binding is operationalized as integrating a sound with a preceding image. However, this does not mean that there is no functional significance of theta or gamma in binding, whether separately or together. As noted earlier, there are many ways to look at synchrony. For this initial study, we chose to look simply at increases in power within particular
frequencies to assess synchrony, as this is a commonly reported synchrony measure, and closely mirrored Schneider et al. (2008), which our study was based on. But in fact, recordings from the CA3 region in rat hippocampus has demonstrated cross-frequency coupling (e.g., the phase of one frequency modulates the activity/phase of another) between theta and gamma which increased with training and performance accuracy (Tort, Komorowski, Manns, Kopell, & Eichenbaum, 2009). We did not assess cross-frequency coupling in this study, rather, we looked at separate latencies for gamma (100-300ms) and theta (500-800ms). Thus, analyses of cross-frequency coupling across the epoch would be fruitful as it could reveal a role for theta and gamma in binding. Future plans involve applying cross-frequency coupling analyses in this study/similar ones.

We now turn to the unexpected increase in theta power that we observed for incongruent compared to congruent words and sound pairs. Some researchers theorize that theta increases may be due to working memory load (Klimesch, 1999). For example, one study showed an increase in frequencies from 6-10 Hz starting at 300ms and continuing until at least 1500ms when subjects were completing a 2-back task as opposed to a 0-back or 1-back task (Krause et al., 2000). Although 6-10 Hz encompasses both theta and alpha, the authors did examine results from 6-8Hz separately from results from 8-10Hz, and observed that the increase remained in the theta band (6-8Hz in this case). If theta is, in fact, indicative of working memory, the reason that we observed more theta for incongruent compared to congruent trials could be due to the increased effort of deciding which object/animal could have made the noise, as they do not have the aid of a congruent visual prime.

Future Directions:

As noted above (see also the Appendix), power increases are only one way to observe
neuronal synchrony. In order to assess more concretely whether gamma and theta do or do not play a role in binding, it may be useful to look at time frequency methods which are more specific than increases in power. Another method is coherence, a fixed pattern of phase-locking signals between two or more oscillators/neuronal groups (Fries, 2005). The idea put forward by Fries (2005) is that neuronal groups can only communicate effectively if there is coherence, because coherence would cause the windows for input and output to be open at the same time for all groups. One could also use cross-frequency coupling, where signals in different frequencies modulate each other, and this modulation could be in phase or amplitude and take on excitatory or inhibitory functions (Roux & Uhlhaas, 2014). In relation to the current study, this could take place via short-range increases in gamma contributing to long-range increases in theta. An improvement on this study would be to examine the oscillatory activity using these methods, with specific hypotheses about which oscillators may be communicating with one another, or which frequency bands may modulate one another. For example, if one looks at the same epoch for both gamma and theta, one may find that increases in gamma activity in more basic sensory processing areas contribute to increases in theta activity in association areas, or parts of the brain that seem to play a role in integrating information from different sensory input.

One last notable open question is whether our study could be improved with a different procedure. The priming paradigm used in Schneider et al. (2008) has benefits compared to a simultaneous paradigm because, by minimizing the amount of basic processing (i.e., initial sensory processing of the visual stimuli) occurring in the target epoch, it reduces the possibility of this basic processing overshadowing the integrative processes that are of interest. However, one could also consider that this “basic processing” may actually be more indicative of integration. More concretely, visual processing occurs on the timescale of milliseconds, and if
one considers integration to be sensory information being bound together, or sensory input affecting perception of another sensory input, the oscillatory activity from visual processing may have returned back to baseline in the 1000ms period between the offset of visual stimulus and onset of auditory stimulus. Therefore, the oscillatory activity produced by visual processing and auditory processing may not overlap in time, thus not allowing one to observe influences of visual input on the auditory input using our particular methodology. In fact, in a study that showed pictures of animals and either congruent or incongruent sounds at the same time, there was increased induced (i.e., not phase-locked to stimulus [by comparison, our analysis examined total power, which includes both evoked/phase-locked and induced/non-phase-locked activity]) gamma-band activity around 260ms post-onset of the stimulus for congruent pairs compared to incongruent pairs (Yuval-Greenberg & Deouell, 2007). Induced activity is thought to be due to higher-order changes in the brain, and may be linked to feature-linking in visual cortex (Başar-Eroğlu, Strüber, Schürmann, Stadler, & Başar, 1996).

In summary, although previous studies suggest a role for gamma-band and theta-band activity in binding information across sensory modalities, we did not find increases in gamma or theta activity for congruent compared to incongruent visual/non-lexical or visual/lexical pairings. While these null results are compatible with there being no role for oscillatory activity in the gamma or theta ranges during the type of integration elicited by our paradigm, there are many different ways to operationalize integration, and many ways of analyzing time-frequency data. Looking at increases of power in certain frequencies is a first step, but analyses that are more specific to certain regions communicating to other regions seems to be a logical way to progress in this investigation. Further, different experimental paradigms may more directly target the influence of oscillatory activity from one sensory input to another sensory input, rather than
integration that may be influenced by recall of an already integrated and stored concept.

Thus, much work remains to be done to explore whether gamma and/or theta band activity supports integration, especially if we wish to understand the neural mechanisms behind binding. From an applied perspective, understanding these mechanisms more fully may lead to advances in techniques to improve reading (which involves binding visual orthography with auditory phonemes) or aid in interventions for populations which present atypical integration processes (e.g., Autism Spectrum Disorder: Iarocci & McDonald, 2006; Schizophrenia: de Gelder, Vroomen, Annen, Masthof, & Hodiamont, 2003). From a fundamental science perspective, understanding how the brain binds information together would aid in improving theories surrounding semantic memory organization and distribution, theories surrounding how episodic memory is preserved, and theories on learning.
Appendix: Types of Synchrony

Broadly, neural synchrony means two (or more) neurons firing at the same time or within a short time interval. One can measure synchrony of multiple neurons by looking at the amplitude of firing. If an input (i.e., stimulus) causes a change in neuronal activity (i.e., processing the stimulus), one should see changes in amplitude compared to no-stimulus or another appropriate control (Buzsáki, 2006). To make an analogy to fMRI, this is similar to seeing increased BOLD response to a stimulus compared to a baseline period.

However, synchrony can mean more than a group of neurons firing close enough in time to produce changes in amplitude that we can see via EEG. There are other forms of synchrony, such as coherence, where instead of one group of neurons, there are two or more groups that have a fixed phased relationship with one another; group A may show increased amplitude when group B shows decreased amplitude, or they may show increases/decreases at the same time, but the relationship needs to be fixed (Buzsáki, 2006). Similarly, phase-locking is another type of synchrony very much like coherence, where there must be a fixed relationship between two or more neuronal groups, but where amplitude does not matter. Rather, the phase of the groups should be “locked” together; if one thinks of two sine waves, the wave produced by group A should be at 90 degrees when the wave from group B is at 90 degrees.

Importantly, phase-locked (sometimes called “evoked” activity) and non-phase-locked activity (sometimes called “induced” activity) should not be confused with time-locked activity. Whether activity is phase-locked or non-phase-locked, it can still be time-locked, in that the onset of a stimulus can produce both types of activity. The difference is that phase-locked activity occurs when across trials, the activity at time point 0 is in the same phase. Both types of activity are produced in response to a stimulus, but non-phase-locked activity is often lost in ERP
averaging (see Fig 1 for visual aid; Cohen, 2014, p. 20, 55). Often, much of the research on neural entrainment focuses on phase-locking synchrony.

Further, synchrony can refer to cross-frequency phase synchrony, where two oscillators/neuronal groups have a fixed phase relationship, but it occurs in two or more frequencies (i.e., neuronal group A may be producing gamma frequencies that have a fixed phase relationship with beta frequencies produced by group B). A system can also have phase modulation of power, where the faster rhythm (i.e., gamma) of neuronal group A can have power fluctuations that vary according to a slower rhythm (i.e., theta) produced by neuronal group B (Buzsáki, 2006).

There are only a few examples of different types of synchrony, but they show the wide variation in use of this term. For my purposes, I will be discussing neural synchrony in what I view is the simplest form: a group or groups of neurons working synchronously enough to cause observable changes in power in EEG sensor space. This paper will not be covering source localization, therefore I will not be making any claims about which areas of the brain are producing which signals, which is why we are unable to use any of the more specific types of measuring synchrony. Rather I will only be providing information of where these power fluctuations occur in sensor space, or the regions of the brain we assume the EEG electrodes cover.
Figure 1. Modified image from Cohen, 2014, p. 20. The left column shows simulated raw data, which includes both phase-locked and non-phase-locked activity, which are all time-locked. The second row only includes phase-locked (and time-locked) activity, which survives ERP averaging, while non-phase-locked activity does not. In the left column, one can see that the dips and valleys are slightly different at time 0, but in the right column, time 0 is almost always at a valley which is starting to rise.
References


semantic violations in online sentence processing. Brain and Language, 96, 90-105.


