THE UNIVERSITY OF CHICAGO

TEMPORAL PREDICTIONS IN PERCEPTION AND LANGUAGE COMPREHENSION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY
GEOFFREY THOMAS BROOKSHIRE

CHICAGO, ILLINOIS
JUNE 2018

For my parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful for the help I've received from a large number of people. To my family and friends, thank you for your support. To my labmates, thank you for your friendship and help working through ideas. To my advisors, thank you for teaching me how to be a scientist. To Daniel, thank you for the years of excitement and encouragement, and for the superlative scientific training.

ABSTRACT

During perception, information is not uniformly distributed over time; some time-points are more informative than others. How does the brain get ready to perceive upcoming bursts of information? When people listen to quasi-periodic stimuli such as speech or music, electrophysiological activity in cerebral cortex synchronizes with pulses in the stimulus. Cortical synchronization helps to align information in the stimulus with periods of maximal neural excitability. However, it remains unknown how the brain uses temporal structure in the stimulus to drive cortical synchronization. In Study 1, I test whether cortical synchronization depends on characteristics of sensory cortex, or on informational characteristics of the stimulus. I develop a metric to quantify visual information over time, and use electroencephalography (EEG) to test for cortical synchronization to temporal structure in sign language. Results show that cortical synchronization depends on the structure of information in the stimulus, not on modality-specific perceptual processes. Cortical synchronization is often thought to rely on oscillatory resonance to rhythms in the stimulus, analogous to harmonic resonance in other physical and biological systems. However, the brain synchronizes to both isochronous and non-isochronous sequences, raising the possibility that temporal expectations may rely on a non-oscillatory mechanism. In Study 2, with two EEG experiments and a computational model, I show that oscillatory resonance can generate predictive neural activity to predictable non-isochronous rhythms. Resonance can lead to behavior that appears "smart", despite being solely stimulus-driven. Together, these studies suggest that oscillatory resonance provides a flexible, domain-general mechanism that can support temporal predictions across sensory modalities, enabling the cerebral cortex to maximize sensitivity to informational peaks in time-varying signals.

CHAPTER 1

INTRODUCTION

Animals must predict the future in order to survive. These predictions are crucial for a wide range of behaviors and a wide range of timescales, from milliseconds to decades. When a dog jumps up to catch a frisbee, for example, it doesn't jump to where the disc currently is, but to where it will soon be. Far from being a peripheral aspect of perception and cognition, prediction has been hypothesized to constitute the brain's core function (Friston, 2005; Clark, 2013). By constantly forming predictions of the sensory consequences of behavior, animals can maximize the efficiency of cognitive processing and the effectiveness of their actions.

In the first section of this dissertation, I use language as a model system to examine the cognitive and neural basis of prediction. When people comprehend language, they predict upcoming syntax, semantics, pragmatics, words, and phonemes (Kuperberg & Jaeger, 2016). Because of this rich and well-studied predictive structure, language provides a useful testbed for hypotheses about prediction. In the second section, I investigate the mechanisms supporting predictions in language.

In addition to predicting *what* will happen next, people make fine-grained predictions of *when* events will happen. For example, when people play double-dutch, they predict when the rope will swing under their feet, and jump in time to avoid being hit. Temporal predictions differ from content predictions in a number of ways. Most importantly, temporal predictions are agnostic about the actual information that will appear – instead, they allow the person to allocate attention at the periods of time when behaviorally-relevant events are likely to occur. The outcome of temporal predictions is not a measure of how the outcome differed from what was expected, as in the predictive coding framework (Friston, 2010; Huang & Rao, 2011); instead, temporal predictions help to ensure that the animal is ready to perceive and act at important time-points. Because temporal predictions prepare an animal for bursts of information, they serve to sharpen perception (Nobre & van Ede,

2018) and improve learning (Heideman, van Ede, & Nobre, 2016; Nobre & van Ede, 2018).

Just as language comprehension depends on predictions of upcoming content, it depends on predictions of the *timing* of that content. For example, manipulating the ongoing speech rate can change which phonemic category listeners perceive a sound as belonging to (reviewed in (Peelle & Davis, 2012)), and it can influence the likelihood that listeners will perceive word boundaries in phonetically ambiguous sentences (Dilley & Pitt, 2010). Additionally, temporal predictions can improve language comprehension by ensuring that the perceptual systems are prepared for bursts of information in speech (Peelle & Davis, 2012; Ding & Simon, 2014).

When people listen to speech, they must attend to sounds that come from speech while ignoring sounds that comprise background noise. The speech signal is therefore more informative at times when the volume of speech is high (during vowels and stressed syllables), compared with times when the volume is low (between syllables and phrases) (Cooke, 2006; Ding & Simon, 2014). To improve comprehension, the brain predicts when bursts of information will appear, ensuring that attention is focused on the moments when the speech signal is the most informative (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Schroeder & Lakatos, 2009; Zion Golumbic, Poeppel, & Schroeder, 2012).

How does the brain predict when the next burst of information will appear? Temporal expectations arise in a wide range of contexts, and the brain most likely uses different neural mechanisms across these different contexts (Ivry & Schlerf, 2008; Nobre & van Ede, 2018). One mechanism is particularly prominent during language comprehension: cortical synchronization to the volume envelope of speech. In cortical synchronization (sometimes called phase-locking or entrainment), electrophysiological activity in the brain locks onto quasi-periodic pulses in the stimulus. When people listen to speech, low-frequency (1-10 Hz) electrophysiological activity in the brain synchronizes to peaks in the volume of the speech envelope (Ahissar et al., 2001; Luo & Poeppel, 2007).

Cortical synchronization may improve speech comprehension by ensuring that the brain's perceptual systems are prepared for bursts of information. Cortical oscillations correspond to fluctuations in the polarization of neurons in a local population (Buzsáki & Draguhn, 2004). The phase of ongoing neural oscillations influences both perceptual sensitivity (Busch, Dubois, & VanRullen, 2009; Mathewson, Gratton, Fabiani, Beck, & Ro, 2009; Mathewson et al., 2012; Ng, Schroeder, & Kayser, 2012; Henry & Obleser, 2012; Cravo, Rohenkohl, Wyart, & Nobre, 2013; Henry, Herrmann, & Obleser, 2014; Spaak, de Lange, & Jensen, 2014) and neural excitability (Lakatos et al., 2005; Jacobs, Kahana, Ekstrom, & Fried, 2007; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Mathewson et al., 2009; Romei, Gross, & Thut, 2012), with the peaks in perceptual sensitivity occurring at the most excitable phase of the oscillation.

Although a large number of studies have shown that cortical activity synchronizes to the volume envelope of speech, the basic neural processes that support cortical synchronization to speech remain unknown. Cortical activity robustly synchronizes to volume in speech, but this process is not unique to either speech comprehension or changes in volume. The brain synchronizes to rhythms in music (Doelling & Poeppel, 2015) as well as to meaningless sequences of periodic sounds (Henry & Obleser, 2012; Xiang, Simon, & Elhilali, 2010; Wang et al., 2011) and images (Busch et al., 2009; Cravo et al., 2013; Spaak et al., 2014). Although many studies focus on synchronization to fluctuations in volume, cortex can also entrain to other aspects of the stimulus, such as its pitch (Picton, Skinner, Champagne, Kellett, & Maiste, 1987; Henry et al., 2014; Henry & Obleser, 2012) or high-level spectral features (Ding, Chatterjee, & Simon, 2014; Zoefel & VanRullen, 2016). These results show that entrainment to the speech envelope is a specific case of a general neural phenomenon in which the brain synchronizes to regularities in any quasi-periodic signal.

Cortical synchronization is hypothesized to improve speech comprehension by aligning bursts of information in speech to times of peak cortical excitability (Schroeder et al., 2008;

Schroeder & Lakatos, 2009; Zion Golumbic et al., 2012; Peelle & Davis, 2012), but the studies described above do not address the functional role of synchronization. Does cortical synchronization boost comprehension?

A number of studies show that the strength of entrainment correlates with speech comprehension. Cortical activity synchronizes more strongly to intelligible speech than to speech that has been made unintelligible by speeding up the recording (Ahissar et al., 2001), adding background noise (Ding & Simon, 2013), or decreasing the spectral resolution (Luo & Poeppel, 2007; Peelle, Gross, & Davis, 2012) (but see Howard & Poeppel, 2010). Furthermore, when people listen to multi-speaker recordings, low-frequency neural activity follows the speech envelope of attended speech more closely than it follows the envelope of ignored speech (Kerlin, Shahin, & Miller, 2010; Ding & Simon, 2012; Zion Golumbic et al., 2013; Horton, Srinivasan, & D'Zmura, 2014)

Because these studies rely on correlations between brain and behavior, however, they cannot tell us whether synchronization plays a causal role in speech comprehension. Subsequent studies have tested whether cortical synchronization is causally involved in speech comprehension by using transcranial alternating current stimulation (tACS) to manipulate neural activity. These studies show that synchronizing neural activity to the volume envelope increases the strength of fMRI BOLD responses to speech (Zoefel, Archer-Boyd, & Davis, 2018), and that synchronization influences listeners' speech recognition performance (Riecke, Formisano, Sorger, Başkent, & Gaudrain, 2017).

In these studies, I examine how cortical synchronization helps animals predict the future. I address two open questions regarding the basis of synchronization.

In the first study, I ask: How does cerebral cortex synchronize to language? To answer this question, I investigate why cortex synchronizes to speech specifically around 1-10 Hz. Speech carries information across a range of frequencies, and in principle, the brain could synchronize at any of these time-scales. By determining why the brain synchronizes at this

frequency, we can learn why it synchronizes at all. I examine two possible answers to this question.

First, the brain may synchronize to speech at 1-10 Hz because these frequencies are important for information in speech. Spoken syllables and oscillations in the volume of speech both occur in this range, with a peak around 4 Hz (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Pellegrino, Coupé, & Marsico, 2011; Edwards & Chang, 2013; Ding et al., 2017). Perhaps the brain synchronizes to speech at this frequency because of the informational properties of speech – because syllables, words, and short phrases lie in this range.

Alternatively, synchronization at these frequencies may reflect a basic processing constraint of auditory perception. When people listen to meaningless auditory stimuli, auditory cortex entrains most strongly below 10 Hz, at approximately the same rate as informational changes due to spoken syllables (Hickok, Farahbod, & Saberi, 2015; Henry et al., 2014; Ng et al., 2012). Perhaps the brain synchronizes to speech at this frequency due to filtering properties of sensory processing.

Because these two different explanations predict entrainment in the same frequency band, it is impossible to distinguish between them by studying how auditory cortex processes speech. To solve this problem, I ask whether – and at what frequencies – visual cortex synchronizes to information in sign language.

Sign language can disentangle these hypotheses because linguistic units in sign language occur at a different rate from the 'preferred' frequency of visual entrainment. Unlike auditory cortex, visual cortex entrains most strongly to stimulation at approximately 10 Hz (Herrmann, 2001; İlhan & VanRullen, 2012). Crucially, this frequency band is higher than the frequencies of informational changes in sign language. Although manual analyses of sign language show that individual signs occur at approximately 2 Hz (Bellugi & Fischer, 1972; Hwang, 2011), no quantitative metric existed to allow us to examine physical variability in

the visual signal of sign language, analogous to the volume envelope of speech. To solve this problem, I developed a tool to quantify the magnitude of visual movement over time in a video, and used it to characterize temporal fluctuations in naturalistic sign language. By testing for cortical entrainment to sign language, I asked whether synchronization is determined by informational properties of the communicative signal, or by filtering properties of sensory cortex.

In a second study, I investigate the underlying neural processes that enable the brain to synchronize to speech. Many researchers assume that cortical synchronization occurs when oscillating networks in the brain resonate along with quasi-periodic stimulation. When multiple oscillators interact, they tend to synchronize; this phenomenon is called *oscillatory resonance*, and is commonly found in many physical and biological systems. For example, a wine glass will resonate if someone sings next to it at the right frequency. Perhaps cortical synchronization is driven by a similar process in the brain: if people perceive a periodic sequence of stimuli at one of the brain's resonant frequencies, then intrinsic cortical oscillations will resonate along with those stimuli.

Oscillatory resonance in a sine-wave oscillator, however, can only create a one-to-one mapping to stimuli that appear in an isochronous rhythm, with the same interval between each stimulus. In study 1, however, I found that the brain synchronizes to visual pulses of movement in sign language even though sign language is not strongly periodic (Brookshire, Lu, Nusbaum, Goldin-Meadow, & Casasanto, 2017). Other studies have reported converging evidence of cortical synchronization to sequences that are not periodic (Andreou, Griffiths, & Chait, 2015; Breska & Deouell, 2017). These studies cast doubt on the hypothesis that oscillatory resonance forms the basis of cortical synchronization.

Can cortical synchronization to non-isochronous sequences be explained by oscillatory resonance? I address this question by using EEG to test how the brain generates predictions to non-isochronous sequences. I then test whether a computational model of oscillatory

resonance can account for our EEG results. This study will address the question of whether resonance can lead to behavior that appears "smart", despite being solely stimulus-driven.

CHAPTER 2

VISUAL CORTEX ENTRAINS TO SIGN LANGUAGE

This chapter has been published as:

Brookshire, G., Lu, J., Nusbaum, H.C., Goldin-Meadow, S., & Casasanto, D. (2017). Visual cortex entrains to sign language. *Proceedings of the National Academy of Sciences.* 114(24), 6352-6357. (`www.pnas.org/content/114/24/6352.full`)

2.1 Abstract

Despite immense variability across languages, people can learn to understand any human language, spoken or signed. What neural mechanisms allow people to comprehend language across sensory modalities? When people listen to speech, electrophysiological oscillations in auditory cortex entrain to slow (<8 Hz) fluctuations in the acoustic envelope. Entrainment to the speech envelope may reflect mechanisms specialized for auditory perception. Alternatively, flexible entrainment may be a general-purpose cortical mechanism that optimizes sensitivity to rhythmic information regardless of modality. Here we test these proposals by examining cortical coherence to visual information in sign language. First, we develop a metric to quantify visual change over time. We find quasi-periodic fluctuations in sign language, characterized by lower frequencies than fluctuations in speech. Next, we test for entrainment of neural oscillations to visual change in sign language, using electroencephalography (EEG) in fluent speakers of American Sign Language (ASL) as they watch videos in ASL. We find significant cortical entrainment to visual oscillations in sign language below 5 Hz, peaking at about 1 Hz. Coherence to sign is strongest over occipital and parietal cortex, in contrast to speech, where coherence is strongest over the auditory cortex. Non-signers also show coherence to sign language, but entrainment at frontal sites is reduced relative to fluent signers. These results demonstrate that flexible cortical entrainment to language does not

depend on neural processes that are specific to auditory speech perception. Low-frequency oscillatory entrainment may reflect a general cortical mechanism that maximizes sensitivity to informational peaks in time-varying signals.

2.2 Introduction

Languages differ dramatically from one another, yet people can learn to understand any natural language. What neural mechanisms allow humans to understand the vast diversity of languages, and to distinguish linguistic signal from noise? One mechanism that has been implicated in language comprehension is neural entrainment to the volume envelope of speech. The volume envelope of speech fluctuates at low frequencies (< 8 Hz), decreasing at boundaries between syllables, words, and phrases. When people listen to speech, neural oscillations in the delta (1–4 Hz) and theta bands (4–8 Hz) become entrained to these fluctuations in volume (Ahissar et al., 2001; Luo & Poeppel, 2007; Gross et al., 2013; Park, Ince, Schyns, Thut, & Gross, 2015).

Entrainment to the volume envelope may represent an active neural mechanism to boost perceptual sensitivity to rhythmic stimuli (Luo & Poeppel, 2007; Schroeder & Lakatos, 2009; Giraud & Poeppel, 2012; Peelle & Davis, 2012). Although entrainment is partly driven by bottom-up features of the stimulus (Howard & Poeppel, 2010; Doelling, Arnal, Ghitza, & Poeppel, 2014; Ding et al., 2014), it also depends on top-down signals to auditory cortex from other brain areas. Auditory entrainment is strengthened when people see congruent visual and auditory information (Luo, Liu, & Poeppel, 2010; Crosse, Butler, & Lalor, 2015), and is modulated by attention (Zion Golumbic et al., 2013) and by top-down signals from frontal cortex (Kayser, Ince, Gross, & Kayser, 2015; Park et al., 2015).

Cortical entrainment is proposed to perform a key role in speech comprehension, such as segmenting out syllables from a continuous speech stream (Ahissar et al., 2001; Ahissar & Ahissar, 2005; Luo & Poeppel, 2007) or optimizing perceptual sensitivity to rhythmic

pulses of sound (Schroeder & Lakatos, 2009; Giraud & Poeppel, 2012; Peelle & Davis, 2012). However, the mechanisms driving entrainment to speech remain unclear. We consider two hypotheses. First, flexible entrainment to quasi-periodic rhythms may be specific to auditory perception (Giraud & Poeppel, 2012); in visual perception, by contrast, cortical oscillations in the alpha band (8-12 Hz) may phase-lock only to consistent stimulus rhythms, without adjusting to variable stimulus rhythms (VanRullen, Zoefel, & İlhan, 2014). Second, low-frequency cortical entrainment may be a general-purpose neural mechanism that helps optimize perception to time-varying stimuli regardless of the perceptual modality. Neural oscillations may allow the brain to rhythmically orient attention to quasi-periodic stimuli (Schroeder & Lakatos, 2009; Landau & Fries, 2012) across sensory systems.

Because previous studies of cortical entrainment to rhythms in language have focused on oral speech, they have been unable to distinguish between these competing hypotheses. Here we test for low-frequency entrainment to a purely visual language: American Sign Language (ASL). Prior studies show that neural and behavioral oscillations in vision are preferentially entrained by stimuli that flicker in the alpha band (Herrmann, 2001; Mathewson et al., 2012; VanRullen & Macdonald, 2012; Spaak et al., 2014). Therefore, if flexible cortical entrainment to oral speech depends on modality-specific properties of auditory processing, then phase-locking to sign language should be concentrated in the alpha band, if it occurs at all (VanRullen et al., 2014). Alternatively, if cortical entrainment is a generalized neural strategy to maximize sensitivity to rhythmic stimuli, then oscillatory activity in visual cortex should entrain at the frequency of informational changes in ASL.

To determine whether human cerebral cortex entrains to rhythmic information in sign language, first we developed a metric for quantifying the amplitude of visual change in sign, by analogy to the acoustic envelope of speech. Next, we characterized visual variability across four sign languages, showing that this variability is quasi-periodic below 8 Hz. Finally, we demonstrated that cerebral cortex entrains to visual variability in sign language, and showed

that entrainment is strongest around the frequencies of phrases and individual signs in ASL.

2.3 Results

2.3.1 Developing a metric for quantifying visual change

In order to examine neural entrainment to visual rhythms in sign language, we must first quantify the 'amplitude envelope' of a visual signal. The acoustic envelope is a highly reduced representation of sound, tracing extreme amplitude values in the time-varying signal. Oscillations in the envelope of speech depend on movements of various components of the vocal tract, including the rhythmic opening and closing of the mandible (Chandrasekaran et al., 2009). Sign language, in contrast, does not involve consistent oscillatory movements by any single effector (Meier, 2002). However, quasi-periodic oscillations in sign language may arise from the coordinated movements of multiple effectors.

Here we present the Instantaneous Visual Change (IVC) as a metric that is conceptually similar to the acoustic envelope, summarizing the amplitude of change at each time point. The IVC is a time-series of aggregated visual changes between frames (Fig. 2.1, *Method and Materials*). This algorithm provides an automatic, objective alternative to human-coded methods of studying temporal structure in sign (Petitto, Holowka, Sergio, Levy, & Ostry, 2004).

The amplitude of the IVC indexes the amount of visual change between two video frames. The largest peaks in the IVC therefore occur during large, quick movements. In the videos we analyzed, these changes corresponded primarily to movements of the signers' hands and arms, but may also reflect movements of the face, head, and torso. For example, a quick arm movement results in a larger number of pixels changing in each adjacent frame – and a higher peak in the IVC – than a slow arm movement. The IVC thus offers a heuristic index of new linguistic information in the visual signal. An example video illustrating the IVC is

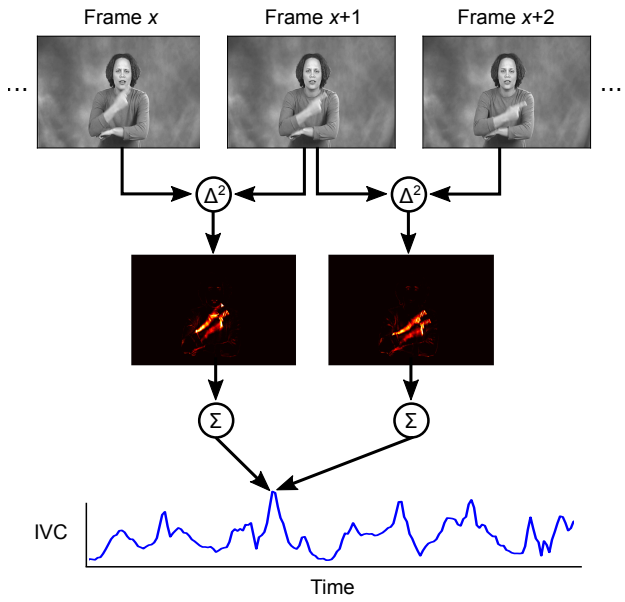

Figure 2.1: Calculation of the Instantaneous Visual Change (IVC). The IVC summarizes total visual change at each point in time. First, the difference between adjacent grayscale video frames (top row) is calculated for each pixel. To aggregate over both increases and decreases in brightness, these pixel-wise differences are then squared (middle row). Finally, the brightness values in all pixels of the squared-difference images are summed to obtain a single value summarizing the magnitude of change between two video frames. Computation of this value for each adjacent pair of frames yields a time-series (bottom).

included in the online Supporting Information (Fig. 2.6).

2.3.2 Characterizing temporal structure in sign language

The IVC allows us to characterize one dimension of the temporal structure in sign language, and to directly compare the spectral signatures of amplitude variability across sign and oral speech. Visual examination of the raw IVC of sign language reveals quasi-periodic oscillations with irregularly-timed peaks (Fig. 2.2*A*). To characterize variability within and across sign languages, we computed the power spectra of the IVC from samples of four different sign languages: American Sign Language (ASL), German Sign Language (Deutsche Gebärdensprache, DGS), British Sign Language (BSL), and Australian Sign Language (Auslan). These languages developed independently of the oral languages spoken in these coun-

tries, and come from three genetically unrelated language families (BSL and Auslan from the British, Australian, and New Zealand sign Language family; ASL from the French Sign Language family; DGS from the German Sign Language family). In all four languages, power in the IVC decreases monotonically with increasing frequency, without any pronounced peaks in the spectrum (Fig. 2.2C). We tested for rhythmic components in the IVC by comparing these spectra against the $1/f$ spectrum characteristic of many signals in the natural world. Power that is higher than the $1/f$ function indicates periodicity at that frequency (Chandrasekaran et al., 2009). The IVC of sign language showed elevated power at approximately 2–8 Hz ($Ps < .01$). Individual signs in sign languages tend to occur at approximately 2–2.5 Hz (Bellugi & Fischer, 1972; Hwang, 2011), on the lower end of the rhythmic components in the IVC. These analyses suggest that sign language involves weak, quasi-periodic rhythms with variable frequencies in the delta and theta range.

To explicitly compare temporal structure between sign and speech, we contrasted the IVC of sign with the broadband envelope of speech (Chandrasekaran et al., 2009). We computed the broadband envelope of samples from nine spoken languages representing five language families: English, French, Portuguese, Dutch, German, Hungarian, Japanese, Arabic, and Mandarin. After resampling the broadband envelopes and IVC signals to a common frequency (30 Hz) and standardizing the amplitude of each recording by dividing out its standard deviation, we compared the average spectra of the IVC of sign and the broadband envelope of speech (Fig. 2.2D). Spoken languages showed stronger modulations than sign languages above 2 Hz. This increased power may reflect modulation due to syllables in speech, which occur at approximately 2–10 Hz (Greenberg, Carvey, Hitchcock, & Chang, 2003; Chandrasekaran et al., 2009). Indeed, peaks from individual syllables are visible in the broadband envelope, and these peaks occur at about 4 Hz (Fig. 2.2B). These results indicate that visual motion in sign language is modulated at lower frequencies than auditory volume in spoken language. This difference is consistent with the slower movements in the

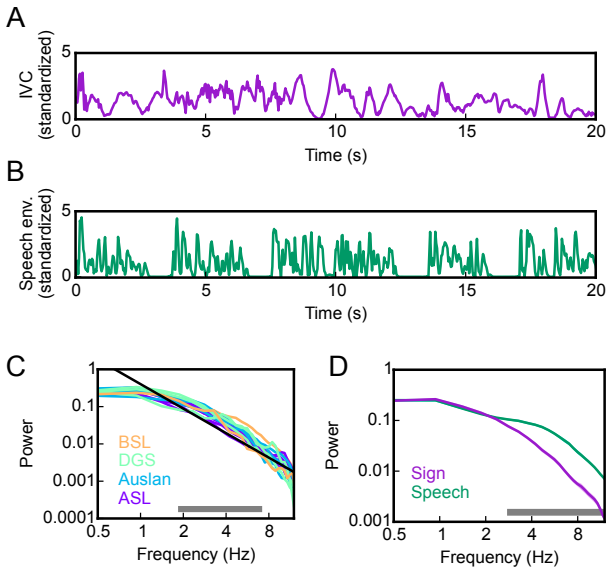

Figure 2.2: Temporal structure in signed and spoken language. *(A)* Example trace of the IVC of ASL. *(B)* Example trace of the broadband envelope of English. *(C)* Spectrum of sign language plotted on log-log axes. Line color denotes language (BSL: British Sign Langauge; DGS: German Sign Language; Auslan: Australian Sign Language; ASL: American Sign Language). Each curve shows the spectrum of a separate video sample ($N = 14$, total duration 1:10:22). The black line shows the best-fit $1/f$ trend across all samples. The gray bar at the base of the plot shows where the IVC spectra are significantly greater than the $1/f$ fit ($P < .01$ by 1-sample t-tests). *(D)* Comparison of the mean spectra across signed and spoken languages. Shaded area depicts the standard error of the mean. The gray bar indicates significant differences between the two curves ($P < .01$ by independent-samples t-tests). Sign language samples are the same as in panel *(C)*. Audio recordings were sampled from speech in 9 languages ($N = 12$, total duration 1:07:28). Amplitude in all analyses has been standardized by each recording's standard deviation.

articulators for sign (the hands) than in the articulators for speech (the vocal tract) (Bellugi & Fischer, 1972; Hwang, 2011).

2.3.3 Cortical coherence to visual rhythms in sign language

We used electroencephalography (EEG) to examine cortical entrainment to quasi-rhythmic fluctuations in visual information in sign language. Fluent speakers of ASL watched videos of ASL stories against a static background. We tested for coherence of low-frequency electrophysiological oscillations to quasi-periodic oscillations in the IVC.

Coherence was calculated separately at each EEG channel in partially overlapping, logarithmically spaced bins centered over 0.5–16 Hz (*Materials and Methods*). Because coherence provides no intrinsic measure of chance performance, we created a null distribution of coherence using a randomization procedure. To obtain each value in the null distribution, we time-shifted the IVC to a randomly-selected starting point, moving the portion of the IVC that remained after the final time point of the EEG signal to the beginning of the recording. This procedure preserves the spectral and temporal characteristics of the EEG and IVC recordings, but eliminates any relationship between these signals. Coherence was then computed between the EEG recordings and the randomly-shifted IVC.

A cluster-based permutation test indicated that coherence between cortical oscillations and the IVC of sign was stronger than would be expected by chance ($P = .0001$). Averaging the coherence spectrum across every EEG channel, coherence was above chance from 0.4–5 Hz, peaking at 1 Hz (Fig. 2.3*A*). Coherence emerged over a similar range of frequencies when we selected only occipital channels (0.8–5 Hz; $P = .0001$), primarily reflecting entrainment in visual cortex (Fig. 2.3*B*). In frontal channels, above-chance coherence was present from 0.4–1.25 Hz ($P = .0001$; Fig. 2.3*C*), revealing top-down control from frontal cortex. Examining the entire scalp distribution, cortical coherence to the IVC of sign language was strongest over central and occipital channels (Fig. 2.3*D*).

To test whether cortical entrainment depends on linguistic knowledge, we examined coherence to sign language in people who did not know any ASL. Like signers, non-signers showed significant coherence to videos of ASL storytelling ($P < .0005$), with the strongest coherence over central and occipital channels from 0.8–3.5 Hz (Fig. 2.5). We then separately analyzed effects of linguistic knowledge on entrainment in occipital and frontal cortex. Although coherence at occipital channels did not significantly differ between groups (Fig. 2.4*B*), coherence at frontal channels was stronger in signers than in non-signers, indicating differences in top-down control based on familiarity with ASL ($P < .05$; Fig. 2.4*A*; Fig. 2.5).

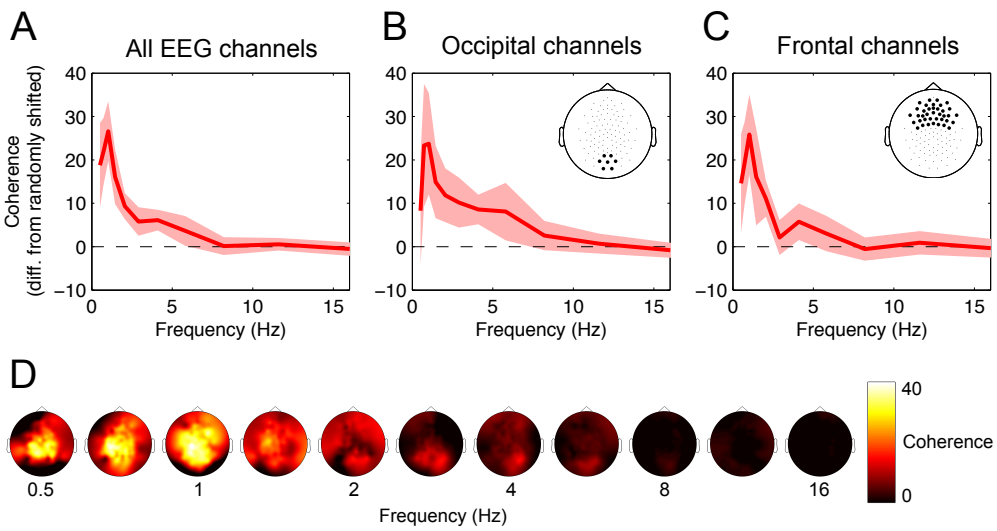

Figure 2.3: Coherence between cortex and the IVC of sign language in fluent signers ($N = 13$). *(A)* Coherence spectrum averaged across all EEG channels. For each participant, we computed the difference in empirical coherence and a distribution of cortical coherence to randomly shifted IVC. The solid line shows the mean difference in empirical and randomized coherence across participants, and the shaded area shows the 95% CI around the mean. The dotted line shows chance levels. *(B)* Coherence spectrum averaged over occipital channels. Inset shows the location of selected channels. *(C)* Coherence spectrum averaged over frontal channels. Inset shows the location of selected channels. *(D)* Scalp topography of coherence in each frequency bin, averaged across participants.

2.4 Discussion

2.4.1 Cortical coherence to sign language

In this study, we find that electrophysiological oscillations in human cerebral cortex become entrained to quasi-periodic fluctuations of visual movement in sign language. In fluent signers, cortical entrainment to sign language was found between 0.4 and 5 Hz, peaking at about 1 Hz, and emerged most robustly over occipital and central EEG channels. These results show that the human brain entrains to low-frequency variability in language whether it is perceived with the ears or eyes. Visual cortex flexibly phase-locks to visible changes in sign language, analogously to the way auditory cortex phase-locks to amplitude changes in oral speech. Our findings argue that flexible entrainment depends on mechanisms that are not

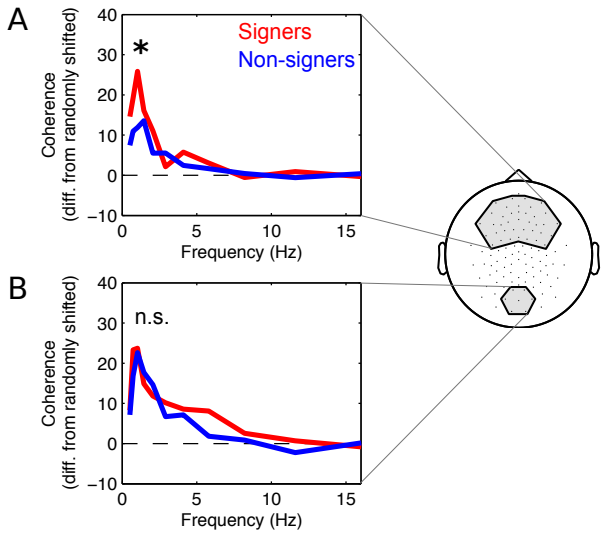

Figure 2.4: Comparison of coherence in fluent signers ($N = 13$) and non-signers ($N = 15$). *(A)* Coherence at frontal channels (channel selection illustrated at right) was stronger in signers than in non-signers. *(B)* Coherence at occipital channels did non differ between groups. Data from signers is the same as in Fig. 2.3. *, $P < .05$; *n.s.*, not statistically significant.

specific to any given effector or sensory modality.

Prior results suggest that auditory and visual perception are differentially modulated by rhythms at different frequencies (VanRullen et al., 2014). Auditory sensitivity varies as a function of the power and phase of spontaneous 2–6 Hz rhythms (Ng et al., 2012), and these oscillations are entrained by sounds modulated at 3 Hz (Henry & Obleser, 2012; Hickok et al., 2015). Visual sensitivity, by contrast, depends on the power and phase of spontaneous alpha rhythms (Hanslmayr et al., 2007; Van Dijk, Schoffelen, Oostenveld, & Jensen, 2008; Mathewson et al., 2009), and electrophysiological oscillations in visual cortex are robustly entrained by periodic stimulation around 10 Hz (Mathewson et al., 2012; Spaak et al., 2014). When humans watch a light flicker at frequencies from 1-100 Hz, visual cortex shows the strongest entrainment around 10 Hz (Herrmann, 2001).

Although these rhythmic preferences are not absolute (visual cortex also shows rhythmic oscillations in the delta/theta range; Busch et al., 2009; Park, Kayser, Thut, & Gross, 2016; Luo et al., 2010; Power, Mead, Barnes, & Goswami, 2012; Cravo et al., 2013; Landau

& Fries, 2012), differences between the sensory modalities are apparent when different frequency bands are directly compared. Auditory detection sensitivity depends on the phase of underlying delta-theta but not alpha oscillations (Ng et al., 2012). In response to aperiodic stimulation, visual cortex oscillates in the alpha band (VanRullen & Macdonald, 2012), whereas auditory cortex does not show consistent oscillatory activity (İlhan & VanRullen, 2012).

If sensory-specific oscillatory preferences determined the spectrum of entrainment, then peak coherence to sign language should be observed at the frequencies preferred by visual cortex: around 10 Hz. Contrary to this prediction, we find that cortical coherence to sign language only emerges below 5 Hz. Cerebral cortex entrains to sign language around the frequencies of words and phrases in ASL.

2.4.2 Coherence across signers and non-signers

We find that cerebral cortex phase-locks to visual changes in ASL both in fluent signers and in people with no knowledge of sign language. In principle, coherence to ASL in non-signers could emerge for two reasons. Coherence could be driven either bottom-up, by sensory stimulation, or top-down, by non-linguistic temporal predictions. Human bodies move in predictable ways, and non-signers could entrain to sign language based on these regularities in human movement.

In frontal areas, fluent signers showed stronger coherence than non-signers. This difference in frontal coherence may reflect top-down sensory predictions based on knowledge of ASL. Alternatively, differences in coherence could reflect differences in attention to the videos. Cortical entrainment to oral speech decreases when people direct attention away from the speech stimulus (Zion Golumbic et al., 2013; O'Sullivan et al., 2015). Reduced coherence in non-signers, therefore, would also be predicted if non-signers do not attend to videos of ASL as strongly as fluent signers do. However, our findings at occipital channels

argue against this possibility. If differences between groups were driven by attention, then occipital coherence should be stronger in signers than in non-signers. However, we find no evidence that occipital coherence depends on linguistic knowledge. Taken together, these results suggest that although linguistic knowledge is not necessary for entrainment, signers may leverage knowledge about ASL to sharpen temporal predictions during language comprehension. These sharpened predictions result in stronger entrainment in the frontal regions that exert top-down control over visual perception.

2.4.3 Specialization for speech?

Syllables in oral speech occur at frequencies that largely overlap with cortical entrainment to the volume envelope. This overlap could be interpreted as evidence for a specialized oscillatory mechanism for speech comprehension. This type of speech-specific mechanism could evolve in at least two ways. First, "the articulatory motor system [may have] structured its output to match those rhythms the auditory system can best apprehend" (Giraud & Poeppel, 2012). Second, auditory mechanisms may have developed to comprehend speech based on the timing of preexisting oral behaviors. Non-human primates create vocalizations and facial displays that fluctuate at frequencies similar to human speech syllables (Ghazanfar, Takahashi, Mathur, & Fitch, 2012), and their attention is preferentially captured by faces that move at these frequencies (Ghazanfar, Morrill, & Kayser, 2013); perhaps auditory processing evolved to fit the timing profile of these behaviors.

The data we report here, however, suggest that entrainment may not have any close evolutionary link to oral speech. Instead, a more general process may underlie cortical phase-locking to variability in language. Previous results are consistent with this interpretation as well. When participants watch videos of speech, entrainment emerges not only in auditory cortex, but also in visual cortex (Park et al., 2016; Luo et al., 2010; Power et al., 2012). Furthermore, cortical rhythms entrain to rhythms in music (Doelling & Poeppel, 2015) and

to other rhythmic stimuli in audition (Henry & Obleser, 2012; Hickok et al., 2015) and vision (Lakatos et al., 2008; Mathewson et al., 2012). These examples of low-frequency cortical entrainment to a broad range of stimuli across sensory modalities suggest that the cortical mechanisms supporting entrainment to the volume envelope of speech may be a specialized case of a general predictive process.

2.4.4 Neural mechanisms of language comprehension across sensory modalities

Previous studies have shown that the functional neuroanatomy of speech largely overlaps with that of sign (MacSweeney, Capek, Campbell, & Woll, 2008). At the coarsest level of anatomical specificity, the left hemisphere is specialized for spoken language. The left hemisphere is also asymmetrically active during sign language perception (Leonard et al., 2012) and production, regardless of which hand people use to sign (Corina, San Jose-Robertson, Guillemin, High, & Braun, 2003; Gutierrez-Sigut et al., 2015). Left hemisphere damage, furthermore, results in linguistic deficits in signing patients (Hickok, Bellugi, & Klima, 1998).

Specific regions within the left hemisphere show similar involvement in processing both speech and sign. Across signed and spoken language, bloodflow increases to the LIFG and left inferior parietal lobe (IPL) during phonemic discrimination (MacSweeney, Waters, Brammer, Woll, & Goswami, 2008; Williams, Darcy, & Newman, 2015) and morphosyntactic processing (Newman, Supalla, Hauser, Newport, & Bavelier, 2010). Similarly, word production in both signed and spoken languages activates LIFG, left IPL, and left temporal areas (Emmorey, Mehta, & Grabowski, 2007).

Differences in the cortical areas involved in sign and speech can often be attributed to differences in the form of these languages. For example, comprehension of sign language activates primary visual but not primary auditory cortex (Leonard et al., 2012). Consistent with the fact that sign language relies on spatial contrasts, inferior and superior parietal cortex

is more strongly active during signed than during spoken language production (Emmorey et al., 2007) and perception (MacSweeney, Waters, et al., 2008).

Our findings go beyond functional neuroanatomy to examine neurophysiological processes that can arise in multiple cortical areas. We show that oscillatory entrainment to low-frequency variability in the stimulus occurs regardless of whether language is being processed using auditory cortex or visual cortex.

Our results differ from previous studies on entrainment to speech primarily in the scalp topography of coherence. Entrainment to auditory speech is strongest over auditory cortex (Luo & Poeppel, 2007; Howard & Poeppel, 2010; Luo et al., 2010; Park et al., 2016) and central frontal sites (Kayser et al., 2015). By contrast, our results show that entrainment to sign language is strongest at occipital and parietal channels, consistent with greater parietal activation during sign compared with speech (Emmorey, McCullough, Mehta, & Grabowski, 2014). This difference likely reflects increased visual and spatial demands of perceiving sign language.

2.4.5 The IVC quantifies temporal structure in visual perception

The Instantaneous Visual Change (IVC) provides a novel method for examining gross temporal structure in natural visual stimuli. Analogously to the way the broadband envelope summarizes early stages of auditory processing, the IVC provides a first approximation of the magnitude of information available to the earliest stages of visual processing. At the first stage of auditory transduction, hair cells in the cochlea extract the narrowband envelope of sounds. Summing these narrowband envelopes together yields the overall auditory responses over time: the broadband envelope. In the retina, center-surround retinal ganglion cells respond to changes in the brightness of specific wavelengths of light. Summing the responses from these cells yields the overall visual responses over time, approximated by the IVC.

The IVC provides a coarse index of visual information in sign language, just as the

broadband envelope provides a coarse index of information in speech. For example, the volume envelope does not reflect small spectral differences that are crucial for discriminating vowels. The IVC, analogously, does not preserve information about which effectors are moving or their trajectories. Nevertheless, sign language comprehenders may use the IVC of sign heuristically, as listeners use the acoustic envelope of speech, to anticipate when important information is likely to appear.

In the current study, we use the IVC to characterize temporal structure in sign language, and to examine responses of the human brain to that temporal structure. The IVC could also be applied to study temporal structure in other domains, such as gesture, biological motion, and movement in natural scenes.

2.4.6 The functional role of entrainment to language

Oscillatory entrainment to language may be a specific case of a general cortical mechanism. In primates, spiking probability varies with the phase of low-frequency oscillations: neurons are most likely to fire at specific points in the phase of ongoing oscillations (Jacobs et al., 2007; Lakatos et al., 2005). Perhaps the cortex strategically resets the phase of ongoing neural oscillations to ensure that perceptual neurons are in an excitable state when new information is likely to appear (Schroeder & Lakatos, 2009; Giraud & Poeppel, 2012; Peelle & Davis, 2012; Kayser et al., 2015). Oscillatory entrainment may constitute a cortical strategy to boost perceptual sensitivity at informational peaks in language. Our findings suggest that the brain can flexibly entrain to linguistic information regardless of the modality in which language is produced or perceived.

2.5 Materials and Methods

Participants watched approximately 20 minutes of naturalistic storytelling in American Sign Language (ASL) while EEG was recorded. Participants were instructed to watch the videos

and remain still and relaxed. All procedures were approved by the Institutional Review Board of the University of Chicago. Detailed methods and analyses are available in the online Supporting Text.

2.5.1 Participants

Participants had corrected-to-normal vision and reported no history of epilepsy, brain surgery, or traumatic brain injuries. Informed consent was obtained before beginning the experiment, and participants were paid $20/hour for their participation. We recorded EEG from 16 fluent signers of American Sign Language (ASL). Data from 2 participants were excluded before analysis due to excessive EEG artifacts, and data from 1 participant were lost due to experimenter error. All participants retained in the analyses began learning ASL by 5 years of age ($N = 13$; 3 female, 10 male; age 24–44; mean age of acquisition 1.1 years). Participants who used hearing aids or cochlear implants removed the devices before beginning the experiment. A fluent speaker of ASL (J.L.) answered participants' questions about the study. We recorded EEG from an additional 16 participants who had no prior exposure to ASL. These participants were recruited from the University of Chicago community through online postings. One participant who was currently learning ASL was excluded before analyses, leaving $N = 15$ non-signing participants (10 female, 5 male; age 18–31).

2.5.2 Instantaneous Visual Change (IVC)

The IVC represents a time-series of aggregated visual change between frames (Fig. 2.1), and is computed as the sum of squared differences in each pixel across sequential frames of video:

$$\mathrm{IVC}(t) = \sum_i [x_i(t) - x_i(t-1)]^2$$

where x is the grayscale value of pixel i at time t. Python code for the IVC is available at `http://github.com/gbrookshire/ivc`.

2.5.3 EEG analysis

See Supporting Text for details about EEG acquisition and preprocessing. To compute coherence, IVC and EEG data were filtered into overlapping log-spaced frequency bins using phase-preserving forward-reverse Butterworth bandpass filters. Bins were centered on values from 0.5–16 Hz, and included frequencies in the range $(0.8f, 1.25f)$, where f is the center frequency $f = 2^n$ for $n \in \{-1, -0.5, 0, \ldots, 4\}$. Instantaneous phase and power were determined with the Hilbert transform. Power was computed as the absolute value of the analytic signal, and phase as the angle of the analytic signal. These power and phase estimates were then used to calculate coherence:

$$Coh = \frac{|\sum_t (e^{i\theta_t} \sqrt{P_{C,t} \cdot P_{V,t}})|}{\sqrt{\sum_t (P_{C,t} \cdot P_{V,t})}}$$

where t is the time point, θ is the phase difference between the IVC and EEG, P_V is power in the IVC, and P_C is power in the EEG recording (Doelling et al., 2014). Statistical significance of coherence was determined by a two-stage randomization procedure. First, the IVC was randomly shifted to obtain a null distribution of coherence between the two signals. Second, statistical significance was determined using cluster-based permutation tests (Maris & Oostenveld, 2007) (Supporting Text).

2.6 Supplementary Information

2.6.1 Spectral analysis of speech and sign

We analyzed temporal structure in multiple spoken and signed languages. These analyses were performed using custom software written in Python.

Spoken language samples comprised audio recordings of stories in 9 languages from 5 languages families (N=12 samples, total duration 1:07:28; Table 2.2). Before computing the spectra, these recordings were trimmed to a maximum of 6 min (mean 5:37, SD 0:52), mixed down to 1 output channel, and downsampled to 22050 Hz.

Sign language samples were chosen to be as similar as possible to the spoken samples (N=14, total duration 1:10:22, mean 5:02, SD 3:19; TableTable 2.1), representing 4 languages from 3 families. All samples comprised forward-facing views of a single speaker. In 12 videos, the speaker told a story; in the remaining 2 videos, the speaker gave an informational speech. The IVC was calculated for each video.

All IVC and broadband envelope recordings were resampled to a common frequency of 30 Hz, and power was normalized by dividing out the SD of each signal. Power spectra were computed using Welch's method. Signals were split into 2.13-s long segments (2^6 samples) that overlapped by 1.07 s (2^5 samples). A Hanning window was applied to each segment, and the linear trend was removed. Fast Fourier transforms (FFT) were then computed for each segment. The spectrum for each signal was obtained by averaging the spectra in all segments.

We fit sign language IVC spectra to a $1/f$ function using least squares regression by transforming the equation $Y = Af^{-\alpha}$ into the linear form $\log Y = \log A - \alpha \log f$, with power Y and frequency f (Chandrasekaran et al., 2009). We tested for deviations from the $1/f$ trend using 1-sample t-tests.

2.6.2 Broadband envelope

The broadband envelope was computed over samples of oral speech by adapting methods from previous studies (Chandrasekaran et al., 2009). First, the raw waveform was band-pass filtered into 25 logarithmically-spaced frequency bands from 0.1–10 kHz using least-squares filters with 501 taps. The narrowband envelope of each filtered signal was then calculated as the absolute value of the Hilbert transform. The narrowband envelopes were summed to obtain the broadband envelope for each recording.

2.6.3 Stimuli

Stimuli comprised two videos of ASL storytelling: "Kondima" and "Little Feet" from the Rosa Lee Show (Timm & Timm, 2008 (Pelican Ave Inc., Fremont, CA) American sign langauge performance art). Each video was approximately 10 min long with a 30 Hz frame-rate, and depicted a native speaker of ASL telling a story against a static background. To ensure that the timing of the video was accurately matched with EEG recordings, a small white square flashed in the corner of the display once every 30 frames of video (out of view of the participant), and was registered by a photodiode connected to the EEG amplifier. In total, the session lasted 60–90 minutes.

2.6.4 EEG acquisition and preprocessing

EEG was recorded at 250 Hz using a 128-channel net (Electrical Geodesics, Eugene, OR). Impedances were reduced to < 50 kΩ before participants watched each story. EEG analyses were performed in Matlab using custom software and the FieldTrip package (Oostenveld, Fries, Maris, & Schoffelen, 2010). Recordings from electrodes along the face, beneath the ears, and at the base of the neck were excluded before any analysis, leaving 103 channels in all analyses. Electrode movement artifacts were manually identified and rejected by replacing the tagged region with zeros and applying a 4000-ms half-Hann taper to each side

of the artifact. Artifacts from blinks and eye-movements were identified and removed using independent component analysis (ICA). To ensure that the IVC of each story was accurately matched to the EEG recordings, we used cubic spline interpolation to warp the IVC to the time-stamps registered by the photodiode. This process simultaneously resampled the IVC from 30 Hz to 250 Hz. Before computing coherence, EEG signals were re-referenced to the average mastoids.

2.6.5 EEG statistical testing

Statistical significance of EEG coherence to the IVC was determined by a two-stage randomization procedure. To obtain a null distribution of coherence, the onset of the IVC was circularly shifted to a randomly selected starting point, and coherence was computed between EEG signals and the shifted IVC. This procedure preserves the spectrotemporal characteristics of both signals, but eliminates any relationship between them. For each subject, we computed 100 randomly shifted baselines. Next, we tested for significant differences between the empirical and randomly shifted coherence using a cluster-based non-parametric permutation test (Maris & Oostenveld, 2007). This test looked for a difference in coherence between the empirical and randomly shifted data, contrasted with N=10,000 permutations in which the 'empirical' data was randomly selected from the group of empirical and randomly-shifted traces. For each frequency and each channel, a T-statistic was computed on the difference between empirical and randomly shifted data using a dependent-samples regression. The test statistic was computed as the maximum cluster size in each permutation (cluster threshold: $\alpha = .01$, two-tailed). The p-value was calculated as the proportion of permuted cluster statistics that were more extreme than the empirical value.

In previous studies, cortical entrainment to speech is often strongest over auditory cortex (Luo & Poeppel, 2007; Howard & Poeppel, 2010; Luo et al., 2010; Park et al., 2016) and frontal cortex (Kayser et al., 2015). We tested coherence to sign language in two analogous

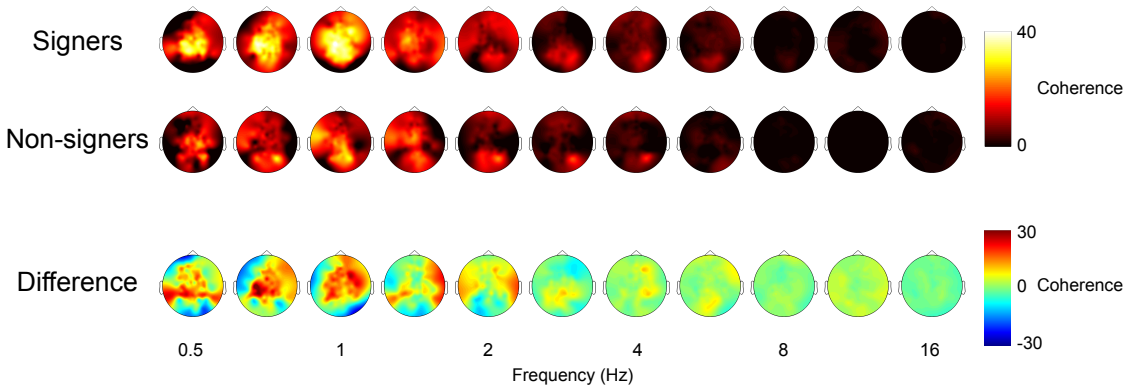

Figure 2.5: Comparison of coherence across signing and non-signing participants. Coherence was robustly present in both groups of participants, regardless of experience with sign language. The top row shows coherence in each frequency bin for fluent signers, and the middle row shows coherence for non-signers. The bottom row shows the difference in coherence between the two groups. Positive values (shown in red) indicate stronger coherence in signers, and negative values (blue) indicate stronger coherence in non-signers.

regions of interest (ROIs): at occipital channels, and at frontal channels.

To test whether knowledge of ASL influenced the strength of entrainment, we performed cluster permutation analyses on the difference in coherence between signers and non-signers at the occipital and frontal ROIs. For each participant, we normalized the data by computing the Z-score of their empirical coherence against the randomly shifted baseline coherence values. Z-scored empirical coherence was then compared between groups using cluster permutation tests.

2.6.6 Data deposition

All EEG data and samples of speech and sign language are available on an Open Science Framework repository at `https://osf.io/5pktz/`.

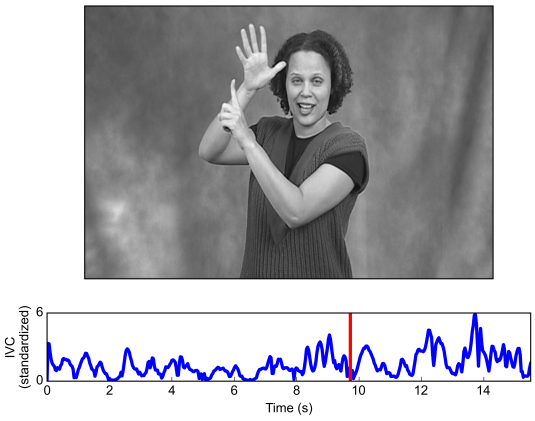

Figure 2.6: See online version of the published paper for example video illustrating how the Instantaneous Visual Change (IVC) tracks visual movement during sign language.

Table 2.1: Samples of sign language.

Sample name	**Language**	**Duration (s)**	**Frame rate**
Scottish Storytelling Centre - British Sign Language storytelling bursary awarded	BSL	758	25
Visual BSL Workshop	BSL	354	25
DIE TAUBE ZEIT-MASCHINE	DGS	120	23.976
Integrationsmaßnahme für taube und schwerhörige Arbeitssuchende oder arbeitslose Personen	DGS	165	25
Ausbildung zur Gebärdensprachdozentin–lehrerin und zum Gebärdensprach-dozentenlehrer	DGS	138	50
Auslan Storybooks - Spider Ants	Auslan	202	25
Auslan Storybooks - A Whole New Star	Auslan	137	25
Auslan Storybooks - Channel 801	Auslan	329	25
Bonnie's Introduction	ASL	279	30
Dummy	ASL	302	30
Mama Speaks	ASL	250	30
CODA Is My Deaf Club	ASL	76	30
Kondima	ASL	590	30
Little Feet	ASL	522	30

Table 2.2: Samples of spoken language.

Sample name	**Language**	**Duration (s)**
121101_fiction_sedaris	English	360
c1	English	360
c10	English	360
alombredesjeunesfillesenfleurs_1_01_proust_64kb	French	360
20000mijl_01_verne_64kb	Dutch	360
fairytale003_02_szelikekiralykisasszony_anonymus_dii_64kb	Hungarian	360
fairytale003_04_desteufelsrussigerbruder_grimm_elli_64kb	German	360
fairytale003_15_sapateiroeosgnomos_brothersgrimm_aw_64kb	Portuguese	247
7thanniversary_14_ookamitonanahikinokodomoyagi_um_64kb	Japanese	360
7thanniversary_15_karasunohokutoshichisei_um_64kb	Japanese	360
tabaialistibdad_08_alkawakibi	Arabic	360
chinese_congcong_zhu_sn_64kb	Mandarin	201

CHAPTER 3

CORTICAL RESONANCE TO NON-ISOCHRONOUS SEQUENCES

This chapter is in preparation for submission to a peer-reviewed journal. Brookshire, G. & Casasanto, D. (In preparation.)

3.1 Abstract

During perception, animals predict both the content and timing of future events. According to common assumptions, temporal predictions rely on cortical oscillatory resonance to roughly isochronous stimulus rhythms. Consistent with this proposal, when people perceive quasi-periodic stimuli such as speech or music, electrophysiological activity synchronizes to bursts in stimulus energy. However, the brain also forms temporal predictions of non-isochronous rhythms, despite the fact that they cannot be matched by a simple oscillation. Can cortical synchronization to non-isochronous sequences be explained by oscillatory resonance, or is some other predictive mechanism necessary to account for these results? We addressed this question in two human electroencephalography (EEG) experiments and a computational model. We computed mutual information between EEG recordings and stimulus onsets while participants were presented with three types of sequences: isochronous, predictable but non-isochronous, and random. We found that although cortical synchronization was strongest during isochronous sequences, synchronization was stronger during predictable non-isochronous sequences than during random sequences. This pattern emerged in EEG activity from the delta band (2-4 Hz) through the beta-band (12-25 Hz), was present during both visual and auditory perception, and could not be attributed to oscillations at a harmonic frequency of the stimuli. The computational model of oscillatory resonance indicated that this pattern of predictive information could be driven by cortical resonance

to the predictable non-isochronous sequences. Together, these results show that oscillatory resonance is sufficient to produce cortical synchronization to non-isochronous sequences, so long as they are predictable. Some neural activity that appears to be predictive may depend on stimulus-driven resonance.

3.2 Significance statement

Temporal predictions are fundamental to many behaviors, yet the neural mechanisms supporting these predictions remain poorly understood. How does the brain predict when a future event will happen? When people perceive quasi-periodic sequences, neural oscillations synchronize to pulses in the sequences. Temporal predictions are often thought to rely on oscillatory resonance in cortical networks, analogous to resonance in other physical and biological systems. However, the brain also synchronizes to non-isochronous sequences, raising the possibility that temporal expectations may rely on a non-oscillatory mechanism. Here we show that resonance in a bank of oscillators is sufficient to generate predictive neural activity to repeating non-isochronous rhythms. Resonance can generate behavior that appears "smart" despite being solely stimulus-driven.

3.3 Introduction

During perception, animals anticipate future events (Rao & Ballard, 1999; Friston, 2005; Clark, 2013). In sensory cortex, predictions of the content of upcoming events rely on top-down modulation in the alpha (8-12 Hz) and beta (12-25 Hz) bands (Bastos et al., 2012; van Kerkoerle et al., 2014), and can be triggered by pattern completion in the hippocampus (Hindy, Ng, & Turk-Browne, 2016).

In addition to predicting *what* events will happen, animals make predictions of *when* those events will happen (Nobre & van Ede, 2018). How does the brain form these temporal

predictions? A growing body of research suggests that temporal expectations may rely on neural oscillations. Both neural excitability and perceptual sensitivity depend on the phase of oscillatory activity across a broad range of frequencies (Jacobs et al., 2007; Haegens, Nácher, Luna, Romo, & Jensen, 2011; Ng et al., 2012; Henry & Obleser, 2012). To maximize sensitivity to temporally predictable events, the brain may adjust the phase of ongoing oscillations to ensure that perceptual gain is highest when important information arrives (Lakatos et al., 2008; Arnal & Giraud, 2012; Giraud & Poeppel, 2012). Consistent with this proposal, when people perceive quasi-periodic stimuli such as language or music, cortical activity in the delta-theta range (1-8 Hz) synchronizes to bursts in stimulus energy (Ahissar et al., 2001; Brookshire et al., 2017; Doelling & Poeppel, 2015; Lakatos et al., 2008; Luo & Poeppel, 2007).

Cortical synchronization is often thought to rely on oscillatory resonance to roughly isochronous stimulus rhythms (Arnal & Giraud, 2012; Lakatos et al., 2013): When people perceive a roughly periodic sequence, networks in the brain begin to resonate at the frequency of fluctuations in the stimulus. However, certain findings call this proposal into question. Activity in the delta band (0.5-4 Hz) synchronizes to predictable visual stimuli even when those stimuli do not follow an isochronous rhythm (Breska & Deouell, 2017). Furthermore, the brain shows a consistent response to the end of predictable non-isochronous sequences, but not to random sequences (Andreou et al., 2015). Finally, visual cortex synchronizes to low-frequency fluctuations of movement in sign language, even though sign language is not strongly periodic (Brookshire et al., 2017). Together, these studies show that the brain synchronizes to non-isochronous sequences.

Because a single periodic oscillator cannot entrain to non-isochronous sequences, these findings appear to contradict the notion that oscillatory resonance drives cortical synchronization. Can cortical synchronization to non-isochronous sequences can be explained by oscillatory resonance alone, or is some additional predictive mechanism necessary? In this

paper, we address this question using two human electroencephalography (EEG) experiments and a computational model. We test for cortical synchronization to visual and auditory sequences in which the intervals between stimuli are (i) isochronous, (ii) non-isochronous but grouped into temporally predictable chunks, and (iii) random. Using an information theoretic framework, we show that cortical activity carries predictive information about the interval until the next stimulus. Results show that this predictive information is strongest for isochronous sequences, and stronger for predictable non-isochronous sequences than for random sequences. Finally, we simulate responses to our stimulus sequences in a bank of resonators, and find that our EEG results closely match the responses generated by oscillatory resonance.

3.4 Materials and Methods

3.4.1 Experimental Design

Human participants attended to sequences that varied in their temporal structure. Separate groups of participants watched sequences of abstract images (Exp. 1) and listened to sequences of frequency-modulated sine tones (Exp. 2).

Participants were presented with sequences that were (1) isochronously periodic, (2) temporally predictable but non-isochronous, and (3) random (Fig. 3.1B). Stimulus onset asynchronies (SOAs) comprised 4 approximately logarithmically-spaced values (100, 167, 300, 500 ms; 210 Hz) spanning a similar frequency band as variability in speech (Chandrasekaran et al., 2009; Edwards & Chang, 2013; Ding et al., 2017; Pellegrino et al., 2011) and sign language (Brookshire et al., 2017; Bellugi & Fischer, 1972; Hwang, 2011). To ensure that these sequences could not be subdivided with an oscillation at the frequency of the shortest SOA, the 4 SOAs were not integer multiples of the shortest SOA. The distribution of SOAs was matched across conditions, allowing us to directly compare neural responses across

conditions.

In the isochronous condition, all stimuli within a block were separated by the same SOA, giving uniform and predictable timing. In the predictable non-isochronous condition, we created a chunk of the 4 SOAs. This chunk was then repeated back-to-back. Because every SOA uniquely predicts the following SOA, each stimulus onset is predictable despite being non-isochronous. In the Random condition, the SOAs were randomly shuffled. The condition order was randomly determined for each participant.

Participants were instructed to attend to sequences of stimuli (images in Exp. 1, sounds in Exp. 2), and to press the space bar whenever a specific catch stimulus appeared. Participants were asked to maintain fixation on a white dot in the center of the computer screen.

Participants were presented with 4 blocks of 700 trials in each condition. Each block lasted for approximately 3 min. In the isochronous condition, each block included only one SOA, and the order of blocks was randomly shuffled between participants. Between each block, participants rested briefly and continued at their own pace.

3.4.2 Participants

All participants reported no history of epilepsy, brain surgery, or traumatic brain injuries. Informed consent was obtained before beginning the experiment, and participants were paid $20 per hour for their participation. All procedures were approved by the local institutional review board.

3.4.3 EEG analysis and preprocessing

EEG analyses were performed in Matlab with custom software and the FieldTrip package (Oostenveld, Fries, Maris, & Schoffelen, 2011). Electrode movement artifacts were manually identified and rejected. Artifacts from blinks and eye-movements were identified and removed using independent component analysis. Electrodes with excessive noise were rejected prior

to further analysis. The first 10 s of each block were excluded from all analysis.

The brain synchronizes to hierarchically-organized groups of stimuli with slow oscillations at the frequency of the group (Zhang & Ding, 2017). We ensured that our results could not be influenced by slow entrainment at the frequency of the predictable non-isochronous SOA-chunks (0.94 Hz) by filtering all EEG data above that frequency (2 Hz). EEG signals were bandpass filtered from 2–40 Hz, and linear trends were removed. For analyses within separate frequency bands, EEG signals were band-pass filtered using Butterworth filters.

3.4.4 Mutual information analysis

We used mutual information (MI) to examine cortical expectations about when the next stimulus would appear. Mutual information is defined as

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)\, p(y)} \right) \tag{3.1}$$

where $p(x, y)$ is the joint distribution of X and Y, $p(x)$ is the marginal distribution of X, and $p(y)$ is the marginal distribution of Y. We examined MI between stimulus onsests and EEG activity in each channel separately. MI was computed at varying lags between the EEG and stimulus sequences, analogous to cross-correlations. The MI at a given lag indicates how much information the EEG signal carries about whether a stimulus will appear at that lag. For example, MI at a lag of -0.1 s shows how much information the EEG signal carries about stimulus onsets 0.1 s in the future. We call this measure *cross-MI*.

Before calculating MI, EEG signals for each channel were discretized into 10 equally populated bins. MI was calculated at lags from -0.5 to 0.5 s. Cross-MI was calculated within each block of stimuli, and then averaged across blocks within the same experimental condition.

3.4.5 Statistical analysis

Experimental conditions were compared using within-subjects permutation tests ($k = 10^5$). In each permutation, the conditions labels were randomly shuffled within each participant, and the mean difference between shuffled data was computed. The empirical between-condition mean difference was then compared with the distribution of permuted mean differences. The p-value is computed as the proportion of permuted values greater than or equal to the empirical value.

3.4.6 Exp. 1: Visual sequences

Participants

We recorded EEG from $N = 12$ participants who were recruited from the University of Chicago community and reported corrected-to-normal vision. We excluded data from one participant due to a faulty recording, leaving $N = 11$ in all analyses.

EEG recording

EEG was digitized at 250 Hz using a 128-channel net (Electrical Geodesics Inc., Eugene, OR). Impedances were reduced to $< 50\text{k}\Omega$ before each block. Stimulus onsets were measured using a photo-diode, and coded as impulses (a sequence of 0s with 1s at each stimulus onset). EEG channels from the face and the back of the neck were excluded before any analysis, leaving 103 channels in all analyses.

Visual stimuli

Participants viewed sequences of 5 different colorful, circular kaleidoscopic images (Fig. 3.1A) created using NodeBox (https://www.nodebox.net/code/index.php/Home) and ImageMagick (http://imagemagick.org/script/index.php). One image was chosen as the catch stimulus

for all participants. Images were presented at the center of the screen and subtended 4° of the visual field. Stimuli were presented against a gray background, and were accompanied by a round fixation point (0.2°) at the center of the screen. Stimulus presentation was controlled using the PsychoPy library (Peirce, 2007) in Python. Participants sat approximately 65 cm from the 21.5-inch LED-LCD computer monitor.

3.4.7 Exp. 2: Auditory sequences

Participants

We recorded EEG from $N = 14$ participants who reported no hearing impairments and were recruited from the Cornell University community. One participant was excluded before any analysis due to excessive noise in the EEG signals, leaving $N = 13$ participants in all analyses.

EEG recording

EEG was recorded at 1024 Hz using a 128-channel system (Biosemi; Amsterdam, Netherlands). Offsets were reduced to < 30 mV before recording.

Auditory stimuli

Participants listened to sequences of frequency-modulated sounds (Fig. 3.3A) created using SuperCollider (https://supercollider.github.io/). One sound was chosen as the catch stimulus for all participants. Sounds were presented through desktop speakers at a comfortable volume (70 dB SPL). To minimize artifacts from eye movements, participants were asked to maintain fixation on a white dot in the center of a gray computer screen.

3.4.8 Computational model of oscillatory resonance

We created a computational model to determine whether a bank of resonating filters could produce predictive activity in response to non-isochronous sequences. This model was implemented using custom code in Python.

We modeled EEG activity as the sum of responses from a bank of real-valued wavelet filters. Wavelets were created for 20 logarithmically-spaced frequencies from 2-40 Hz, matching the range of frequencies included in our EEG analysis. Each wavelet was modeled as a sine wave convolved by the probability density function of the gamma distribution

$$g(x; k, \theta) = \frac{x^{k-1} e^{\frac{-x}{\theta}}}{\Gamma(k)\theta^k} \tag{3.2}$$

where x is the time-step and $\Gamma(k)$ is the gamma function. The shape parameter $k = 2$ provides a tapered decay, and the scale parameter was set to $\theta = 0.2\frac{w}{f}$, where w describes the window length (the number of oscillations in the main impulse response; $w = 2$ in all analyses) and f is the oscillator frequency. The length of the window therefore varied as a function of frequency, with each wavelet encompassing 2 strong oscillations. To ensure that the convolved activity appeared after stimulus onsets, gamma wavelets were padded to double their initial length with zeros at the beginning of the window.

To obtain the simulated resonance response, each gamma wavelet was convolved with the sequences of impulse responses for stimulus onsets, and the convolved time-courses were averaged at each time-point. A small amount of brownian noise was added to the resonance responses to ensure that the responses could be split into equally-sized bins. Cross-MI was then computed between the simulated resonance response and stimulus onsets as in the EEG analysis.

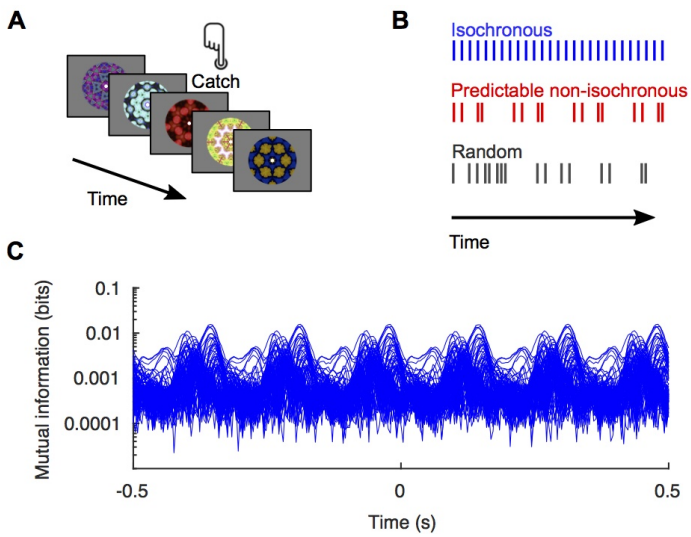

Figure 3.1: Task design. **A**, Participants viewed sequences of abstract kaleidoscopic images, and were instructed to press a button in response to a specific memorized image. **B**, Schematic illustration of the three stimulus conditions. Each vertical line corresponds to a stimulus onset. **C**, Example of cross-mutual-information (cross-MI) while one participant watched isochronous stimuli appearing at 6 Hz. Each line shows cross-MI for a separate EEG channel.

3.5 Results

Participants in Experiment 1 watched sequences of flashing abstract images (Fig. 3.1A) that varied in their periodicity and temporal predictability. Each participant saw three types of sequences: isochronous, predictable but non-isochronous, and random (Fig. 3.1B). We quantified neural predictions of upcoming stimuli using mutual information (MI) between EEG activity and stimulus onsets. We examined MI as a function of the lag between EEG and stimulus onsets, analogous to cross-correlations (cross-MI). The MI at a given lag indicates how much information the EEG signal carries about whether a stimulus will appear at that lag (Fig. 3.1C).

3.5.1 Predictive information during non-isochronous sequences

Across electrodes and pre-stimulus time-points, we found that although MI was strongest during isochronous sequences, MI was stronger during predictable non-isochronous sequences than during random sequences (Fig. 3.2A). To focus on cortical activity that reflects temporal expectations about future stimulus events, we averaged MI over pre-stimulus lags (Fig. 3.2B). Collapsing over all EEG channels, this predictive information during isochronous sequences was stronger than during either predictable non-periodic ($p = .0005$) or random sequences ($p = .0005$). Importantly, MI was also stronger during predictable non-isochronous sequences than during random sequences ($p = .0009$; Fig. 3.2C). This result shows that cortical activity encodes temporal predictions for non-isochronous sequences in addition to isochronous sequences.

To test whether temporal predictions of isochronous and non-isochronous sequences rely on different areas of the brain, we compared the scalp topography of predictive information across the three stimulus conditions. In all three stimulus conditions, predictive information emerged over a broad region of the scalp, with the highest MI at occipital electrodes (Fig. 3.2D). Furthermore, the differences in scalp topography are centered over occipital channels both when comparing isochronous to predictable non-isochronous sequences, and when comparing predictable non-isochronous to random sequences. The similarity of patterns across conditions suggests that temporal predictions of isochronous and non-isochronous sequences may rely on a common neural substrate. Two caveats limit the interpretability of this result, however. First, this result may constitute a null effect due to the poor spatial resolution of EEG. Second, these scalp topographies may reflect predictive activity over visual cortex that arises following temporal processing in separate brain regions for the three conditions.

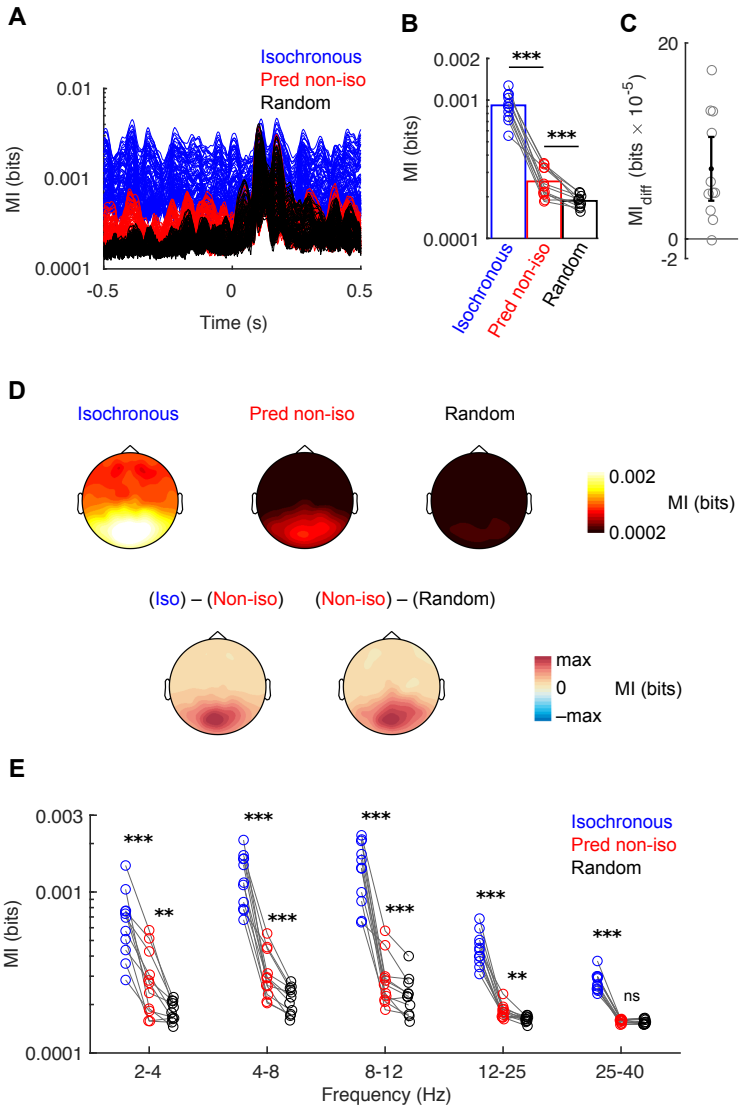

Figure 3.2: Cortical activity carries broadband predictive information about visual stimulus onsets. **A**, Cross-mutual-information (MI) between EEG activity and stimulus onsets averaged across participants. Each trace shows cross-MI from a separate EEG channel. **B**, Predictive information. Cross-MI averaged over all electrodes and all pre-stimulus lags (-0.5–0 s). Each point shows the value for one participant, and black lines connect observations from the same subject. Barplot heights show the average across participants. **C**, Difference in predictive information between predictable non-isochronous sequences and random sequences, with one point for each participant. **D**, Scalp topographies of predictive information shown separately for each condition, and for the difference between conditions. Difference plots are scaled to the maximum difference in that comparison. **E**, Predictive information after filtering EEG signals into separate frequency bands. Each point shows data from one participant, as in **B**. Asterisks indicate pairwise differences between conditions. * $p < .05$; ** $p < .005$; *** $p < .001$; *ns*, not significant.

3.5.2 Predictive information in a wide range of frequency bands

To determine whether cortical phase-locking depends on neural activity within a particular frequency band, we computed MI after band-pass filtering the EEG signals into separate frequency bands (delta: 2-4 Hz; theta: 4-8 Hz; alpha: 8-12 Hz; beta: 12-25 Hz; low gamma: 25-40 Hz). We observed the same pattern in all frequencies below the gamma band: pre-stimulus MI was strongest to isochronous sequences (all ps $< .001$), and was stronger to predictable non-isochronous sequences than to random sequences (all ps $< .005$; Fig. 3.2E). Above 25 Hz, cortical predictive information was strongest for isochronous sequences (ps $< .0006$), but did not distinguish between predictable and random non-isochronous sequences ($p = .19$). Predictive information in the brain, therefore, does not occur exclusively in a specific frequency band, but occurs across a wide range of frequencies.

3.5.3 Predictive information to auditory sequences

How does the brain encode temporal expectations across sensory modalities? Visual scenes are relatively stable in time, which may allow visual perception to rhythmically sample the environment without losing information; auditory scenes, by contrast, are dynamic over time, and therefore require adaptive sampling to match bursts of information in the environment (Zoefel & VanRullen, 2017). Perhaps visual perception, therefore, is strongly entrained by isochronous sequences, whereas auditory perception is equally entrained by all predictable sequences, regardless of whether they are isochronous. In Experiment 2, we tested this prediction by presenting a new group of participants with sequences of synthesized sounds (Fig. 3.3A) with temporal structure identical to that used in in Experiment 1. We found that results were similar across visual and auditory perception: predictive information during isochronous sequences was stronger than during predictable non-isochronous sequences ($p = .0001$) and stronger than during random sequences ($p = .0001$). Importantly, MI during predictable non-isochronous sequences was also stronger than during random sequences ($p =$

.0006; Fig. 3.3B-C). This result shows that, across sensory modalities, cortex generates temporal expectations to predictable sequences even in the absence of isochronous sequences. This pattern was independently present in the theta ($p = .004$), alpha ($p = .001$), and beta bands ($p = .046$), but not in the delta ($p = .22$) or low gamma bands ($p = .28$; Fig. 3.3E). Unlike MI during perception of visual sequences, overall predictive information was strongest over central and lateral temporal channels (Fig. 3.3D), suggesting that predictive information relating to temporal expectations arises in the perceptual cortex used to process the sensory input.

3.5.4 Computational model of oscillatory resonance

Our EEG experiments demonstrate that cortical activity encodes temporal expectations during predictable non-isochronous sequences. A sine-wave oscillator, however, cannot synchronize to non-isochronous sequences in which SOAs are not integer multiples of a single interval. Can oscillatory resonance account for these data? We developed a computational model of oscillatory resonance to determine whether a stimulus-driven bank of oscillators could generate temporal predictions to predictable non-isochronous sequences. We modeled each oscillator as a sine-wave convolved by a gamma window (gamma wavelets; Fig. 3.4A), and determined the response of each oscillator by convolving these gamma wavelets with the sequences of impulse responses from our EEG experiments. The total resonance response was taken as the average response across oscillators (Fig. 3.4B). Temporal expectations were measured using cross-MI as in the EEG experiments.

This resonance model generated cross-MI that closely matched the EEG data. Although oscillatory resonance produced the highest predictive information during isochronous sequences, predictive information was stronger during predictable non-isochronous sequences than during random sequences (Fig. 3.4C). In the EEG experiments, temporal expectations to predictable non-periodic sequences emerged across a wide range of frequency bands.

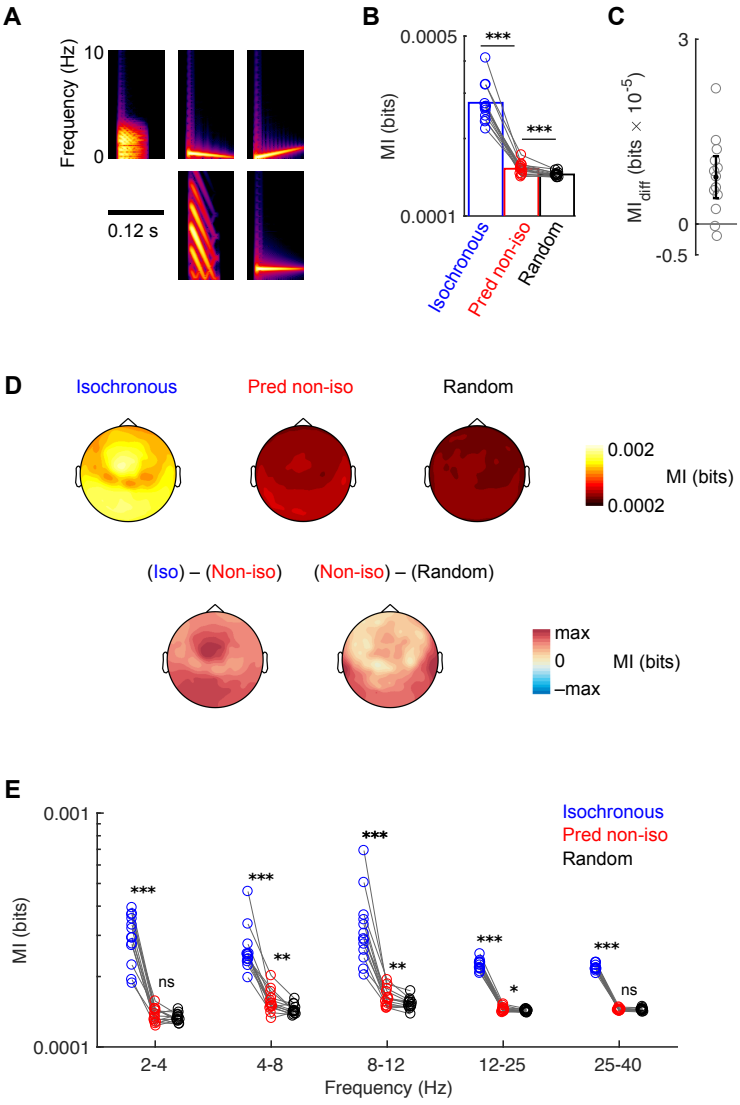

Figure 3.3: Cortical predictive information during auditory sequences. **A**, Auditory stimuli used in the experiment. Spectrograms show power of each stimulus. **B**, Predictive information. Cross-MI averaged over all electrodes and all pre-stimulus lags (-0.5 to 0 s). Each point shows the value for one participant, and black lines connect observations from the same subject. Barplot heights show the average across participants. **C**, Difference in predictive information between predictable non-isochronous sequences and random sequences, with one point for each participant. **D**, Scalp topographies of predictive information shown separately for each condition, and for the difference between conditions. Difference plots are scaled to the maximum difference in that comparison. **E**, Predictive information after filtering EEG signals into separate frequency bands. Each point shows data from one participant, as in **B**. Asterisks indicate pairwise differences between conditions. * $p < .05$; ** $p < .005$; *** $p < .001$; *ns*, not significant.

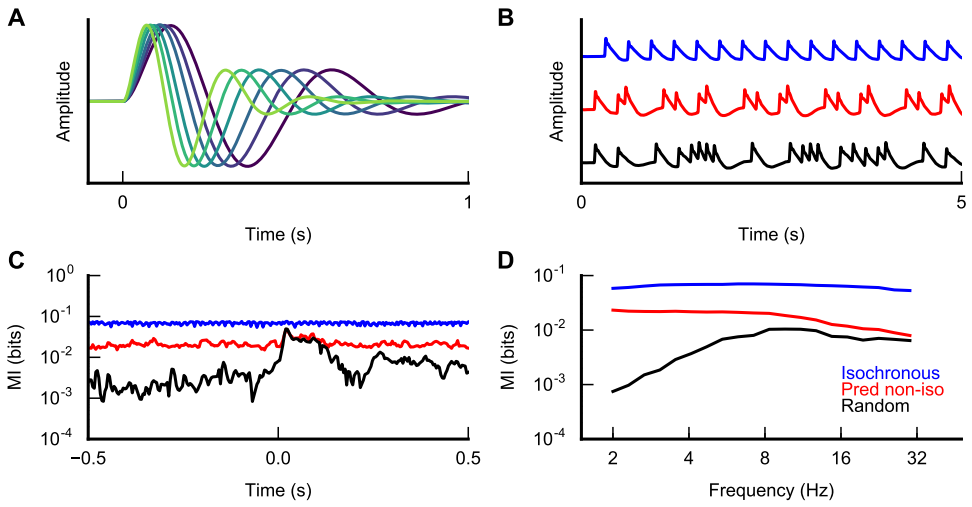

Figure 3.4: Resonance model of temporal expectations. **A**, Examples of gamma wavelets. Resonance responses were obtained by convolving these wavelets with a sequence of stimulus onsets (impulses). **B**, Example responses of the resonance model to the three types of temporal structure. To improve visibility, the traces have been separated on the y-axis. **C**, Cross-MI between responses of the resonance model and stimulus onsets. **D**, Predictive MI as a function of frequency. Cross-MI averaged over pre-stimulus lags in each condition, computed separately for responses from oscillators at each frequency.

To test whether oscillatory resonance can explain synchronization across frequency bands, we examined predictive information in the responses of individual gamma wavelets, computed identically to the EEG analyses. At all frequencies tested (2-40 Hz), we found the same qualitative pattern: predictive information was strongest during isochronous sequences, but stronger during predictable non-isochronous sequences than during random sequences (Fig. 3.4D). Consistent with our EEG results, the difference between predictable non-isochronous and random sequences was pronounced at lower frequencies, and attenuated at higher frequencies. The behavior of this model shows that oscillatory resonance to stimulus onsets is sufficient to account for our findings of cortical synchronization to predictable non-isochronous sequences, without positing any additional mechanism of temporal prediction.

3.6 Discussion

In this study, we tested whether oscillatory resonance can account for cortical predictions of non-isochronous intervals. We found that EEG activity before the onset of a stimulus carried information about when the next stimulus would appear. Importantly, we showed that brain activity encodes temporal expectations not only during isochronous sequences, but also during non-isochronous sequences. We then showed that this type of predictive cortical activity is captured in a computational model of stimulus-driven oscillatory resonance. Oscillatory resonance is therefore sufficient to account for cortical synchronization and temporal predictions to some complex rhythms, even when those rhythms are not isochronous or reliant on subdivision at a higher frequency. Even though oscillatory resonance is a simple, stimulus-driven process, it can account for some neural activity that appears flexible and predictive.

3.6.1 Temporal expectation across frequency bands

Some theories suggest that cortical activity in different frequency bands reflects different aspects of underlying neural computation. For example, a growing body of evidence indicates that alpha oscillations reflect rhythmic bursts of inhibitory activity (Jensen & Mazaheri, 2010). Another theory proposes that alpha/beta band activity encodes top-down predictions, whereas theta and gamma activity encodes bottom-up sensory signals (Bastos et al., 2012; van Kerkoerle et al., 2014). Do temporal expectations of non-isochronous intervals rely on activity in a particular frequency band?

During both visual and auditory perception, we found evidence of temporal expectations across a wide range of frequencies, from the delta band to the beta band. This finding suggests that oscillatory resonance can support temporal expectations across a wide range of frequencies. Oscillatory resonance may act in parallel to any other functional roles of oscillations in particular frequency bands.

3.6.2 Temporal expectation across sensory modalities

Electrophysiological activity in the brain resonates most strongly at particular frequencies, and some studies suggest that different sensory modalities may differ in their 'preferred' resonant frequencies. Oscillations in visual cortex are more strongly entrained by isochronous sequences at 10 Hz than at other frequencies (Herrmann, 2001). In auditory cortex, by contrast, perceptual sensitivity has been found to depend on the phase of 2-6 Hz oscillations, but not on the phase of 10 Hz oscillations (Ng et al., 2012). In addition to resonating at different frequency bands, some studies suggest that the sensory modalities differ in their propensity to resonate at all. Visual cortex resonates at 10 Hz in response to random fluctuation in brightness (VanRullen & Macdonald, 2012), whereas auditory cortex was not found to strongly oscillate in response to random fluctuations in volume (İlhan & VanRullen, 2012).

In the current experiments, by contrast, we found similar patterns of cortical synchronization to auditory and visual sequences. In all frequency bands tested, both sensory modalities showed the strongest predictive information during isochronous sequences. Both sensory modalities also showed stronger predictive information during predictable non-isochronous sequences than during random sequences, and this pattern was present independently in the theta, alpha, and beta bands. In the delta band, we only observed this pattern for visual sequences, but not for auditory sequences. We did not predict the absence of delta-band auditory temporal expectations for predictable non-isochronous intervals. The lack of an effect in this frequency band is difficult to interpret, and seems to conflict with other findings. First, the computational model of resonance predicts strong non-isochronous temporal predictions in the delta-band (Fig. 3.4D), as we found during visual sequences. Second, cortical activity in the delta band robustly synchronizes to non-isochronous sequences in speech and sign language.

Building on findings of cortical synchronization to naturalistic language, our results are

consistent with the idea that cortical synchronization may not depend on filtering properties of sensory cortex, but on informational characteristics of the stimulus (Brookshire et al., 2017). In visual cortex, delta- and theta-band activity synchronizes to movements of the mouth when people watch videos of speech (Luo et al., 2010; Park et al., 2016), to aggregated visual movement when people watch videos of sign language (Brookshire et al., 2017), and to non-linguistic sequences when the dominant temporal structure occurs at those frequencies (Cravo et al., 2013). Considered alongside the results of our resonance model, these results suggests that oscillatory resonance may support temporal expectations at the frequencies of information in the stimulus, across sensory modalities.

3.6.3 How does the brain form temporal expectations?

In many theories of neural entrainment, an ongoing neural oscillation adjusts its phase and frequency to match a sequence of events, with a certain phase of the oscillation predicting the occurrence of each new event (Large & Jones, 1999; Arnal & Giraud, 2012; Lakatos et al., 2013). Although single-oscillator theories succinctly describe cortical synchronization to isochronous sequences, they cannot account for synchronization to the types of sequences in our experiments: predictable non-isochronous sequences in which SOAs are not related by integer ratios. What underlying neural mechanisms support predictions of non-isochronous intervals? An answer to this question has two parts: *where* in the brain temporal expectations are generated, and *how* those temporal expectations are generated.

Our EEG experiments provide only limited evidence about *where* in the brain temporal expectations are computed. We found that the scalp topography of predictive information differed between visual and auditory sequences; in each experiment, predictive information was consistent with activity in perceptual cortex for that sensory modality. This finding shows that temporal predictions correspond to activity in perceptual cortex. However, these data leave open a further question: Are temporal predictions initially computed in the

perceptual cortex that processes the sensory input, or in some other brain area that operates across sensory modalities (Ivry & Schlerf, 2008; Nobre & van Ede, 2018)? Future studies with higher spatial resolution will be necessary to answer this question.

Turning to the question of *how* the brain coordinates temporal expectations, we consider four types of neural processes that may support expectations of predictable non-isochronous intervals: (i) non-oscillatory interval-based timing; (ii) phase offset in multiple oscillators at a single frequency; (iii) fluctuations in oscillatory power; (iv) aggregate activity in a multi-frequency bank of resonators.

Non-oscillatory interval-based timing

First, the brain may use a non-oscillatory mechanism to predict each interval as it appears (Breska & Deouell, 2017). Whereas cerebral cortex and the basal ganglia have been implicated in entrainment to isochronous rhythms (Nobre & van Ede, 2018), the cerebellum plays a crucial role in representing non-rhythmic timing (Ivry & Schlerf, 2008; Breska & Ivry, 2016). Our modeling results show that oscillatory resonance can account for cortical synchronization for sequences more complex than simple isochronous rhythms. A cerebellar non-oscillatory process may represent timing in situations where oscillatory resonance could not, such as isolated events with no consistent rhythmic context (Breska & Ivry, 2016).

Phase offset in multiple oscillators at a single frequency

Second, the brain may encode expectations of complex repeating rhythms by using multiple oscillations at the same frequency but different phase offsets (Andreou et al., 2015). In this case, the oscillations occur at the frequency of repeating rhythmic chunks, and the phase of peak excitability for each oscillation coincides with a different event. Although this model is consistent with prior findings, it cannot account for the results of this study. We found predictive information across a wide range of frequencies, including frequencies that are

approximately 20 times higher than the frequency of the rhythmic chunks in our predictable non-isochronous sequences. Our findings, therefore, suggest that cortical synchronization could rely on oscillators at different frequencies.

Modulations of oscillatory power

Third, the brain may implement temporal predictions by modulating the power of higher-frequency oscillations. Oscillatory power in the alpha and beta bands is reduced before a predictable event occurs (Zanto et al., 2011; van Ede, de Lange, Jensen, & Maris, 2011; Rohenkohl & Nobre, 2011; Fujioka, Trainor, Large, & Ross, 2012), suggesting that alpha/beta desynchronization may play an important role in coordinating temporal expectations. Results from our experiments, however, do not reflect changes in oscillatory power, since mutual information between stimulus onsets and activity in a particular frequency band indicates a consistent EEG potential at a particular lag. Phase-locked activity across a wide range of frequencies may play a role in temporal expectation beyond the role of oscillatory desynchronization.

Resonators at multiple frequencies

Finally, cortical synchronization to predictable non-isochronous sequences may occur via oscillatory resonance in a bank of oscillators with different frequencies. Here we provide evidence that oscillatory resonance in such a bank of oscillators produces activity similar to EEG activity in human participants. We find that aggregated resonance across a bank of wavelet filters, like our EEG data, can generate temporal expectations during surprisingly complex sequences.

CHAPTER 4

CONCLUSION

In four EEG experiments, a corpus analysis, and a computational model, I examined how the brain uses cortical synchronization to predict when new information will appear. By demonstrating that visual cortex synchronizes to videos of sign language at < 5 Hz, I show that synchronization to language does not depend on processes that are specific to audition or spoken language. Phase-locking may be a general cortical mechanism that maximizes sensitivity to informational peaks in time-varying signals. Then, I show that electrophysiological activity in the brain encodes temporal predictions to complex, non-isochronous rhythms, and that this pattern of neural activity is captured by a simple model of oscillatory resonance. Oscillatory resonance can account for temporal predictions of surprisingly complex rhythms.

In the natural world, animals must anticipate the timing of events on a wide range of timescales, and with a wide range of temporal structure. Environmental stimuli can be quasi-periodic, aperiodic, or they can require temporal predictions of isolated intervals. This diversity of contexts most likely requires a diversity of neural strategies, with different strategies specialized for different situations. Cortical synchronization is one tool that allows animals to accurately anticipate future events, and it may rely on oscillatory resonance. Outside of the nervous system, oscillatory resonance governs the behavior of many biological and physical systems. These results suggest that resonance may also be able to account for complex behavior in the human brain.

REFERENCES

Ahissar, E., & Ahissar, M. (2005). Processing of the temporal envelope of speech. In R. König, P. Heil, E. Budinger, & H. Scheich (Eds.), *The auditory cortex: A synthesis of human and animal research* (p. 295-313). London: Lawrence Erlbaum Associates, Inc.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367–13372.

Andreou, L.-V., Griffiths, T. D., & Chait, M. (2015). Sensitivity to the temporal structure of rapid sound sequencesan meg study. *NeuroImage*, *110*, 194–204.

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398.

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*(4), 695–711.

Bellugi, U., & Fischer, S. (1972). A comparison of sign language and spoken language. *Cognition*, *1*(2), 173–200.

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS Biology*, *15*(2), e2001665.

Breska, A., & Ivry, R. B. (2016). Taxonomies of timing: where does the cerebellum fit in? *Current Opinion in Behavioral Sciences*, *8*, 282–288.

Brookshire, G., Lu, J., Nusbaum, H. C., Goldin-Meadow, S., & Casasanto, D. (2017). Visual cortex entrains to sign language. *Proceedings of the National Academy of Sciences*, *114*(24), 6352–6357.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing eeg oscillations predicts visual perception. *Journal of Neuroscience*, *29*(24), 7869–7876.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926–1929.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*(7), e1000436.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573.

Corina, D. P., San Jose-Robertson, L., Guillemin, A., High, J., & Braun, A. R. (2003). Language lateralization in a bimanual language. *Journal of Cognitive Neuroscience*, *15*(5), 718–730.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *Journal of Neuroscience*, *33*(9), 4002–4010.

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *The Journal of Neuroscience*, *35*(42), 14195–14204.

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664–1670.

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, *88*, 41–46.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*.

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

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

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768.

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences*, *112*(45), E6233–E6242.

Edwards, E., & Chang, E. F. (2013). Syllabic ($\sim$ 2–5 hz) and fluctuation ($\sim$ 1–10 hz) ranges in speech and auditory processing. *Hearing Research*, *305*, 113–134.

Emmorey, K., McCullough, S., Mehta, S., & Grabowski, T. J. (2014). How sensory-motor systems impact the neural organization for language: direct contrasts between spoken and signed language. *Frontiers in Psychology*, *5*, 484.

Emmorey, K., Mehta, S., & Grabowski, T. J. (2007). The neural correlates of sign versus word production. *NeuroImage*, *36*(1), 202–208.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological sciences*, *360*(1456), 815–836.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127.

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *Journal of Neuroscience*, *32*(5), 1791–1802.

Ghazanfar, A. A., Morrill, R. J., & Kayser, C. (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Sciences*, *110*(5), 1959–1963.

Ghazanfar, A. A., Takahashi, D. Y., Mathur, N., & Fitch, W. T. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology*, *22*(13), 1176–1182.

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

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech: a syllable-centric perspective. *Journal of Phonetics*, *31*(3), 465–485.

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

Gutierrez-Sigut, E., Daws, R., Payne, H., Blott, J., Marshall, C., & MacSweeney, M. (2015). Language lateralization of hearing native signers: A functional transcranial doppler sonography (ftcd) study of speech and sign production. *Brain and Language*, *151*, 23–34.

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). α-oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences*, *108*(48), 19377–19382.

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., & Bäuml, K.-H. (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage*, *37*(4), 1465–1473.

Heideman, S. G., van Ede, F., & Nobre, A. C. (2016). Early behavioural facilitation by temporal expectations in complex visual-motor sequences. *Journal of Physiology-Paris*, *110*(4), 487–496.

Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences*, *111*(41), 14935–14940.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, *109*(49), 20095–20100.

Herrmann, C. S. (2001). Human eeg responses to 1–100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, *137*(3-4), 346–353.

Hickok, G., Bellugi, U., & Klima, E. S. (1998). The neural organization of language: evidence from sign language aphasia. *Trends in cognitive sciences*, *2*(4), 129–136.

Hickok, G., Farahbod, H., & Saberi, K. (2015). The rhythm of perception: entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychological Science*, *26*(7), 1006–1013.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665.

Horton, C., Srinivasan, R., & D'Zmura, M. (2014). Envelope responses in single-trial eeg indicate attended speaker in a 'cocktail party'. *Journal of Neural Engineering*, *11*(4), 046015.

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, *104*(5), 2500–2511.

Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593.

Hwang, S.-O. K. (2011). *Windows into sensory integration and rates in language processing: Insights from signed and spoken languages* (Unpublished doctoral dissertation). University of Maryland.

İlhan, B., & VanRullen, R. (2012). No counterpart of visual perceptual echoes in the auditory system. *PloS one*, *7*(11), e49287.

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception.

Trends in Cognitive Sciences, *12*(7), 273–280.

Jacobs, J., Kahana, M. J., Ekstrom, A. D., & Fried, I. (2007). Brain oscillations control timing of single-neuron activity in humans. *Journal of Neuroscience*, *27*(14), 3839–3844.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in Human Neuroscience*, *4*, 186.

Kayser, S. J., Ince, R. A., Gross, J., & Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *The Journal of Neuroscience*, *35*(44), 14691–14701.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *Journal of Neuroscience*, *30*(2), 620–628.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

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*(5872), 110–113.

Lakatos, P., Musacchia, G., OConnel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*(4), 750–761.

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*(3), 1904–1911.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, *22*(11), 1000–1004.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119.

Leonard, M. K., Ramirez, N. F., Torres, C., Travis, K. E., Hatrak, M., Mayberry, R. I., & Halgren, E. (2012). Signed words in the congenitally deaf evoke typical late lexicosemantic responses with no early visual responses in left superior temporal cortex. *The Journal of Neuroscience*, *32*(28), 9700–9705.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*(8), e1000445.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–1010.

MacSweeney, M., Capek, C. M., Campbell, R., & Woll, B. (2008). The signing brain: the neurobiology of sign language. *Trends in Cognitive Sciences*, *12*(11), 432–440.

MacSweeney, M., Waters, D., Brammer, M. J., Woll, B., & Goswami, U. (2008). Phonological processing in deaf signers and the impact of age of first language acquisition. *NeuroImage*, *40*(3), 1369–1379.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: prestimulus α phase predicts visual awareness. *Journal of Neuroscience*, *29*(9), 2725–2732.

Mathewson, K. E., Prudhomme, C., Fabiani, M., Beck, D. M., Lleras, A., & Gratton, G. (2012). Making waves in the stream of consciousness: entraining oscillations in eeg alpha and fluctuations in visual awareness with rhythmic visual stimulation. *Journal of Cognitive Neuroscience*, *24*(12), 2321–2333.

Meier, R. P. (2002). Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In R. P. Meier, K. Cormier, & D. Quinto-Pozos (Eds.), *Modality and structure in signed and spoken languages* (p. 1-26). Cambridge: Cambridge University Press.

Newman, A. J., Supalla, T., Hauser, P., Newport, E. L., & Bavelier, D. (2010). Dissociating

neural subsystems for grammar by contrasting word order and inflection. *Proceedings of the National Academy of Sciences*, *107*(16), 7539–7544.

Ng, B. S. W., Schroeder, T., & Kayser, C. (2012). A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *Journal of Neuroscience*, *32*(35), 12268–12276.

Nobre, A. C., & van Ede, F. (2018). Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience*, *19*(1), 34.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., ... Lalor, E. C. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial eeg. *Cerebral Cortex*, *25*(7), 1697–1706.

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

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

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320.

Peelle, J. E., Gross, J., & Davis, M. H. (2012). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378–1387.

Peirce, J. W. (2007). Psychopypsychophysics software in python. *Journal of Neuroscience*

Methods, *162*(1-2), 8–13.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). Across-language perspective on speech information rate. *Language*, *87*(3), 539–558.

Petitto, L. A., Holowka, S., Sergio, L. E., Levy, B., & Ostry, D. J. (2004). Baby hands that move to the rhythm of language: hearing babies acquiring sign languages babble silently on the hands. *Cognition*, *93*(1), 43–73.

Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J., & Maiste, A. C. (1987). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *The Journal of the Acoustical Society of America*, *82*(1), 165–178.

Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2012). Neural entrainment to rhythmically presented auditory, visual, and audio-visual speech in children. *Frontiers in Psychology*, *3*(216), 10–3389.

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*(1), 79.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2017). Neural entrainment to speech modulates speech intelligibility. *Current Biology*.

Rohenkohl, G., & Nobre, A. C. (2011). Alpha oscillations related to anticipatory attention follow temporal expectations. *Journal of Neuroscience*, *31*(40), 14076–14084.

Romei, V., Gross, J., & Thut, G. (2012). Sounds reset rhythms of visual cortex and corresponding human visual perception. *Current Biology*, *22*(9), 807–813.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106–113.

Spaak, E., de Lange, F. P., & Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *Journal of Neuroscience*, *34*(10), 3536–3544.

Timm, R. L., & Timm, D. (2008 (Pelican Ave Inc., Fremont, CA) American sign langauge performance art). *The Rosa Lee Show: 2004-2008* [DVD].

Van Dijk, H., Schoffelen, J.-M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *The Journal of Neuroscience*, *28*(8), 1816–1823.

van Ede, F., de Lange, F., Jensen, O., & Maris, E. (2011). Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha-and beta-band oscillations. *Journal of Neuroscience*, *31*(6), 2016–2024.

van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., Van Der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences*, *111*(40), 14332–14341.

VanRullen, R., & Macdonald, J. S. (2012). Perceptual echoes at 10 Hz in the human brain. *Current Biology*, *22*(11), 995–999.

VanRullen, R., Zoefel, B., & İlhan, B. (2014). On the cyclic nature of perception in vision versus audition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1641), 20130214.

Wang, Y., Ding, N., Ahmar, N., Xiang, J., Poeppel, D., & Simon, J. Z. (2011). Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: Meg evidence. *Journal of Neurophysiology*, *107*(8), 2033–2041.

Williams, J. T., Darcy, I., & Newman, S. D. (2015). Modality-independent neural mechanisms for novel phonetic processing. *Brain Research*, *1620*, 107–115.

Xiang, J., Simon, J., & Elhilali, M. (2010). Competing streams at the cocktail party: exploring the mechanisms of attention and temporal integration. *Journal of Neuroscience*, *30*(36), 12084–12093.

Zanto, T. P., Pan, P., Liu, H., Bollinger, J., Nobre, A. C., & Gazzaley, A. (2011). Age-related changes in orienting attention in time. *Journal of Neuroscience*, *31*(35), 12461–12470.

Zhang, W., & Ding, N. (2017). Time-domain analysis of neural tracking of hierarchical linguistic structures. *NeuroImage*, *146*, 333–340.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... Schroeder, C. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, *77*(5), 980–991.

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

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*.

Zoefel, B., & VanRullen, R. (2016). Eeg oscillations entrain their phase to high-level features of speech sound. *NeuroImage*, *124*, 16–23.

Zoefel, B., & VanRullen, R. (2017). Oscillatory mechanisms of stimulus processing and selection in the visual and auditory systems: state-of-the-art, speculations and suggestions. *Frontiers in Neuroscience*, *11*, 296.