Electrophysiological Signatures of Multisensory Temporal Processing in the Human Brain

By

David Michael Simon

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

May 11, 2018

Nashville, Tennessee

Approved:

Geoffrey Woodman, Ph.D.

Ramnarayan Ramachandran, Ph.D.

Frank Tong, Ph.D.

Mark Wallace, Ph.D.

Acknowledgements

This work would not have been possible without the thoughtful discussion and input of numerous faculty, friends, staff, and students at Vanderbilt University. In particular, I would like to thank my fellow graduate students: Aaron Nidiffer, Jacob Feldman, and Jean-Paul Noel for fruitful discussions and scientific advice. I would also like to thank my advisor, Dr. Mark Wallace, for his mentorship and for providing the opportunity to work on such interesting scientific endeavors. Additionally, I would like to thank Dr. Geoffrey Woodman, Dr. Ramnarayan Ramachandran, Dr. Frank Tong, and Dr. Vivien Casagrande for their invaluable insight and input as members of my thesis committee. I would also like to thank my previous scientific mentor, Dr. Blythe Corbett for fostering my interests in scientific research and encouraging me to pursue a graduate degree in neuroscience. I would also like to thank the research participants, as without their dedication and effort works such as this one would not be possible.

Table of Contents

List of Tables

List of Figures

Page

List of Abbreviations

ANOVA ..... Analysis of Variance
ASD ..... Autism Spectrum Disorder
BOLD ..... Blood Oxygen Level Dependent
EEG ..... Electroencephalography
ERP ..... Event Related Potential
FMRI ..... Functional Magnetic Resonance Imaging
ICA ..... Independent Component Analysis
IPS ..... Intraparietal Sulcus
ITC ..... Inter-Trial Coherence
LFP ..... Local Field Potential
LIP ..... lateral intraparietal area
MEG ..... Magnetoencephalography
MT/V5 ..... Middle Temporal Area
PCA ..... Principal Component Analysis
PET ..... Positron Emission Tomography
PSS ..... Point of Subjective Simultaneity
SC ..... Superior Colliculus
SJ ..... Simultaneity Judgment
SOA ..... Stimulus Onset Asynchrony
STS ..... Superior Temporal Sulcus
SWS ..... Sine Wave Speech
TACS ..... Transcranial Alternating Current Stimulation
TBW ..... Temporal Binding Window
TD ..... Typically Developing
TFR ..... Time-Frequency Representation
TMS ..... Transcranial Magnetic Stimulation
TOJ ..... Temporal Order Judgment

WPLI... Weighted Phase Lag Index

CHAPTER I

INTRODUCTION

A Multisensory World

We live in a world in which we are continuously bombarded by sensory inputs originating from objects and events in our environment. At any given moment, myriad visual, auditory, somatosensory, gustatory and olfactory inputs are available to the peripheral senses. These sensory inputs carry rich information about the world and are fundamentally important for cognition and behavior, whether it be in the evolutionary context of the brain or in modern life. Everyday activities such as watching television, eating a meal, or speaking with a co-worker all include inputs to multiple sensory systems. Sensory inputs are also important for survival, such as viewing an approaching car, or perhaps hearing its engine. Both of these stimuli carry crucially important information regarding the car's position in space, and both inputs might indicate when it would be unwise to cross a street. Importantly, the auditory and visual inputs in this example are interdependent due to commonalities in their physical origin. Due to common origin, the car grows closer in perceived visual distance in approximate correspondence with increases in the loudness of the engine's roar. The first sight of the car and first moment hearing its engine might also occur simultaneously, as the vehicle turns a corner. Making efficient and correct use of these relationships constitutes an important way for the brain as it seeks to form a single unified and coherent representation of the world around it. This perceptual representation

then becomes the basis of cognition and action, which in this case might be opting to stay on the sidewalk. This example is specific, but combining sensory inputs appropriately based on similar cues is an important general function of the nervous system. This process of combining inputs across sensory modalities is known as multisensory integration, and it has been recognized as a fundamental and important contributor to numerous facets of perception and behavior (G. Calvert, Spence, & Stein, 2004; Murray & Wallace, 2012; Stein, 2012; Stein & Meredith, 1993). With this recognized importance has come the impetus for understanding the fundamental nature of sensory integration in the human brain.

As exemplified by the example of an approaching car, inputs arriving at the sensory periphery frequently have useful commonalities. For the car, several useful relationships are present in the form of spatial and temporal correspondences between the visual and auditory inputs. In the temporal dimension, this takes the form of alignment between the auditory and visual inputs. In other words, the onsets and changes in magnitude of the two sensory inputs happen at approximately the same time. Additionally, due to differences in the velocities of light and sound across space, the visual input also arrives to the retina slightly ahead of the auditory input in this example. This relative structuring between the sensory inputs carries important information about how they are related to each other, and forms an important tool for the nervous system in disambiguating complex environments. The slight temporal lead of the visual input, for example, might suggest to a naïve observer that the visual stimulus plays a causal role for the auditory stimulus, or it might be used to inform distance. Informative cross-modal temporal relationships such as this are ubiquitous in the everyday environment and are deeply important to the brain's ability to construct an appropriate representation of the world. Everyday examples of events with such cross-modal temporal correspondence include seeing an object strike a hard

surface and the sound of the blow, the aforementioned approaching car, and watching mouth movements during speech (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Schwartz & Savariaux, 2014). In each of these cases, the integration of this information across modalities offers substantial facilitation of behavior; In the case of the object strike, reaction times to the impulse like (i.e. rapid onset and offset) inputs are faster (Diederich & Colonius, 2004) and motion perception may be enhanced (Kim, Peters, & Shams, 2012). In the case of the car, approaching audiovisual inputs generate faster responses than their stationary counterparts do, and this might keep a pedestrian out of harm's way (Cappe, Thut, Romei, & Murraya, 2009). Lastly, in the case of audiovisual speech, intelligibility is substantially improved by temporally correlated visual speech compared to auditory speech alone, and this effect is greatly magnified in conditions with acoustic interference (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby, 1954).

Informative stimulus relationships are not limited to the temporal dimension, however, and are frequently observed in other relational qualities of sensory inputs. For example, when the object strikes a surface not only does the auditory input occur in relatively close temporal proximity to the visual strike, but also in close *spatial* proximity – the sound and visual inputs occur from the same point in space. Emanation from a spatially coincident source is similarly true for the aforementioned examples of an approaching car and audiovisual speech. Similar to temporal correspondence, spatial correspondence is an important factor which shapes perception and behavior (L. K. Harrington & Peck, 1998; Slutsky & Recanzone, 2001; Spence & Mcdonald, 2004). Stimuli can also share informative features beyond spatial and temporal correspondence. A finger sliding across a corrugated surface, for example, generates auditory and tactile inputs with a common frequency. Similarly, angular objects generate irregular (i.e. more "rough")

sounds when encountering surfaces. These correspondences are clearly important for the appropriate integration of inputs, as they serve to bias whether inputs should be combined or segregated. Different sensory systems have strengths in different domains, such as vision's superiority to audition in terms of spatial precision, and auditions superior temporal precision (Welch, DuttonHurt, & Warren, 1986). These relative strengths mean that correspondences between the senses must also be evaluated in the context of the sensory modalities involved to maximize information about the environment. Given the contingencies present in even these simplified examples, and the combinatorial nature of stimulus properties, the exact contribution of integration at the level of perception and action is extremely complex. It is thus important to identify the properties of sensory inputs that result in integration (i.e. what are the 'rules' by which inputs are combined) and identify the physiological bases of integrative processes.

A Brief Introduction to the Principles of Multisensory Integration

As alluded above, informative relationships between sensory inputs almost invariably occur across multiple stimulus dimensions. These dimensions include the aforementioned space, time, and frequency dimensions, but numerous other correspondences are possible. Further, the combinatorial nature of stimulus properties mean that stimulus space is effectively infinite. How then can the nervous system decide which stimuli are integrated and which are not? Early neural and behavioral studies focused heavily on this question. These works identified that spatial proximity, temporal proximity, and the relative strength of sensory inputs serve as some of the strongest determinants of integration. The critical role these factors play has given rise to the terms principle when addressing their contributions to how inputs are integrated.

Space

One of the primary determinants of whether stimuli are integrated is the spatial concordance between the sources of sensory signals. As alluded in the above example of an object striking a hard surface, sensory signals originating from a single location are likely to have a common source located at that point in space, and thus it makes sense to afford a preference for integrating such signals. A pertinent example of the spatial principle at work is the dependence on spatial overlap of response speed facilitation for visual targets an additional auditory input (Frens, Vanopstal, & Vanderwilligen, 1995). Similarly, the ventriloquism illusion, in which auditory spatial localization is captured by visual inputs, is highly dependent on spatial proximity. Only when visual angle between the auditory and visual signals is sufficiently small does the visual stimulus capture the location of the auditory stimulus (Slutsky & Recanzone, 2001).

Time

Another important factor illustrated by many of the previous examples is that temporal concordance between stimuli is an equally important determinant of multisensory integration. Many stimuli commonly encountered in the environment have highly informative temporal structures across sensory modalities, and this motivates integration of stimuli when they are temporally aligned. An example of such a sensory input is audiovisual speech, which has an obligatory temporal correlation between the visual and auditory streams at low frequencies (Chandrasekaran et al., 2009; Schwartz & Savariaux, 2014). Object strikes (i.e. clapping) are similarly associated with a temporally aligned auditory stimulus, and these temporally aligned

sounds improve sensitivity for visual motion (Staufenbiel, van der Lubbe, & Talsma, 2011).

Inverse Effectiveness

Complimenting spatial and temporal contributions to multisensory integration is the principle of inverse effectiveness. Inverse effectiveness states that integration is most effective when the individual sensory inputs are weak. For example, a soft sound and a dim light would have a stronger predisposition towards integration than a loud sound and bright light with equivalent spatiotemporal overlap. This increased integrative efficacy is sometimes referred to as multisensory gain, which serves to quantify the contribution of integrative effects relative to unisensory processing (Stevenson, Ghose, et al., 2014). Increased gain for weak stimuli has been observed at the level of single neurons (Meredith & Stein, 1983), cortical populations measured with Electroencephalography (EEG) (Stevenson, Bushmakin, et al., 2012), as well as directly in human behavior (Senkowski, Saint-Amour, Hofle, & Foxe, 2011). The relative strength of sensory inputs is thus a crucial determinant of integration.

Principles of Integration Interact

Importantly, individual factors (i.e. principles) of integration are not evaluated in isolation (i.e. independently). Instead, interaction between these factors directly affects multisensory integration. Time and space both contribute to whether stimuli are perceptually unified and small disparities in one stimulus dimension can be overridden by strong concordance in another (Stevenson, Fister, Barnett, Nidiffer, & Wallace, 2012; Wallace, Roberson, et al., 2004). Similarly, stimuli are more likely to be perceived as simultaneous when their origin in space is closer (Zampini, Guest, Shore, & Spence, 2005; Zampini, Shore, & Spence, 2003). Interactions

between inverse effectiveness and spatial and temporal alignment have been demonstrated by the way stimulus intensity modulates perceived timing (Fister, Stevenson, Nidiffer, Barnett, & Wallace, 2016) and spatial interactions (Nidiffer, Stevenson, Fister, Barnett, & Wallace, 2016). An even more striking demonstration of the interaction between these factors is the ventriloquist illusion (Vroomen & De Gelder, 2004), in which sufficient spatiotemporal overlap is associated with integration of auditory and visual speech signals which do not have a common source (Slutsky & Recanzone, 2001). Perception and behavior are thus shaped by synergistic interactions between individual factors.

Convergence and Integration of Sensory Inputs in the Nervous System

The acceptance that cross-modal stimulus relationships make substantial contributions to perception and behavior has spurred questions regarding the neural instantiations of these processes. The abundance of modern neuroscience techniques with varying spatiotemporal strengths (Sejnowski, Churchland, & Movshon, 2014) has facilitated diverse approaches to characterizing the anatomical and physiological bases of integrative processes.

Anatomical Substrates of Multisensory Convergence

Characterizing the structural connectivity between sensory systems formed the basis of many early investigations and provided some of the earliest evidence of sensory convergence in the central nervous system. Studies using this approach have revealed that cross modal interactions are supported by anatomical convergence in numerous association regions and, strikingly, even in primary sensory cortices. Higher association areas in particular have long been recognized

play a crucial role in the mechanistic bases of integration (E. G. Jones, Powell, T.P.S., 1970). These association areas are frequently positioned at transition points between primary sensory areas, and invasive physiology has indicated that even within these regions there are gradients in terms of the overall proportion of neurons that respond to respective sensory modalities. A thorough example of this can be found in (Olcese, Iurilli, & Medini, 2013), in which the authors found the proportion of visual and tactile responsive neurons to be graded between the respective visual and tactile brain areas. Importantly, multisensory neurons were also found to be most abundant at the center of this gradient. This is consistent with models of cortical transitions between sensory systems, which posit that multisensory neurons should be most frequent within cortical border regions (Wallace, Ramachandran, & Stein, 2004). A similar transitional structure is suggested by anatomical and electrophysiological studies of the primate superior temporal sulcus (STS), a region in which multisensory neurons are found in abundance (Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Benevento, Fallon, Davis, & Rezak, 1977; Bruce, Desimone, & Gross, 1981; Dahl, Logothetis, & Kayser, 2009; Hikosaka, Iwai, Saito, & Tanaka, 1988). These studies indicate that unisensory inputs and neurons are interleaved with multisensory neurons (Dahl et al., 2009; Seltzer & Pandya, 1994), a structural organization that is likewise present in humans (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Gentile, van Atteveldt, De Martino, & Goebel, 2017). Similar anatomical convergence of sensory inputs occurs in parietal (Avillac, Ben Hamed, & Duhamel, 2007) and frontal (Sugihara, Diltz, Averbeck, & Romanski, 2006) association areas.

In addition to association areas, regions long thought to be primarily unisensory have also recently been recognized to be a crucial component of the anatomical networks contributing to multisensory integration. For example, the core and parabelt regions of the primate auditory

cortex project directly to eccentric (i.e. slightly off fovea) regions of the primary visual cortex (Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003) and similar projections are present in the cat (Hall & Lomber, 2008). Similar connections in the reverse direction from visual to auditory cortical areas are also present in primates (Falchier et al., 2010; Musacchia & Schroeder, 2009; Smiley et al., 2007). Anatomical inputs to auditory and visual primary cortices also originate from somatosensory cortex (Budinger, Heil, Hess, & Scheich, 2006; Cappe & Barone, 2005; Hackett, Smiley, et al., 2007). These anatomical studies compliment physiological approaches indicating neurons which respond to somatosensory inputs can be found in primary and secondary auditory and visual processing areas (Fu et al., 2003; Schroeder & Foxe, 2002; Schroeder et al., 2001). It is notable, however, that direct activation (i.e. action potentials) by visual cortex neurons is not present for auditory inputs (Y. Wang, Celebrini, Trotter, & Barone, 2008) (reviewed in: (Cappe, Rouiller, & Barone, 2009)). It is thus clear that the anatomical bases for multisensory interactions are in place even in putatively 'primary' sensory regions.

In addition to cortico-cortico circuits there are also subcortical circuits which provide anatomical convergence between sensory systems. One of the best studied areas is the superior colliculus (SC), a midbrain structure involved in processes related to orienting, target selection and the transformation of sensory inputs into motor commands (Huerta, 1984; Munoz & Guitton, 1985; Sprague, 1996). The superior colliculus is a laminar structure, and neurons in the superficial layers classically are believed to respond primarily to visual inputs (Casagrande, Harting, Hall, Diamond, & Martin, 1972) (although see (Ghose, Maier, Nidiffer, & Wallace, 2014) for a more recent perspective). The deep layers, however, contain relatively common neurons that respond to visual, auditory, and somatosensory inputs, indicating that convergent

input from all three of these sensory systems is present (May, 2006; Meredith & Stein, 1986a). There are also thalamic connections that are less direct, but could still serve to provide extremely rapid inputs to early sensory regions from other brain areas (Cappe, Morel, Barone, & Rouiller, 2009; Hackett, De La Mothe, et al., 2007; Henschke, Noesselt, Scheich, & Budinger, 2015). These thalamic connections are also not merely relays, but directly serve integrative functions as well (Komura, Tamura, Uwano, Nishijo, & Ono, 2005), and thus provide an additional anatomical route for rapid feed forward integration of sensory inputs.

Physiological Manifestations of Integration

Complementing anatomical work, physiological approaches such as single unit recording and laminar neurophysiology are able to examine how neural activity is shaped by interactions between the senses. Some of the earliest invasive physiological work identified that subcortical brain regions such as the superior colliculus not only contain neurons which respond to more than one type of sensory input, but that these neurons *integrate* information from these convergent inputs. The most powerful example of this integration at the single neuron level is that many of these neurons respond super additively, in terms of the sheer number of action potentials, when inputs occur in more than one sensory modality (Meredith & Stein, 1986b). Additivity in neural responses can be easily quantified by comparing the sum of unisensory responses to the multisensory response using the formula ([AV / (A+V)] *100), with values greater than 100 percent (i.e. AV > A+V) indicating a 'super-additive' multisensory response relative to the sum of unisensory inputs (Stein & Stanford, 2008; Stevenson, Ghose, et al., 2014). Similarly, values well below 100 percent indicate neural response suppression, and this 'sub-additive' suppression is particularly striking when the multisensory response is smaller than the

response to either sensory input in isolation (i.e. AV < A || V). The non-linear nature of both sub- and super-additivity also serve as a critical indicator of multisensory *integration*, rather than just multisensory *convergence*, as they indicate neurons are performing non-linear computations rather than merely summating inputs. Employing this analytical approach in a series of seminal studies Wallace, Stein, Meredith and colleagues charted the nature of multisensory integration in the SC and its dependence on factors such as input from other cortical regions, developmental stage, and sensory experience (Jiang, Wallace, Jiang, Vaughan, & Stein, 2001; Meredith & Stein, 1983, 1985, 1986b; Wallace & Stein, 1994, 2000). This line of work established several critical findings which have informed the broader field since that time.

First, the observation of multisensory interactions in a subcortical structure that was classically believed to be primarily visual established that, as predicted by anatomical connectivity, physiological manifestations of integration are present throughout the nervous system. This finding has since been extended to include numerous subcortical regions such as the medial and lateral geniculate nuclei of the thalamus (Noesselt et al., 2010) the striatum (Reig & Silberberg, 2014), and the basal ganglia (Nagy, Eordegh, Paroczy, Markus, & Benedek, 2006). Other subcortical regions such as the cochlear nucleus have also been shown to be critical to multisensory interactions, as their structural connectivity with thalamic and cortical areas predicts the strength of multisensory behavioral interactions (van den Brink et al., 2014). Multisensory convergence and integration thus involve putatively 'low level' processing areas that are often positioned 'early' in the cortical processing hierarchy. This contrasted with early views that integrative processes were primarily instantiated in 'association cortex' areas such as the superior temporal sulcus (STS), intraparietal sulcus (IPS), and frontal lobe (E. G. Jones, Powell, T.P.S., 1970). It is worth noting, however, that midbrain integration (i.e. non-linear

interactions), but not convergence, was later shown to be dependent on input from higher multimodal association cortex, specifically the anterior ectosylvian sulcus (AES) (Wallace & Stein, 1994). This suggests that these views are not entirely dichotomous, and that integration in the midbrain is part of a complex cortico-tectal network which functions to transform sensory inputs into perceptual representations and motor commands (Stitt et al., 2015). This dependency also demonstrates the bidirectional nature of interactions between midbrain structures and higher cortical association areas classically recognized to have multisensory functions (Benevento et al., 1977; Bruce et al., 1981; Hikosaka et al., 1988; Hyvarinen & Shelepin, 1979). These network level interactions serve as a mechanism for top down regulation of putatively "low level" integration by higher cognitive systems.

This line of research also began to establish direct neural instantiations regarding how the relationships between stimuli affect whether sensory signals are integrated. This includes some of the first demonstrations that temporal relationships (Meredith, Nemitz, & Stein, 1987) and spatial relationships (Meredith & Stein, 1986a, 1996) contribute to the magnitude of integrative effects observed in the neural activity of individual multisensory neurons. Specifically, neuronal super-additivity was found to be strongest when stimuli were temporally coincident and to decrease monotonically for temporal misalignment (Meredith et al., 1987). Importantly, the slope of this monotonic decrease was also found to be asymmetric, which is in line with the asymmetry found in perceptual thresholds and tolerance for asynchrony in audiovisual stimuli (reviewed below). More recent demonstrations of timing sensitivity for audiovisual stimuli in the cortex of rodent models have found similar temporal tuning functions (Schormans et al., 2016). Individual SC neurons thus directly implement the temporal principle of multisensory integration through non-linear summation of synaptic inputs. This line of work also characterized neural

instantiations of the spatial principle for the first time, by demonstrating that integration in the SC both requires spatial concordance between stimuli, and that these effects depend on the receptive field architecture of individual SC neurons. Specifically, each multimodal neuron in the SC has separate unisensory receptive fields (i.e. a trimodal neuron would have a visual receptive field, a separate auditory receptive field, and a third separate somatosensory receptive field) (Carriere, Royal, & Wallace, 2008; Royal, Carriere, & Wallace, 2009; Wallace & Stein, 1996). These spatial receptive fields are frequently coincident, but are not necessarily the same size or shape. In particular, auditory spatial receptive fields are generally very large in size, which corresponds with audition's poor spatial precision (Wallace, Meredith, & Stein, 1993; Wallace & Stein, 1996). Overlap, or co-registration, between the respective receptive fields is thus incomplete, and a single neuron might have spatial receptive field overlap for some regions of physical space, but only a single receptive field for other physical locations. Critically, when stimulation occurred within more than one receptive field, super additive multisensory interactions were observed. Stimulation within only a single receptive field, however, either failed to produce super additivity or even produced suppressive (i.e. sub-additive) effects (Kadunce, Vaughan, Wallace, Benedek, & Stein, 1997; Kadunce, Vaughan, Wallace, & Stein, 2001; Meredith & Stein, 1986a, 1996). Notably, enhancement effects dependent on spatial receptive field overlap are also specific to cross-modal stimuli and do not occur for two concurrent stimuli of the same modality (Alvarado, Vaughan, Stanford, & Stein, 2007), reviewed in: (Stein & Stanford, 2008). This indicates that they index true spatially dependent sensory integration. This line of work thus extended the notion of spatial receptive fields first demonstrated by Hubel and Wiesel (Hubel & Wiesel, 1998) by demonstrating that cross-modal co-registration of sensory receptive fields functions as an important neural mechanism

instantiating the spatial rule of multisensory integration at the level of the single neuron.

Similar to the spatial and temporal principles, inverse effectiveness was also found to be instantiated at the level of the individual neuron in the SC. The non-linear interactions indicative of integration reviewed above are at their strongest, in terms of relative multisensory gain, when neural responses are relatively weak to individual sensory inputs (Meredith & Stein, 1986b; Stanford & Stein, 2007). For neurons that are relatively unresponsive to unisensory inputs, this can result in non-linear multisensory response gain of over an order of magnitude (Stein & Stanford, 2008). The presence of such massive levels of response enhancement indicates that neurons downstream of these SC neurons receive fundamentally different inputs when weak sensory signals are integrated across modalities. Together with the instantiations of the temporal and spatial principles, this indicates that all the primary principles of multisensory integration are instantiated at the level of individual computational units in the midbrain. Further, lesion of the deep layers of the SC not only results in loss of integration at the neural level, but loss of multisensory facilitation of spatial localization (Burnett, Stein, Chaponis, & Wallace, 2004; Burnett, Stein, Perrault, & Wallace, 2007). In concert with deactivation studies (Wallace & Stein, 1994), SC neurons thus instantiate multisensory principles via a distributed processing network that incorporates input and modulation from cortical association areas. These multisensory interactions then contribute to ecologically important behaviors long known to depend on the SC, such as orientation, target selection, and initiation of oculomotor actions such as saccades (Casagrande et al., 1972; Sparks & Hartwich-Young, 1989). Combined with anatomical evidence, multisensory integration thus appears to be ubiquitous in the brain, spanning from early sensory relay sites such as the thalamus and cochlear nucleus, midbrain structures such as the SC, and cortical structures spanning the entire hierarchy from primary

sensory cortices to higher association areas. These changes in neural encoding when sensory inputs occur in multiple modalities then form the basis of behavioral and perceptual facilitation effects such as improved target detection and speeded spatial orienting.

Mechanisms of Cortical Integration

As suggested by the necessity of association cortex for multisensory integration in the midbrain, convergence and integration in cortical regions might play a particularly nuanced and important role in multisensory processing. This is exemplified by studies indicating that cortical association regions, frequently found on the border between two neighboring sensory regions, contain large populations of neurons which integrate sensory inputs (Olcese et al., 2013; Schormans et al., 2016; Wallace, Ramachandran, et al., 2004). There has thus been a substantial interest in identifying population level mechanisms of multisensory integration and their role in cortical circuits, which are frequently less stereotyped and more flexible than midbrain circuits. This interest is reinforced by apparent functional differences between midbrain and cortical signatures of integration, such as the paucity of suppressive local circuit interactions for simple stimuli in cortex (Stein & Wallace, 1996). Additionally, it has been recognized that cortical population signals such as the phase and amplitude of the local field potential (LFP) or correlations in trial-by-trial spike rates between neurons (also known as spike correlation) carry information that is distinct and complimentary to that found in the activity of individual neurons (for a review see: (Panzeri, Macke, Gross, & Kayser, 2015)). This highlights that cortex might have its own unique mechanisms of multisensory integration that leverage information encoded in population signals such as neural oscillations.

One of the most important of these population level interactions is phase reset, a process by

which inputs to a system can ‘prime’ it for processing of a new input by changing the relative timing of sub threshold oscillatory activity. Evidence for the importance of this mechanism comes from experiments employing laminar neurophysiology, which have indicated that a sensory input in one modality (e.g. a somatosensory input) results in increased inter-trial coherence (ITC) in primary cortical regions for other sensory modalities (e.g. primary auditory cortex) (Lakatos et al., 2007). ITC measures the concentration of phase angles across stimulus

events, with values close to 0 indicating that phase is randomly distributed and a value of 1 indicating perfect phase consistency (**Figure 1-1)**. Importantly, this increased phase consistency occurs in the absence of changes in oscillatory amplitude or multiunit activity. In other words, the magnitude of ongoing activity is unaffected, but the relative timing (phase) of ongoing activity shifts to match the input (**Figure 1-2**). A new input to this region would thus fall at a consistent phase and activity level and be processed in a manner differing from a stimulus that

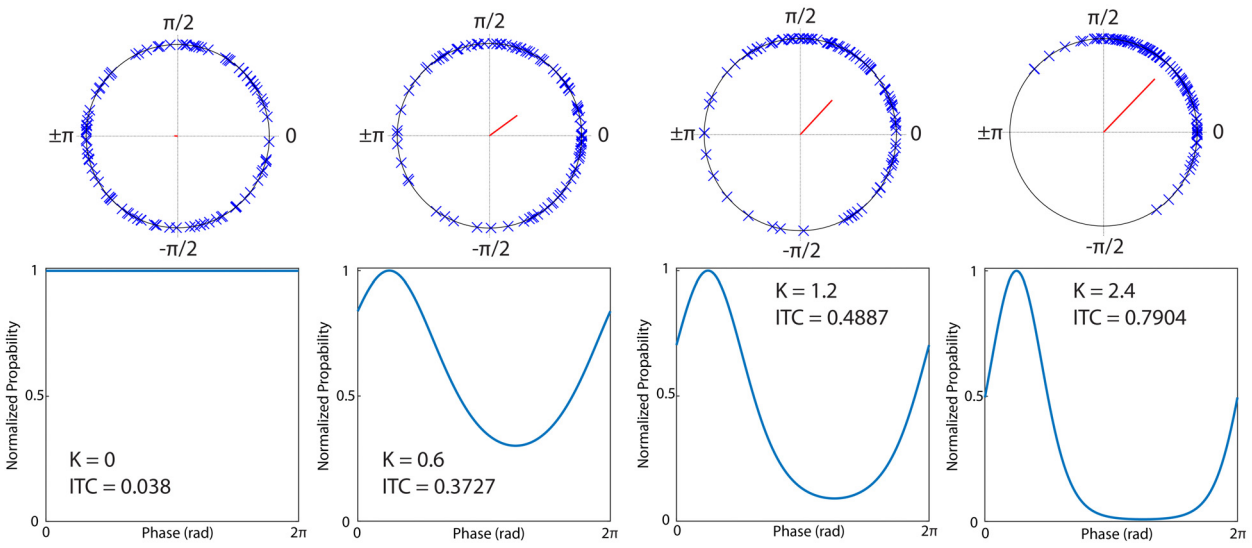

Figure 1-1 Inter-trial Phase Coherence as a Measure of Phase Consistency
Inter-trial phase coherence (ITC) measures the dispersion of phase angles across trials. A value of 0 indicates a uniform distribution (i.e. phase is completely random) and a value of 1 indicates perfect consistency (i.e. phase is identical on all trials). To illustrate the ITC metric, 200 trials (blue X's) were drawn at random for each of four von Mises distributions with varying levels of concentration (K) and a mean angle of $\pi/4$. The mean resultant vector length (MRVL) for each sample distribution was then calculated, which is the ITC. Top row: the phase angles of the individual trials in each sample (blue X's) and the mean resultant vector (orange line). Bottom row: the normalized probability (maximum = 1) of each phase angle in the von Mises distributions from which samples were drawn. The concentration parameter K controls the dispersion of the von Mises distribution.

was not preceded by an 'off primary' sensory stimulus. If the second stimulus were to be delayed, these changes in processing would fluctuate rhythmically depending on what phase the new input landed on. Rhythmic amplification of this nature is directly observable in the rhythmic structure of LFP and multiunit responses to a second stimulus administered with varying temporal delays relative to the 'priming' cross modal stimulus (Lakatos et al., 2007). Additionally, this mechanism has been demonstrated to be attention dependent, establishing it as

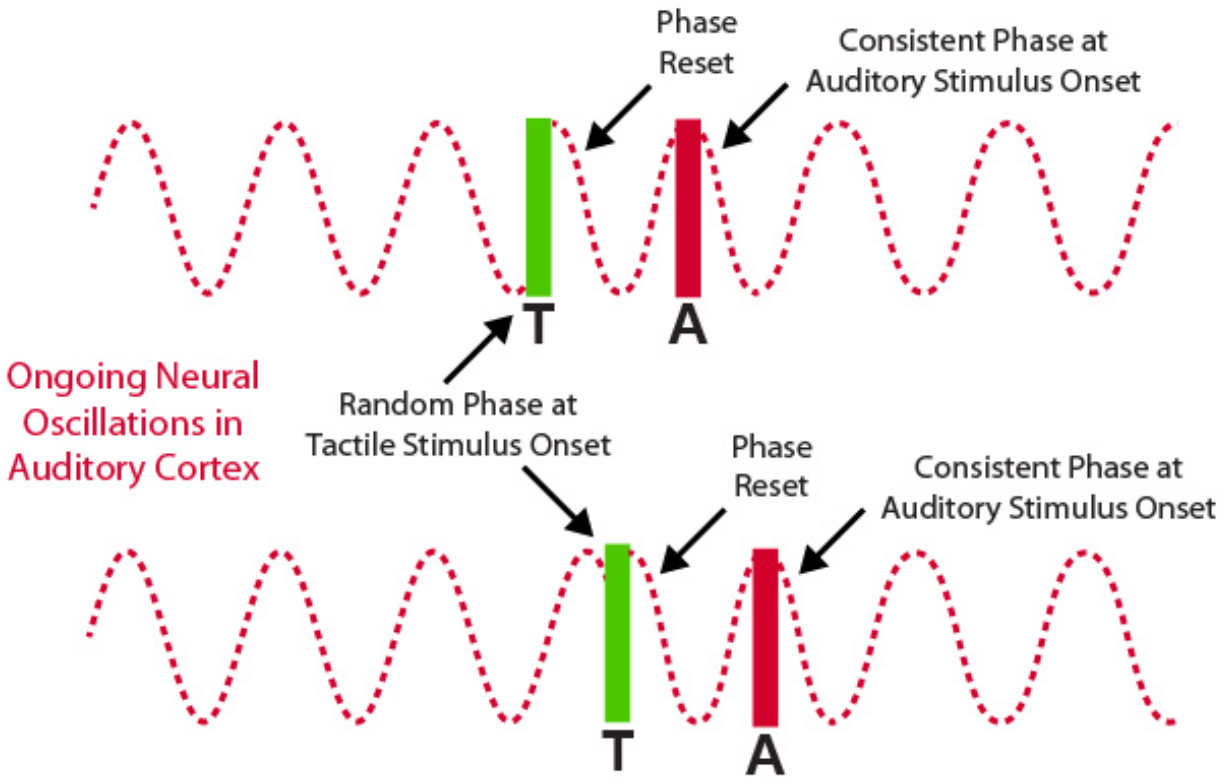

Figure 1-2 Canonical Oscillatory Phase Reset in Auditory Cortex
Endogenous neural oscillations (dotted red line) are constantly occurring in auditory cortex. When a new stimulus occurs without temporal structure, the phase it falls upon is random. Here, tactile stimuli (green bars) occur at random phases. Importantly, however, the tactile stimulus resets the phase of the ongoing oscillation. Auditory stimuli (red bars) following with a consistent temporal delay thus fall on identical oscillatory phases and are processed consistently. For an empirical demonstration of this effect, see (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007).

a flexible mechanism for cross-modal physiological interaction (Lakatos et al., 2009). In other words, phase reset is able to accomplish task or attention dependent modulation of neural activity, consistent with the flexibility of multisensory integration that has been observed at the behavioral level (N. M. van Atteveldt, Peterson, & Schroeder, 2014; N. van Atteveldt, Murray, Thut, & Schroeder, 2014). Such flexibility in cortical processing also allows task demands to influence subcortical processing via the cortico-tectal circuits previously described. Furthermore, the presence of phase reset in the human brain and its conferrence of behavioral facilitation is directly supported by a number of works utilizing non-invasive approaches such as EEG (Naue et al., 2011; Thorne, De Vos, Viola, & Debener, 2011). Phase reset thus offers a promising mechanism of flexible integration in cortex beyond direct instantiations of the principles as observed in midbrain neurons. Recent evidence suggests that subcortical regions may also participate in phase dependent sensory processing through interactions with cortical networks (Stitt et al., 2015). Lastly, phase reset of ongoing activity is a crucial step in the establishment of consistent phase relationships *between* brain regions. This phenomenon, known as neural coherence, is believed to be an important contributor to multisensory integration and is discussed more extensively below in the context of non-invasive approaches.

Multisensory Integration in Humans

In addition to invasive anatomical and physiological approaches, the neural bases of integration are also well studied in the human brain. In particular, the utilization of non-invasive techniques such as electroencephalography and magnetoencephalography (EEG / MEG) and Functional Magnetic Resonance Imaging (fMRI) has yielded substantial information regarding anatomical

hubs and physiological mechanisms of multisensory integration in humans.

Anatomical Hubs of Multisensory Integration in the Human Brain

In regards to the structural substrates of multisensory integration, non-invasive studies have identified that the human brain has important anatomical hubs that specifically process cross-modal stimuli and which might perform important multisensory perceptual functions such as binding. An example of such a region is the superior temporal sulcus (STS), a region positioned between auditory and visual cortical areas which has been identified to be a critical hub for audiovisual integration. The human STS has a patchy organization (Beauchamp, Argall, et al., 2004) in which auditory and visual sub-regions are intermingled with neural patches that respond to both types of stimuli. This intermingled organization provides for the transitional boundaries which have been identified to be important in animals models of multisensory integration (i.e. (Olcese et al., 2013; Schormans et al., 2016; Wallace, Ramachandran, et al., 2004)). Numerous studies have indicated that the STS is active during integration of multisensory inputs, and is particularly involved when inputs are audiovisual in nature (Balz, Keil, et al., 2016; N. van Atteveldt, Formisano, Goebel, & Blomert, 2004). The STS has also been pinpointed in studies targeting audiovisual temporal processing (Bushara et al., 2003) and audiovisual speech processing, making this region highly relevant to the current work (Baum, Martin, Hamilton, & Beauchamp, 2012; Beauchamp, Lee, Argall, & Martin, 2004; Beauchamp, Yasar, Frye, & Ro, 2008; Bishop & Miller, 2009; G. A. Calvert, Hansen, Iversen, & Brammer, 2001; Macaluso, George, Dolan, Spence, & Driver, 2004; L. M. Miller & D'Esposito, 2005; Nath & Beauchamp, 2011, 2012; Nath, Fava, & Beauchamp, 2011; Schepers, Schneider, Hipp, Engel, & Senkowski, 2013; Stevenson, Altieri, Kim, Pisoni, & James, 2010; Stevenson, Geoghegan, & James, 2007;

Stevenson, Kim, & James, 2009; Stevenson, VanDerKlok, Pisoni, & James, 2011; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003). The crucial role of the STS for audiovisual speech integration has also been directly tested by the application of transcranial Magnetic stimulation (TMS) during the McGurk illusion (Beauchamp, Nath, & Pasalar, 2010). This study found that TMS of the STS attenuated perception of the illusion, indicating that this brain area play an important role in the binding of auditory and visual stimuli with speech like properties into a unified percept.

Frontal cortex and parietal cortex have also been noted as hubs of multisensory integration. For example, audiovisual sensory evidence, as indexed by the fMRI signal, appears to accumulate in the inferior frontal sulcus (Noppeney, Ostwald, & Werner, 2010), which is one of the primary nodes of the fronto-parietal networks that contribute to perceptual decisions (Heekeren, Marrett, & Ungerleider, 2008). Integration of auditory and visual inputs for sentence comprehension also occurs in inferior frontal cortex (Homae, Hashimoto, Nakajima, Miyashita, & Sakai, 2002), as does voice-gesture integration (Xu, Gannon, Emmorey, Smith, & Braun, 2009). Invasive work has similarly demonstrated the presence of integrative processes in prefrontal cortex (Romanski, 2007; Sugihara et al., 2006) (for a review see: (Romanski, 2012)). Parietal cortex is also intimately involved in multisensory convergence and integration, and a number of studies have indicated that information from multiple sensory systems converges in parietal cortex. For example, (Bremmer et al., 2001) found that motion signals activated IPS for multiple sensory modalities, and (Macaluso & Driver, 2001) found similar IPS convergence for tactile and visual stimulation. It is noteworthy, however, that conjuctivity analysis of this nature only indicates convergence of information rather than integration (Noppeney, 2012). There is, however, evidence that parietal cortex is involved with true multisensory integration effects

during speeded response tasks. Specifically the superior parietal lobule demonstrates non-linear interactions when measured via electrocorticography (Moran, Molholm, Reilly, & Foxe, 2008) and structural connectivity in parietal regions play a crucial role in multisensory reaction time facilitation (i.e. violation of the miller inequality (J. Miller, 1982)) (Brang, Taich, Hillyard, Grabowecky, & Ramachandran, 2013). Regions of the parietal cortex have also been shown to be implementing the integration of cues based on their sensory reliability (Rohe & Noppeney, 2016) and causal inference (Rohe & Noppeney, 2015), both of which are important factors determining how multisensory signals are integrated. Additionally, in a paradigm similar to the aforementioned study using TMS to disrupt audiovisual speech integration in the STS, targeted disruption of posterior parietal cortex has been shown to impair visuotactile integration (Pasalar, Ro, & Beauchamp, 2010). Lastly, the multisensory role of parietal cortex has been confirmed across species by the presence of visuotactile interactions in the ventral intraparietal cortex of the macaque (Avillac et al., 2007). Importantly, parietal areas are known to play an important role in aligning sensory signals into a common reference frame (Cohen & Andersen, 2002; Mullette-Gillman, Cohen, & Groh, 2005). Multisensory interactions in parietal cortex may thus make considerable contributions to the formation of a unified spatial representation of a multimodal environment.

Evoked Electroencephalography and Magnetoencephalography

In contrast to fMRI's spatial resolution, EEG and MEG are well suited to studying multisensory integration in humans due to their high temporal resolution and more direct relationship to physiological activity. One of the most straightforward approaches to doing so employs the same non-linear effects analysis used in invasive physiological studies (i.e. AV ≠ A+V). Utilizing this

approach, a host of non-linear interactions have been identified in evoked potentials (or the equivalent evoked fields for MEG) when sensory inputs occur in more than one sensory modality (Foxe et al., 2000; Miniussi, Girelli, & Marzi, 1998; Murray, Foxe, Higgins, Javitt, & Schroeder, 2001; Murray et al., 2005; Raij et al., 2010; Sperdin, Cappe, Foxe, & Murray, 2009). When utilizing simple stimuli such as impulses (i.e. flashes, brief touches, or beeps) studies have found that non-linear interactions are present in the form of super-additivity, which emerges as early as ~50 ms after stimulus onset (Giard & Peronnet, 1999; Molholm et al., 2002; Vidal, Giard, Roux, Barthelemy, & Bruneau, 2008). These interactions are sometimes found to be somewhat delayed (to ~80-100ms) using more complex paradigms designed to conservatively eliminate potential duplication of cognitive event related potential (ERP) components (Gondan & Roder, 2006; Teder-Salejarvi, McDonald, Di Russo, & Hillyard, 2002). The low latency relative to stimulus onset and estimated cortical sources of these super-additive interactions indicates that both early sensory cortices and more mid-level cortical structures exhibit super-additive multisensory interactions. Rapid cross modal activation has also been confirmed using Electrocorticography (also known as intracranial EEG) (Brang et al., 2015), which provides confidence in the estimated primary and secondary sensory cortical generators.

Contrasting the super-additive effects seen in these studies of simple stimuli (i.e. AV > A+V), for stimuli consisting of audiovisual speech, which is sometimes considered to be a 'complex' stimulus, non-linear interactions take the form of sub-additivity (i.e. AV < A+V) (Baart, 2016; Besle, Fort, Delpuech, & Giard, 2004; Klucharev, Mottonen, & Sams, 2003; Mottonen, Schurmann, & Sams, 2004; van Wassenhove, Grant, & Poeppel, 2005). This sub-additivity is also true for other ecologically valid events with extended time courses (Stekelenburg & Vroomen, 2007) and for stimulus pairings which are not audio-visual (Treille,

Cordeboeuf, Vilain, & Sato, 2014). Furthermore, in (Stekelenburg & Vroomen, 2007) it is demonstrated that for ecologically valid events without anticipatory visual information this sub-additivity seems to be greatly attenuated, at least for early auditory ERP components such as the N1 and P2. This reversal from super to sub-additivity suggests that the presence of stimulus features in one modality which provide *anticipatory* information regarding upcoming events in another modality fundamentally change the way sensory integration in the brain occurs for naturalistic events. In other words, relative timing of stimuli matters a great deal in terms of neural effects, at least for naturalistic stimuli with extended time courses (Stekelenburg & Vroomen, 2007). This anticipatory information is particularly important for speech due to the complex correlational structure between auditory and visual speech and the fact that visual mouth movements generally precede the auditory signal (Chandrasekaran et al., 2009; Schwartz & Savariaux, 2014). Changes in auditory neural processing by anticipatory visual speech information have thus been proposed to be a form of predictive coding (van Wassenhove et al., 2005) (for a review of predictive coding at the circuit level see: (Bastos et al., 2012)). Sub-additive speech interactions have also been observed in a portion of the multiunit activity in the auditory cortex of non-human primates (Kayser, Petkov, & Logothetis, 2008). This suggests that reductions in EEG signal amplitude may be well correlated with changes in multiunit spiking activity and LFP magnitude in auditory cortex.

Importantly, these non-invasive measurements of human neural activity also demonstrate instantiations of the principles of integration which are remarkably similar to those found in invasive physiology. For example, the temporal correspondence of inputs regulates the magnitude of oscillatory neural responses (Senkowski, Talsma, Grigutsch, Herrmann, & Woldorff, 2007). The magnitude of non-linear interactions in ERPs also depends on the relative

strength of the individual inputs (Stevenson, Bushmakin, et al., 2012). These non-linear interactions also incorporate finer stimulus correspondences. An example of this is found for looming sounds, in which increases in sound amplitude over time cause the impression that the source is approaching. Responding to such a stimulus (i.e. an approaching predator or vehicle) might be particularly crucial for survival. Fitting with this evolutionary hypothesis, non-linear ERP effects have been shown to be stronger for looming sounds (Cappe, Thut, Romei, & Murray, 2010; Cappe, Thut, et al., 2009). This is particularly important given the bottom up nature of the ERP differences in this task, which occur very early and originate from regions that include primary sensory cortices. Together these findings indicate that non-invasive measurement of multisensory interactions in the human brain exhibit similar response properties and flexible cue dependent integration to that seen in invasive physiological models.

Oscillatory Contributions to Multisensory Integration

In addition to the non-linear interactions in evoked activity, neural oscillations have also been proposed to play a crucial role in multisensory integration due to their ability coordinate activity across disparate neural networks. This coordination invokes the concept of non-stationary neural coherence, in which the momentary relationship between oscillatory activity in two or more regions determines the rate and nature of information flow (Hipp, Hawellek, Corbetta, Siegel, & Engel, 2012; Siegel, Donner, & Engel, 2012). Coherence can occur along multiple oscillatory dimensions within a frequency (i.e. phase coupling (Lachaux, Rodriguez, Martinerie, & Varela, 1999), amplitude coupling (Bruns, Eckhorn, Jokeit, & Ebner, 2000)), as well as between frequencies (i.e. phase amplitude coupling (Canolty & Knight, 2010) and cross-frequency envelope modulation (Hipp et al., 2012)). The diversity of these so called intrinsic coupling

modes (Engel, Gerloff, Hilgetag, & Nolte, 2013) allows for potentially multiplexed and multidirectional information flow within a single oscillatory network (Akam & Kullmann, 2014; Helfrich et al., 2016; Hillebrand et al., 2016). Coupling across distinct neural populations using these oscillatory mechanisms directly influences both spiking in individual neurons, as well as correlation in the magnitude of LFPs in spatially distinct local circuits (Womelsdorf et al., 2007). Oscillatory networks are thus believed to be a crucial factor underpinning the formation of task specific neural networks and information integration in the brain (Engel, Fries, & Singer, 2001; Engel & Singer, 2001; Fries, 2005; Singer & Gray, 1995; Womelsdorf et al., 2007).

This selective information integration has been specifically proposed to be important for multisensory interactions (Senkowski, Schneider, Foxe, & Engel, 2008). This is supported by studies of simple multisensory stimuli, which have indicated that high frequency oscillations in the beta (13-30 Hz) and gamma (> 30 Hz) bands are more strongly recruited for multisensory stimuli than for their unisensory counterparts (reviewed in: (Senkowski et al., 2008)). This indicates that phase coherence (illustrated in figure **1-3**) within localized neural populations is likely enhanced when stimulation occurs in multiple sensory modalities. This oscillatory enhancement has also been shown to be supremely sensitive to the aforementioned integration

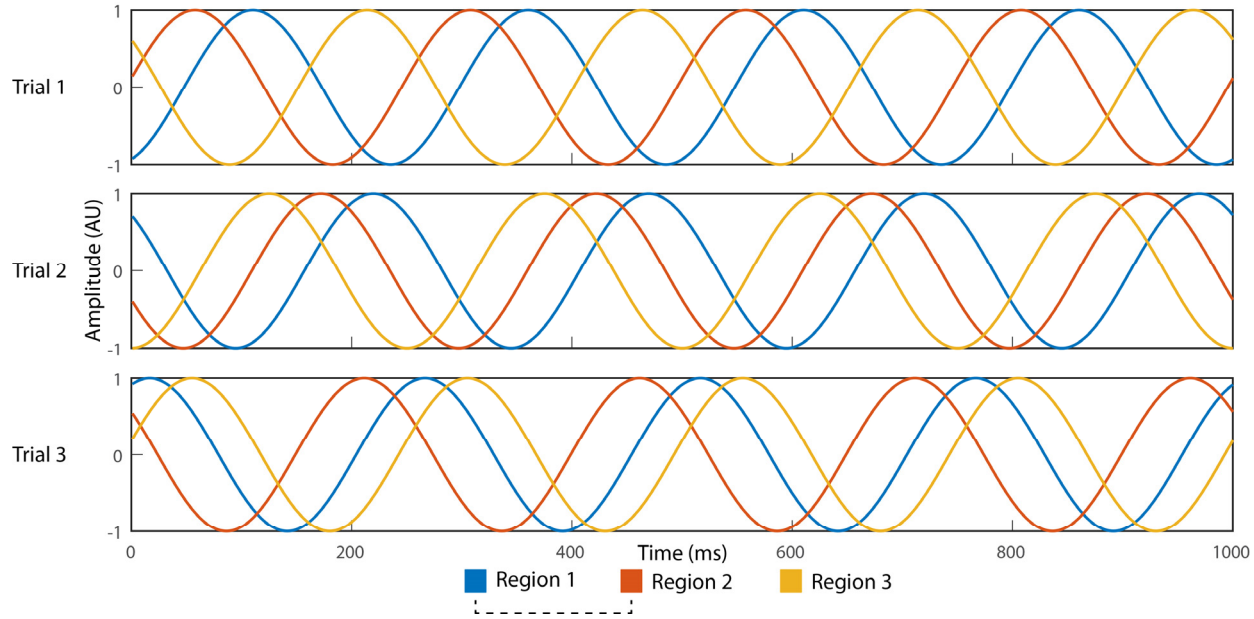

Figure 1-3 Canonical Oscillatory Phase Coherence
This figure illustrates a stationary version of simple phase coherence. In each trial oscillatory activity was generated for 3 hypothetical brain regions. In each trial the phase difference between regions 1 and 2 is relatively consistent, with the blue oscillation (region 1) lagging the red oscillation (region 2). These two circuits are thus coherent (indicated by the dashed line in the key), in the sense that their phases are related. In contrast, the oscillation for region 3 (yellow) neither consistently lags nor leads the oscillations for regions 1 or 2. This region is thus incoherent with regions 1 and 2. Phase coherence can be found both within localized circuits, as well as between disparate brain regions. Data were simulated using a 4 Hz Sinewave with 0.1 radians of uniformly distributed noise for regions 1 and 2 (red and blue) and 2π radians of uniformly distributed noise for region 3 (yellow). Note that high phase coherence at small anatomical scales often results in differences in the power of the LFP or EEG signal due to signal summation.

determinants such as temporal correspondence (Senkowski et al., 2007). A number of studies have also demonstrated that high frequency oscillations are associated with the process of perceptually binding stimuli together, specifically in the beta band (Schepers et al., 2013; Senkowski, Molholm, Gomez-Ramirez, & Foxe, 2006; von Stein, Rappelsberger, Sarnthein, & Petsche, 1999) and the gamma band (Balz, Keil, et al., 2016; Bhattacharya, Shams, & Shimojo, 2002; Mishra, Martinez, Sejnowski, & Hillyard, 2007; Sakowitz, Quiroga, Schurmann, & Basar, 2001). Other studies have implicated both the beta and gamma bands (Kisley & Cornwell, 2006), while lower frequency bands have also been reported in some cases (Sakowitz, Quian Quiroga, Schurmann, & Basar, 2005; Sakowitz, Schurmann, & Basar, 2000). Together these studies, which by no means constitute an exhaustive list of the oscillatory literature, suggest that changes in oscillatory power occur across the frequency spectrum and are involved in the process of perceptually combining sensory inputs.

Coupling of these oscillations across cortical regions to form functional networks has also been shown contribute to multisensory interactions in a number of paradigms including the rubber hand illusion (Kanayama, Sato, & Ohira, 2007) and the stream bounce illusion (Hipp, Engel, & Siegel, 2011). In these studies, formation of a coherent network is associated with the illusory percept, which serves as an index of trial-by-trial multisensory binding. In other words, formation of a complex interregional brain network mechanistically based in oscillatory coherence predicts whether multisensory perceptual interactions occur. Similarly, in an audiovisual duration judgment task, the formation of a coherent brain network in the alpha band predicted whether audiovisual cross-modal interactions occurred, despite the frequently detrimental nature of the second sensory input to participant performance (van Driel, Knapen, van Es, & Cohen, 2014). These studies clearly indicate that rhythmic synchronization between

brain areas forms a flexible mechanism of integration across sensory systems. Invasive physiological investigations with finer spatial resolution have also yielded similar results, although their correspondence to perception is less well established. One example of such a finding is increased gamma band neural coherence between auditory cortex and STS for looming (i.e. approaching) sensory cues, which are known to be preferentially integrated even in non-human primates (Maier, Chandrasekaran, & Ghazanfar, 2008; Maier, Neuhoff, Logothetis, & Ghazanfar, 2004). These studies are thus consistent with the notion that large-scale synchronization forms the backbone of perception (Rodriguez et al., 1999; Siegel et al., 2012), and indicate that multisensory specific synchronization may underlie many perceptual phenomena.

Studies of Multisensory Temporal Processing in Humans

A common theme to many of the most striking demonstrations of audiovisual integration, such as audiovisual speech intelligibility (Ross, Saint-Amour, Leavitt, Javitt, et al., 2007; Sumby, 1954), the McGurk illusion (McGurk & MacDonald, 1976), the stream bounce illusion (Sekuler, Sekuler, & Lau, 1997) and the sound induced flash illusion (Shams, Kamitani, & Shimojo, 2000) is the substantial temporal alignment between the auditory and visual streams. This temporal concordance and its contribution to the final percept strongly suggests that temporal alignment is not only a crucial factor in physiological interactions, but also perceptual ones. This has led researchers to interrogate the contribution of temporal relationships to multisensory processing in a myriad of tasks and experimental designs, which has yielded not only critical insights into integrative processes, but also helped to establish how the unifying principles of multisensory

integration translate to perception and behavior. One of the crucial findings synthesized from this body of research is the concept of the temporal binding window (TBW), a temporal psychometric function that has been empirically derived in numerous experiments and which describes both the magnitude and probabilistic nature of multisensory interactions (Vroomen & Keetels, 2010; Wallace & Stevenson, 2014). Importantly, the TBW bears striking resemblance to the temporal tuning functions of individual audiovisual responsive neurons found in frontal cortex and in the SC (Benevento et al., 1977; Meredith et al., 1987). The TBW also frequently displays an asymmetric shape with a slight visual bias (Alcala-Quintana & Garcia-Perez, 2013; Stevenson & Wallace, 2013; van Wassenhove, Grant, & Poeppel, 2007). This asymmetry is consistent with the statistics of the natural world, in which visual events frequently precede auditory signals due to the faster travel speed of light across space. It is also consistent with the causal structure of many events such as object impact, where visual motion precedes sound generation, and audiovisual speech, where mouth movements frequently precede sounds. These commonalities with physiology and the environment have led researcher to conclude that this construct strongly captures temporal dynamics in multisensory processing, and to characterize it in a number of settings.

Behavioral Characterizations of Multisensory Temporal Processing

Armed with the knowledge that the TBW likely represents, to some degree, the underlying tuning of the nervous system to temporal correspondence, researchers have systematically probed the importance of temporal relationships for myriad multisensory interactions. One of the most important tools for doing so is the utilization of psychophysical tasks consisting of synchrony judgments (SJ, also known as simultaneity judgment) (Stevenson & Wallace, 2013;

Zampini, Guest, et al., 2005) and temporal order judgments (TOJ) (Dixon & Spitz, 1980; Hirsh & Sherrick, 1961; Sternberg; Zampini et al., 2003). In both of these tasks, participants are asked to explicitly judge the temporal relationships of multisensory stimuli by answering whether stimuli occurred simultaneously or not (SJ) or which stimulus came first (TOJ). The stimuli utilized in this task range from simple flashes of light and auditory beeps, which are real world approximations of impulses for vision and audition respectively, to audiovisual speech, which has a substantially more complex spectro-temporal structure within each sensory modality. Despite the binary response nature of these tasks, and the resultant conversion of the TBW to a probabilistic window, these tasks have been demonstrated to reliably index temporal acuity in human observers and have thus become important tools for interrogating multisensory temporal function. These tasks have been used to elucidate a number of critical factors in temporal perception that are pertinent to the current work and formed the basis of the current research hypotheses. Synchrony judgment in particular is highly important to the scope of the current work, as it was the primary psychophysical task in all of our EEG experiments.

The Window of Integration Depends on Stimuli and Task

One of the most important characteristics of the TBW is that it is not constant across different experimental stimuli (**Figure 1-4**). This inter-stimulus variability is extremely important because it indicates that temporal perception is not rooted solely in biophysical constraints. As illustrated, a straightforward demonstration of stimulus dependency is comparing audiovisual SJ and TOJ studies utilizing impulse stimuli (flashes for vision and beeps or noise bursts for audition) with the same task demands using audiovisual speech. Comparisons of these studies resoundingly indicates that the TBW for speech stimuli is substantially larger than the TBW for impulse

stimuli (examples of speech stimuli include: (Conrey & Pisoni, 2006; Dixon & Spitz, 1980; J. A. Jones & Jarick, 2006; van Wassenhove et al., 2007; Vatakis, Ghazanfar, & Spence, 2008; Vatakis & Spence, 2006a)) (examples of impulses: (Hirsh & Sherrick, 1961; Keetels & Vroomen, 2005; Zampini, Guest, et al., 2005; Zampini et al., 2003). This larger window of integration is particularly striking given the information content of the signals, as speech carries a substantially more refined set of spectro-temporal information than impulses, yet the nervous system seems to leverage this increased information content to actually reduce multisensory temporal acuity. Several non-exclusive explanations for this effect have been proposed,

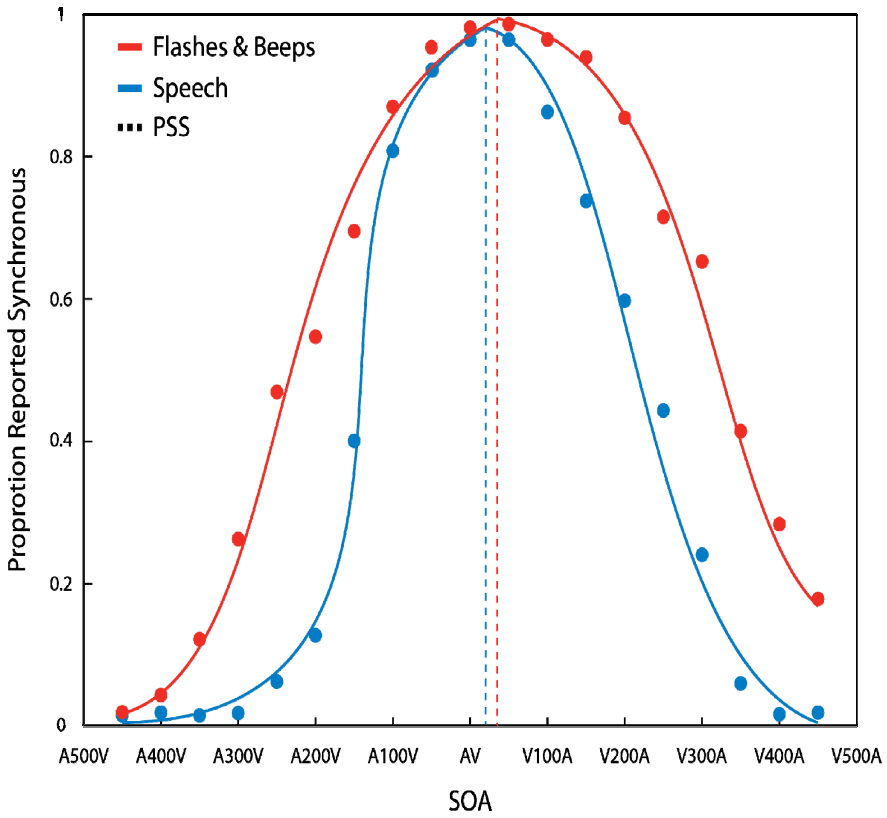

Figure 1-4 Example Temporal Binding Windows
Simulated simultaneity judgement data for a single participant for stimuli consisting of flashes and beeps (blue) and speech (orange). Data was generated for illustrative purposes. The solid lines indicate the logistic psychometric functions fit. Dotted lines indicate the point of subjective simultaneity (PSS). Note the asymmetry, in which visually leading stimuli (right side of figure) are more likely to be judged synchronous than auditory leading stimuli at the same absolute temporal offset (stimulus onset asynchrony – SOA). Similarly, the PSS is at values for which vision leads audition.

including the notion that stimulus 'complexity' drives temporal acuity down (Stekelenburg & Vroomen, 2007; Vatakis & Spence, 2006a, 2006b) or the presence of a 'unity effect' in which prior experience with speech generates a predisposition (i.e. a perceptual prior) for aggregating sensory streams in time (L. H. Chen & Vroomen, 2013; Y. C. Chen & Spence, 2017; Vatakis et al., 2008). The latter unity hypothesis is particularly appealing given the substantial experience most adult research participants have with audiovisual speech and Bayesian views of optimal integration in which experience might shape priors (Alais & Burr, 2004; Ernst & Banks, 2002) (for a review of probabilistic Bayesian inference in the brain see: (Knill & Pouget, 2004)). This unity explanation has been critically tested by using sine-wave speech (SWS), a spectro-temporally impoverished speech signal that is variably perceived as speech or non-speech depending on participant's prior experience (Davis & Johnsrude, 2007; Remez, Rubin, Pisoni, & Carrell, 1981). Critically, this study indicated that SWS signals perceived as non-speech are perceived with the same temporal acuity as SWS signals perceived as speech (Vroomen & Stekelenburg, 2011). This strongly suggests that the majority of differences in temporal acuity for speech and non-speech stimuli are rooted in low-level stimulus properties such as the duration of envelope fluctuations. Perception of synchrony is also affected by the overall duration of the stimuli, which further supports that this might be the case (Boenke, Deliano, & Ohl, 2009; Kuling, van Eijk, Juola, & Kohlrausch, 2012). Additional studies have further indicated that audiovisual stimuli with intermediate levels of 'complexity' and duration, such as two natural objects striking one another, have temporal acuity levels between that of impulses and speech (Vatakis & Spence, 2006a), further reinforcing links between low level stimulus properties and temporal acuity.

In addition to differences rooted in stimulus properties, multisensory temporal acuity also

varies based on task demands. This contrast is most obvious when contrasting the SJ and TOJ tasks directly with one another for physically identical stimuli. An example of such a comparison can be found in (van Eijk, Kohlrausch, Juola, & van de Par, 2008), in which TOJ, SJ, and SJ3 (a version of SJ in which participants may respond "sound-first," "synchronous," or "light-first") tasks were used for near impulse stimuli consisting of a bouncing ball and it's corresponding impact sound. This study found that participant performance (in terms of the point of subjective simultaneity; PSS) for SJ and SJ3 was tightly coupled, but that TOJ performance was strongly decoupled from either SJ task. This finding has also been replicated using a much broader array of stimuli (Love, Petrini, Cheng, & Pollick, 2013). Because the physical stimuli are identical in these experiments, the observed differences in cross-modal timing perception must be accounted for in a top-down (i.e. task dependent) manner. Proposals for this top down regulation include prior entry (Spence & Parise, 2010; Spence, Shore, & Klein, 2001; Titchener, 1908; Zampini, Shore, & Spence, 2005) and differences in participant interpretation which yield response bias (i.e. TOJ instructions might often be interpreted by participants to indicate that perceptual segregation of the stimuli is desired, while the opposite might be inferred for SJ) (van Eijk et al., 2008; Vroomen & Keetels, 2010). In an attempt to disentangle these differences additional task instruction such as 'perceptual fusion' have been employed, in which participants rate stimuli as one or two events (Stevenson & Wallace, 2013). This same study also utilized stimuli of varying complexity to provide a more comprehensive view of the roles both tasks and stimuli play in influencing temporal integration. The primary finding of this study, aside from an important replication of the decoupling between SJ and TOJ tasks, was the finding of a consistent correlation across stimuli for participant performance not just in terms of PSS, but also in terms of TBW size (i.e. the range of stimulus onset asynchronies (SOAs) reported as predominantly

synchronous in SJ). This study thus indicates that the substantial variability in individual temporal acuity (i.e. TBW width) (Powers, Hillock, & Wallace, 2009) not only reflects task and stimulus dependencies, but also represents a trait level variable. Further this study also firmly established that differing methodologies for determining TBW width, such as the distance between ambiguous SOAs (50% rate of reported synchrony) or the distance between predominantly synchronous SOAs (75% or 76% rate of reported synchrony) yield tightly correlated results. This indicates that differences in procedures for fitting psychometric functions to behavioral data, which vary across the literature, have minimal impact when interpreting studies.

The Window of Integration is Plastic in Development and Adulthood

The aforementioned decoupling between individual temporal acuity and dependencies on task and stimuli suggests that variability in multisensory temporal acuity might reflect individualized processing of multisensory temporal cues. One piece of evidence for this is the stability across measurements of inter-individual variability in the audiovisual PSS (Freeman et al., 2013; Grabot & van Wassenhove, 2017; Ipser et al., 2017; J. V. Stone et al., 2001). A second piece of evidence is the extremely large magnitude of inter-individual differences, with differences in TBW size of over two fold being observed. (i.e. (Powers et al., 2009)). Such large inter-individual differences are difficult to account for based on gross biological substrates alone. Such differences are instead better accounted for by experience, or potentially by differences in top down drive. The critical role of experience in temporal acuity is further emphasized by the strong developmental contribution to acuity elucidated by studies charting its trajectory in childhood (Hillock-Dunn, Grantham, & Wallace, 2016; Hillock-Dunn & Wallace, 2012; Hillock, Powers, & Wallace, 2011;

Kaganovich, 2016; Noel, De Niear, Van der Burg, & Wallace, 2016). A lifetime's worth of experience is clearly able to substantially shape the window of integration based on these studies, and this immediately raises the question of how much experience is required to change multisensory temporal perception. An initial approach to this involved extensive multisensory feedback training utilizing impulse stimuli. After approximately five hours of training spaced over the course of a week, participants demonstrated substantial reductions in TBW size for visual leading stimuli which persisted for at least a week (Powers et al., 2009). Substantial experience with artificial stimuli and appropriate feedback is thus able to shape temporal acuity, particularly for stimuli in which vision leads audition. This training has later been extended by targeting both environmentally invalid auditory leading acuity and environmentally valid visual leading acuity (Cecere, Gross, & Thut, 2016). This approach indicated that despite specifically targeting auditory leading temporal acuity, training only improves participant's acuity for visually leading stimuli. As this is the natural configuration of most environmental stimuli these studies together strongly suggest that, in adults, plasticity based on extensive experience is possible, but is restricted to ecologically valid stimulus configurations.

Malleability in temporal perception is not restricted to the longer time scales examined in these perceptual learning studies, however, as anyone who has watched a movie with temporally misaligned audio and video might corroborate. In such a movie, the mouth movements of the actors and the auditory speech streams are often substantially temporally misaligned. Over time, however, these temporal incongruities become less noticeable and eventually disappear. Psychophysical investigation of this adaptation phenomenon, termed temporal recalibration, has revealed that a few minutes of exposure to asynchrony causes participant's temporal perception to recalibrate (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, &

Bertelson, 2004). This adaptation takes the form of the TBW shifting towards the direction of the repeated asynchrony, resulting in the presented stimuli quite literally becoming the 'new synchronous'. Plasticity in temporal perception can thus clearly be driven by the statistics of the environment over time scales measured in minutes.

The environment, however, is dynamic on time scales far faster than minutes and it would thus be advantageous to adapt to asynchrony at extremely high speeds. Investigations specifically targeting rapid adaptation (often referred to as rapid recalibration) have shown that it occurs at the level of single trials for both impulse (Van der Burg, Alais, & Cass, 2013) and speech stimuli (Van der Burg & Goodbourn, 2015) in a manner that can be strongly dissociated from the aforementioned prolonged recalibration (Van der Burg, Alais, & Cass, 2015). Crucially, this rapid adaptation also only occurs for audiovisual stimulus pairings, which have substantially more temporal dynamic range in the natural environment than stimulus pairing which include somatosensory inputs (Van der Burg, Orchard-Mills, & Alais, 2014). Multisensory temporal perception is thus highly plastic over a broad range of timescales ranging from days to seconds, and the ability to dissociate these processes highlights that multiple discrete mechanisms of plasticity contribute to perception of this type.

The Window of Integration Applies Beyond Explicit Temporal Judgments

An important limitation of the work regarding temporal perception thus far discussed is the reliance on explicit temporal judgments with limited ecological validity. Moving away from such tasks is thus an important demonstration that the TBW is a generalized principle of multisensory perception rather than a manifestation of task demands. An excellent way to approach this problem is to employ perceptual illusions and examine their dependency on temporal

relationships. A pair of strong examples of this approach utilized the aforementioned McGurk illusion and probed whether changes in audiovisual temporal relationships affected the rate of the illusory percept (Munhall, Gribble, Sacco, & Ward, 1996; van Wassenhove et al., 2007). These studies revealed that perception of the illusion is affected by temporal concordance with a structure strikingly similar to that found for synchrony judgments, including the anticipated visual first bias common to multisensory temporal tuning. Another illusion leading to a similar conclusion for impulse stimuli is examining the effect of audiovisual synchrony on temporal ventriloquism (Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Temporal ventriloquism is an illusion in which cross modal interactions distort the perceived timing of audiovisual stimuli towards the timing of the auditory sensory stimulus. This has been explained as occurring due to audition's dominance over vision in the temporal domain (modality appropriateness, (Welch et al., 1986)), and more recently and precisely as optimal integration given the superior temporal precision of the auditory system (Fendrich & Corballis, 2001; Repp & Penel, 2002). When the synchrony between the auditory and visual stimuli is manipulated the strength of temporal ventriloquism, defined as the difference between veridical timing and perceived timing, diminishes with increasing degrees of temporal offset (Kuling, Kohlrausch, & Juola, 2013). The sound induced flash illusion similarly relies on temporal alignment and diminishes with increased temporal spacing between the visual flash and auditory stimulus (Shams, Kamitani, & Shimojo, 2002). The generalization of the TBW to tasks in which temporal concordance is not explicitly evaluated serves as strong evidence that the TBW is a generalized manifestation of the temporal principle of multisensory integration. This generality is even further reinforced by the correlation between illusory perception rates for the McGurk illusion, the sound induced flash illusion, and

individual TBW size (Stevenson, Zemtsov, & Wallace, 2012), indicating that individual differences in the temporal dependency of multisensory integration are stable across all these tasks and can be indexed with the SJ task.

Physiological Studies of Multisensory Temporal Processing in Humans

The extensive behavioral characterizations of multisensory temporal integration have, unsurprisingly, motivated researchers to investigate the neural instantiations of these processes. Surprisingly, despite the incredibly rich body of behavioral work reviewed above, our understanding of how these timing processes unfold within the brain is in its relative infancy. Proposals that sensory timing might arise from neural circuits not intrinsically dedicated to the process highlight that the neural basis of temporal perception might be particularly difficult to study, even for unisensory inputs (Johnston & Nishida, 2001; Karmarkar & Buonomano, 2007; van Wassenhove, 2009). Some authors have even gone as far to propose that processing of temporal information in the brain might even be described as degenerate (Friston & Price, 2003), a concept in which multiple non-redundant circuits play the same role of a highly crucial biological function (Lewis & Meck, 2012; Merchant, Harrington, & Meck, 2013). Degeneracy would allow multiple distinct, yet functionally equivalent, neural architectures for temporal perception tasks (Wiener, Matell, & Coslett, 2011). The brain might also be able to keep time in the absence of a *dedicated* anatomical or functional "clock" circuit by comparing temporally evolving network states (Bueno et al., 2017; Karmarkar & Buonomano, 2007), but this has not been explicitly demonstrated for cross-modal timing. The available hypotheses and evidence for neural instantiations of temporal processing, including for audiovisual inputs, are reviewed below, but it is noteworthy that empirical evidence for these mechanisms is relatively sparse

(Aghdaee, Battelli, & Assad, 2014).

Temporal Coding Through Latency

In regards to computing relative timing, one hypothesis that has been investigated is whether temporal characteristics of stimuli can be encoded in the actual temporal dynamics of the neural activity generated by sensory inputs. In other words, temporal order is coded through the absolute latency of neural activity in this framework. This possibility was investigated by (Vibell, Klinge, Zampini, Spence, & Nobre, 2007) using EEG to elucidate a modest acceleration of ERPs to attended sensory modalities, on the order of 4-17 ms depending on the participant and the ERP peak examined. In the context of prior entry theory (Spence & Parise, 2010; Zampini, Shore, et al., 2005) this acceleration was interpreted quite literally as attention accelerating the temporal dynamics of neural activity and the perceived temporal ordering of events in a consistent manner. Notably, this effect is not present for spatial cuing (McDonald, Teder-Salejarvi, Di Russo, & Hillyard, 2005), which instead demonstrated a relationship between changes in ERP amplitude and explicit timing. Other investigations of temporal processing have revealed that increased auditory ERP magnitude, hypothesized to represent better stimulus encoding, might play a role in differences in temporal acuity across individuals (Kaganovich & Schumaker, 2016). The strength of ERPs believed to originate from early auditory regions and possibly the STS also appear to be related to subjective temporal perception of speech (Huhn, Szirtes, Lorincz, & Csepe, 2009). These findings are congruent with the notion that auditory inputs are dominant for temporal processing (Repp & Penel, 2002; Welch et al., 1986) and similarly agree with the proposed temporal organization hypothesis, in which an ERP of greater magnitude would reach a given amplitude threshold more quickly after stimulus onset.

Importantly, a strengthened response should also carry a more precise temporal estimate and yield an unbound percept (i.e. stimuli reported as occurring at different times), which is what these studies generally report. Together these studies indicate that absolute strength and timing of phase and time locked neural responses may encode crucial temporal information about the stimulus, and that this form of encoding might be particularly viable for auditory inputs. Further support for the notion that the temporal dynamics of neural activity might themselves form the mechanism of temporal encoding comes from recent work using representational similarity analysis (Cecere, Gross, Willis, & Thut, 2017). In this study, the authors tested the hypothesis that neural network engagement differs based on stimulus ordering, and the confirmation of this theory means that different neural circuits (and thus different neural timing characteristics) are engaged depending on which stimulus occurs first. This study thus allows the temporal coding hypothesis to correctly account for the marked visual leading asymmetry seen in behavioral TBWs (reviewed above) even though absolute latencies of evoked components are relatively consistent for different stimulus orders.

Temporal Coding Using an Oscillatory Clock

Another potential mechanism for representation of time is neural oscillations, which could provide a cyclic internal clock against which stimuli could be indexed. This clock could also contribute to the eventual percept by binding stimulus pairings which occur within a single oscillatory cycle while segregating those that do not. This clock has been proposed to exist both locally (i.e. within sensory cortices), but also proposed to reside in subcortical hubs connected to multiple regions such as the basal ganglia (Kononowicz & van Wassenhove, 2016), an important integrator of cortical oscillatory activity (Coull, Cheng, & Meck, 2011). These oscillations are

clearly able to encode or entrain to the temporal structure of stimuli (Cravo, Rohenkohl, Wyart, & Nobre, 2011; Giraud & Poeppel, 2012; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008) and are believed to drive well documented temporal attention affects (M. R. Jones, Moynihan, MacKenzie, & Puente, 2002). Their role in *perceived* timing is less clear, however. Recently there have been attempts to reconcile coupled oscillator models of slow neural oscillations and working memory as a mechanism of interval timing (Gu, van Rijn, & Meck, 2015; Jensen, Gips, Bergmann, & Bonnefond, 2014). These studies suggest that there is substantial promise in oscillatory timing models, but a direct demonstration of a role for phase coupling over the short timing intervals used in the popular SJ and TOJ tasks remained un-assessed until the current thesis work, which in part aimed to remedy this empirical gap.

Support for the presence and utility of an oscillatory clock at low frequencies can be found in MEG work demonstrating that phase shifts in auditory cortex, and thus phase differentials between auditory and visual cortices, explains a substantial portion of sustained audiovisual temporal recalibration (Kosem, Gramfort, & van Wassenhove, 2014). In this study, exposure to asynchrony shifted the phase of ongoing activity in auditory and visual cortices and the magnitude of the shift in auditory cortex predicted the degree of shift in participant's temporal perception. Additional indirect evidence of oscillatory contributions to multisensory temporal perception come from experiments extending findings of oscillatory speed (i.e. frequency) as a contributor to the temporal resolution of vision (Samaha & Postle, 2015) to the multisensory domain (although see (Baumgarten, Schnitzler, & Lange, 2017) for evidence refuting this effect for somatosensory inputs). An example of such a study used the sound induced flash illusion and found that individual differences in oscillatory speed map onto individual susceptibility to the illusion (Cecere, Rees, & Romei, 2015). Specifically, individuals with a faster oscillatory clock

had narrower intervals (i.e. TBWs) over which stimuli elicited the illusion. This study then drove changes in individual illusion susceptibility using transcranial alternating current stimulation (TACS), causally linking participant's perceptual thresholds with oscillatory speed. In other words, experimentally accelerating the participant's internal oscillatory clock made them more sensitive to temporal asynchrony and thus less susceptible to the illusion.

Notably, none of the aforementioned studies indicate a role for absolute oscillatory *phase* in audiovisual time perception, even when studies have been specifically designed *a-priori* to examine phase (Grabot, Kosem, Azizi, & van Wassenhove, 2017). This lack of an absolute phase effect is interesting given the importance of phase for temporal perception within sensory modalities using nearly identical perceptual tasks such as tactile simultaneity judgment (Baumgarten, Schnitzler, & Lange, 2015) and the profound temporal rhythmicity phase instantiates in primary cortices (Lakatos et al., 2007). It is possible that audiovisual phase effects are more complex or individualized due to the cross-modal nature of processing (i.e. a phase co-modulation effect, see (Henry, Herrmann, & Obleser, 2014) for an auditory example). Such a relationship is suggested by work showing oscillatory phase effects for auditory speech perception, which are highly consistent with audiovisual phase reset (Ten Oever & Sack, 2015). In this study, theta phase shaped syllable perception for ambiguous syllables, but the absolute phase angles were strongly individualized. In a true audiovisual experiment, potentially individualized phase differences for each modality might compound, greatly obscuring the influence of phase. Phase individualization in multisensory processing is also suggested by the results reported in (Kambe, Kakimoto, & Araki, 2015), which found increases in post-stimulus phase consistency (quantified via inter-trial phase coherence [ITC]) in the beta band at central electrodes when flashes and beeps were perceived as simultaneous, but no relationship to a

specific phase angle. It is also possible that individual variability in audiovisual simultaneity thresholds (Powers et al., 2009; Stevenson & Wallace, 2013), which has been suggested to be a trait level variable (Grabot & van Wassenhove, 2017), obscures phase dependency effects. As an example of such a phenomenon, individualized and inverting differences were found for alpha power between correct and incorrect judgments in the auditory cortex during audiovisual TOJ (Grabot et al., 2017). In other words, the sign of oscillatory power difference flips depending on the participant's inherent response bias for audiovisual order, and a similar flip for phase may be at work. Clarifying the role of phase in explicit multisensory timing remains an important goal.

Importantly, latency and oscillatory timing models are unlikely to be exclusive or exhaustive, and work is ongoing to characterize the way temporal structure and stimulus encoding interact (for examples see: (Lakatos et al., 2008; Marchant & Driver, 2013; D. M. Simon, Wallace, M. T., 2017)). The current work also makes contributions to this topic by beginning to disentangle the relative timing and nature of temporal encoding in the human brain. Combined with recently developed animal models of audiovisual temporal processing and perception (Schormans et al., 2016) there are now promising targets for interrogating how the brain encodes cross modal temporal structure.

Explicit Timing Engages Distributed Neural Networks

Regardless of the underlying physiological timing mechanisms, it is clear that a broadly distributed network of regions is involved in the processing of timing information. The majority of evidence for this comes from fMRI and positron emission tomography (PET) studies, which have superior spatial precision to EEG/MEG. These studies have shown that audiovisual synchronicity modulates brain responses in areas including the STS, SC, anterior insula, and

intraparietal sulcus (Bushara, Grafman, & Hallett, 2001; Macaluso et al., 2004; L. M. Miller & D'Esposito, 2005). Activity in the temporal-parietal junction and dorsolateral prefrontal cortex have also been reported (Adhikari, 2013). Additionally, the Blood Oxygen Level Dependent (BOLD) response in some of these regions, in particular the STS, demonstrates a tuning function with substantial similarity to the TBW found in behavioral studies (Stevenson et al., 2010). Differences in activity in these areas also emerge for stimulus trains when temporal correspondence is manipulated (Marchant, Ruff, & Driver, 2012; Noesselt et al., 2007), indicating that they are sensitive to temporal correspondence across modalities rather than just stimulus ordering. Notably, the brain regions recruited when judging audiovisual synchronicity (i.e. SJ task, occurrence at the same time) only partially overlap with those recruited by cross-modal interval timing for auditory and visual stimuli, which includes the basal ganglia and thalamus in addition to the expected superior temporal cortex (D. L. Harrington, Castillo, Fong, & Reed, 2011). This serves as evidence that perceptual processes might be more strongly localized in cortical areas, while the fundamental circuitry for sensory interval timing might be housed in subcortical circuits such as the basal ganglia, as proposed by the oscillator model (Kononowicz & van Wassenhove, 2016).

Building upon this, a limited number of these studies have attempted to disambiguate processing of synchrony (i.e. physical stimulus characteristics) from processing of subjective perception (i.e. whether the percept is integrated and reported as occurring at the same time). One such study found that STS activation in particular is seemingly more related to subjective temporal perception than physical timing (Stevenson et al., 2011). Specifically, this study found that the fMRI BOLD response in STS for a mildly asynchronous but perceptually fused stimulus was indistinguishable from a truly synchronous stimulus. Similar results were found contrasting

a mildly asynchronous but perceptually unfused stimulus with a highly asynchronous stimulus. Additional similar results linking the fMRI BOLD response in STS to subjective perception have also been demonstrated for audiovisual motion (Bushara et al., 2003). The STS is also hypothesized to be one of the cortical generators contributing to differences in evoked auditory responses to speech between fused and unfused percepts (Huhn et al., 2009). These studies establish that activity in STS in particular may be more closely related to perception than to stimulus processing, an important distinction given the generally accepted integrative functions of STS.

Lastly, just as SJ and TOJ tasks provide different behavioral results (Love et al., 2013; Stevenson & Wallace, 2013; van Eijk et al., 2008), the neural networks engaged by the two tasks have recently been elucidated to differ. In an fMRI study, (Binder, 2015), the author found that additional parietal and parieto-occipital brain regions were activated for TOJ that were not activated for SJ. The reverse was not true, however, as no SJ specific regions were found. This result suggests that additional cognitive resources are utilized in the TOJ task, possibly for forming multiple discrete sensory representations *before* comparing timing. This study also shows, for the first time, that top down task demands drive the selection of brain networks for processing audiovisual temporal relationships, providing a ready explanation for why SJ and TOJ produce distinct behavioral results. Importantly, this result has also been replicated to a fair degree in the unisensory domain using tactile TOJ and SJ (Miyazaki et al., 2016), although the authors did find SJ specific activation in the insula. Top down network construction for temporal processing can thus be conceived of as distinct from stimulus properties, a notion that was further examined in the current work.

Disruptions of Audiovisual Temporal Processing in Clinical Populations

Multisensory temporal function has recently emerged as an important topic of study in a number of clinical populations that present changes in sensory function. In particular, alterations in sensory function, including disruptions in temporal function, have been hypothesized to disrupt the development of higher order cognitive functions in developmental disabilities (Wallace & Stevenson, 2014). In autism spectrum disorder (ASD), for example, there are a number of behavioral studies indicating that auditory and visual cues for ecologically important speech stimuli are not integrated in the same manner as they are in typically developing (TD) individuals (Foxe et al., 2015; Magnee, de Gelder, van Engeland, & Kemner, 2008; Smith & Bennetto, 2007; Stevenson, Siemann, Woynaroski, et al., 2014a; Williams, Massaro, Peel, Bosseler, & Suddendorf, 2004; Woynaroski et al., 2013). Generalized disruptions in sensory function have also recently gained increased prominence in terms of diagnostic criteria for ASD (American Psychiatric Association, 2013). A key finding stemming from investigation using the SJ task is that children with ASD have audiovisual temporal acuity comparable to that of TD children for impulse stimuli and intermediate stimuli consisting of a hammer strike and the associated sound. For audiovisual speech stimuli, however, children with ASD have a significantly larger TBW (Stevenson, Siemann, Schneider, et al., 2014). In other words, for ecologically important speech stimuli children with ASD are less able to segregate stimuli with substantial temporal offsets than their TD peers. The speech specificity of this finding is also consistent with previous work indicating differences in processing between speech and non-speech stimuli in ASD (Mongillo et al., 2008).

Similar enlargement of the TBW in ASD has also been found in a number of other approaches to multisensory temporal processing (Bebko, Weiss, Demark, & Gomez, 2006; de

Boer-Schellekens, Eussen, & Vroomen, 2013; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011), including for audiovisual illusions (Foss-Feig et al., 2010; Woynaroski et al., 2013). For example, by parametrically varying the SOA between the flash and the second beep it was discovered that the temporal tolerance of the sound induced flash illusion is substantially elevated in individuals with ASD (Foss-Feig et al., 2010). In natural environments, such as a classroom, this increased window of temporal integration might result in incorrect or inappropriate integration of stimuli that should otherwise be perceptually segregated, such as visual and auditory speech originating from different speakers (Stevenson, Segers, et al., 2017). Additional evidence of audiovisual temporal dysfunction in ASD is provided by the fact that, during the SJ task, rapid recalibration (i.e. single trial adaptation) also does not occur or is atypical in individuals with ASD (Noel, De Niear, Stevenson, Alais, & Wallace, 2016; Turi, Karaminis, Pellicano, & Burr, 2016). This lack of adaptation indicates that not only is the overall precision of temporal acuity reduced in this population for some stimuli, but that adaptive processes believed to incorporate the temporal statistics of the natural environment are also impaired or function differently.

Little is currently known about the physiological bases of temporal processing deficits in ASD, although differences in neural measures of early multisensory integration for simple audiovisual and audiotactile stimuli have been observed in ASD (Brandwein et al., 2013; Russo et al., 2010). It has been hypothesized, however, that multisensory temporal function might be exceedingly vulnerable to disruptions in oscillatory synchronization processes (D. M. Simon & Wallace, 2016). These local (i.e. power) and interregional (i.e. coherence) synchronization processes have previously been shown to be associated with a host of unisensory integrative deficits in ASD (Peiker et al., 2015; Stroganova et al., 2012; Sun et al., 2012). Individuals with

ASD also show atypical neural coherence during resting state (reviewed in: (J. Wang et al., 2013)) and after sensory stimulation (reviewed in: (D. M. Simon & Wallace, 2016)), strongly suggested the possibility of such disruption. Differences in local and regional circuit synchronization are thus likely contributors to multisensory temporal dysfunction in ASD.

In developmental disabilities other than ASD, such as schizophrenia and dyslexia, similar reductions in multisensory temporal acuity have been found. For example, individuals with dyslexia continue to benefit from auditory flankers during visual-visual TOJ at far greater intervals than their TD peers (Hairston, Burdette, Flowers, Wood, & Wallace, 2005). This flanking effect is a form of temporal ventriloquism (Morein-Zamir et al., 2003) and these differences indicate that auditory stimuli which are perceptually segregated in TD individuals are still 'pulling' the temporal anchors of visual stimuli for individuals with dyslexia. More recently, utilization of the SJ task in adults with dyslexia found modest enlargement of the TBW for both speech and non-speech stimuli (Francisco, Jesse, Groen, & McQueen, 2017). These studies suggest that there may be a generalized deficit in multisensory temporal integration in dyslexia, although additional work in this population is clearly needed given the sparsity of findings. Individuals with schizophrenia similarly present an enlarged TBW for audiovisual stimuli when tested using both the sound induced flash illusion (Hass et al., 2017) and the SJ task (Martin, Giersch, Huron, & van Wassenhove, 2013). Multisensory temporal function has also been shown to be associated with other more classical schizophrenia symptoms such as hallucinations (Stevenson, Park, et al., 2017). Importantly, generalized multisensory dysfunction in schizophrenia is better studied than in dyslexia, supporting that dysfunctional temporal processing is but one aspect of more generalized cross-modal processing deficits in this population (Balz, Romero, et al., 2016; Foucher, Lacambre, Pham, Giersch, & Elliott, 2007; Roa

Romero, Keil, Balz, Gallinat, & Senkowski, 2016; Romero et al., 2016; Stekelenburg, Maes, Van Gool, Sitskoorn, & Vroomen, 2013; D. B. Stone, Coffman, Bustillo, Aine, & Stephen, 2014; Tseng et al., 2015).

The commonality of disruptions in multisensory temporal processing across multiple developmental disorders provides one of the primary motivations for the investigations pursued in the current work. Specifically, it is believed that disruptions in multisensory processes make contributions to the development of deficits in higher order social and communicative processes that serve as the hallmark of autism spectrum disorder in particular (American Psychiatric Association, 2013; Wallace & Stevenson, 2014). It is hoped that a better understanding of the neural bases of multisensory temporal processing in adults might lay the foundation for studying multisensory temporal processing in clinical disorders and development. Such work is hoped to eventually provide tools for identification of sensory dysfunction which might serve an important role in early diagnosis and treatment of these disorders.

Electroencephalography as a Tool for Studying Multisensory Temporal Processing

Temporal Resolution

Electroencephalography (EEG) is one of many non-invasive approaches available for interrogating multisensory processing in the human brain. Its primary strength is its high temporal resolution, which yields a well resolved measure of neural processing (Nunez & Srinivasan, 2006). This high temporal resolution affords the ability to both resolve activity separately for each of the component sensory stimuli and examination of neural oscillations on time scales as short as 10-20 milliseconds. These neural oscillations (Buzsaki & Draguhn, 2004;

D. M. Simon & Wallace, 2016) have been increasingly recognized as important contributors to processing of sensory inputs and perception. These oscillations at the level of the scalp are believed to index synchronization of postsynaptic activity in localized cortical circuits on the scale of millimeters or centimeters (Buzsaki, Anastassiou, & Koch, 2012). Successful perception of near threshold stimuli has been shown to depend on the phase of these neural oscillations at stimulus onset, which links them to local network excitability for visual, auditory, and somatosensory stimuli (Ai & Ro, 2014; Henry et al., 2014; Henry & Obleser, 2012; Mathewson, Gratton, Fabiani, Beck, & Ro, 2009; Spaak, de Lange, & Jensen, 2014; Strauss, Henry, Scharinger, & Obleser, 2015). This consistency across modalities suggests that rhythmicity and resonance are intrinsic properties of the nervous system contributing to perception, and are not sensory system specific or epiphenomenal of highly localized activity (i.e. action potentials). This recognized importance motivates the use of tools with the ability to accurately measure these fluctuations, including EEG, MEG, and invasive physiological recordings. Furthermore, rhythmicity in neural processing is also believed the play an important role in communication between anatomically and functionally distinct cortical networks. Synchronization between these networks is an important tool that the brain utilizes to multiplex information (Akam & Kullmann, 2014) and information flow between neural networks, controlled through differential synchronization, is believed to be an important factor in virtually every facet of human cognition (Rodriguez et al., 1999; Siegel, Buschman, & Miller, 2015; Siegel et al., 2012). Given the obligatory communication between sensory systems during multisensory integration, utilization of tools able to capture transient interregional synchronization is important for understanding the mechanistic basis of integration. Advances in signal processing over the last decade have also provided a myriad of improved approaches for investigating neural synchronization with EEG.

These approaches allow resolving of connectivity with a degree of fidelity that was not previously possible, and have permitted new functional insights into neural mechanisms of information transfer (for examples of recent advances in the phase and amplitude envelope domains see (Hipp et al., 2012; Vinck, Oostenveld, van Wingerden, Battaglia, & Pennartz, 2011)). In the context of temporal processing tasks that typically utilize discrete stimuli which generate powerful evoked responses and obscure connectivity, these sophisticated techniques offer the opportunity to partially resolve the brain networks performing these computations.

Physiological Relevance

As suggested by the oscillatory excitability framework, EEG directly indexes physiological processes related to population level neural activity known as the local field potential (LFP; a measure of voltage change in proximity to a recording electrode) (Buzsaki et al., 2012). This contrasts activity such as the blood oxygen level dependent (BOLD) signal indexed by functional magnetic resonance imaging (fMRI) which has less direct physiological correlates (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). The LFP is known to indirectly measure activity of individual neurons participating in networks, as membrane voltage fluctuations in individual neurons are strongly coupled to LFP changes and thus correspond with the degree of input required to fire an action potential (X. J. Wang, 2010; X. J. Wang & Buzsaki, 1996). This correspondence is particularly strong in the high gamma band which does not penetrate the skull easily (Ray & Maunsell, 2011), but high gamma is also known to be coupled to lower frequency activity readily indexed by scalp EEG (Canolty et al., 2006; Canolty & Knight, 2010). This strong EEG to LFP coupling thus allows the EEG signal to be interpreted somewhat more directly in terms of network excitability and yields somewhat more focal inference about the

functional role of activation.

Translational Promise

Lastly, EEG has a substantial degree of portability and ease of use that allows relatively easy translation of research findings to pediatric, clinical, and, potentially, non-verbal participants. While such populations are not the focus of the current work, translating these findings to clinical and pediatric populations was considered an important motivation for the current work and thus contributed to the selection of the approach. For example, it is currently known that children integrate important multisensory cues such as visual speech to a lesser degree than adults, and that integration of these signals has a relatively smooth developmental trajectory (Foxe et al., 2015). Investigations utilizing EEG have likewise indicated that neural markers of audiovisual speech integration have a protracted developmental time course (Kaganovich & Schumaker, 2014; Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014). Extensibility to developmental population is thus an important factor in the selection of EEG for the current investigations, given behavioral evidence for a protracted developmental trajectory for multisensory temporal acuity (Hillock-Dunn et al., 2016; Hillock-Dunn & Wallace, 2012; Hillock et al., 2011; Noel, De Niear, Van der Burg, et al., 2016). Extension to pediatric populations also offers to elucidate developmental changes in information transfer underlying multisensory integration, which to date has not been investigated. Additionally, dysfunction of multisensory temporal processing and audiovisual speech integration has been noted in a number of clinical populations such as ASD (Foxe et al., 2015; Stevenson, Siemann, Schneider, et al., 2014), and schizophrenia (Hass et al., 2017; Ross, Saint-Amour, Leavitt, Molholm, et al., 2007; van Wassenhove et al., 2005). These same clinical populations have been identified to have

deficits in the formation of local and long range oscillatory neural networks readily assayed with EEG or MEG (D. M. Simon & Wallace, 2016; P. J. Uhlhaas & Singer, 2006; P. J. Uhlhaas, Singer W., 2007; P. J. Uhlhaas & Singer, 2012). Preliminary work also indicates that numerous aspects of the EEG signal might be of substantial utility for early diagnosis of ASD in particular (Bosl, Tierney, Tager-Flusberg, & Nelson, 2011; Damiano C.R. et al., 2017; Gabard-Durnam, Tierney, Vogel-Farley, Tager-Flusberg, & Nelson, 2015; Righi, Tierney, Tager-Flusberg, & Nelson, 2014; D. M. Simon, Damiano, et al., 2017). The ability to extend the current approaches to clinical populations was thus a crucial factor in the experimental designs used and the selection of EEG.

Introduction to the Current Dissertation Work

The current dissertation work is motivated by a growing body of literature indicating that multisensory temporal processing in particular is disrupted in developmental disabilities such as ASD (Foss-Feig, Heacock, & Cascio, 2012; Kwakye et al., 2011; Noel, De Niear, Stevenson, et al., 2016; Stevenson, Siemann, Schneider, et al., 2014; Stevenson, Siemann, Woynaroski, et al., 2014b), schizophrenia (Hass et al., 2017; Martin et al., 2013) and dyslexia (Francisco et al., 2017; Hairston et al., 2005). These same clinical populations also show disruptions in multisensory integration for signals where temporal cues are of high importance, such as audiovisual speech (Foxe et al., 2015; Ross, Saint-Amour, Leavitt, Molholm, et al., 2007; Stevenson, Siemann, Woynaroski, et al., 2014a; Woynaroski et al., 2013). This body of work also strongly suggests that dysfunction in multisensory temporal processing may make substantial contributions to clinically relevant core symptoms of these disorders such as

disrupted social communication in ASD (Wallace & Stevenson, 2014; Woynaroski et al., 2013). As multisensory temporal training has been proposed as a method for ameliorating these deficits (Wallace & Stevenson, 2014), identification of neural markers for specific stages of multisensory processing is also an important step towards neurophysiological validation to interventions.

Unfortunately, limitations in our understanding of how the typically developing brain processes multisensory temporal information impairs our ability to construct investigations into the neural basis and developmental time course of these deficits. A better understanding of these processes is thus the primary focus of this volume, which represents the first comprehensive investigation of temporal perception of audiovisual speech using EEG. To approach the objective of comprehensively characterizing the tuning, top-down regulation, and plasticity of audiovisual temporal processing in adults we designed a series of experiments utilizing the SJ task. In this task, participants report whether sensory stimuli presented with varying degrees of temporal offset occurred synchronously or asynchronously. This task has been utilized extensively in the literature in both typically and atypically developing individuals to characterize TBWs, and is strongly believed to index multisensory temporal processing (Vroomen & Keetels, 2010; Wallace & Stevenson, 2014). Crucially, this task is highly flexible, in that sensory stimuli of varying content such as flashes and beeps or audiovisual speech can be presented within the same experimental framework. Each experiment was designed to isolate and evaluate one or more individual stages of audiovisual temporal processing through manipulations to the temporal alignment of the physical stimuli and changing psychophysical task demands. We collected physiologically interpretable measures of neural processing by concurrently recording EEG as participants performed these tasks. Critically, we also examined how neural processing related to participant's perceptual judgments, thus establishing direct links between neural activity and

participant's behavior.

The Neural Correlates of Temporal Integration of Audiovisual Speech

In our first experiment (experiment 1, chapter 2), we examined the degree to which neural responses to auditory speech change with the addition of visual speech. Previous work has shown that neural responses to auditory speech are of smaller magnitude in the presence of congruent visual speech (Baart, 2016; Besle et al., 2004; van Wassenhove et al., 2005). Additionally it is also known that this sub-additive effect is absent given a substantial auditory leading misalignment (Pilling, 2009). The magnitude of this suppression effect across a full range of temporal alignments, however, has not previously been explored. To determine the limits of temporal tolerance for this response suppression effect, we employed a combination of audiovisual speech stimuli and manipulated the temporal relationship between the auditory and visual streams. This experiment served to construct, for the first time, an accurate neural description of temporal tolerance in audiovisual speech integration with remarkable similarities to the TBW found in behavioral studies. We also identified, for the first time, that low theta band power occurring relatively late after stimulus onset depends on temporal alignment between the sensory inputs. Critically, we were then able to relate the strength of this theta band activity to individual differences in multisensory temporal acuity. This study thus simultaneously elucidated the effects of temporal integration in early cortical regions while also identifying the first known correlate of individual temporal acuity for audiovisual speech.

Top-Down Control of Multisensory Information Flow

In our next study (experiment 2, chapter 3), we sought to extend our previous work by

determining whether the theta band correlate of temporal processing isolated in experiment 1 depended on directed processing of audiovisual temporal structure. We combined a reduced experimental design from experiment 1 with an additional speeded response condition, and alternated participants randomly between these task demands. This approach allowed us to contrast physically identical stimuli during which the top down demand for processing of multisensory temporal structure was either present or absent. We demonstrate for the first time that the low theta power we previously linked to participant performance is modulated by temporal relationships only when participants are actively attending temporal structure. We further demonstrate that phase coupling between the power modulated local network and other brain regions differs based on task demands. Both local and interregional network synchronization thus serve to support active processing of audiovisual temporal concordance. These results establish a novel and critical role of top down regulation of neural coherence in multisensory temporal acuity tasks.

Single Trial Adaptation as a Manifestation of Plasticity in Sensory Evidence Accumulation

Lastly, in a third experiment (experiment 3, chapter 4), we then sought to investigate the neural basis of single trial plasticity in audiovisual temporal perception (Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015; Van der Burg et al., 2014). Specifically, we combined the experimental approach employed in experiment 1 with an analytical approach previously utilized to elucidate the physiological bases of single trial plasticity (D. M. Simon, Noel, & Wallace, 2017). Relative to the previous work, we utilized audiovisual speech stimuli, which have a substantially larger TBW than flashes and beeps (Stevenson & Wallace, 2013). This larger binding window allows for an accentuated degree of plasticity in individual temporal

recalibration, which has been shown to correlate with TBW width (D. M. Simon, Noel, et al., 2017; Van der Burg et al., 2013). Further, by employing a speeded design we were able to model the evolving the decisional process using the drift diffusion model (Ratcliff & McKoon, 2008; Vandekerckhove & Tuerlinckx, 2007, 2008), which posits that choice and response time are based on the temporal evolution of an underlying decision variable which integrates sensory evidence over time. Lastly, audiovisual speech stimuli have fewer sharp transients in stimulus energy than flashes and beeps, affording a uniquely temporally resolved view of the unfolding decisional process that is obscured for impulse stimuli (Kelly & O'Connell, 2013; O'Connell, Dockree, & Kelly, 2012). Together these strengths allowed us to elucidate that single trial plasticity affects neural processes associated with decision making, rather than early sensory processing. Importantly, we were able draw links between changes in these physiological processes and drift rate, which quantifies the speed and direction of the internal decision variable. This study thus established for the first time that single trial plasticity in temporal perception occurs because the internal decision signal which determines participant choice incorporates the sensory past into its trajectory.

Together these three experiments serve to elucidate multiple aspects of the neural basis of multisensory temporal processing in the human brain. These aspects range the full span from changes in processing in initial auditory cortical circuits, to transmission of timing information between synchronized brain networks, to the accumulation of sensory evidence controlling behavioral responses. These stages of neural processing span the full breadth of the sensorimotor hierarchy and highlight the importance of sensory integration at every stage of the transformation from sensory inputs to motor outputs. In sum, they offer a novel and unique perspective into how temporal factors shape neural activity and contribute to appropriate and robust integration of

audiovisual sensory inputs. Together these studies also offer, for the first time, a substantial number of promising targets for neurophysiological study in clinical populations with multisensory temporal processing deficits, such as ASD, schizophrenia, and dyslexia.

References

Adhikari, B.M.; Goshorn, E.S.; Lamichhane, B.; Dhamala, M.;. (2013). Temporal-order judgment of audiovisual events involves network activity between parietal and prefrontal cortices. *Brain Connectivity, 3*(5), 536-545.

Aghdaee, S. M., Battelli, L., & Assad, J. A. (2014). Relative timing: from behaviour to neurons. *Philos Trans R Soc Lond B Biol Sci, 369*(1637), 20120472.

Ai, L., & Ro, T. (2014). The phase of prestimulus alpha oscillations affects tactile perception. *J Neurophysiol, 111*(6), 1300-1307.

Akam, T., & Kullmann, D. M. (2014). Oscillatory multiplexing of population codes for selective communication in the mammalian brain. *Nat Rev Neurosci, 15*(2), 111-122.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol, 14*(3), 257-262.

Alcala-Quintana, R., & Garcia-Perez, M. A. (2013). Fitting model-based psychometric functions to simultaneity and temporal-order judgment data: MATLAB and R routines. *Behav Res Methods, 45*(4), 972-998.

Alvarado, J. C., Vaughan, J. W., Stanford, T. R., & Stein, B. E. (2007). Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *J Neurophysiol, 97*(5), 3193-3205.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: American Psychiatric Association.

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension - 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology, 50*(1-2), 157-163.

Avillac, M., Ben Hamed, S., & Duhamel, J. R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *J Neurosci, 27*(8), 1922-1932.

Baart, M. (2016). Quantifying lip-read-induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays. *Psychophysiology, 53*(9), 1295-1306.

Balz, J., Keil, J., Romero, Y. R., Mekle, R., Schubert, F., Aydin, S., . . . Senkowski, D. (2016). GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. *Neuroimage, 125*, 724-730.

Balz, J., Romero, Y. R., Keil, J., Krebber, M., Niedeggen, M., Gallinat, J., & Senkowski, D. (2016). Beta/Gamma Oscillations and Event-Related Potentials Indicate Aberrant Multisensory Processing in Schizophrenia. *Frontiers in Psychology, 7*.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci, 17*(3), 377-391.

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.

Baum, S. H., Martin, R. C., Hamilton, A. C., & Beauchamp, M. S. (2012). Multisensory speech perception without the left superior temporal sulcus. *Neuroimage, 62*(3), 1825-1832.

Baumgarten, T. J., Schnitzler, A., & Lange, J. (2015). Beta oscillations define discrete perceptual cycles in the somatosensory domain. *Proc Natl Acad Sci U S A, 112*(39), 12187-12192.

Baumgarten, T. J., Schnitzler, A., & Lange, J. (2017). Beyond the Peak - Tactile Temporal Discrimination Does Not Correlate with Individual Peak Frequencies in Somatosensory Cortex. *Frontiers in Psychology, 8*.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci, 7*(11), 1190-1192.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*(5), 809-823.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci, 30*(7), 2414-2417.

Beauchamp, M. S., Yasar, N. E., Frye, R. E., & Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *Neuroimage, 41*(3), 1011-1020.

Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal

synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry, 47*(1), 88-98.

Benevento, L. A., Fallon, J., Davis, B. J., & Rezak, M. (1977). Auditory-Visual Interaction in Single Cells in Cortex of Superior Temporal Sulcus and Orbital Frontal Cortex of Macaque Monkey. *Experimental Neurology, 57*(3), 849-872.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension - 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology, 50*(1-2), 147-155.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci, 20*(8), 2225-2234.

Bhattacharya, J., Shams, L., & Shimojo, S. (2002). Sound-induced illusory flash perception: role of gamma band responses. *Neuroreport, 13*(14), 1727-1730.

Binder, M. (2015). Neural Correlates of Audiovisual Temporal Processing - Comparison of Temporal Order and Simultaneity Judgments. *Neuroscience, 300*, 432-447.

Bishop, C. W., & Miller, L. M. (2009). A Multisensory Cortical Network for Understanding Speech in Noise. *Journal of Cognitive Neuroscience, 21*(9), 1790-1804.

Boenke, L. T., Deliano, M., & Ohl, F. W. (2009). Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. *Experimental Brain Research, 198*(2-3), 233-244.

Bosl, W., Tierney, A., Tager-Flusberg, H., & Nelson, C. (2011). EEG complexity as a biomarker for autism spectrum disorder risk. *BMC Med, 9*, 18.

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2013). The development of multisensory integration in high-functioning autism: high-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cereb Cortex, 23*(6), 1329-1341.

Brang, D., Taich, Z. J., Hillyard, S. A., Grabowecky, M., & Ramachandran, V. S. (2013). Parietal connectivity mediates multisensory facilitation. *Neuroimage, 78*, 396-401.

Brang, D., Towle, V. L., Suzuki, S., Hillyard, S. A., Di Tusa, S., Dai, Z. T., . . . Grabowecky, M. (2015). Peripheral sounds rapidly activate visual cortex: evidence from electrocorticography. *Journal of Neurophysiology, 114*(5), 3023-3028.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K., . . . Fink, G. R.

(2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron, 29*(1), 287-296.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol, 46*(2), 369-384.

Bruns, A., Eckhorn, R., Jokeit, H., & Ebner, A. (2000). Amplitude envelope correlation detects coupling among incoherent brain signals. *Neuroreport, 11*(7), 1509-1514.

Budinger, E., Heil, P., Hess, A., & Scheich, H. (2006). Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience, 143*(4), 1065-1083.

Bueno, F. D., Morita, V. C., de Camargo, R. Y., Reyes, M. B., Caetano, M. S., & Cravo, A. M. (2017). Dynamic representation of time in brain states. *Sci Rep, 7*, 46053.

Burnett, L. R., Stein, B. E., Chaponis, D., & Wallace, M. T. (2004). Superior colliculus lesions preferentially disrupt multisensory orientation. *Neuroscience, 124*(3), 535-547.

Burnett, L. R., Stein, B. E., Perrault, T. J., Jr., & Wallace, M. T. (2007). Excitotoxic lesions of the superior colliculus preferentially impact multisensory neurons and multisensory integration. *Exp Brain Res, 179*(2), 325-338.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci, 21*(1), 300-304.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience, 6*(2), 190-195.

Buzsaki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents - EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience, 13*(6), 407-420.

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

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14*(2), 427-438.

Calvert, Gemma, Spence, Charles, & Stein, Barry E. (2004). *The handbook of multisensory processes*. Cambridge, Mass.: MIT Press.

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., . . . Knight,

R. T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science, 313*(5793), 1626-1628.

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends Cogn Sci, 14*(11), 506-515.

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience, 22*(11), 2886-2902.

Cappe, C., Morel, A., Barone, P., & Rouiller, E. M. (2009). The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cereb Cortex, 19*(9), 2025-2037.

Cappe, C., Rouiller, E. M., & Barone, P. (2009). Multisensory anatomical pathways. *Hear Res, 258*(1-2), 28-36.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J Neurosci, 30*(38), 12572-12580.

Cappe, C., Thut, G., Romei, V., & Murraya, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia, 47*(4), 1045-1052.

Carriere, B. N., Royal, D. W., & Wallace, M. T. (2008). Spatial heterogeneity of cortical receptive fields and its impact on multisensory interactions. *Journal of Neurophysiology, 99*(5), 2357-2368.

Casagrande, V. A., Harting, J. K., Hall, W. C., Diamond, I. T., & Martin, G. F. (1972). Superior colliculus of the tree shrew: a structural and functional subdivision into superficial and deep layers. *Science, 177*(4047), 444-447.

Cecere, R., Gross, J., & Thut, G. (2016). Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *Eur J Neurosci, 43*(12), 1561-1568.

Cecere, R., Gross, J., Willis, A., & Thut, G. (2017). Being First Matters: Topographical Representational Similarity Analysis of ERP Signals Reveals Separate Networks for Audiovisual Temporal Binding Depending on the Leading Sense. *J Neurosci, 37*(21), 5274-5287.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive

crossmodal illusory perception. *Curr Biol, 25*(2), 231-235.

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

Chen, L. H., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention Perception & Psychophysics, 75*(5), 790-811.

Chen, Y. C., & Spence, C. (2017). Assessing the Role of the 'Unity Assumption' on Multisensory Integration: A Review. *Frontiers in Psychology, 8*.

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nat Rev Neurosci, 3*(7), 553-562.

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *J Acoust Soc Am, 119*(6), 4065-4073.

Coull, J. T., Cheng, R. K., & Meck, W. H. (2011). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology, 36*(1), 3-25.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2011). Endogenous modulation of low frequency oscillations by temporal expectations. *J Neurophysiol, 106*(6), 2964-2972.

Dahl, C. D., Logothetis, N. K., & Kayser, C. (2009). Spatial organization of multisensory responses in temporal association cortex. *J Neurosci, 29*(38), 11924-11932.

Damiano C.R., Woynaroski T.G., Simon D.M., Ibañez L.V., Murias M., Kirby A., . . . C.J., Cascio. (2017). Developmental Sequelae and Neurophysiologic Substrates of Sensory Seeking in Infant Siblings of Children with Autism Spectrum Disorder. *Developmental Cognitive Neuroscience*.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*(1-2), 132-147.

de Boer-Schellekens, L., Eussen, M., & Vroomen, J. (2013). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Front Integr Neurosci, 7*, 8.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept Psychophys, 66*(8), 1388-1404.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience, 2*(10), 704-716.

Engel, A. K., Gerloff, C., Hilgetag, C. C., & Nolte, G. (2013). Intrinsic Coupling Modes: Multiscale Interactions in Ongoing Brain Activity. *Neuron, 80*(4), 867-886.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci, 5*(1), 16-25.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci, 22*(13), 5749-5759.

Falchier, A., Schroeder, C. E., Hackett, T. A., Lakatos, P., Nascimento-Silva, S., Ulbert, I., . . . Smiley, J. F. (2010). Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cereb Cortex, 20*(7), 1529-1538.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63*(4), 719-725.

Fister, J. K., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P., & Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing. *Neuropsychologia, 88*, 92-100.

Foss-Feig, J. H., Heacock, J. L., & Cascio, C. J. (2012). Tactile responsiveness patterns and their association with core features in autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*(1), 337-344.

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Exp Brain Res, 203*(2), 381-389.

Foucher, J. R., Lacambre, M., Pham, B. T., Giersch, A., & Elliott, M. A. (2007). Low time resolution in schizophrenia Lengthened windows of simultaneity for visual, auditory and bimodal stimuli. *Schizophrenia Research, 97*(1-3), 118-127.

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H. P., Russo, N. N., Blanco, D., . . . Ross, L. A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with Autism Spectrum Disorder (ASD) and their resolution during early adolescence. *Cereb Cortex, 25*(2), 298-312.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res Cogn Brain Res, 10*(1-2), 77-83.

Francisco, A. A., Jesse, A., Groen, M. A., & McQueen, J. M. (2017). A General Audiovisual Temporal Processing Deficit in Adult Readers With Dyslexia. *J Speech Lang Hear Res, 60*(1), 144-158.

Freeman, E. D., Ipser, A., Palmbaha, A., Paunoiu, D., Brown, P., Lambert, C., . . . Driver, J. (2013). Sight and sound out of synch: fragmentation and renormalisation of audiovisual integration and subjective timing. *Cortex, 49*(10), 2875-2887.

Frens, M. A., Vanopstal, A. J., & Vanderwilligen, R. F. (1995). Spatial and Temporal Factors Determine Auditory-Visual Interactions in Human Saccadic Eye-Movements. *Perception & Psychophysics, 57*(6), 802-816.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9*(10), 474-480.

Friston, K. J., & Price, C. J. (2003). Degeneracy and redundancy in cognitive anatomy. *Trends Cogn Sci, 7*(4), 151-152.

Fu, K. M., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., . . . Schroeder, C. E. (2003). Auditory cortical neurons respond to somatosensory stimulation. *J Neurosci, 23*(20), 7510-7515.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci, 7*(7), 773-778.

Gabard-Durnam, L., Tierney, A. L., Vogel-Farley, V., Tager-Flusberg, H., & Nelson, C. A. (2015). Alpha asymmetry in infants at risk for autism spectrum disorders. *J Autism Dev Disord, 45*(2), 473-480.

Gentile, F., van Atteveldt, N., De Martino, F., & Goebel, R. (2017). Approaching the Ground Truth: Revealing the Functional Organization of Human Multisensory STC Using Ultra-High Field fMRI. *J Neurosci, 37*(42), 10104-10113.

Ghose, D., Maier, A., Nidiffer, A., & Wallace, M. T. (2014). Multisensory response modulation in the superficial layers of the superior colliculus. *J Neurosci, 34*(12), 4332-4344.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci, 11*(5), 473-490.

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

Gondan, M., & Roder, B. (2006). A new method for detecting interactions between the senses in event-related potentials. *Brain Research, 1073*, 389-397.

Grabot, L., Kosem, A., Azizi, L., & van Wassenhove, V. (2017). Prestimulus Alpha Oscillations and the Temporal Sequencing of Audiovisual Events. *J Cogn Neurosci, 29*(9), 1566-1582.

Grabot, L., & van Wassenhove, V. (2017). Time Order as Psychological Bias. *Psychological Science, 28*(5), 670-678.

Gu, B. M., van Rijn, H., & Meck, W. H. (2015). Oscillatory multiplexing of neural population codes for interval timing and working memory. *Neurosci Biobehav Rev, 48*, 160-185.

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J Comp Neurol, 502*(6), 924-952.

Hackett, T. A., Smiley, J. F., Ulbert, I., Karmos, G., Lakatos, P., de la Mothe, L. A., & Schroeder, C. E. (2007). Sources of somatosensory input to the caudal belt areas of auditory cortex. *Perception, 36*(10), 1419-1430.

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res, 166*(3-4), 474-480.

Hall, A. J., & Lomber, S. G. (2008). Auditory cortex projections target the peripheral field representation of primary visual cortex. *Exp Brain Res, 190*(4), 413-430.

Harrington, D. L., Castillo, G. N., Fong, C. H., & Reed, J. D. (2011). Neural underpinnings of distortions in the experience of time across senses. *Front Integr Neurosci, 5*, 32.

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp Brain Res, 122*(2), 247-252.

Hass, K., Sinke, C., Reese, T., Roy, M., Wiswede, D., Dillo, W., . . . Szycik, G. R. (2017). Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion. *Cogn Neuropsychiatry, 22*(2), 145-158.

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience, 9*(6), 467-479.

Helfrich, R. F., Knepper, H., Nolte, G., Sengelmann, M., Konig, P., Schneider, T. R., & Engel, A. K. (2016). Spectral fingerprints of large-scale cortical dynamics during ambiguous

motion perception. *Hum Brain Mapp*.

Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc Natl Acad Sci U S A, 111*(41), 14935-14940.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci U S A, 109*(49), 20095-20100.

Henschke, J. U., Noesselt, T., Scheich, H., & Budinger, E. (2015). Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. *Brain Struct Funct, 220*(2), 955-977.

Hikosaka, K., Iwai, E., Saito, H., & Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *J Neurophysiol, 60*(5), 1615-1637.

Hillebrand, A., Tewarie, P., van Dellen, E., Yu, M. C., Carbo, E. W. S., Douw, L., . . . Stam, C. J. (2016). Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proceedings of the National Academy of Sciences of the United States of America, 113*(14), 3867-3872.

Hillock-Dunn, A., Grantham, D. W., & Wallace, M. T. (2016). The temporal binding window for audiovisual speech: Children are like little adults. *Neuropsychologia, 88*, 74-82.

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Dev Sci, 15*(5), 688-696.

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia, 49*(3), 461-467.

Hipp, J. F., Engel, A. K., & Siegel, M. (2011). Oscillatory Synchronization in Large-Scale Cortical Networks Predicts Perception. *Neuron, 69*(2), 387-396.

Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., & Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat Neurosci, 15*(6), 884-890.

Hirsh, I. J., & Sherrick, C. E., Jr. (1961). Perceived order in different sense modalities. *J Exp Psychol, 62*, 423-432.

Homae, F., Hashimoto, R., Nakajima, K., Miyashita, Y., & Sakai, K. L. (2002). From perception to sentence comprehension: the convergence of auditory and visual information of language in the left inferior frontal cortex. *Neuroimage, 16*(4), 883-900.

Hubel, D. H., & Wiesel, T. N. (1998). Early exploration of the visual cortex. *Neuron, 20*(3), 401-

412.

Huerta, M.F., Harting J.K. (1984). The mammalian superior colliculus studies of its morphology and connections. In H. Vanegas (Ed.), *Comparative neurology of the optic tectum* (pp. 687–773). New York: Plenum.

Huhn, Z., Szirtes, G., Lorincz, A., & Csepe, V. (2009). Perception based method for the investigation of audiovisual integration of speech. *Neurosci Lett, 465*(3), 204-209.

Hyvarinen, J., & Shelepin, Y. (1979). Distribution of visual and somatic functions in the parietal associative area 7 of the monkey. *Brain Res, 169*(3), 561-564.

Ipser, A., Agolli, V., Bajraktari, A., Al-Alawi, F., Djaafara, N., & Freeman, E. D. (2017). Sight and sound persistently out of synch: stable individual differences in audiovisual synchronisation revealed by implicit measures of lip-voice integration. *Sci Rep, 7*, 46413.

Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends Neurosci, 37*(7), 357-369.

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., & Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of Neurophysiology, 85*(2), 506-522.

Johnston, A., & Nishida, S. (2001). Time perception: brain time or event time? *Curr Biol, 11*(11), R427-430.

Jones, E.G., Powell, T.P.S. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain, 93*(4), 793-820.

Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: the relationship between space and time. *Exp Brain Res, 174*(3), 588-594.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol Sci, 13*(4), 313-319.

Kadunce, D. C., Vaughan, J. W., Wallace, M. T., Benedek, G., & Stein, B. E. (1997). Mechanisms of within- and cross-modality suppression in the superior colliculus. *Journal of Neurophysiology, 78*(6), 2834-2847.

Kadunce, D. C., Vaughan, J. W., Wallace, M. T., & Stein, B. E. (2001). The influence of visual and auditory receptive field organization on multisensory integration in the superior colliculus. *Exp Brain Res, 139*(3), 303-310.

Kaganovich, N. (2016). Development of sensitivity to audiovisual temporal asynchrony during midchildhood. *Dev Psychol, 52*(2), 232-241.

Kaganovich, N., & Schumaker, J. (2014). Audiovisual integration for speech during mid-childhood: Electrophysiological evidence. *Brain Lang, 139C*, 36-48.

Kaganovich, N., & Schumaker, J. (2016). Electrophysiological correlates of individual differences in perception of audiovisual temporal asynchrony. *Neuropsychologia, 86*, 119-130.

Kambe, J., Kakimoto, Y., & Araki, O. (2015). Phase reset affects auditory-visual simultaneity judgment. *Cogn Neurodyn, 9*(5), 487-493.

Kanayama, N., Sato, A., & Ohira, H. (2007). Crossmodal effect with rubber hand illusion and gamma-band activity. *Psychophysiology, 44*(3), 392-402.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron, 53*(3), 427-438.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cereb Cortex, 18*(7), 1560-1574.

Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Exp Brain Res, 167*(4), 635-640.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *J Neurosci, 33*(50), 19434-19441.

Kim, R., Peters, M. A. K., & Shams, L. (2012). 0+1 > 1: How Adding Noninformative Sound Improves Performance on a Visual Task. *Psychological Science, 23*(1), 6-12.

Kisley, M. A., & Cornwell, Z. M. (2006). Gamma and beta neural activity evoked during a sensory gating paradigm: Effects of auditory, somatosensory and cross-modal stimulation. *Clinical Neurophysiology, 117*(11), 2549-2563.

Klucharev, V., Mottonen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Res Cogn Brain Res, 18*(1), 65-75.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci, 27*(12), 712-719.

Knowland, V. C., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. (2014). Audio-visual speech perception: a developmental ERP investigation. *Dev Sci, 17*(1), 110-124.

Komura, Y., Tamura, R., Uwano, T., Nishijo, H., & Ono, T. (2005). Auditory thalamus integrates visual inputs into behavioral gains. *Nature Neuroscience, 8*(9), 1203-1209.

Kononowicz, T. W., & van Wassenhove, V. (2016). In Search of Oscillatory Traces of the Internal Clock. *Front Psychol, 7*, 224.

Kosem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage, 92*, 274-284.

Kuling, I. A., Kohlrausch, A., & Juola, J. F. (2013). Quantifying temporal ventriloquism in audiovisual synchrony perception. *Attention Perception & Psychophysics, 75*(7), 1583-1599.

Kuling, I. A., van Eijk, R. L. J., Juola, J. F., & Kohlrausch, A. (2012). Effects of stimulus duration on audio-visual synchrony perception. *Experimental Brain Research, 221*(4), 403-412.

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Front Integr Neurosci, 4*, 129.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum Brain Mapp, 8*(4), 194-208.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

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., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron, 64*(3), 419-430.

Lewis, P. A., & Meck, W. H. (2012). Time and the sleeping brain. *Psychologist, 25*(8), 594-597.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157.

Love, S. A., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A Psychophysical Investigation of

Differences between Synchrony and Temporal Order Judgments. *Plos One, 8*(1).

Macaluso, E., & Driver, J. (2001). Spatial attention and crossmodal interactions between vision and touch. *Neuropsychologia, 39*(12), 1304-1316.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage, 21*(2), 725-732.

Magnee, M. J. C. M., de Gelder, B., van Engeland, H., & Kemner, C. (2008). Audiovisual speech integration in pervasive developmental disorder: evidence from event-related potentials. *Journal of Child Psychology and Psychiatry, 49*(9), 995-1000.

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology, 18*(13), 963-968.

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus monkeys. *Neuron, 43*(2), 177-181.

Marchant, J. L., & Driver, J. (2013). Visual and audiovisual effects of isochronous timing on visual perception and brain activity. *Cereb Cortex, 23*(6), 1290-1298.

Marchant, J. L., Ruff, C. C., & Driver, J. (2012). Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities. *Hum Brain Mapp, 33*(5), 1212-1224.

Martin, B., Giersch, A., Huron, C., & van Wassenhove, V. (2013). Temporal event structure and timing in schizophrenia: preserved binding in a longer "now". *Neuropsychologia, 51*(2), 358-371.

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *J Neurosci, 29*(9), 2725-2732.

May, P. J. (2006). The mammalian superior colliculus: laminar structure and connections. *Prog Brain Res, 151*, 321-378.

McDonald, J. J., Teder-Salejarvi, W. A., Di Russo, F., & Hillyard, S. A. (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nat Neurosci, 8*(9), 1197-1202.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and

estimation of time. *Annu Rev Neurosci, 36*, 313-336.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci, 7*(10), 3215-3229.

Meredith, M. A., & Stein, B. E. (1983). Interactions among Converging Sensory Inputs in the Superior Colliculus. *Science, 221*(4608), 389-391.

Meredith, M. A., & Stein, B. E. (1985). Descending Efferents from the Superior Colliculus Relay Integrated Multisensory Information. *Science, 227*(4687), 657-659.

Meredith, M. A., & Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res, 365*(2), 350-354.

Meredith, M. A., & Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol, 56*(3), 640-662.

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol, 75*(5), 1843-1857.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol, 14*(2), 247-279.

Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience, 25*(25), 5884-5893.

Miniussi, C., Girelli, M., & Marzi, C. A. (1998). Neural site of the redundant target effect: Electrophysiological evidence. *Journal of Cognitive Neuroscience, 10*(2), 216-230.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J Neurosci, 27*(15), 4120-4131.

Miyazaki, M., Kadota, H., Matsuzaki, K. S., Takeuchi, S., Sekiguchi, H., Aoyama, T., & Kochiyama, T. (2016). Dissociating the neural correlates of tactile temporal order and simultaneity judgements. *Sci Rep, 6*, 23323.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res, 14*(1), 115-128.

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal*

of Autism and Developmental Disorders, 38(7), 1349-1358.

Moran, R. J., Molholm, S., Reilly, R. B., & Foxe, J. J. (2008). Changes in effective connectivity of human superior parietal lobule under multisensory and unisensory stimulation. *Eur J Neurosci, 27*(9), 2303-2312.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research, 17*(1), 154-163.

Mottonen, R., Schurmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neurosci Lett, 363*(2), 112-115.

Mullette-Gillman, O. A., Cohen, Y. E., & Groh, J. M. (2005). Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *J Neurophysiol, 94*(4), 2331-2352.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys, 58*(3), 351-362.

Munoz, D. P., & Guitton, D. (1985). Tectospinal Neurons in the Cat Have Discharges Coding Gaze Position Error. *Brain Research, 341*(1), 184-188.

Murray, M. M., Foxe, J. J., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2001). Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: a high-density electrical mapping study. *Neuropsychologia, 39*(8), 828-844.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., . . . Foxe, J. J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex, 15*(7), 963-974.

Murray, M. M., & Wallace, M.T. (2012). *The Neural Bases of Multisensory Processes* (M. M. Murray, Wallace, M.T. Ed.). Boca Raton, FL: CRC Press.

Musacchia, G., & Schroeder, C. E. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hear Res, 258*(1-2), 72-79.

Nagy, A., Eordegh, G., Paroczy, Z., Markus, Z., & Benedek, G. (2006). Multisensory integration in the basal ganglia. *Eur J Neurosci, 24*(3), 917-924.

Nath, A. R., & Beauchamp, M. S. (2011). Dynamic Changes in Superior Temporal Sulcus

Connectivity during Perception of Noisy Audiovisual Speech. *Journal of Neuroscience, 31*(5), 1704-1714.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage, 59*(1), 781-787.

Nath, A. R., Fava, E. E., & Beauchamp, M. S. (2011). Neural Correlates of Interindividual Differences in Children's Audiovisual Speech Perception. *Journal of Neuroscience, 31*(39), 13963-13971.

Naue, N., Rach, S., Struber, D., Huster, R. J., Zaehle, T., Korner, U., & Herrmann, C. S. (2011). Auditory event-related response in visual cortex modulates subsequent visual responses in humans. *J Neurosci, 31*(21), 7729-7736.

Nidiffer, A. R., Stevenson, R. A., Fister, J. K., Barnett, Z. P., & Wallace, M. T. (2016). Interactions between space and effectiveness in human multisensory performance. *Neuropsychologia, 88*, 83-91.

Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2016). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Res*.

Noel, J. P., De Niear, M., Van der Burg, E., & Wallace, M. T. (2016). Audiovisual Simultaneity Judgment and Rapid Recalibration throughout the Lifespan. *PLoS One, 11*(8), e0161698.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci, 27*(42), 11431-11441.

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H. J., & Driver, J. (2010). Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *J Neurosci, 30*(41), 13609-13623.

Noppeney, U. (2012). Characterization of Multisensory Integration with fMRI: Experimental Design, Statistical Analysis, and Interpretation. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton (FL).

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *J Neurosci, 30*(21), 7434-7446.

Nunez, Paul L., & Srinivasan, Ramesh. (2006). *Electric fields of the brain : the neurophysics of EEG* (2nd ed.). Oxford ; New York: Oxford University Press.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat Neurosci, 15*(12), 1729-1735.

Olcese, U., Iurilli, G., & Medini, P. (2013). Cellular and synaptic architecture of multisensory integration in the mouse neocortex. *Neuron, 79*(3), 579-593.

Panzeri, S., Macke, J. H., Gross, J., & Kayser, C. (2015). Neural population coding: combining insights from microscopic and mass signals. *Trends Cogn Sci, 19*(3), 162-172.

Pasalar, S., Ro, T., & Beauchamp, M. S. (2010). TMS of posterior parietal cortex disrupts visual tactile multisensory integration. *Eur J Neurosci, 31*(10), 1783-1790.

Peiker, I., David, N., Schneider, T. R., Nolte, G., Schottle, D., & Engel, A. K. (2015). Perceptual Integration Deficits in Autism Spectrum Disorders Are Associated with Reduced Interhemispheric Gamma-Band Coherence. *J Neurosci, 35*(50), 16352-16361.

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *J Speech Lang Hear Res, 52*(4), 1073-1081.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jaaskelainen, I. P., Letham, B., . . . Belliveau, J. W. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur J Neurosci, 31*(10), 1772-1782.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput, 20*(4), 873-922.

Ray, S., & Maunsell, J. H. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol, 9*(4), e1000610.

Reig, R., & Silberberg, G. (2014). Multisensory integration in the mouse striatum. *Neuron, 83*(5), 1200-1212.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech-Perception without Traditional Speech Cues. *Science, 212*(4497), 947-950.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology-Human Perception and Performance, 28*(5), 1085-1099.

Righi, G., Tierney, A. L., Tager-Flusberg, H., & Nelson, C. A. (2014). Functional connectivity in the first year of life in infants at risk for autism spectrum disorder: an EEG study. *PLoS*

One, 9(8), e105176.

Roa Romero, Y., Keil, J., Balz, J., Gallinat, J., & Senkowski, D. (2016). Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. *J Neurophysiol, 116*(3), 1396-1407.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *Int J Psychophysiol, 50*(1-2), 19-26.

Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature, 397*(6718), 430-433.

Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol, 13*(2), e1002073.

Rohe, T., & Noppeney, U. (2016). Distinct Computational Principles Govern Multisensory Integration in Primary Sensory and Association Cortices. *Curr Biol, 26*(4), 509-514.

Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cereb Cortex, 17 Suppl 1*, i61-69.

Romanski, L. M. (2012). Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. *Proc Natl Acad Sci U S A, 109 Suppl 1*, 10717-10724.

Romero, Y. R., Keil, J., Balz, J., Niedeggen, M., Gallinat, J., & Senkowski, D. (2016). Alpha-Band Oscillations Reflect Altered Multisensory Processing of the McGurk Illusion in Schizophrenia. *Frontiers in Human Neuroscience, 10*.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. *Cerebral Cortex, 17*(5), 1147-1153.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C., & Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr Res, 97*(1-3), 173-183.

Royal, D. W., Carriere, B. N., & Wallace, M. T. (2009). Spatiotemporal architecture of cortical receptive fields and its impact on multisensory interactions. *Experimental Brain Research, 198*(2-3), 127-136.

Russo, N., Foxe, J. J., Brandwein, A. B., Altschuler, T., Gomes, H., & Molholm, S. (2010). Multisensory processing in children with autism: high-density electrical mapping of auditory-somatosensory integration. *Autism Res, 3*(5), 253-267.

Sakowitz, O. W., Quian Quiroga, R., Schurmann, M., & Basar, E. (2005). Spatio-temporal frequency characteristics of intersensory components in audiovisually evoked potentials. *Brain Res Cogn Brain Res, 23*(2-3), 316-326.

Sakowitz, O. W., Quiroga, R. Q., Schurmann, M., & Basar, E. (2001). Bisensory stimulation increases gamma-responses over multiple cortical regions. *Cognitive Brain Research, 11*(2), 267-279.

Sakowitz, O. W., Schurmann, M., & Basar, E. (2000). Oscillatory frontal theta responses are increased upon bisensory stimulation. *Clinical Neurophysiology, 111*(5), 884-893.

Samaha, J., & Postle, B. R. (2015). The Speed of Alpha-Band Oscillations Predicts the Temporal Resolution of Visual Perception. *Curr Biol, 25*(22), 2985-2990.

Schepers, I. M., Schneider, T. R., Hipp, J. F., Engel, A. K., & Senkowski, D. (2013). Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. *Neuroimage, 70*, 101-112.

Schormans, A. L., Scott, K. E., Vo, A. M., Tyker, A., Typlt, M., Stolzberg, D., & Allman, B. L. (2016). Audiovisual Temporal Processing and Synchrony Perception in the Rat. *Front Behav Neurosci, 10*, 246.

Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res, 14*(1), 187-198.

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *J Neurophysiol, 85*(3), 1322-1327.

Schwartz, J. L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Comput Biol, 10*(7), e1003743.

Sejnowski, T. J., Churchland, P. S., & Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature Neuroscience, 17*(11), 1440-1441.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*(6614), 308.

Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J Comp Neurol, 343*(3), 445-463.

Senkowski, D., Molholm, S., Gomez-Ramirez, M., & Foxe, J. J. (2006). Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: A high-density electrical mapping study. *Cerebral Cortex, 16*(11), 1556-1565.

Senkowski, D., Saint-Amour, D., Hofle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage, 56*(4), 2200-2208.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci, 31*(8), 401-409.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia, 45*(3), 561-571.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature, 408*(6814), 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*(1), 147-152.

Siegel, M., Buschman, T. J., & Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science, 348*(6241), 1352-1355.

Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience, 13*(2), 121-134.

Simon, D. M., Damiano, C. R., Woynaroski, T. G., Ibanez, L. V., Murias, M., Stone, W. L., . . . Cascio, C. J. (2017). Neural Correlates of Sensory Hyporesponsiveness in Toddlers at High Risk for Autism Spectrum Disorder. *J Autism Dev Disord*.

Simon, D. M., Noel, J. P., & Wallace, M. T. (2017). Event Related Potentials Index Rapid Recalibration to Audiovisual Temporal Asynchrony. *Front Integr Neurosci, 11*, 8.

Simon, D. M., & Wallace, M. T. (2016). Dysfunction of sensory oscillations in Autism Spectrum Disorder. *Neurosci Biobehav Rev, 68*, 848-861.

Simon, D.M., Wallace, M. T. (2017). Rhythmic Modulation of Entrained Auditory Oscillations

by Visual Inputs. *Brain Topogr*.

Singer, W., & Gray, C. M. (1995). Visual Feature Integration and the Temporal Correlation Hypothesis. *Annual Review of Neuroscience, 18*, 555-586.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12*(1), 7-10.

Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol, 502*(6), 894-923.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry, 48*(8), 813-821.

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

Sparks, D. L., & Hartwich-Young, R. (1989). The deep layers of the superior colliculus. *Rev Oculomot Res, 3*, 213-255.

Spence, C., & Mcdonald, J. (2004). The Cross-Modal Consequences of the Exogenous Spatial Orienting of Attention. In C. G.A., C. Spence & B. E. Stein (Eds.), *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Spence, C., & Parise, C. (2010). Prior-entry: a review. *Conscious Cogn, 19*(1), 364-379.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *J Exp Psychol Gen, 130*(4), 799-832.

Sperdin, H. F., Cappe, C., Foxe, J. J., & Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front Integr Neurosci, 3*, 2.

Sprague, J. M. (1996). Neural mechanisms of visual orienting responses. *Prog Brain Res, 112*, 1-15.

Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: putting the computation in context. *Neuroreport, 18*(8), 787-792.

Staufenbiel, S. M., van der Lubbe, R. H. J., & Talsma, D. (2011). Spatially uninformative sounds increase sensitivity for visual motion change. *Experimental Brain Research, 213*(4), 457-464.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the

perspective of the single neuron. *Nat Rev Neurosci, 9*(4), 255-266.

Stein, B. E., & Wallace, M. T. (1996). Comparisons of cross-modality integration in midbrain and cortex. *Prog Brain Res, 112*, 289-299.

Stein, Barry E. (2012). *The new handbook of multisensory processing*. Cambridge, Mass.: MIT Press.

Stein, Barry E., & Meredith, M. Alex. (1993). *The merging of the senses*. Cambridge, Mass.: MIT Press.

Stekelenburg, J. J., Maes, J. P., Van Gool, A. R., Sitskoorn, M., & Vroomen, J. (2013). Deficient multisensory integration in schizophrenia: an event-related potential study. *Schizophr Res, 147*(2-3), 253-261.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci, 19*(12), 1964-1973.

Sternberg, S., Knoll, R.L. (1973). *The perception of temporal order: Fundamental issues and a general model* (K. S. Ed.). New York: Academic Press.

Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., & James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *Neuroimage, 49*(4), 3308-3318.

Stevenson, R. A., Bushmakin, M., Kim, S., Wallace, M. T., Puce, A., & James, T. W. (2012). Inverse Effectiveness and Multisensory Interactions in Visual Event-Related Potentials with Audiovisual Speech. *Brain Topography, 25*(3), 308-326.

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research, 219*(1), 121-137.

Stevenson, R. A., Geoghegan, M. L., & James, T. W. (2007). Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research, 179*(1), 85-95.

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., . . . Wallace, M. T. (2014). Identifying and Quantifying Multisensory Integration: A Tutorial Review. *Brain Topography, 27*(6), 707-730.

Stevenson, R. A., Kim, S., & James, T. W. (2009). An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio,

visual, and haptic sensory streams using fMRI. *Experimental Brain Research, 198*(2-3), 183-194.

Stevenson, R. A., Park, S., Cochran, C., McIntosh, L. G., Noel, J. P., Barense, M. D., . . . Wallace, M. T. (2017). The associations between multisensory temporal processing and symptoms of schizophrenia. *Schizophrenia Research, 179*, 97-103.

Stevenson, R. A., Segers, M., Ncube, B. L., Black, K. R., Bebko, J. M., Ferber, S., & Barense, M. D. (2017). The cascading influence of multisensory processing on speech perception in autism. *Autism*, 1362361317704413.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *J Neurosci, 34*(3), 691-697.

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014a). Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. *J Autism Dev Disord, 44*(6), 1470-1477.

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014b). Evidence for diminished multisensory integration in autism spectrum disorders. *J Autism Dev Disord, 44*(12), 3161-3167.

Stevenson, R. A., VanDerKlok, R. M., Pisoni, D. B., & James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *Neuroimage, 55*(3), 1339-1345.

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Exp Brain Res, 227*(2), 249-261.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *J Exp Psychol Hum Percept Perform, 38*(6), 1517-1529.

Stitt, I., Galindo-Leon, E., Pieper, F., Engler, G., Fiedler, E., Stieglitz, T., & Engel, A. K. (2015). Intrinsic coupling modes reveal the functional architecture of cortico-tectal networks. *Sci Adv, 1*(7), e1500229.

Stone, D. B., Coffman, B. A., Bustillo, J. R., Aine, C. J., & Stephen, J. M. (2014). Multisensory stimuli elicit altered oscillatory brain responses at gamma frequencies in patients with schizophrenia. *Frontiers in Human Neuroscience, 8*.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., . . . Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society B-Biological Sciences, 268*(1462), 31-38.

Strauss, A., Henry, M. J., Scharinger, M., & Obleser, J. (2015). Alpha phase determines successful lexical decision in noise. *J Neurosci, 35*(7), 3256-3262.

Stroganova, T. A., Orekhova, E. V., Prokofyev, A. O., Tsetlin, M. M., Gratchev, V. V., Morozov, A. A., & Obukhov, Y. V. (2012). High-frequency oscillatory response to illusory contour in typically developing boys and boys with autism spectrum disorders. *Cortex, 48*(6), 701-717.

Sugihara, T., Diltz, M. D., Averbeck, B. B., & Romanski, L. M. (2006). Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci, 26*(43), 11138-11147.

Sumby, W.H., Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *J. Acoust. Soc. Am., 26*(2), 212-215.

Sun, L., Grutzner, C., Bolte, S., Wibral, M., Tozman, T., Schlitt, S., . . . Uhlhaas, P. J. (2012). Impaired gamma-band activity during perceptual organization in adults with autism spectrum disorders: evidence for dysfunctional network activity in frontal-posterior cortices. *J Neurosci, 32*(28), 9563-9573.

Teder-Salejarvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14*(1), 106-114.

Ten Oever, S., & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proc Natl Acad Sci U S A, 112*(52), 15833-15837.

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *J Neurosci, 31*(10), 3853-3861.

Titchener, Edward Bradford. (1908). *Lectures on the elementary psychology of feeling and attention*. New York: Macmillan.

Treille, A., Cordeboeuf, C., Vilain, C., & Sato, M. (2014). Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions. *Neuropsychologia, 57*, 71-77.

Tseng, H. H., Bossong, M. G., Modinos, G., Chen, K. M., McGuire, P., & Allen, P. (2015). A

systematic review of multisensory cognitive-affective integration in schizophrenia. *Neuroscience and Biobehavioral Reviews, 55*, 444-452.

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports, 6*.

Uhlhaas, P. J., & Singer, W. (2006). Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron, 52*(1), 155-168.

Uhlhaas, P. J., Singer W. (2007). What Do Disturbances in Neural Synchrony Tell Us About Autism. *Biological Psychiatry, 62*(3), 190-191.

Uhlhaas, P. J., & Singer, W. (2012). Neuronal Dynamics and Neuropsychiatric Disorders: Toward a Translational Paradigm for Dysfunctional Large-Scale Networks. *Neuron, 75*(6), 963-980.

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron, 43*(2), 271-282.

van Atteveldt, N. M., Peterson, B. S., & Schroeder, C. E. (2014). Contextual control of audiovisual integration in low-level sensory cortices. *Hum Brain Mapp, 35*(5), 2394-2411.

van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron, 81*(6), 1240-1253.

van den Brink, R. L., Cohen, M. X., van der Burg, E., Talsma, D., Vissers, M. E., & Slagter, H. A. (2014). Subcortical, modality-specific pathways contribute to multisensory processing in humans. *Cereb Cortex, 24*(8), 2169-2177.

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *J Neurosci, 33*(37), 14633-14637.

Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Sci Rep, 5*, 14526.

Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proc Biol Sci, 282*(1804), 20143083.

Van der Burg, E., Orchard-Mills, E., & Alais, D. (2014). Rapid temporal recalibration is unique to audiovisual stimuli. *Exp Brain Res*.

van Driel, J., Knapen, T., van Es, D. M., & Cohen, M. X. (2014). Interregional alpha-band synchrony supports temporal cross-modal integration. *Neuroimage, 101*, 404-415.

van Eijk, R. L., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. *Percept Psychophys, 70*(6), 955-968.

van Wassenhove, V. (2009). Minding time in an amodal representational space. *Philos Trans R Soc Lond B Biol Sci, 364*(1525), 1815-1830.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A, 102*(4), 1181-1186.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychon Bull Rev, 14*(6), 1011-1026.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: a DMAT primer. *Behav Res Methods, 40*(1), 61-72.

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *J Vis, 8*(9), 14 11-11.

Vatakis, A., & Spence, C. (2006a). Audiovisual synchrony perception for music, speech, and object actions. *Brain Res, 1111*(1), 134-142.

Vatakis, A., & Spence, C. (2006b). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neurosci Lett, 393*(1), 40-44.

Vibell, J., Klinge, C., Zampini, M., Spence, C., & Nobre, A. C. (2007). Temporal order is coded temporally in the brain: early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *J Cogn Neurosci, 19*(1), 109-120.

Vidal, J., Giard, M. H., Roux, S., Barthelemy, C., & Bruneau, N. (2008). Cross-modal processing of auditory-visual stimuli in a no-task paradigm: a topographic event-related potential study. *Clin Neurophysiol, 119*(4), 763-771.

Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., & Pennartz, C. M. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage, 55*(4), 1548-1565.

von Stein, A., Rappelsberger, P., Sarnthein, J., & Petsche, H. (1999). Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cerebral Cortex, 9*(2), 137-150.

Vroomen, J., & De Gelder, B. (2004). Perceptual effects of cross-modal stimulation: ventriloquism and the freezing phenomenon. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 141-150): MIT Press.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys, 72*(4), 871-884.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res, 22*(1), 32-35.

Vroomen, J., & Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: not that special. *Cognition, 118*(1), 75-83.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1993). Converging Influences from Visual, Auditory, and Somatosensory Cortices onto Output Neurons of the Superior Colliculus. *Journal of Neurophysiology, 69*(6), 1797-1809.

Wallace, M. T., Ramachandran, R., & Stein, B. E. (2004). A revised view of sensory cortical parcellation. *Proc Natl Acad Sci U S A, 101*(7), 2167-2172.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Exp Brain Res, 158*(2), 252-258.

Wallace, M. T., & Stein, B. E. (1994). Cross-Modal Synthesis in the Midbrain Depends on Input from Cortex. *Journal of Neurophysiology, 71*(1), 429-432.

Wallace, M. T., & Stein, B. E. (1996). Sensory organization of the superior colliculus in cat and monkey. *Extrageniculostriate Mechanisms Underlying Visually-Guided Orientation Behavior, 112*, 301-311.

Wallace, M. T., & Stein, B. E. (2000). Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. *Journal of Neurophysiology, 83*(6), 3578-3582.

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64C*, 105-123.

Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y., & Sweeney, J. A. (2013). Resting state EEG abnormalities in autism spectrum disorders. *J Neurodev Disord, 5*(1),

24.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol Rev, 90*(3), 1195-1268.

Wang, X. J., & Buzsaki, G. (1996). Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *Journal of Neuroscience, 16*(20), 6402-6413.

Wang, Y., Celebrini, S., Trotter, Y., & Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci, 9*, 79.

Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Percept Psychophys, 39*(4), 294-300.

Wiener, M., Matell, M. S., & Coslett, H. B. (2011). Multiple mechanisms for temporal processing. *Front Integr Neurosci, 5*, 31.

Williams, J. H. G., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual-auditory integration during speech imitation in autism. *Research in Developmental Disabilities, 25*(6), 559-575.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science, 316*(5831), 1609-1612.

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *J Autism Dev Disord, 43*(12), 2891-2902.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex, 13*(10), 1034-1043.

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proc Natl Acad Sci U S A, 106*(49), 20664-20669.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Percept Psychophys, 67*(3), 531-544.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Exp Brain Res, 152*(2), 198-210.

Zampini, M., Shore, D. I., & Spence, C. (2005). Audiovisual prior entry. *Neurosci Lett, 381*(3), 217-222.

CHAPTER II

INTEGRATION AND TEMPORAL PROCESSING OF ASYNCHRONOUS AUDIOVISUAL SPEECH

The contents of this chapter are drawn from a published manuscript in the Journal of Cognitive Neuroscience:

Simon, D.M., Wallace M.T., Integration and Temporal Processing of Asynchronous Audiovisual Speech

Abstract

Multisensory Integration of visual mouth movements with auditory speech is known to offer substantial perceptual benefits, particularly under challenging (i.e., noisy) acoustic conditions. Previous work characterizing this process has found that event related potentials (ERPs) to auditory speech are of shorter latency and smaller magnitude in the presence of visual speech. We sought to determine the dependency of these effects on the temporal relationship between the auditory and visual speech streams using electroencephalography (EEG). We found that reductions in ERP latency and suppression of ERP amplitude are maximal when the visual signal precedes the auditory signal by a small interval, and that increasing amounts of asynchrony reduce these effects in a continuous manner. Time-Frequency analysis revealed that these effects

are found primarily in the theta (4-8 Hz) and alpha (8-12 Hz) bands, with a central topography consistent with auditory generators. Theta effects also persisted in the lower portion of the band (3.5-5 Hz), and this late activity was more frontally distributed. Importantly, the magnitude of these late theta oscillations not only differed with the temporal characteristics of the stimuli, but also served to predict participant's task performance. Our analysis thus reveals that suppression of single trial brain responses by visual speech depends strongly on the temporal concordance of the auditory and visual inputs. It further illustrates that processes in the lower theta band, which we suggest as an index of incongruity processing, might serve to reflect the neural correlates of individual differences in multisensory temporal perception.

Introduction

Audiovisual Integration of Speech Signals

We live in a complex environment in which events frequently generate signals in multiple sensory modalities. Multisensory integration, the process of combining these sensory inputs to form a single coherent percept, has been shown to offer numerous behavioral and perceptual advantages in a variety of tasks (Murray & Wallace, 2012). A particularly striking and ecologically important example of this process is the integration of visual speech (i.e. mouth movements) with auditory speech. In acoustically challenging environments, the integration of these signals has been shown to substantially facilitate speech comprehension (Cherry, 1953; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). The presence of visual inputs is also known to play an important role in assisting in the process of stream segregation, in which features of selectively attended auditory signals are grouped together while unattended signals are filtered

out (Shinn-Cunningham, 2008).

An important factor facilitating these integrative processes is that auditory and visual speech signals share an obligatory temporal correlation due to the nature of speech production. This correlation is highly intuitive and well quantified in natural speech (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Schwartz & Savariaux, 2014) – when the mouth is open the speech envelope (i.e. the sound amplitude) is large, and when the mouth is closed the speech envelope is small. This temporal correlation is primarily found at relatively low frequencies (1-5 Hz), which correspond with the temporal structure imposed by the basic units of speech, syllables and words (Chandrasekaran et al., 2009). Given this seemingly useful temporal structure, a number of studies have aimed to investigate the degree to which temporal concordance is important for multisensory speech processing (Munhall, Gribble, Sacco, & Ward, 1996; Ten Oever, Sack, Wheat, Bien, & van Atteveldt, 2013; van Wassenhove, Grant, & Poeppel, 2007). These studies have elucidated that integration, as measured through a number of psychophysical tasks, occurs with a degree of temporal tolerance around true simultaneity. To capture the temporal interval within which auditory and visual signals can be perceptually integrated and bound, the construct of a temporal binding window (TBW) has been put forth (Wallace & Stevenson, 2014). The TBW has been shown to be asymmetric for speech stimuli and larger for temporal asynchronies in which vision leads audition. This is both consistent with the statistics of the natural environment, in which sound travels more slowly than light, as well as the causal structure of speech signals in which articulatory movements of the vocal apparatus generally precede sounds (Chandrasekaran et al., 2009; Schwartz & Savariaux, 2014).

Neural Manifestations of Audiovisual Speech Integration

This recognized ecological importance of vision for bolstering the processing of speech signals has spurred investigations regarding how the presence of visual inputs modifies neural processing of the acoustic signal. Early investigations revealed that the presence of visual speech attenuates the magnitude of early brain responses to auditory speech (Besle, Fort, Delpuech, & Giard, 2004; van Wassenhove, Grant, & Poeppel, 2005) and may reduce the latency of individual processing stages (van Wassenhove et al., 2005). A meta-analytical approach has indicated that these changes in processing magnitude and latency are present over a wide range of experimental designs and task demands (Baart, 2016), reinforcing that reduction in neural response and reduced onset latency is a robust and generalized indicator of audiovisual speech integration. A notable feature of this integrative effect is the sub-additive nature of the interaction, in which the absolute magnitude of the multisensory neural response is substantially smaller than its constituent components (i.e. AV < A+V). This sub-additivity contrasts starkly with the super-additive neural responses often seen for non-speech stimuli (Cappe, Thut, Romei, & Murray, 2010). This suggests important mechanistic differences in the processing and integration of multisensory stimuli with informative anticipatory information, such as speech (Stekelenburg & Vroomen, 2007).

The presence of visual speech has also been shown to contribute to the ability of the brain to entrain to the auditory speech envelope. Speech is constructed of elements (syllables and words) that are produced in a semi-rhythmic stream. The brain has been shown to utilize this information by locking the phase of neural activity to the phase of the speech rhythm. This process, known as entrainment, is believed to form the backbone of temporal attention (Jones, Moynihan, MacKenzie, & Puente, 2002), processing ‘to the beat’ (Breska & Deouell, 2016), and for speech serves to selectively amplify future speech signals occurring at the correct phase

(Giraud & Poeppel, 2012). The presence of visual speech has also been shown to facilitate this entrainment process when multiple speakers are present (Zion Golumbic, Cogan, Schroeder, & Poeppel, 2013; E. M. Zion Golumbic et al., 2013), as well as to directly entrain neural oscillations (Park, Kayser, Thut, & Gross, 2016), indicating that the visual rhythm can be used to disambiguate which portions of the acoustic envelope should be entrained to (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008). Importantly, the phase of these rhythms has been shown to causally affect speech perception under challenging conditions. When the auditory signal is perceptually ambiguous, the final speech percept depends on the phase of these ongoing spontaneous oscillations (Ten Oever & Sack, 2015). In a further entrainment experiment, it was then established that acoustic entrainment generates rhythmic fluctuations in perceptual outcomes at the entrained frequency (Ten Oever & Sack, 2015). These experiments establish low frequency oscillations as more than simple neural resonance or a byproduct of meaningful acoustic processing, and further indicate that visual influences on these oscillations likely have perceptual consequences.

For audiovisual integration these cortical entrainment processes have been shown to occur over an extended window of time (>400ms) (Crosse, Di Liberto, & Lalor, 2016). Entrainment also exhibits the greatest amount of audiovisual integration at the low frequencies capable of capturing the slow cycle times present in naturalistic visual speech (Crosse, Butler, & Lalor, 2015). Importantly, this neural integration time is substantially greater than the temporal integration time seen for auditory speech processing under challenging acoustic conditions (~200 ms) (Ding & Simon, 2013a). This extended audiovisual window strongly suggests that audiovisual speech integration functions as a temporally privileged operation, which may be critical given that audiovisual speech has a variable temporal structure for the constituent

auditory and visual components (Schwartz & Savariaux, 2014; Ten Oever et al., 2013).

Motivations for the Current Study

Despite the robustness of the finding that visual speech reduces the magnitude of neural responses to auditory speech, the degree to which this depends on the temporal structure of the auditory and visual signals has not been fully explored. Previous work has established that an ecologically implausible 200 ms auditory lead precludes this effect (Pilling, 2009), but did not systematically determine how changes in temporal structure influenced the effect. Given the strong characterizations of temporal tolerance in the behavioral domain (Vroomen & Keetels, 2010; Wallace & Stevenson, 2014), we hypothesized that a similar temporal window for response reduction would be present in the associated neural measures. We further hypothesized that these effects would be strongest at a small visual lead rather than true synchrony. This would be consistent with the natural statistics of audiovisual speech signals, predictive coding accounts of audiovisual integration (Talsma, 2015), and numerous behavioral accounts in which participants judge slight visual leads to be 'most synchronous', a point often referred to as the point of subjective simultaneity (PSS) (Vroomen & Keetels, 2010).

We thus sought to determine the degree to which temporal coincidence between auditory and visual inputs mediates the multisensory integration of speech signals measured via reduction in neural response amplitude. To do so, we recorded EEG from human participants while they performed a psychophysical simultaneity judgment (SJ) task featuring audiovisual speech stimuli. Our results indicate that reductions in response amplitude afforded by the presence of visual speech operate within an asymmetric temporal window with striking similarities to the TBW reported in behavioral studies. The width of this window also strongly aligns with the

cycle times of the frequencies with strongest temporal correlation in natural audiovisual speech. We further identify a novel theta band incongruity signal present in later stages of processing, which is of greatest amplitude in conditions in which temporal misalignment is present. Crucially, we link the strength of this signal to participant's ability to identify asynchronous stimuli correctly. Our results shed new light on the neural correlates of the temporal tolerance for audiovisual speech integration by elucidating the nature of temporal integration in multiple frequency bands and time windows corresponding with distinct stages of cortical processing.

Methods and Materials

Participants

Twenty-eight typically developing adults participated in the study. All participants reported that they were right handed, had normal or corrected-to-normal vision, and normal hearing. Two participants were excluded from analysis due to behavioral performance indicating they did not correctly perform the task and 1 participant did not complete the task, leaving a total of 25 analyzed participants (16 women) with a mean age of 22.08 years (± 4.21). The study was conducted in accordance with the declaration of Helsinki, and informed written consent was obtained from all participants. All procedures were approved by the Vanderbilt University institutional review board.

Psychophysical Task

Participants performed a speeded 2 alternative forced choice SJ task (**Fig 2-1**). The experimental stimuli consisted of an audiovisual movie of a woman saying the syllable 'BA', including all pre-articulatory movements, with a resolution of 720 x 1280 and a duration of 2000ms. We selected 'BA' as our stimulus because it is a highly visually specified syllable (i.e. it is easily lip read) which may generate stronger integration effects than less visually specified syllables (van Wassenhove et al., 2005). The movie was presented on a 24-inch monitor (ASUS VG248QE)

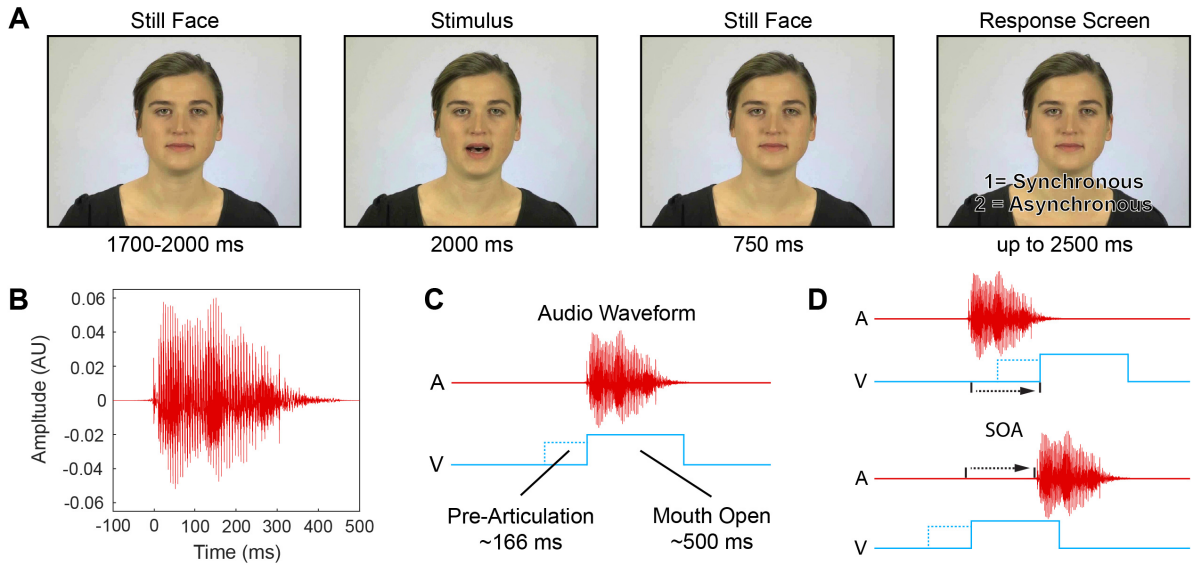

Figure 2-1 Speeded Simultaneity Judgment Paradigm

A) Experimental Timeline. Trials began with a 1700-2000 ms period of a still face consisting of the first video frame, followed by the 2s video. Following the movie there was a 750 ms period of additional still face consisting of the last movie frame. If participants had not yet responded, a response screen appeared for up to 2500ms. Participants were explicitly told to respond as quickly and accurately as possible and that the response screen was an indicator they were responding too slowly.
B) Auditory signal waveform for the 'BA' syllable used for the experiment
C) Temporal alignment between the auditory signal and visual mouth movements for the 'BA' syllable used for the experiment. Pre-articulatory movements began 166ms before sound onset.
D) Method for creation of stimulus onset asynchronies (SOAs). To create auditory leads the visual stimulus was pushed backwards in time by padding the movie with additional still frames. Similarly, visual leads were created by pushing the auditory stimulus backwards in time through zero padding of the sound waveform. Total movie stimulus duration was kept constant at 2 seconds by removing still frames from the end of the movie for auditory leads.

with a refresh rate of 60 Hz at a distance of 1 meter. The woman's face was central on the monitor and occupied an area approximately 12 cm high by 8.5 cm wide (approximately 6.8° x 4.8° of visual angle), while the open mouth occupied an area approximately 1.75 cm high by 3 cm wide (approximately 1° x 1.7° of visual angle). The auditory portion of the movie was presented at normal conversational volume (~65dB) through bilateral speakers 1 meter from the participant's head. Trials began with presentation of a still face consisting of the first video frame for between 1700 and 2000 ms with a uniform distribution. This was followed by the audiovisual movie, with a duration of 2000 ms. Following the movie, a still face consisting of the last video frame was presented for 750 ms. If no response was given by the end of the still face period a response screen appeared for a maximum of 2500 ms or until a response was given. Participants were instructed to fixate on the mouth and to use their right hand to indicate whether the stimuli were perceived to occur at the same time (i.e., synchronously) or at different times (i.e., asynchronously) via keyboard button press. Participants were also explicitly told to respond as quickly and accurately as possible, and that the appearance of the response screen was an indicator that their responses were too slow. All participants completed a practice block before the main experiment.

To create the experimental temporal asynchronies, we manipulated the audiovisual stimulus by delaying either the visual stimulus (to create an AV trial) or delaying the auditory stimulus (to create a VA trial). We created 6 asynchronies ranging from audition leading vision by 450 ms (A450V) to vision leading audition by 450 ms (V450A) in steps of 150 ms, resulting in a total of 7 conditions including the original movie featuring synchronized stimuli. Blocks consisted of 105 stimuli presented in a random order and participants completed 13 or 14 blocks, for a total of 1365 or 1470 trials. Stimulus onset for all stimuli was considered relative to the leading stimulus.

That is, for auditory leads stimulus onset was at the time of auditory onset, while for visual leads stimulus onset was the onset of the video frame associated with auditory onset in the original video. These events occurred simultaneously in the synchronous video. In other words, time 0 corresponded with the first point at which task relevant information was present.

Behavioral Data Analysis

We began data analysis by first excluding trials in which no response was given and trials in which response times were less than 150 ms. We then excluded, on a per condition and participant basis, trials with response times more than 3 standard deviations above that participant's mean response time. We also excluded EEG data using these same response time criteria. Together these procedures resulted in an exclusion of 10.36 ± 3.19 trials per participant. We then calculated for each participant the percentage of trials in each condition in which they indicated the stimulus occurred synchronously. For each participant we then fit a Gaussian distribution to the reported rate of synchronies in all 7 conditions using the Matlab fit.m function with free parameters of amplitude, mean, and standard deviation. The standard deviation of this distribution was taking as the temporal binding window (TBW) width and its mean as the point of subjective simultaneity (PSS). We note that TBWs are known to be asymmetric, but we utilized a symmetric Gaussian fit given the limited number of data points (7 total SOAs), which impairs the reliability of asymmetric fitting procedures. We also calculated the full width at 75 percent maximum (FW75M), equivalent to 1.517 standard deviations, for each Gaussian distribution.

Additionally, we calculated mean participant response time for each condition. Response time was calculated from onset of the auditory stimulus for conditions with auditory leads

(A450V, A300V, and A150) and synchronous stimulus onset (AV). For visual leads, response times were calculated from onset of the video frame where auditory onset would have occurred, had it not been delayed. In all conditions, response times thus began at the point at which task relevant information was first available. We compared response times across conditions using repeated measures analysis of variance (ANOVA) with follow-up paired sample t-tests.

EEG Recording and Processing

Continuous EEG was recorded from 128 electrodes referenced to the vertex (Cz) using a Net Amps 400 amplifier and Hydrocel GSN 128 EEG cap (EGI systems Inc.). Data were acquired with NetStation 5.3 with a sampling rate of 1000 Hz and were further processed using MATLAB and EEGLAB (Delorme & Makeig, 2004). Continuous EEG data were band-pass filtered from 0.15 to 50 Hz with a 6 dB roll-off of 0.075 to 50.075 Hz using the EEGLab firfiltnew.m function, which implements a bi-directional zero-phase finite impulse response filter. Epochs 3s long from 1000 ms before to 2000 ms after onset of the first stimulus were then extracted. Artifact contaminated trials and bad channels were identified and removed through a combination of automated identification of trials in which any channel exceeded +/- 100 μV and rigorous visual inspection. Data were then recalculated to the average reference and submitted to independent component analysis (ICA) using the Infomax algorithm (Jung, Makeig, Humphries, et al., 2000; Jung, Makeig, Westerfield, et al., 2000) (0.5E-7 stopping weight, 768 maximum steps). Lastly bad channels were reconstructed using spherical spline interpolation (Perrin, Pernier, Bertrand, Giard, & Echallier, 1987) and data were re-inspected for residual artifacts. Overall a mean of 1081 (79% ± 9.5%) of trials were retained, while 4.17 (SD ± 2.42) channels and 10.56 (SD ± 4.14) Independent components were removed per participant. There was no

difference in the number of trials accepted per condition across participants ($F_{6,144} = 1.46$, $p = 0.196$).

ERP Analysis

Event related potentials were calculated by averaging trials for each condition in the time domain. To reduce the possibility of brain responses to the pre-articulatory mouth movements contaminating the baseline, we baseline corrected ERPs using period from 300 ms to 100 ms before onset of the first stimulus. We focused our ERP analysis on auditory event related potentials based on previous literature showing they are moderated by concurrent visual speech (Baart, 2016; Besle et al., 2004; van Wassenhove et al., 2005).

Peak Amplitude of Auditory Event Related Potentials.

We extracted the amplitude of event related potentials by defining windows based on canonical brain responses and averaging amplitude within those windows. For the N1 component, we used a window of 90 to 130 ms after auditory stimulus onset. For the P2 component, we used a window of 160 to 240 ms after auditory stimulus onset. For the N2 component, we used a window of 250 to 350 ms after auditory stimulus onset. For peak-to-peak voltage differences and topographies, we subtracted either the N1 or N2 component from the P2 component. Positive values thus indicate greater prominence of the P2 relative to the preceding or following ERP component.

Peak Latency of Auditory Event Related Potentials

We calculated the latency of the N1 and P2 peaks based on previous work indicating they should

occur earlier in the presence of a visual stimulus (Baart, 2016; van Wassenhove et al., 2005). We identified peaks using the findpeaksG.m function (mathworks.com/matlabcentral/fileexchange/11755) with slope threshold = 0.001, amplitude threshold = -1.5, and a 5-point boxcar smoothing kernel. The N1 peak was searched in the range of 70 to 150 ms and the P2 peak was searched in the range of 160 to 240 ms. In cases where more than 1 peak was found or the peak was within 5 ms of the edges of the search range the primary peak was selected if unambiguous by the first author (DS) and otherwise treated as missing data. For statistical analysis of peak latency, missing data was imputed using the MDI toolbox (Folch-Fortuny, Arteaga, & Ferrer, 2016) via trimmed data regression, and analyses were repeated with only complete records to confirm results.

Frequency Domain Analysis

We further examined whether visual inputs modulated brain responses without signal averaging using time-frequency analysis. Time Frequency decomposition of single trial EEG data was accomplished using convolution with Morlet wavelets with frequencies from 3.5 to 35 Hz in 0.5 Hz steps. Wavelets had 2.5 cycles at the lowest frequency rising to 9.3 cycles at the highest and convolution was performed with a temporal resolution of 10 ms. Power was then decibel transformed relative to the -600 ms to -200 ms pre stimulus period.

Re-alignment to Auditory Stimulus Onset

We present event related potentials (ERPs) and time-frequency representations (TFRs) aligned such that time 0 is onset of the leading stimulus and task relevant information. For statistical analysis of auditory locked brain responses to stimuli in which visual inputs occurred first, we

employed a stimulus re-alignment procedure. Data were re-aligned with conditions in which audition occurred first by subtracting the temporal delay from the data time points (i.e. for the V150A condition 150 ms was subtracted from the time point values for all data points). This results in a new time series in which onset of the auditory stimulus occurs at time 0. Note that this re-alignment procedure was performed *after* baseline correction to prevent potential baseline contamination, as these re-aligned time series then had visual stimuli occurring before time 0.

Statistical Analysis

We performed statistical analysis using paired sample t-tests and repeated measures ANOVAs when time windows of interest were known. For analysis of time series data without a pre-defined window of interest, we utilized nonparametric randomization testing to control for multiple comparisons. For the amplitude of evoked potentials, we utilized repeated measures ANOVA and performed follow up paired sample t-tests when appropriate. For time-frequency representations of single trial data we collapsed in the theta and alpha bands, then performed time series testing using non-parametric randomization testing with cluster based correction for multiple comparisons (Maris & Oostenveld, 2007). We used the implementation in Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) with 10,000 randomizations, cluster inclusion alpha = 0.05, and a permutation alpha = 0.025. The test statistic used for cluster inclusion was dependent sample F multivariate, although we note that the choice of test statistic and cluster inclusion threshold do not directly impact the likelihood of a significant finding compared to the permutation distribution (Maris & Oostenveld, 2007). Data from 600 ms before to 1050 ms after auditory stimulus onset were included in this process. For the theta band, we then performed follow-up repeated measures ANOVA testing by collapsing in two separate time windows and

then using follow-up paired sample t-tests as appropriate.

The first time window was the *a-priori* selected canonical N1 to P2 time region (90-240ms). The second time window was 300-650 ms and was selected to account for the remaining significant theta power differences while reducing data contamination from the earlier period due to the temporal precision of the wavelet transform. We restricted follow-up analysis of this second time period to the lower portion of the theta band (3.5–5 Hz) based on visual inspection of TFRs which indicated that the majority of the effect was found in the lowest frequencies. For completeness, we further analyzed whether temporal windows of 350-650ms or 400-650ms yielded different theta band results than our initial analytical selection. For induced alpha power, we took an additional step by calculating the first temporal derivative of alpha power to isolate the auditory locked power increase from slow induced power decreases clearly attributable to the presence of a visual stimulus.

To estimate the temporal extent of early power effects we fit inverted Gaussian functions to the group averaged power data using the Matlab fit.m function with free parameters of amplitude, mean, and standard deviation and report the temporal window within 1.517 standard deviations (equivalent to the point at which power attenuation effects were at 75% of maximum, the FW75M). Finally, we analyzed brain behavior correlations using linear regression (Pearson correlation) between participant behavior and time-frequency data averaged over both time and frequency.

Results

Behavioral Results

We began our analysis by compiling reported rates of synchrony for each of the 7 SOAs. Consistently with previous experiments, synchronously presented stimuli (i.e., SOA of 0 ms) were reported as synchronous most frequently, and increasing rates of asynchrony led to increasing rates of reported asynchrony (Conrey & Pisoni, 2006). Gaussian functions were then fit to the reported rates of synchrony and were found to fit the individual participant data well (96.11 ± 3.33% of variance explained). Mean TBW size across participants, measured as the standard deviation of these individual Gaussian functions, was 296.1 ± 97.7 ms. In terms of full width 75 percent maximum (FW75M) this window is 449.2ms in width. We also replicated the frequently reported asymmetric nature of the temporal binding window as stimuli with a visually lead were reported as synchronous more frequently than their auditory first counterparts (mean PSS 75.2 ± 33.5 ms) (t-test vs 0; $t_{24} = 11.226$, $p = 4.902 \times 10^{-11}$) **(Fig 2-2A)**. Importantly, the

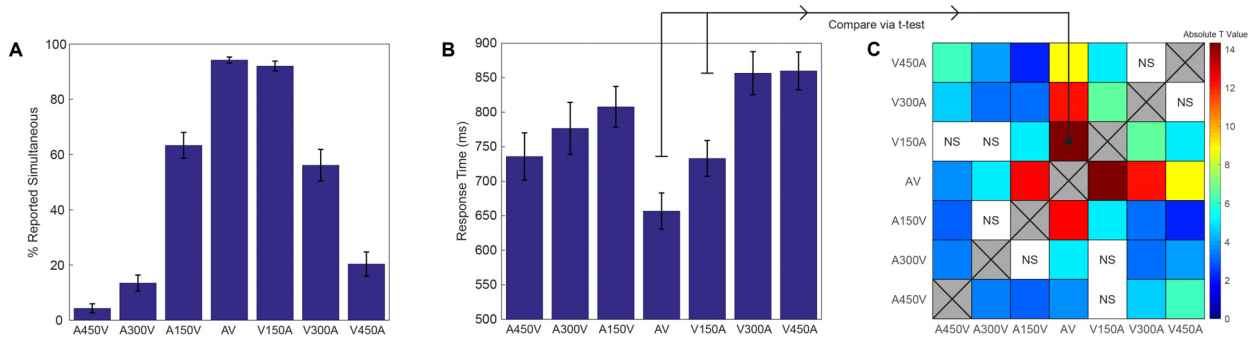

Figure 2-2 Behavioral results

A) Mean rate of synchrony reported in each condition. Error bars indicate standard error of the mean.
B) Mean response times for each condition. Error bars indicate standard error of the mean. Response time was measured from onset of task relevant information. For auditory leads and the synchronous stimulus, this was the onset of the auditory stimulus. For visual leads, this was the video frame associated with auditory stimulus onset in the un-edited movie stimulus.
C) Pairwise comparisons of response times between all 7 conditions. Absolute value of the T statistic is shown and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown. The black line indicate how an individual portion of the grid is derived from a paired sample t-test. NS indicates non-significant ($p >= 0.05$) pairwise comparisons.

smallest visual lead (V150A) was perceived as synchronous on most trials by the majority of participants. Response times were found to differ significantly across conditions (repeated measures ANOVA, $F_{6,144} = 24.93$, $p = 3.8288 \times 10^{-20}$) (**Fig 2-2B**) and were faster for synchronous trials when compared with any of the other asynchronous conditions (all 6 pairwise comparisons $p < 0.0015$). Response times also changed asymmetrically across conditions – increasing the asynchrony for visually leading trials resulted in increasing response times, while increasing the asynchrony for auditory leading trials resulted in decreasing response times. We present all pairwise response time comparisons, masked for significance of $p < 0.05$ uncorrected in **Fig 2-2C.**

Event Related Potentials

The first step in our EEG analysis focused on the time domain by averaging the signal to derive

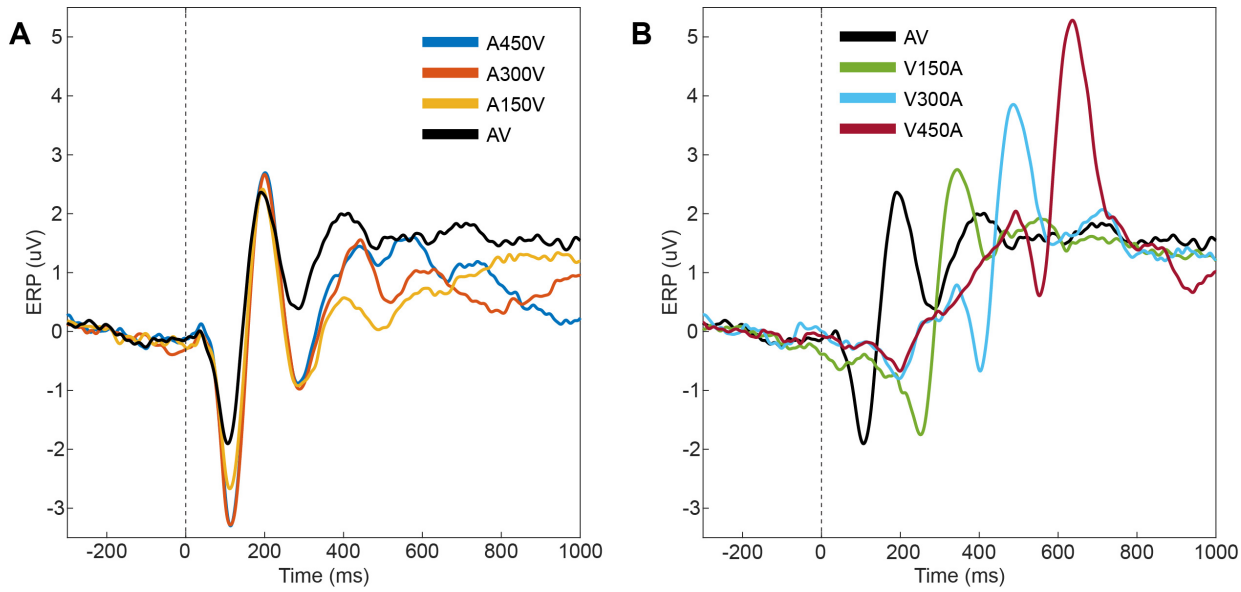

Figure 2-3 Event Related Potentials

A) Event related potentials at electrode CZ for the synchronous condition (AV, black) and the 3 levels of auditory lead (A450V, A300V, A150V).
B) Event related potentials at electrode CZ for the synchronous condition (AV, black) and the 3 levels of visual lead (V150A, V300A, V450A). Note that the auditory ERP is delayed by the SOA and that strong positive voltage buildup can be observed.

ERPs (**Fig 2-3**). We chose to focus our analysis on electrode Cz as this electrode is optimally positioned to capture auditory event related potentials, and proved to be relatively isolated from contamination due to ongoing ERPs from the visual stimulus. Distinct auditory ERPs were clearly present in all 7 SOA conditions at this electrode (**Fig 2-3A and 2-3B)**. Despite its relative isolation from visual ERPs, strong positive voltage trends, likely a result of parietal voltage buildup related to decisional processes, were clearly present in conditions with visual leads (**Fig 2-3B**). We thus began our analysis by focusing on the 4 conditions without a visual lead (A450, A300V, A150V, and AV). Across these 4 conditions the amplitude of the N1 ERP component (90-130 ms) was significantly different (repeated measures ANOVA $F_{3,72} = 25.54$, $p = 2.3218 \times 10^{-11}$) (**Fig 2-4A**). Follow up paired sample t-tests (**Fig 2-4B**) indicated that differences occurred because the smallest auditory lead, A150V, had a smaller N1 (i.e. a less negative voltage) than the two larger auditory leads (A450V vs A150V: $t_{24} = 3.412$, $p = 0.0023$; A300V vs A150V: $t_{24} = 5.056$, $p = 3.6 \times 10^{-5}$). Furthermore, this effect increased for these large auditory leads compared to synchronous presentation (A450V vs AV: $t_{24} = 5.915$, $p = 4.12 \times 10^{-6}$; A300V vs AV $t_{24} = 6.431$, $p = 1.19 \times 10^{-6}$). Importantly, the two larger auditory leads (A450V and A300V) were consistent with one another ($t_{24} = 0.439$, $p = 0.665$) (Pilling, 2009), while the small auditory lead (A150V) represented an intermediate step between these SOAs and the synchronous condition (A150V vs AV $t_{24} = 3.833$, $p = 0.0008$). Voltage for the P2 ERP component (160-240 ms) was found to not be significantly different across conditions (repeated measures ANOVA $F_{3,72} = 0.4574$, $p = 0.7129$). These findings were found to be qualitatively similar when using N1 and P2 amplitudes drawn from individualized peaks (see peak latency analysis below) rather than averaging over pre-determined windows (N1 amplitude $F_{3,72} = 27.2$, $p = 7.19 \times 10^{-12}$; P2 amplitude $F_{3,72} = 1.22$, $p = 0.3072$). We thus not only replicate previous

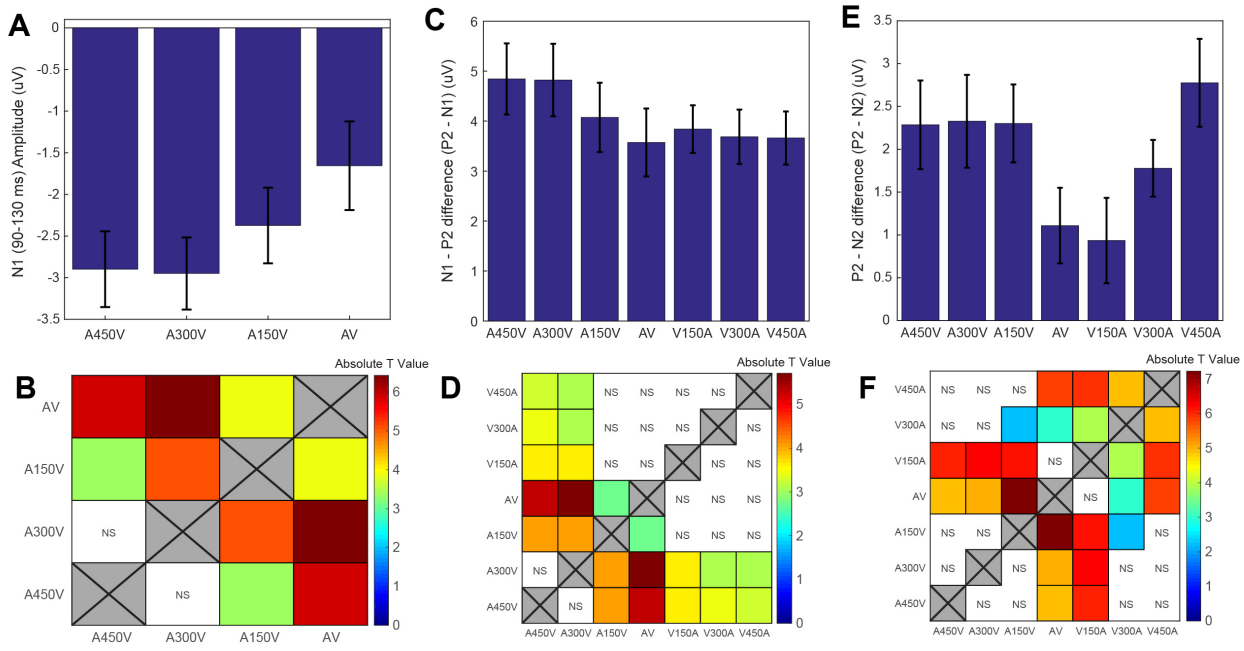

Figure 2-4 Event Related Potential Analysis

A) Auditory N1 voltage (90-130 ms after auditory stimulus onset) for the 3 auditory leads (A450V, A300V, A150V) and synchronous presentation (AV). The N1 is significantly reduced for a small auditory lead and synchrony. Error bars indicate standard error of the mean.

B) Pairwise comparisons of the N1 between the 4 conditions depicted in C. Absolute value of the T statistic is shown and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown.

C) Auditory N1 (90 – 130ms after auditory stimulus onset) to P2 (160 – 240ms after auditory stimulus onset) voltage change for all 7 conditions (P2 – N1). Large auditory leads have a significantly larger voltage shift. Error bars indicate standard error of the mean.

D) Pairwise comparisons of the P2-N1 voltage between all 7 conditions. Absolute value of the T statistic is shown, and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown.

E) Auditory P2 (160 – 240ms after auditory stimulus onset) to N2 (250 – 350ms after auditory stimulus onset) voltage change for all 7 conditions (P2 – N2). Large auditory leads and large visual leads have a significantly larger voltage shift. Error bars indicate standard error of the mean.

F) Pairwise comparisons of the P2-N2 voltage between all 7 conditions. Absolute value of the T statistic is shown and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown.

findings of reduced N1 amplitude due to the presence of visual speech and the lack of this effect for a sufficiently large auditory lead, but we also establish the level of temporal asynchrony

associated with an intermediate level of N1 amplitude suppression.

We then extended our ERP analysis to include all 7 SOAs by focusing on peak-to-peak voltage differences (i.e. peak prominence), which has been previous applied in similar studies (Pilling,

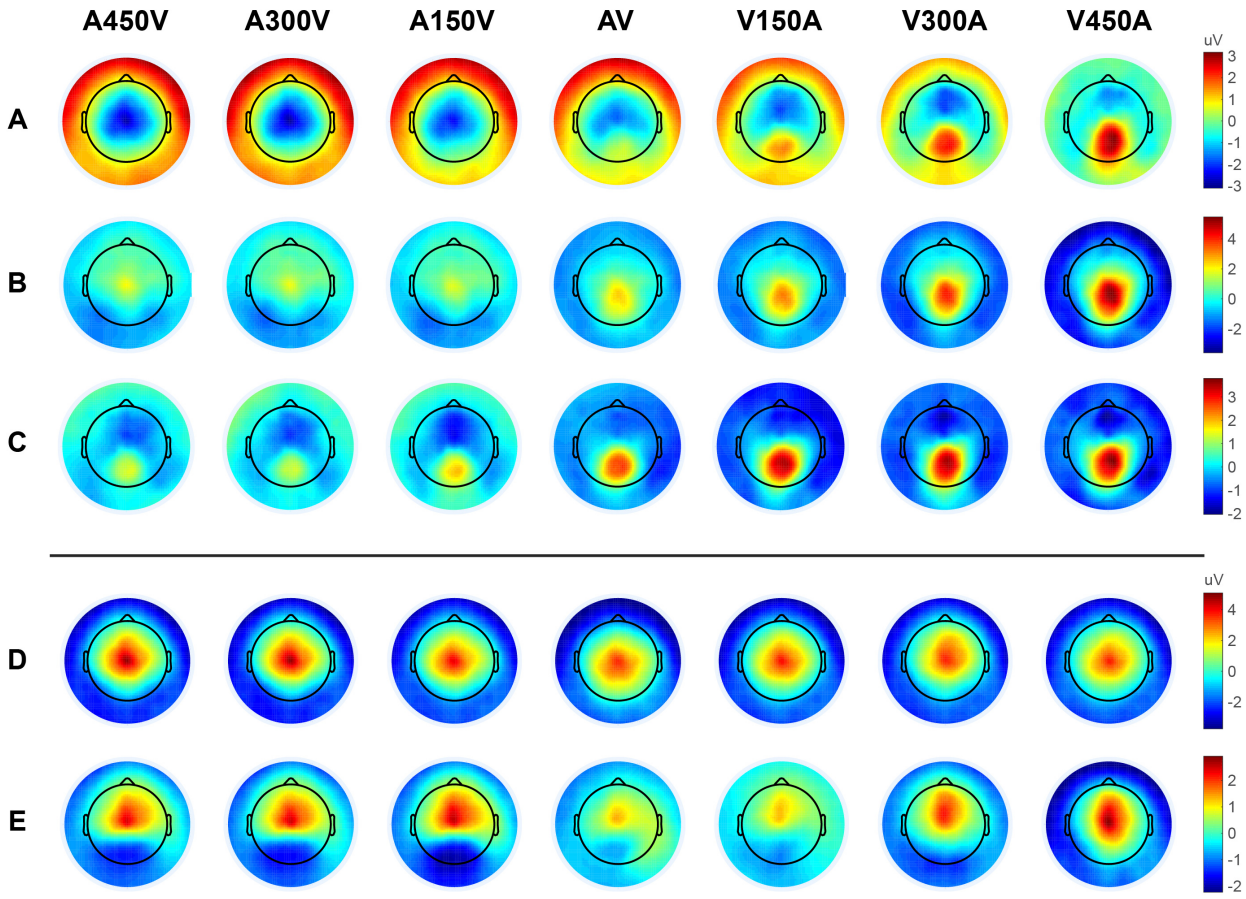

Figure 2-5 Topographic Representation of Event Related Potentials.

A) Topographic representation of the auditory N1 (90-130 ms after auditory stimulus onset) for all 7 conditions.
B) Topographic representation of the auditory P2 (160-240 ms after auditory stimulus onset) for all 7 conditions.
C) Topographic representation of the auditory N2 (250-350 ms after auditory stimulus onset) for all 7 conditions.
D) Topographic representation of N1 to P2 voltage shift (P2 – N1) for all 7 conditions (i.e. Panel B – Panel A). Positive values indicate higher relative prominence of the P2.
E) Topographic representation of P2 to N2 voltage shift (P2 – N2) for all 7 conditions (i.e. Panel B – Panel C). Positive values indicate higher relative prominence of the P2.

2009). One advantage of this peak prominence measure is that it functions to cancel the observed parietal buildup effects. We first subtracted N1 voltage from P2 voltage (P2 – N1) (**Fig 2-4C**). This voltage difference was found to be largest for the large auditory leads and differed significantly across conditions (repeated measures ANOVA $F_{6,144} = 8.089$, $p = 1.5396 \times 10^{-7}$). Similar to single ERP peak analysis, this peak-to-peak effect was found to be robust when using individualized ERP peak amplitudes ($F_{6,144} = 11.27$, $p = 2.679 \times 10^{-10}$). Large auditory leads (A450V, A300V) had a significantly larger N1 to P2 voltage change than all other conditions, while the smallest auditory lead (A150V) had a larger N1 to P2 voltage change than synchronous presentation, but not larger than visual leads. Due to the number of unique pairwise comparisons (21) we present absolute T value for all comparisons, masked for significance of $p < 0.05$ uncorrected (**Fig 2-4D**). We then repeated this analytical approach for peak-to-peak differences between the P2 and N2 (P2 – N2) (**Fig 2-4E**). Voltage differences between these peaks were also found to be significantly different across conditions (repeated measures ANOVA $F_{6,144} = 15.166$, $p = 2.0512 \times 10^{-13}$). Auditory leads (A450V, A300V, and A150V) and the largest visual lead (V450A) were found to have the largest P2 to N2 voltage differences, while this difference was small for synchronously presented stimuli and for the small visual lead condition (V150A). We present all pairwise comparisons, masked for significance of $p < 0.05$ uncorrected in **Fig 2-4F**. Topographies for the ERP components of interest (N1, P2, and N2) as well as difference topographies (P2-N1 and P2-N2) are presented in **Fig 2-5**.

We further analyzed the latency of the N1 and P2 peaks based on previous literature that has shown these components to be accelerated by visual speech (Baart, 2016; van Wassenhove et al., 2005). For the N1 analysis, 19/25 participants had identifiable peak latencies in all 7 conditions, while the latency of 10 individual ERP peaks (out of a total of 175 peaks [25 participants x 7

conditions] or 5.71%) could not be identified. We imputed these 10 data points using the MDI toolbox (Folch-Fortuny et al., 2016) via trimmed scores regression with 4 principle components. We then compared N1 latencies across conditions and found that they differed significantly (repeated measures ANOVA $F_{6,144} = 13.03$, $p = 9.6727 \times 10^{-12}$) (**Fig 2-6A**). Follow-up t--tests

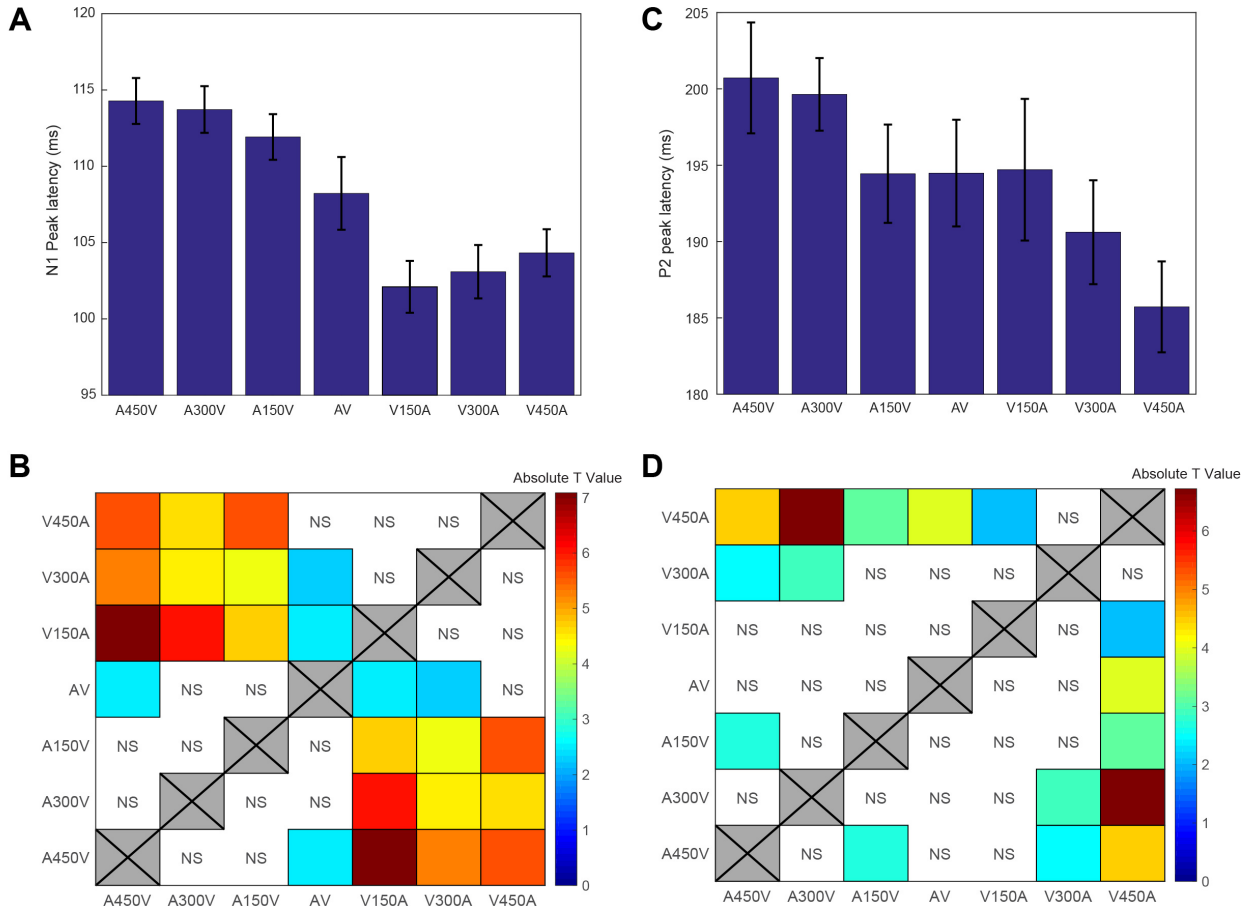

Figure 2-6 N1 and P2 Component Latency

A) N1 latency for all 7 conditions. The latency of the N1 decreased in the presence of a temporally aligned visual stimulus and was lowest for a small visual lead. Error bars indicate standard error of the mean.
B) Pairwise comparisons of the N1 latency between all 7 conditions. Absolute value of the T statistic is shown and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown.
C) P2 latency for all 7 conditions. Reductions in P2 latency were monotonic and the lowest latency occurred at large visual leads. Error bars indicate standard error of the mean.
D) Pairwise comparisons of the P2 latency between all 7 conditions. Absolute value of the T statistic is shown and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown.

indicated this occurred because visual leads reduced the latency of the N1, with the shortest latency occurring at the smallest visual lead (V150A). We present all follow up t-tests, masked for $p < 0.05$ uncorrected in (**Fig 2-6B**). We repeated this analysis excluding the 6 participants with incomplete data and found that the differences were still significant ($F_{6,114} = 9.73$, $p = 1.1219 \times 10^{-8}$). For the analysis of the P2 component, 22/25 participants had identifiable peaks in all conditions, while a total of 3 individual ERP peaks (1.71%) could not be identified and were imputed using trimmed scores regression and 3 principle components. We then compared P2 latencies across conditions (repeated measures ANOVA $F_{6,144} = 4.95$, $p = 0.0001$) (**Fig 2-6C**). Effects for the P2 illustrated a monotonically decreasing latency as the amount of time from visual onset increased. We present all follow up t-tests, masked for $p < 0.05$ uncorrected in (**Fig 2-6D**). We repeated this analysis excluding the 3 participants with incomplete data and found that differences were still significant ($F_{6,126} = 3.71$, $p = 0.002$).

Time-Frequency Analysis

In order to recapture dynamics lost during signal averaging (i.e. induced brain oscillations), and to better separate auditory linked activity from ongoing visual and/or cognitive activity [i.e. the centro-parietal positivity (O'Connell, Dockree, & Kelly, 2012), which is clearly visible in the ERPs and occurs at a lower frequency (~1-2 Hz)], we utilized time-frequency analysis. In this analysis, the wavelets effectively function as band-pass filters, removing slow activity related to decisional buildup. Time frequency representations (TFRs) resulting from wavelet convolution at electrode Cz for all 7 conditions are presented in **Figure 2-7.** We focused our analysis on the theta (4-8 Hz) and alpha (8-12 Hz) frequency bands as these frequency ranges have established functional significance and exhibited clear activity in the TFRs. We did not analyze the

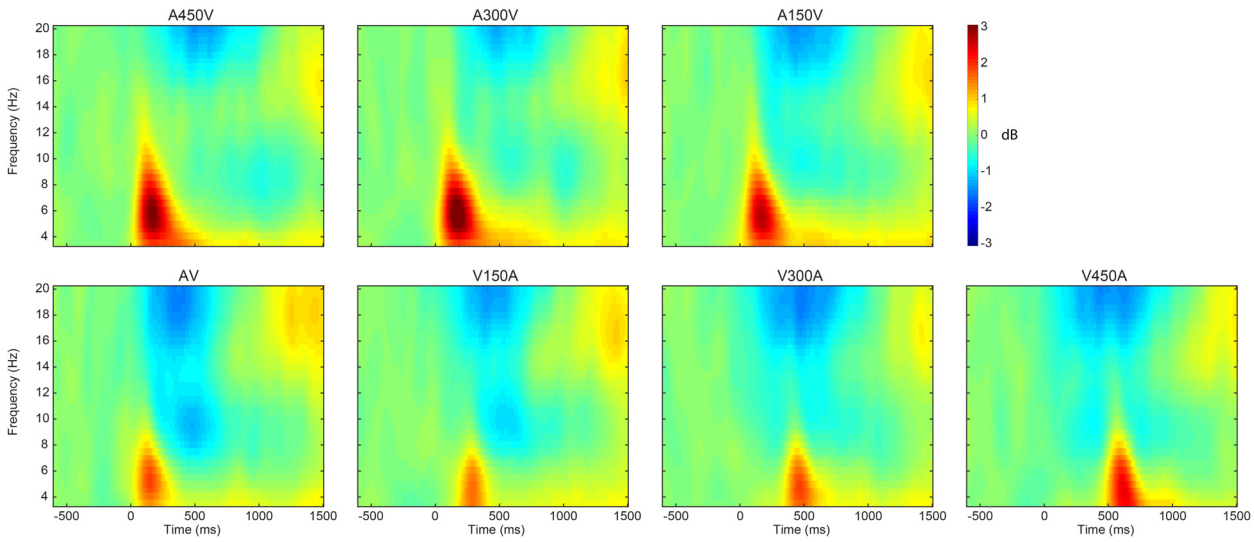

Figure 2-7 Time-Frequency Representations for electrode Cz.

- Top Row: TFRs for auditory leads ranging from large to small (left to right)
- Bottom Row: TFRs for synchronous presentation (AV) and visual leads ranging from small to large (left to right).

pronounced beta desynchronization as it exhibited a left lateralized topography consistent with initiation of motor responses. We averaged data within the alpha and theta bands and then used non-parametric randomization testing with cluster-based correction for multiple comparisons to determine the time period during which conditions differed. In the theta band, a sustained period of significant difference was found from 60 to 660 ms ($p < 0.0001$, randomization test, **Fig 2-8A**). In the alpha band we noted that slow drifts in induced alpha power were present beginning with the onset of the visual stimuli. These appeared to be distinct from the more localized sharp deflections associated with auditory stimulus onset (**Fig 2-8B**). We corrected for these drifts and associated spurious statistical differences by taking the first temporal derivative (i.e. the rate of change) of alpha power, thus isolating sharp deflections in power. These differences were found

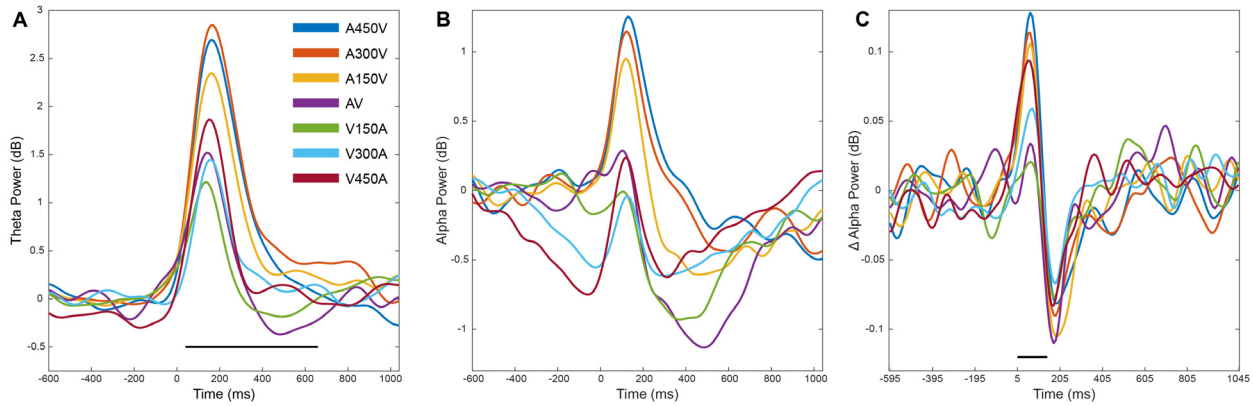

Figure 2-8 Time-Frequency Activity Averaging Within Frequency Bands

A) Time course representation of theta (4-8 Hz) power for all 7 conditions. Time courses were re-aligned such that time 0 corresponds with auditory stimulus onset in all conditions. The black underline indicates the period of sustained significance (60-660ms, $p < 0.0001$, randomization test).

B) Time course representation of alpha (8-12 Hz) power for all 7 conditions. Time courses were re-aligned such that time 0 corresponds with auditory stimulus onset in all conditions. Note the strong negative drift attributable to the presence of a visual stimulus.

C) First temporal derivative of alpha (8-12 Hz) power for all 7 conditions. Time courses were re-aligned such that time 0 corresponds with auditory stimulus onset in all conditions. The black underline indicates the period of sustained significance (5-145ms, $p < 0.0001$, randomization test). A time point of 5 indicates change in power between time points 0 and 10 in panel B and so on.

to be significant from 5-145 ms via randomization testing ($p < 0.0001$, randomization test, **Fig 2-8C**). Importantly, the duration of significant differences in theta power was far more sustained than for alpha power. Further, theta power demonstrated a topography that evolved between distinct central and fronto-central distributions. While statistical significance in the theta band was continuous, changes in topography and bandwidth suggested that theta activity might represent multiple consecutive processes. We thus opted to analyze theta power in an early window (90 – 240 ms, corresponding with the beginning of our *a-priori* N1 window and ending with our *a-priori* P2 window) and a late window (300 – 650 ms). Given the continuous nature of significant time points, we also confirmed that alternative selections for the late time window (i.e. 350-650 ms or 400-650 ms) were consistent with our initial selection (see below results

section). Change in alpha power was averaged across the entire 5-145 ms period deemed significant by the randomization test.

Topographies for early theta power were found to have a central topography consistent with auditory cortical generators and are presented in **Fig 2-9A** while topographies for the alpha power were similarly centrally distributed and are presented in fig **2-9B**. Early theta power differed significantly across conditions ($F_{6,144} = 24.23$, $p = 1.0513 \times 10^{-19}$), with the highest power found for large auditory leads (A450V and A300V). These large auditory leads were not different from one another ($t_{24} = -1.238$, $p = 0.228$), but were different from all other conditions with a more temporally proximate visual stimulus. Strikingly, with the sole exception of V150 compared to V300A, all individual steps of changing temporal synchrony were significantly different from their nearest neighbors (**Fig 2-10A**). We present all pairwise comparisons for early theta power in **Fig 2-10B.** Change in alpha power was also found to be significantly different across conditions ($F_{6,144} = 13.25$, $p = 6.447 \times 10^{-12}$). Similar to theta power, change in alpha power was high for auditory leads, attenuated by synchronous visual stimuli and small visual leads, and recovered for larger visual leads (**Fig 2-10C**). We present all pairwise comparisons for changes in alpha power in **Fig 2-10D**.

We next sough to determine the overall temporal tolerance of our participants, as indexed by the change in suppression across SOAs. To estimate the temporal extent of the effects at the group level, we thus fit a Gaussian distribution (i.e. a TBW) to group averaged power in each frequency band. For theta power, a Gaussian function fit the group data extremely well (96.26% of variance explained) and had a standard deviation of 325.4ms, a mean of 193.25ms, and a FW75M of 493.6ms. For alpha power, a Gaussian function was similarly found to fit the group data well (95.80% of variance explained) and had a standard deviation of 257.9ms, a mean of

102ms, and a FW75M of 391.2ms. Our time-frequency analysis of the initial auditory responses thus strongly indicates that the temporal structure of the audiovisual stimulus relationship mediates the magnitude to which visual speech suppresses single trial auditory speech responses. Furthermore, this process occurs in a highly continuous manner with a temporal tuning function remarkably similar to that found in participant behavior, although we note that we did not find a direct relationship between neural response suppression and participant temporal acuity.

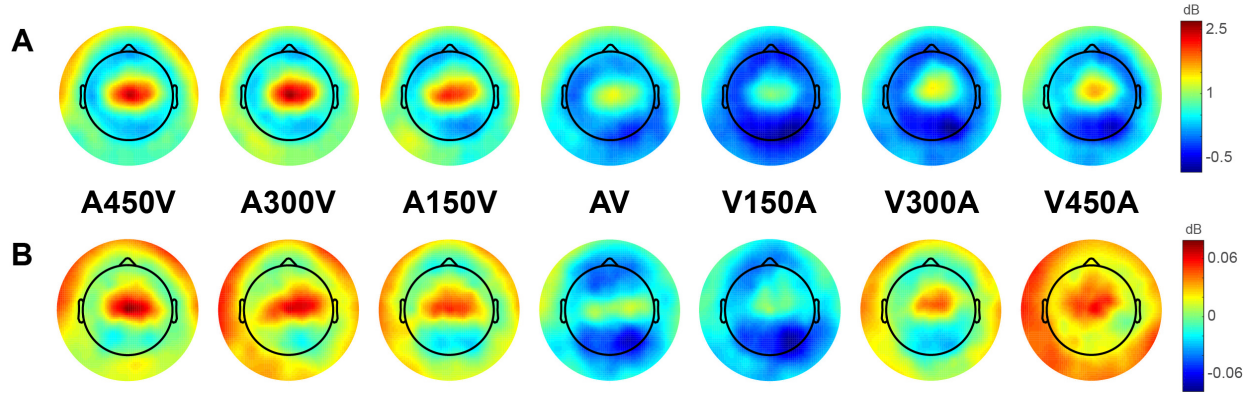

Figure 2-9 Early Time-Frequency Topographies

A) Topographic representation of early theta (4-8 Hz) power (90 – 240 ms). The topography is in full accord with auditory cortical generators.
B) Topographic representation of the first temporal derivative of alpha (8-12 Hz) power (5 – 145 ms). The topography is in fair accord with auditory cortical generators.

Having examined the frequency representation of the initial auditory response we then proceeded to analyze the later period of theta band significance. Visual inspection of results indicated that these late theta power differences were primarily driven by frequencies at the lower bound of the TFRs (3.5-5 Hz). We thus initially restricted our analysis of late (300-650

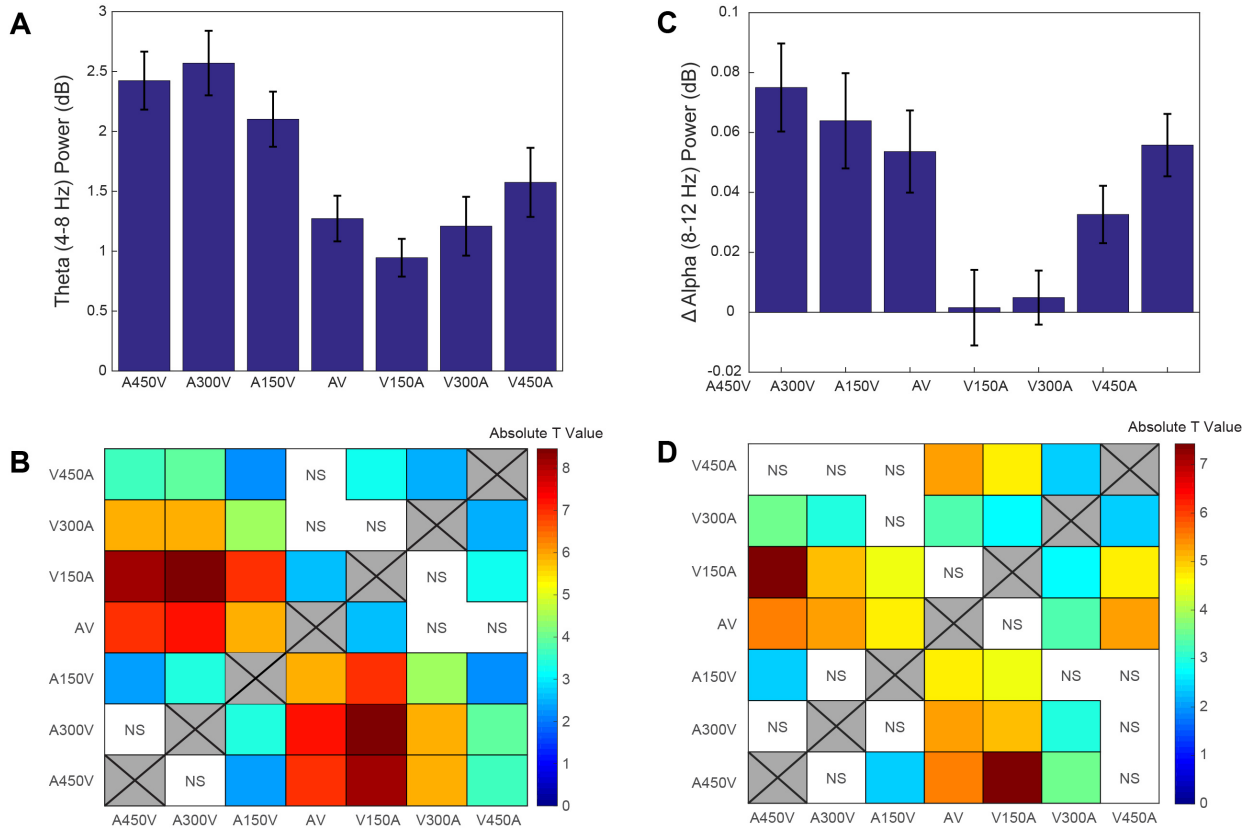

Figure 2-10 Early Time-Frequency Analysis

A) Early theta power for all 7 conditions. Theta power is highest for large auditory leads, decreases with increased temporal alignment between the visual and auditory stimuli, and recovers as the visual lead increases. Error bars indicate standard error of the mean.
B) Pairwise comparisons of early theta power between all 7 conditions. Absolute value of the T statistic is shown and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown.
C) First temporal derivative of alpha power for all 7 conditions. Change in alpha power is highest for large auditory leads, decreases with increased temporal alignment between the visual and auditory stimuli, and recovers as the visual lead increases. Error bars indicate standard error of the mean.
D) Pairwise comparisons of the first temporal derivative of alpha power between all 7 conditions. Absolute value of the T statistic is shown and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown.

ms) effects to these frequencies (see below for analysis in other theta band portions). Topographies of low theta (3.5-5 Hz) activity averaged across this period are presented in **Fig 2-11A**. Notably, the topography in this time-frequency range presented a more fronto-central distribution than observed for early alpha or theta band activity, less consistent with auditory cortical generators. This late theta activity also demonstrated differences across conditions ($F_{6,144} = 11.87$, $p = 8.5026 \times 10^{-11}$), but notably these differences were less continuous than for the early effects (i.e. within 'categories' of auditory lead, synchronous or small visual lead, and large visual lead no significant differences were found) (**Fig 2-11B**). We present all pairwise comparisons for late theta power in figure **2-11C**. The fronto-central distribution of this late theta activity is consistent with previous reports of neural responses to visual stimulus incongruence (Hanslmayr et al., 2008) and with more recent reports of cross-modal stimulus incongruence (Roa Romero, Keil, Balz, Gallinat, & Senkowski, 2016). To determine if this potentially congruence related activity corresponded with participant's ability to perceive temporal incongruence (i.e. stimulus asynchrony) in our experiment we correlated late theta power with individual perceptual report separately for each condition. We found that for small and moderate auditory leads the level of individual theta power negatively correlated with the rate at which participants reported synchrony in each condition (A150V: $r = -0.5920$, $p = 0.0018$; A300V: $r = -0.4349$, $p = 0.0298$) (**Fig 2-12A & 2-12B**). For the largest auditory lead (A450V) this correlation approached significance ($r = -0.3890$, $p = 0.0546$). This weaker correlation for the A450V condition is potentially explained by the high number of participants rarely reporting simultaneity (9/25 participants < 1% reported rate of synchrony, 19/25 participants < 5% reported rate of synchrony). Correlations between reported rate of synchrony and late theta oscillations were similarly negative for synchronous or visually leading conditions, but were not

significant (all $p > 0.163$). We further correlated late theta oscillations in each of these conditions to overall temporal binding window size. This correlation was found to be significant for theta power in the A450V ($r = -0.4397$, $p = 0.0279$) and A150V ($r = -0.5318$, $p = 0.0062$) conditions, but was not significant for the A300V condition ($r = -0.3141$, $p = 0.126$). Similar to correlations for perceptual accuracy, correlations between theta power and binding window size were found to be non-significant for synchronous and visually leading stimuli (all $p > 0.221$).

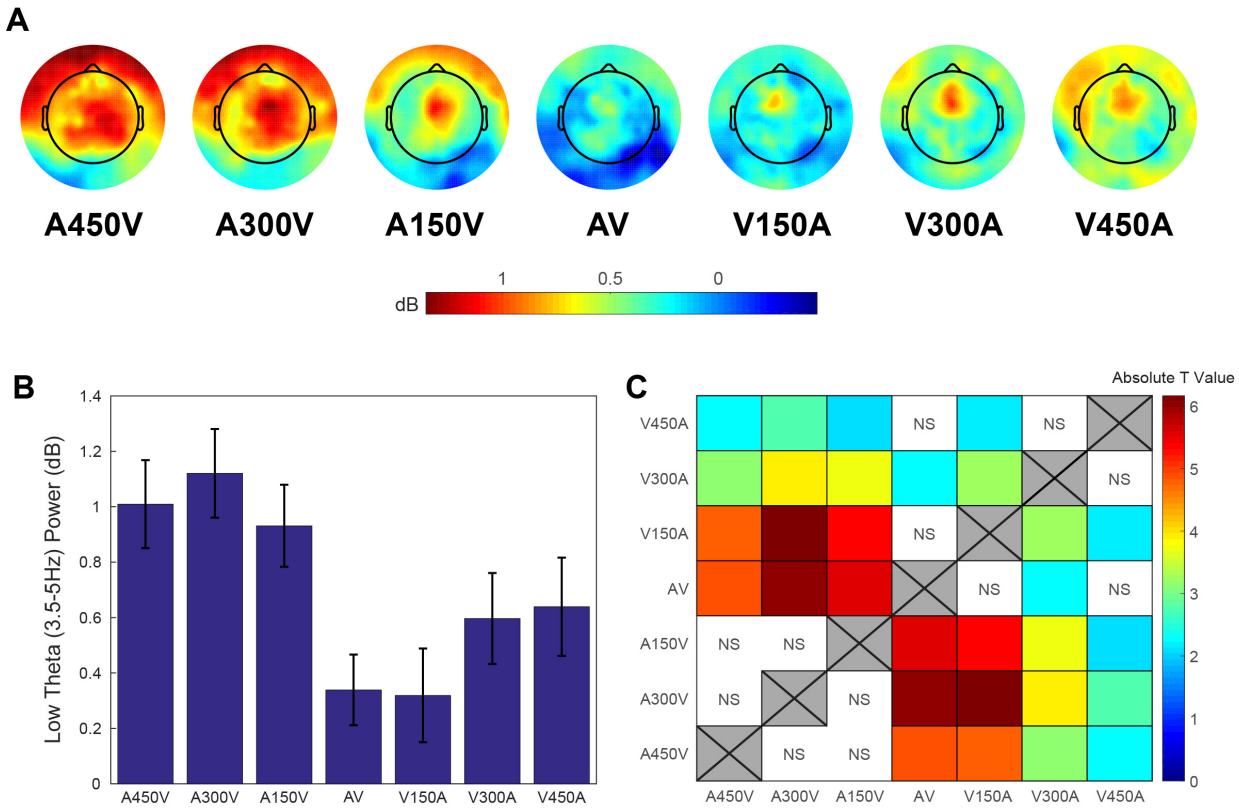

Figure 2-11 Late Theta Activity

A) Topographic representation of low theta (3.5 – 5 Hz) activity occurring 300 – 650 ms after auditory stimulus onset. The fronto-central distribution is inconsistent with auditory cortical generators.

B) Low theta (3.5 – 5 Hz) power averaged from 300 – 650 ms at electrode Cz. Synchronously presented stimuli (AV) and small visual leads (V150A) have the lowest power. Error bars indicate standard error of the mean.

C) Pairwise comparisons for late theta power between all 7 conditions. Absolute value of the T statistic is shown and results are masked so that only comparisons with $p < 0.05$ uncorrected are shown. Note the near categorical nature of significant comparisons.

For completeness, we repeated these analyses in the upper portion of the theta band (5.5 – 8 Hz) and further examined whether relationships held for the earlier (90 – 240 ms) time window. In the upper portion of the theta band (5.5 - 8 Hz) correlations between perceptual accuracy and theta power were substantially weaker, reaching only limited trend levels of significance and which are potentially attributable to spectral imprecision (A450V; r = -0.2277, p = 0.2737, A300V r = -0.3977, p = 0.049; A150V r = -0.3736, p = 0.0658). Theta power in the early time window (90 – 240 ms) were not found to correlate with perceptual accuracy in any condition for low theta (3.5 – 5 Hz; all p > 0.109) or high theta (5.5 – 8 Hz, all p > 0.32). Additionally, we investigated the degree to which the temporal window selected for 'late' oscillations contributed to the results. We found that utilizing narrower, and thus less conservative, analytical windows yielded qualitatively similar results. For example, in the A150V condition a 400 – 650 ms window yielded a correlation between reported synchrony and theta power of r = -0.5943, p = 0.0017. Given the strong similarity across windows and the need to make a temporal division

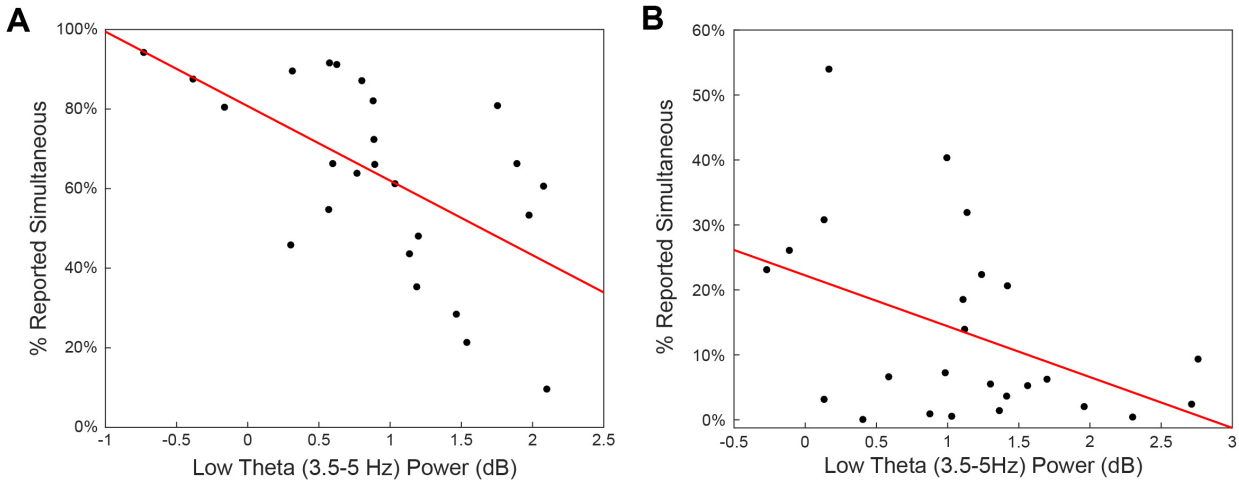

Figure 2-12 Relationship between Late Theta Activity and Behavior

A) Brain behavior correlation between late theta power and reported rate of synchrony in the A150V condition (r = -0.5920, p = 0.0018).
B) Brain behavior correlation between late theta power and reported rate of synchrony in the A300V condition (r = -0.4349, p = 0.0298).

between early and late power, we opted to utilize the most conservative temporal window and do not report on the narrower analytical windows further. Lastly, we confirmed that results in the later time window were not entirely phase locked by repeating the TFR analysis using the ERPs instead of single trial data. This analysis yielded qualitatively dissimilar results, which indicates that the findings in this window require consideration of single trials, and are likely at least partially oscillatory. Results in the early time window were found to be phase locked, and thus primarily represent the frequency domain version of the ERP.

Discussion

We sought to elucidate the effects of temporal concordance between visual speech and auditory speech on amplitude suppression, a well-established neural measure of multisensory integration for speech signals (Baart, 2016; Besle et al., 2004; Pilling, 2009; van Wassenhove et al., 2005). To do so we utilized a simultaneity judgment task that requires participants to attend to both vision and audition and we manipulated the temporal relationship between the sensory inputs. In terms of event related potentials, we partially replicated previous work (Baart, 2016; Pilling, 2009; van Wassenhove et al., 2005) by demonstrating suppression of the N1 component by synchronous visual signals. Importantly we also demonstrate the presence of an intermediate step, in which a sufficiently small auditory lead results in partial amplitude suppression. This intermediate step offers confirmation audiovisual speech integration operates in an efficient manner in which any visual input makes contributions proportional to its information content. Furthermore, through time-frequency analyses designed to reduce the influence ongoing visual and, in particular, decisional activity occurring at lower frequencies, we established that this

amplitude suppression exhibits substantial temporal tuning. Effects were found to be maximal at synchrony and small visual leads, both of which correspond with the greatest reports of perceptual synchrony. We also found that effects in the alpha band occurred more rapidly (~50ms earlier) than those seen in the theta band. These same time-frequency analyses also further indicated that, for auditory leads, low theta (3.5-5 Hz) oscillations persisted well after the auditory stimulus. These more persistent oscillations presented with a topographical pattern consistent with congruence processing and, most importantly, were found to correspond with task performance.

Changes in ERP Amplitude Are Limited to the N1 and N2 Components

When comparing ERP amplitudes for auditory leading stimuli and for true AV synchronous presentations, a highly significant reduction in absolute N1 amplitude was present when visual speech was presented synchronously or with a 150 ms delay relative to the auditory signal (A150V). This replicates and extends previous results indicating that a reduction in the magnitude of auditory cortical responses by visual speech can occur even when vision is slightly lagged and the pre-articulatory motion occurs concurrent with the auditory signal. Importantly, we also establish that the neural response to a stimulus with a small visual lag differs from both audiovisual responses with large visual lags and those to truly synchronous stimuli, thus establishing for the first time an intermediate level of suppression. This partial reduction by concurrent pre-articulation is consistent with the predictive coding account of audiovisual integration (Talsma, 2015), in which the articulatory movements are still informative but to a lesser degree.

Interestingly, we did not replicate findings of reduced P2 amplitude in the presence of

audiovisual speech. It is possible that the lack of P2 amplitude reduction in our results is due to the nature of the task participants were performing. The simultaneity judgment task requests that participants segregate stimuli in an effort to keep their timing as separate as possible. Previous work has established that the P2 may be less automatic and more amenable to top down regulation than the earlier P1 and N1 cortical responses. Specifically, in a multisensory speech task featuring the McGurk illusion (McGurk & MacDonald, 1976), P2 modulation by visual speech was reduced by the presence of an 'incongruent context' before stimulus onset (Ganesh, Berthommier, Vilain, Sato, & Schwartz, 2014). In this study, if the period preceding the experimental stimulus contained mismatched auditory and visual stimuli then the level of P2 integration (i.e. the amount of P2 suppression) for the experimental stimulus was reduced. This indicates that the top down factor of whether stimulus modalities are appropriate to integrate can modulate the degree of integration measured in the P2. We thus speculate that the lack of P2 reduction we observe may be a manifestation of the task demands, which asks our participants to segregate the stimuli as much as possible. Establishing empirically whether this is the case will require additional future work.

Despite the lack of P2 attenuation, peak-to-peak measures indicated that highly asynchronous visual speech accentuates the relative prominence of the P2 compared to the N2. This indicates enhanced N2 negativity when the stimulus has a large degree of temporal asynchrony. The N2 has previously been associated with conflict processing in a number of tasks (Iannaccone et al., 2015; Larson, Clayson, & Clawson, 2014; Yeung, Botvinick, & Cohen, 2004), and the enhanced negativity here indicates the N2 has sensitivity to *temporal* congruence. Time and frequency domain representations of error processing, which may have similar monitoring circuit substrates, have been shown to be partially independent and carry complimentary information

(Munneke, Nap, Schippers, & Cohen, 2015). Given the substantial temporal overlap between ERPs and our time-frequency results, we thus discuss temporal misalignment as a form of stimulus conflict further in the context of theta oscillations below.

Visual speech accelerates the N1 and P2

Previous reports have indicated that the presence of visual speech also accelerates the onset of early ERP components (and thus, presumably, their associated processing stages). We partially replicated this result and found that for the N1 this acceleration was greatest when vision slightly precedes audition (V150A). Although an ~11ms acceleration might seem small, it represents a roughly 10% speeding of peak latency. However, given the highly visually specified token that was used (the syllable 'ba', which is easily lip read and carries relatively well specified temporal information), the acceleration present for synchronous presentation (~5ms) is only about half of that expected based on previous reports linking the amount of acceleration to visual intelligibility (van Wassenhove et al., 2005). This relative reduction may be a result of task demands, similar to our lack of P2 amplitude reduction, or may reflect that even subtle differences in experiments seem to change whether this effect is found (Baart, 2016). We also found acceleration of the P2 component, despite not finding amplitude reduction in this same component. Our P2 finding, however, consisted of a monotonic latency reduction which reached significance in only the largest visual leads and does not replicate findings of fairly substantial (~20 ms) P2 latency facilitation afforded by the presence of a synchronous and readily recognized viseme (van Wassenhove et al., 2005). Given the apparent lack of temporal tuning, and the vulnerability of peak latency measures to effects such as entrainment of alpha oscillations by visual inputs, we do not believe our P2 acceleration result serves as an indicator of multisensory integration in this

task. Rather we interpret this finding similarly to our P2 amplitude finding, to indicate that the degree of P2 latency acceleration present for audiovisual speech may depend on task demands and context.

Temporal Constraints on Multisensory Speech Integration

Time-frequency analysis indicated that reductions in the magnitude of early brain responses occurred in both the theta and alpha frequency bands, and were strongly mediated by the degree of temporal concordance between the auditory and visual signals. As expected, large auditory leads resulted in a robust brain response consistent with auditory only processing. As the temporal lag of the visual stimulus decreased, there were corresponding and continuous decreases in response magnitude, with the smallest neural response at a small visual lead (V150A). Critically, further increases in visual lead resulted in a recovery of response magnitude. This neural distribution, with its visual lead bias and highly Gaussian shape bears striking resemblance to the TBW found both in our behavioral results, as well as in other similar reports (for reviews see: (Vroomen & Keetels, 2010; Wallace & Stevenson, 2014)). This extended temporal window for audiovisual speech integration is also consistent with reports that the auditory system integrates speech signals over a relatively protracted period of time, particularly in challenging acoustic conditions (Ding & Simon, 2013c). Further, these results are also highly consistent with recent reports of a similarly extended temporal integration window in reconstruction of continuous audiovisual speech (Crosse et al., 2016). Lastly, the extent of the temporal integration window we find (~500 ms) corresponds well with the cycle time of the ~2 Hz lower frequency bound in which auditory speech temporally correlates with visual speech (Chandrasekaran et al., 2009). Similar slow frequencies have also been shown to offer the most

robust audiovisual gain in speech stimulus reconstruction (Crosse et al., 2015), suggesting that not only is integration occurring over a large time window, but that integration over longer temporal epochs may result in more robust processing and encoding.

Intriguingly, we observed notable differences in the modulatory effects of visual speech on power in the alpha and theta bands. Specifically, differences in the alpha band emerged substantially earlier, while theta band effects were found later and to be more sensitive in representing differences between small ecologically plausible temporal offsets (i.e. AV vs V150A). Additionally, the positive deflection in alpha power was completely removed by a synchronous visual stimulus, while theta power was only attenuated. These differences are unlikely to be related to limitations in spectral resolution as symmetric spectral transforms such as Morlet wavelets spread lower frequencies further backwards in time. Instead, they likely reflect neural activity originating in functionally distinct but anatomically overlapping cortical circuits in auditory cortical regions. This is particularly relevant for the alpha band, which has been associated with a "working" frequency that determines the temporal resolution of the visual system (Samaha & Postle, 2015), and similarly affects audiovisual multisensory processing (Cecere, Rees, & Romei, 2015). Cortical alpha activity has also been proposed as an important tool for selective inhibition in challenging listening conditions, when integration of visual speech is most valuable (Strauss, Wostmann, & Obleser, 2014). Lastly, phase reset across sensory systems has been shown to be a generalized context and attention sensitive mechanism for multisensory neural interaction (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Lakatos et al., 2009) (for a review see: (van Atteveldt, Murray, Thut, & Schroeder, 2014)). In light of this previous work, the observed differences in timing and effect magnitude indicate that integrative processing of audiovisual speech stimuli may impact the efficacy of alpha reset mechanisms

earlier than neural circuits with responses in the theta band, resulting in rapid selective inhibition of neural populations that might otherwise contribute to the response. Such rapid dampening may then propagate to the slower theta band, which tracks the speech envelope and in our results carries a more continuous and precise representation of the temporal offset. This proposed interaction between frequency bands in speech processing is also supported by evidence that intrinsic theta oscillations shape syllable perception while endogenous alpha oscillations do not (Ten Oever & Sack, 2015). Our findings of band specific latency differences can thus be well accounted for by a two-stage model of multisensory temporal integration for speech signals. In this model, neural populations operating at higher alpha frequencies activate more rapidly, but carry less precise temporal information. These fast alpha circuits then refine cortical responses occurring in slower theta frequencies and thus allow these theta circuits to carry a more precise temporal and envelope representation. Such a model is well aligned with invasive physiological work indicating that sub additive neural interactions are associated with enhanced information content in cortical signals (Angelaki, Gu, & DeAngelis, 2009). We thus suggest that frequency domain analyses such as those conducted here are able to, at least partially, disentangle the temporal dynamics of visual influences on auditory cortical responses.

Theta Oscillations as a Marker of Cross-modal Incongruence Processing

In addition to early theta band effects, we also noted theta oscillations enduring long after auditory stimulus onset. These persistent theta oscillations occurred primarily at the lower end of the theta band (3.5 – 5 Hz) and had a more fronto-central distribution when compared with the early alpha and theta band effects. Not only did these oscillations differ across levels of temporal asynchrony, virtually vanishing in conditions where participants report the stimuli as

synchronous, but they directly corresponded with the accuracy of participant's perceptual report in auditory leading conditions. Given the nature of the task, in which participants are asked to detect temporal incongruence in the stimulus, we believe that these oscillations index processing of cross modal temporal incongruence in the brain. This is consistent with previous work which has indicated that similar theta oscillations are active during reconciliation of incongruent stimulus features (also known as conflict detection or stimulus-stimulus conflict) (Cohen, 2014). A well-known example of such processing is the Stroop task, in which written color words and the color they are written in are mismatched (e.g., the word 'blue' written in red). During performance of this task, fronto-central theta oscillations are observed on trials with conflicting information (Hanslmayr et al., 2008). Another example of such oscillations is during trials with conflicting information in a flanker task (Nigbur, Ivanova, & Sturmer, 2011). The topography of theta oscillations seen in such tasks, as well as in our experiment, is also consistent with anterior cingulate and other medial frontal generators, which have previous been linked to stimulus error processing (Cavanagh & Frank, 2014). The late timing of differences is also consistent with recent work examining the formation of large scale functional brain networks associated with multisensory speech perception (Kumar et al., 2016), and may indicate that frontal monitoring circuits play a critical role in such networks. Additionally, we believe that the presence of this relationship for high level (i.e. late frontal) and its absence for low level (i.e. early auditory cortex) may stem from the dissociation between perceived simultaneity and low level multisensory integration (Harrar, Harris, & Spence, 2017). Our report thus additionally serves as evidence that perceptual simultaneity may emerge from higher cognitive processes emerging relatively late in time.

Importantly, it has been shown that similar theta oscillations with medial frontal generators

are attenuated during cross-modal conflict in populations with reduced top down error control such as patients with schizophrenia (Roa Romero et al., 2016). Atypical multisensory temporal processing is increasingly being recognized in a number of neurological and neuropsychiatric disorders, including schizophrenia (Martin, Giersch, Huron, & van Wassenhove, 2013) and autism (Stevenson et al., 2014). This finding suggests that feed forward activity from sensory processing regions to error monitoring systems may form a neurophysiological basis for multisensory temporal dysfunction in these disorders. Additionally, we believe it is important to highlight the consistency of this relationship across conditions with auditory leads. This consistency indicates that superior temporal acuity is associated with stronger incongruence signaling regardless of relative perceptual difficulty. In other words, stronger conflict signaling during multisensory temporal incongruence is an individual trait and at least somewhat independent of perceptual threshold. Future work relating conflict signaling to individual differences in multisensory integration may yield further insights into the importance of this process.

That we do not find a relationship between theta oscillations linked to incongruence processing and behavior in synchronous or visually leading trials is not surprising. For synchrony and the smallest visual lead there is both little behavioral variability across participants and little perceived conflict to be signaled, as participants report the stimulus as occurring synchronously the vast majority of the time. In conditions with larger visual leads theta band conflict signaling would be expected to occur 300-650 ms after onset of the leading visual stimulus, during which frontal theta activity is obscured by the much larger auditory cortical response. Alternatively, temporal incongruence in visually leading conditions might be processed by different neural networks than auditory leading stimuli. This possibility is specifically raised by recent work

elucidating that the neural networks engaged during simultaneity judgment depend on stimulus ordering (Cecere, Gross, Willis, & Thut, 2017).

Multisensory Temporal Integration as a Fundamental Feature of Speech Processing

Taken together our time and frequency domain analysis point to a substantial temporal window in which visual speech reduces the amplitude and speeds the onset of the neural processing associated with speech signals. This window forms a substrate for the integration of relatively slowly occurring mouth movements and envelope fluctuations and further supports accounts that delta band (1-4 Hz, cycle time 250 ms – 1s) brain activity may serve a role in integrating temporal information in speech signals (Schroeder et al., 2008). Given the nature of audiovisual speech, in which individual syllables have variable visual-auditory onset timing, the presence of such a tolerant mechanism may form a fundamental component of the ability to correctly incorporate visual speech to enhance auditory perception. Our findings are also consistent with work indicating that multisensory temporal integration may serve as a gain control mechanism (Crosse et al., 2016). Our theta band tuning profile in particular, while quite broad, is also quite deep (~1.5 decibels), giving it a great deal of dynamic range to impact processing of signals differently depending on the degree of temporal alignment with visual inputs. This temporal weighting may serve to facilitate neural entrainment in particular, by providing strong weighting to near concurrent events and a more moderate weighting to events with ambiguous temporal concordance. In a rich visual environment, such a process would serve as a temporal "filter" on visual inputs. Lastly, we establish that temporal discordance between stimuli generates activity consistent with systems that respond to stimulus incongruence. In the context of naturalistic speech, continuous monitoring of temporal congruity and appropriate feedback to sensory

systems may make crucial contributions to sharpening the influences visual inputs have on auditory speech processing. The combination of these factors indicates that temporal integration is a fundamental feature of speech processing, and that neural systems are strongly adapted to take advantage of temporal structure in speech signals.

Conclusion

We establish that the temporal relationship between auditory and visual stimuli is critically important to the degree to which visual inputs attenuate auditory brain responses and accelerate the onset of early ERP components. Importantly, this attenuation operates in an asymmetric temporal window, in strong agreement with both behavioral and physiological measures of multisensory temporal integration for speech signals. Furthermore, the perceived temporal relationship of the stimuli is more categorically reflected by late theta oscillations, in that these oscillations are present for stimuli frequently reported as asynchronous and virtually absent for stimuli predominantly reported to be synchronous. In conditions in which audition leads, the strength of these categorical theta oscillations directly corresponded with participant performance, and this relationship was particularly strong when the stimulus was perceptually ambiguous. These findings contribute to a growing body of literature indicating auditory and visual speech signals are integrated over a surprisingly wide window of time, while further indicating that temporal mismatch between sensory modalities is processed in a manner similar to other types of stimulus conflict. The band specific nature of the neural processing differences also suggests that distinct neural populations contribute to temporally distinct stages of integration of audiovisual speech signals. Further investigation of these temporal dynamics may

make substantial contributions to the refinement of circuit models that account for visual enhancement of information content in cortical speech representations.

While we believe this work sheds important light on multisensory temporal processing it is not without limitations. The evoked design, which by nature is highly repetitive, is less naturalistic than normal speech. Similarly, the simultaneity judgment task is somewhat removed from normal speech task demands such as comprehension and stream segregation, and in particular our lack of P2 modulation and unusual P2 latency facilitation may be specific to our experimental design. Lastly, because substantial phase resetting of ongoing neural processes contributes to evoked responses (Makeig et al., 2002), and the inherent non-independence of phase and amplitude measures in noisy signals such as EEG (Ding & Simon, 2013b), our design precludes robust analysis of neural phase. Neural phase is known to play a fundamental role in speech processing (Giraud & Poeppel, 2012; Schroeder et al., 2008), multisensory timing (Kosem, Gramfort, & van Wassenhove, 2014), and the maintenance of ongoing oscillatory dynamics at speech frequencies (Herrmann, Henry, Haegens, & Obleser, 2016; D. M. Simon, Wallace, M. T., 2017), and the inability to analyze phase limits our ability to assess a potentially important processing dynamic contributing to multisensory integration.

The current study suggests several avenues of potential future research. One such approach is examining temporal modulation of neural responses to audiovisual speech across development. Multisensory temporal integration is known to have a developmental trajectory (Hillock-Dunn, Grantham, & Wallace, 2016; Hillock, Powers, & Wallace, 2011), and utilizing a similar approach in children may serve to extend existing findings of reduced audiovisual speech integration in childhood (Kaganovich & Schumaker, 2014; Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014) by determining the degree to which temporal integration sharpens

during maturation. Similarly, given the known deficits in multisensory temporal integration in a number of disorders such as autism spectrum disorder (Stevenson et al., 2014), dyslexia (Hairston, Burdette, Flowers, Wood, & Wallace, 2005), and schizophrenia (Ross, Saint-Amour, Leavitt, Molholm, et al., 2007), extension of this study to these populations may offer to shed light on the nature of dysfunctional multisensory temporal processing. Recent approaches to temporal processing have also elucidated that neural processing of time is highly variable based on existing temporal context (D. M. Simon, Noel, & Wallace, 2017). Given the associations that we establish between temporal acuity and congruity processing, examining trial-by-trial variability may yield important insights into information transfer between sensory systems and performance monitoring circuits in the brain. The use of such approaches may serve to elucidate the integrity and developmental trajectory of multisensory temporal integration in both the typically and atypically developing brain.

References

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Curr Opin Neurobiol, 19*(4), 452-458.

Baart, M. (2016). Quantifying lip-read-induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays. *Psychophysiology, 53*(9), 1295-1306.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci, 20*(8), 2225-2234.

Breska, A., & Deouell, L. Y. (2016). When Synchronizing to Rhythms Is Not a Good Thing: Modulations of Preparatory and Post-Target Neural Activity When Shifting Attention Away from On-Beat Times of a Distracting Rhythm. *J Neurosci, 36*(27), 7154-7166.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J Neurosci,*

30(38), 12572-12580.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences, 18*(8), 414-421.

Cecere, R., Gross, J., Willis, A., & Thut, G. (2017). Being First Matters: Topographical Representational Similarity Analysis of ERP Signals Reveals Separate Networks for Audiovisual Temporal Binding Depending on the Leading Sense. *J Neurosci, 37*(21), 5274-5287.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Curr Biol, 25*(2), 231-235.

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

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am, 25*(1), 975-979.

Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends Neurosci, 37*(9), 480-490.

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *J Acoust Soc Am, 119*(6), 4065-4073.

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *J Neurosci, 35*(42), 14195-14204.

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye Can Hear Clearly Now: Inverse Effectiveness in Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal Integration. *J Neurosci, 36*(38), 9888-9895.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21.

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

Ding, N., & Simon, J. Z. (2013b). Power and phase properties of oscillatory neural responses in the presence of background activity. *J Comput Neurosci, 34*(2), 337-343.

Ding, N., & Simon, J. Z. (2013c). Robust cortical encoding of slow temporal modulations of

speech. *Adv Exp Med Biol, 787*, 373-381.

Folch-Fortuny, A., Arteaga, F., & Ferrer, A. (2016). Missing Data Imputation Toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems, 154*, 93-100.

Ganesh, A. C., Berthommier, F., Vilain, C., Sato, M., & Schwartz, J. L. (2014). A possible neurophysiological correlate of audiovisual binding and unbinding in speech perception. *Front Psychol, 5*, 1340.

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

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res, 166*(3-4), 474-480.

Hanslmayr, S., Pastotter, B., Bauml, K. H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *J Cogn Neurosci, 20*(2), 215-225.

Harrar, V., Harris, L. R., & Spence, C. (2017). Multisensory integration is independent of perceived simultaneity. *Exp Brain Res, 235*(3), 763-775.

Herrmann, B., Henry, M. J., Haegens, S., & Obleser, J. (2016). Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. *Neuroimage, 124*(Pt A), 487-497.

Hillock-Dunn, A., Grantham, D. W., & Wallace, M. T. (2016). The temporal binding window for audiovisual speech: Children are like little adults. *Neuropsychologia, 88*, 74-82.

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia, 49*(3), 461-467.

Iannaccone, R., Hauser, T. U., Staempfli, P., Walitza, S., Brandeis, D., & Brem, S. (2015). Conflict monitoring and error processing: New insights from simultaneous EEG-fMRI. *Neuroimage, 105*, 395-407.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol Sci, 13*(4), 313-319.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol, 111*(10), 1745-1758.

Kaganovich, N., & Schumaker, J. (2014). Audiovisual integration for speech during mid-childhood: Electrophysiological evidence. *Brain Lang, 139C*, 36-48.

Knowland, V. C., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. (2014). Audio-visual speech perception: a developmental ERP investigation. *Dev Sci, 17*(1), 110-124.

Kosem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage, 92*, 274-284.

Kumar, G. V., Halder, T., Jaiswal, A. K., Mukherjee, A., Roy, D., & Banerjee, A. (2016). Large Scale Functional Brain Networks Underlying Temporal Integration of Audio-Visual Speech Perception: An EEG Study. *Front Psychol, 7*, 1558.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron, 64*(3), 419-430.

Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: A theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology, 93*(3), 283-297.

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. *Science, 295*(5555), 690-694.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods, 164*(1), 177-190.

Martin, B., Giersch, A., Huron, C., & van Wassenhove, V. (2013). Temporal event structure and timing in schizophrenia: preserved binding in a longer "now". *Neuropsychologia, 51*(2), 358-371.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys, 58*(3), 351-362.

Munneke, G. J., Nap, T. S., Schippers, E. E., & Cohen, M. X. (2015). A statistical comparison of EEG time- and time-frequency domain representations of error processing. *Brain Research, 1618*, 222-230.

Murray, M. M., & Wallace, M.T. (2012). *The Neural Bases of Multisensory Processes* (M. M. Murray, Wallace, M.T. Ed.). Boca Raton, FL: CRC Press.

Nigbur, R., Ivanova, G., & Sturmer, B. (2011). Theta power as a marker for cognitive interference. *Clinical Neurophysiology, 122*(11), 2185-2194.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat Neurosci, 15*(12), 1729-1735.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci, 2011*, 156869.

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

Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr Clin Neurophysiol, 66*(1), 75-81.

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *J Speech Lang Hear Res, 52*(4), 1073-1081.

Roa Romero, Y., Keil, J., Balz, J., Gallinat, J., & Senkowski, D. (2016). Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. *J Neurophysiol, 116*(3), 1396-1407.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. *Cerebral Cortex, 17*(5), 1147-1153.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C., & Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr Res, 97*(1-3), 173-183.

Samaha, J., & Postle, B. R. (2015). The Speed of Alpha-Band Oscillations Predicts the Temporal Resolution of Visual Perception. *Curr Biol, 25*(22), 2985-2990.

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

Schwartz, J. L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Comput Biol, 10*(7), e1003743.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn Sci, 12*(5), 182-186.

Simon, D. M., Noel, J. P., & Wallace, M. T. (2017). Event Related Potentials Index Rapid Recalibration to Audiovisual Temporal Asynchrony. *Front Integr Neurosci, 11*, 8.

Simon, D.M., Wallace, M. T. (2017). Rhythmic Modulation of Entrained Auditory Oscillations by Visual Inputs. *Brain Topogr*.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci, 19*(12), 1964-1973.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *J Neurosci, 34*(3), 691-697.

Strauss, A., Wostmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Front Hum Neurosci, 8*, 350.

Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Front Integr Neurosci, 9*, 19.

Ten Oever, S., & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proc Natl Acad Sci U S A, 112*(52), 15833-15837.

Ten Oever, S., Sack, A. T., Wheat, K. L., Bien, N., & van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Front Psychol, 4*, 331.

van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron, 81*(6), 1240-1253.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A, 102*(4), 1181-1186.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys, 72*(4), 871-884.

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64C*, 105-123.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*(4), 931-959.

Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". *J Neurosci, 33*(4), 1417-1426.

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

CHAPTER III

THETA POWER AND COHERENCE SUPPORT MULTISENSORY TEMPORAL PROCESSING

The contents of this chapter are drawn from a manuscript in preparation:

Simon, D.M., Wallace M.T., Theta Power and Coherence Support Multisensory Temporal Processing

Abstract

Environmental events emanating from a common source frequently generate sensory inputs in more than one sensory modality. The temporal concordance between these inputs serves as an important cue for the nervous system to appropriately integrate sensory information and form coherent perceptual representations. Multisensory integration is highly flexible and dependent upon contextual cues and task demands, but the neurophysiological basis of this flexibility, particularly in the temporal domain, has not been fully explored. In the current study, we investigated the degree to which top down task demands affect physiological measures of multisensory integration. To do so, we employed a combination of electroencephalography (EEG) and a pair of psychophysical tasks in which participants directly attended or ignored the temporal relationship between the auditory and visual portions of audiovisual speech stimuli.

Consistent with previous work, our results indicate that oscillatory power in the lower portion of the theta band (3.5-5 Hz) is sensitive to the temporal concordance of audiovisual speech events. Importantly, however, this is only the case when temporal relationships are directly attended by participants. We further demonstrate that phase synchronization differs based on task demands. This suggests that selective oscillatory synchronization in the theta band plays an important role in top down regulation of multisensory temporal processing. Our findings thus indicate that low frequency oscillations may encode multisensory temporal information during deliberate processing of temporal structure.

Introduction

Temporal Processing is Fundamental to Multisensory Integration

Events in the environment frequently generate sensory signals in multiple sensory modalities. Integration of these inputs is an important step in forming perceptual representations of the world, and the presence of inputs from multiple senses often conveys numerous behavioral and perceptual benefits (Murray & Wallace, 2012). A salient and ecologically important example of such a multisensory signal is speech, which consists of both acoustic elements as well as visual mouth movements. Integration of these inputs can provide substantial perceptual benefits in term of speech intelligibility (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). An important cue for the appropriate integration of these signals is their relative timing, which has been shown to affect integration at both the behavioral (van Wassenhove, Grant, & Poeppel, 2007) and neurophysiological (D. M. Simon & Wallace, 2017) level. Computing the temporal relationship between stimuli from different sensory modalities thus represents a fundamental step in the

process of multisensory integration. To interrogate the nature of these temporal constraints on multisensory function, researchers have frequently relied on explicitly asking participants to make judgements on the temporal order or simultaneity of stimuli (reviewed in: (Vroomen & Keetels, 2010; Wallace & Stevenson, 2014)). Such questions are highly artificial and ask participants to actively interrogate a stimulus dimension, temporal structure, for which participants have little prior experience and which is rarely actively attended in everyday life. Furthermore, temporal acuity in these tasks varies based on the top down task demands imposed (i.e. synchrony judgement vs temporal order judgment) (Stevenson & Wallace, 2013; van Eijk, Kohlrausch, Juola, & van de Par, 2008). These differences could well be explained by the possibility that these temporal tasks recruit specialized task specific neural networks not observed during 'normal' processing. Examining the nature of these networks may yield important information on how the brain reconfigures multisensory information flow to meet myriad potential task demands.

Task Dependency in Neural Processing

Flexibility in processing is one of the most important features of the brain, and underpins its ability to function in an ever-changing environment. Mechanistically, this flexibility is believed to be rooted in the construction of functional networks (Varela, Lachaux, Rodriguez, & Martinerie, 2001). These networks are constructed in a flexible manner, and examples of this flexible network reconfiguration that serve to meet task demands include inhibitory control (Spielberg, Miller, Heller, & Banich, 2015), and hub based changes in connectivity across tasks (Cole et al., 2013). Recent work has begun to probe the mechanistic nature of these task related networks. Phase coupling between anatomically distinct neural circuits is believed to be an

important contributor to this process (Engel, Gerloff, Hilgetag, & Nolte, 2013; Fries, 2005; Womelsdorf et al., 2007), including during sensory processing (Arnal & Giraud, 2012). Construction of these phase-coupled networks has been shown to occur extremely rapidly in response to task demands, and the fidelity of the constructed network predicts future performance on a trial by trial basis (Phillips, Vinck, Everling, & Womelsdorf, 2014).

Particularly relevant to the current investigation, synchronization in the theta band (4-8 Hz) is known to contribute to the maintenance (Klimesch et al., 2006) and retrieval (Womelsdorf, Vinck, Leung, & Everling, 2010) of choice relevant information, and is believed to play a generalized role in information prioritization, information transfer, and error monitoring (Cavanagh & Frank, 2014; Cooper et al., 2015; Sauseng et al., 2006). Task specific network construction might be particularly important for integration of information across the different senses, which, like processing within a given sensory modality, is known to be flexible in the face of dynamic task directives (van Atteveldt, Murray, Thut, & Schroeder, 2014). Multisensory integration is dependent on neural interactions across anatomically distinct sensory processing circuits, and these interactions are believed to be heavily dependent on neural coherence (Hipp, Engel, & Siegel, 2011; Senkowski, Schneider, Foxe, & Engel, 2008). This cross circuit coherence is thus highly amenable to top down control when enhancing or attenuating information flow is desirable. Multisensory integration during duration judgments has previously been demonstrated to be affected by the degree of neural coherence (van Driel, Knapen, van Es, & Cohen, 2014), formally implicating fluctuations in network strength in trial-by-trial fluctuations in multisensory temporal perception. Top down regulation of neural coherence in low frequencies, which are particularly suitable for long range communication (von Stein & Sarnthein, 2000), and have been proposed as critical to organizing inputs in time (Kosem,

Gramfort, & van Wassenhove, 2014), may thus be an important biological substrate for task dependent modulation of temporal information processing between sensory systems.

Rationale for the Current Study

We previously demonstrated that temporal misalignment between the auditory and visual streams is associated with increased power in the lower portion of the theta band (when compared with the processing of temporally aligned stimuli). Crucially, we also demonstrated that this increased power, which we posited could represent a potential temporal processing signal, corresponded directly with participant task performance (D. M. Simon & Wallace, 2017). While this study strongly suggested that this error signal was directly related to the processing of the temporal structure of the stimuli, it did not determine whether this neural activity is automatic or driven by task demands. In an automatic framework, this signal would be present any time sensory inputs across modalities were temporally misaligned, and would represent an obligatory processing stage evaluating stimulus temporal concordance. Alternatively, as we hypothesize here, and representing the core question of the current study, the error signal could be dependent on participant's attention to the temporal relationship of the auditory and visual sensory streams. We further hypothesized that when task relevant, this signal would synchronize to other cortical regions to transfer the task relevant timing information. To test these hypotheses, we utilized a combination of EEG and two psychophysical tasks in which participants passively viewed or actively attended the temporal structure of audiovisual speech stimuli with varying levels of asynchrony. Our results demonstrate that active processing of multisensory temporal structure recruits both local and distributed theta band neural networks.

Methods and Materials

Participants

Twenty-two typically developing adults participated in the study. All participants reported that they were right handed, had normal or corrected-to-normal vision, and normal hearing. One participant was excluded for analysis due to falling asleep, leaving 21 analyzed participants (11 female) with a mean age of 22 years (SD ± 4.72). The study was conducted in accordance with the declaration of Helsinki, and informed written consent was obtained from all participants. All procedures were approved by the Vanderbilt University institutional review board.

Psychophysical Task

Participants performed two different psychophysical tasks in a blocked design. Task 1 consisted of a speeded two alternative forced choice simultaneity judgement (SJ) regarding an audiovisual speech stimulus. Task 2 consisted of a speeded response (SR) to flashes or beeps embedded within the same speech stimuli. The experimental stimuli consisted of an audiovisual movie of a woman saying the syllable 'BA', including all pre-articulatory movements, with a resolution of 720 x 1280 and a duration of 2000 ms (**Fig 3-1A**). The movie was presented on a 24-inch monitor (ASUS VG248QE) with a refresh rate of 60 Hz at a distance of 1 meter. The woman's face was central on the monitor and occupied an area approximately 12 cm high by 8.5 cm wide (approximately 6.8° x 4.8° of visual angle), while the open mouth occupied an area approximately 1.75 cm high by 3 cm wide (approximately 1° x 1.7° of visual angle). The auditory portion of the movie was presented at normal conversational volume (~65dB) through a pair of bilateral speakers1 meter from the participant's head. Trials began with presentation of a

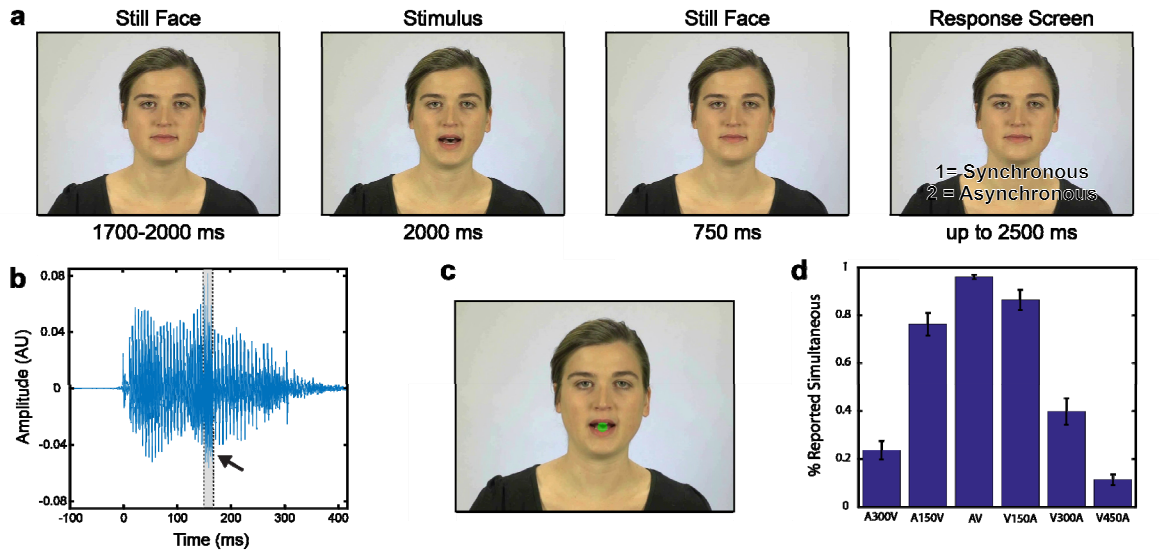

Figure 3-1 Behavioral Task and Behavioral Results

A) Experimental Timeline. Trials began with a 1700-2000 ms period of a still face consisting of the first video frame, followed by the 2 s stimulus movie. Following the movie there was a 750 ms period of additional still face consisting of the last movie frame. During the SJ task if participants had not yet responded a response screen appeared for up to 2500ms. Participants were explicitly told to respond as quickly and accurately as possible and that the response screen was an indicator they were responding too slowly.
B) Example of an auditory catch trial for the SR task. A 16.7 ms burst of Gaussian white noise (gray shading) was inserted beginning 150 ms after stimulus onset.
C) Example of a visual catch trial for the SR task. A 1 cm x 1 cm translucent green circle appeared within the mouth with a duration of 50 ms, beginning 150 ms after stimulus onset when the mouth was fully open.
D) Rate that stimuli were reported as synchronous in the SJ task for each of the six SOAs.

still face consisting of the first video frame for between 1700 and 2000 ms with a uniform distribution. This was followed by the audiovisual movie, with a duration of 2000 ms. Following the movie, a still face consisting of the last video frame was presented for 750 ms.

In task 1 (simultaneity judgment; SJ) participants were instructed to fixate on the mouth and to use their right hand to indicate whether the stimuli were perceived to occur at the same time (i.e., synchronously) or at different times (i.e., asynchronously) via keyboard button press. Participants were explicitly told to respond as quickly and accurately as possible.

In task 2 (speeded response; SR) participants were instructed to fixate on the mouth and to use their right hand to indicate the presence of a green flash or auditory click via keyboard button press. When no flash or click was present, participants were to withhold response. Participants were explicitly told to respond as quickly and accurately as possible.

To create the temporal asynchronies of the stimulus bank, we manipulated the audiovisual stimulus by delaying the visual stimulus (to create an AV trial) or delaying the auditory stimulus (to create a VA trial). We created 5 asynchronies ranging from audition leading vision by 300 ms (A300V) to vision leading audition by 450 ms (V450A) in steps of 150 ms, resulting in a total of 6 conditions, including the original movie featuring synchronized stimuli. For the SR task we additionally created visual and auditory catch trials. For auditory catch trials, we added a 16.7 ms burst of Gaussian white noise at ~75dB to the auditory stimulus (**Fig 3-1B**). For visual catch trials, we added a 50 ms circular green flash 1 cm (approximately 0.6°) in diameter occurring during mouth opening to the visual stimulus (**Fig 3-1C**). We created both auditory and visual catch trials for all 6 temporal conditions (and thus 12 types of catch trials) and weighted their appearance rate to approximately coincide with the weighting of regular trials. The level of asynchrony in the audiovisual speech stimulus was thus not predictive in regards to the presence of absence of the catch stimulus.

As our planned primary analysis involved auditory leading and synchronous stimuli, blocks were weighted to increase the number of trials collected in these conditions. These conditions (A300V, A150V, and AV) each occurred 28 times per block (25.7% of trials per condition). The V150A condition occurred 5 times per block (4.6% of trials), and the V300A and V450A conditions occurred 10 times per block (9.2% of trials per condition). This higher weighting of V300A and V450A served to minimize sustained recalibration (Fujisaki, Shimojo, Kashino, &

Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004) over the course of the experiment. In the SR condition we additionally added 2 visual and 2 auditory catches in the conditions of interest (A300V, A150V, AV) and 1 visual and 1 auditory catch in the visual leading conditions (V150A, V300A, V450A).

Blocks thus consisted of 109 stimuli (SJ) or 127 stimuli (SR) presented in a random order and participants completed 6 blocks of each task, for a total 1416 trials. Block order was random, with the constraint that no more than two repetitions of a given task could occur in a row. Stimulus onset for all stimuli was considered relative to the leading stimulus. That is, for auditory leads stimulus onset was at the time of auditory onset, while for visual leads stimulus onset was the onset of the video frame associated with auditory onset in the original video. These events occurred simultaneously in the synchronous video. In other words, time 0 in our analysis corresponds with the first point at which task relevant information was present. All participants completed a practice block for each of the two tasks before the main experiment, and an example of synchronous speech was re-demonstrated for participants every 2 blocks.

Behavioral Data Analysis

In the SJ task, we excluded any trial in which no response was given. For each SOA we also separately excluded trials in which the response time was more than 2.5 standard deviations greater than an individual participant's mean response time for that SOA. This resulted in the omission of 9.87 (SD ± 3.08) trials per participant, which were also excluded from EEG analysis (see below). We then compiled the reported rate of synchrony and mean reaction time for each participant separately for each SOA. For the SR task, responses were examined in terms of hit rate and false alarm rate to confirm that participants actively attended the task.

EEG Acquisition and Pre-processing

EEG acquisition and processing was performed in line with previous work (D. M. Simon & Wallace, 2017). Continuous EEG was recorded from 128 electrodes referenced to the vertex (Cz) using a Net Amps 400 amplifier and Hydrocel GSN 128 EEG cap (EGI systems Inc.). Data were acquired with NetStation 5.3 with a sampling rate of 1000 Hz and were further processed using MATLAB, EEGLAB (Delorme & Makeig, 2004), and fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). Continuous EEG data were band-pass filtered from 0.15 to 50 Hz with a 6 dB roll-off of 0.075 to 50.075 Hz using the EEGLab firfiltnew.m function, which implements a bi-directional zero-phase finite impulse response filter. Epochs 3s long from 1000 ms before to 2000 ms after onset of the first stimulus were then extracted. Artifact contaminated trials and bad channels were identified and removed through a combination of automated identification of trials in which any channel exceeded ± 100 μV and rigorous visual inspection. Data were then recalculated to the average reference, reduced to 64 dimensions using principal component analysis (PCA) and submitted to independent component analysis (ICA) using the Infomax algorithm (Jung, Makeig, Humphries, et al., 2000; Jung, Makeig, Westerfield, et al., 2000) (0.5E-7 stopping weight, 768 maximum steps). Lastly, bad channels were reconstructed using spherical spline interpolation (Perrin, Pernier, Bertrand, Giard, & Echallier, 1987) and data were re-inspected for residual artifacts. Overall a mean of 1154 (81.8% SD ± 10.0%) of trials were retained, while 4.43 (SD ± 1.80) channels and 6.66 (SD ± 2.39) Independent components were removed per participant. For the SJ task, we also excluded trials using the same behavioral criteria used for behavioral data analysis. For EEG analysis of the SR task, we excluded any trial in which participants committed a false alarm by responding when there was neither a flash nor a

noise burst.

All EEG analysis was focused on the auditory leading conditions (A300V and A150V) and true synchrony (AV) for which we collected a substantial number of trials (see below). A 3 (SOA) x 2 (task) ANOVA for number of trials retained in these conditions indicated that there was no difference in number of trials retained per task ($F_{1,20} = 0.0302$, $p = 0.8637$), but that trials were rejected slightly more often in the AV and A150V conditions than in the A300V condition ($F_{2,40} = 4.366$, $p = 0.0193$). In practice, this indicated the highest value of 138.8 trials per participant were retained in the A300V condition per task, while the lowest value of 135.3 trials were retained in the AV condition per task. We believe this ~2.5% difference in trial retention rate is unlikely to contribute to differences in our results, and note that we use connectivity metrics that compensate for sample size bias.

Time-Frequency Analysis

We employed time-frequency analysis to examine EEG activity in the frequency domain. Time Frequency decomposition of single trial EEG data was accomplished using convolution with Morlet wavelets with frequencies from 3.5 to 35 Hz in 0.5 Hz steps. Wavelets had 2.5 cycles at the lowest frequency rising to 9.3 cycles at the highest and convolution was performed with a temporal resolution of 10 ms. Power was then baseline corrected relative to the -600 to -200 ms pre stimulus period and decibel transformed. Frequency bands were selected for further analysis *a-priori* based on previous results from a separate experiment utilizing physically identical stimuli. We selected the low theta band (3.5 – 5 Hz) and an analytical window of 300-500ms due to our previous demonstration that power in this band is related to perception of asynchrony during simultaneity judgment (D. M. Simon & Wallace, 2017). We further selected the canonical

theta band (4-8 Hz) and an analytical window of 50-250ms to rule out temporal blurring of the initial auditory cortical response as a primary contributor to our results. We then averaged power within time-frequency ROIs and contrasted tasks and conditions using a 3 (SOA) x 2 (task) repeated measures ANOVA. For topographic representations of multiple electrodes, we corrected for multiple comparisons with FDR using $q = 0.05$ and display only electrodes meeting this criterion.

Phase Coupling Analysis

For connectivity analysis we computed the de-biased version of the weighted phase lag index (WPLI) (Vinck, Oostenveld, van Wingerden, Battaglia, & Pennartz, 2011) between all electrode pairs using data from 150-649 ms after stimulus onset (400ms ± 250ms, equivalent to the center of the time-frequency window for power). The WPLI is uniquely suited to low frequency connectivity in relative proximity to evoked responses as it strongly discounts spurious connectivity attributable to volume conduction, thus monotonically representing true neural coherence (Ewald, Aristei, Nolte, & Abdel Rahman, 2012; Haufe, Nikulin, Muller, & Nolte, 2013; Vinck et al., 2011). Data was Hann windowed before transformation to the frequency domain via FFT and thus had a frequency smoothing of ± 2Hz. As de-biased WPLI values are not normally distributed and equivalent to squared WPLI values, we used fisher's Z transform on the square root of de-biased WPLI values to make them approximately normal and used non-parametric statistical testing. The de-biasing procedure may also result in coherence value below zero, and we replaced all negative coherence values with zeros before fisher Z transforming the data. For analysis of connectivity, we then selected only the six right central electrodes showing both a significant FDR corrected main effect of SOA and a significant FDR corrected task x

SOA interaction effect (E80, E104, E105, E110, E111, E129). Connectivity was averaged across these sensors to improve signal to noise ratio. These connectivity values were then submitted to randomization testing with cluster based correction for multiple comparisons (Maris & Oostenveld, 2007). Using randomization testing, we tested for main effects and interaction effects separately by either averaging across SOAs to test for the main effect of task, averaging across tasks to test for the main effect of SOA, or subtracting data in the SR task from the SJ task (SJ – SR) at each SOA individually to test for an interaction. For the main effect of task, we used a dependent samples t-test to determine cluster inclusion. For the main effect of SOA and the interaction effect, we used the dependent samples F multivariate test found in FieldTrip (http://www.fieldtriptoolbox.org/) (Oostenveld et al., 2011) to determine cluster inclusion. In all cases we set a cluster inclusion threshold of alpha = 0.05, a permutation significance of alpha = 0.05, and used 10,000 randomizations. We note that this analysis was hypothesis based, and we initially selected a singular frequency of interest (4 Hz) and time window which best approximated our *a-priori* analytical window for power. In the interest of being thorough, we also analyzed connectivity using canonical frequency bands for theta (4-8 Hz), alpha (9-13 Hz), low beta (13-18 Hz) and high beta (18-30 Hz). Values within significant clusters were then averaged and subjected to 3 (SOA) x 2 (task) repeated measures ANOVA.

We then sought to infer the directionality of information transfer between connected electrodes. To do so we calculated the phase relationship between each cluster and the six right central electrodes. Phase difference was calculated by subtracting, for each possible electrode pairing and trial, the phase at the central electrode from the phase at the temporal or occipital cluster (i.e. $\Delta\theta = \theta_{cluster} - \theta_{central}$). For example, connectivity between clusters with six and nine electrodes would result in 54 phase angle differences per trial. We then took the weighted

circular mean of these values using the absolute value of the sine of the phase angle difference as the sample weighting (i.e. sample weight = $|\sin(\Delta\theta)|$). Weighting these values serves to weight against volume conduction effects, which drive $\Delta\theta$ towards 0° or 180° (Nolte et al., 2004) and is analogous to the weighted calculation performed in the WPLI (Vinck et al., 2011). We took this weighted circular mean across all trials, excluding self-pairings, as the mean phase angle difference for the participant and cluster pairing in each condition. Phase angle differences across participants were statistically analyzed using the circular statistics toolbox (Berens, 2009). Lags and leads less than 90° ($\pi/2$ rad) yield relatively robust information about directionality, while lags and leads between 90° and 180° are interpreted slightly more cautiously. We also calculated the mean resultant vector length (MRVL) across participants, which indicates the consistency of leads and lags between individuals. A MRVL of 1 indicates perfect consistency across participants in terms of cluster phase relationships, while an MRVL of 0 indicates a uniform circular distribution of phase relationships.

Results

Behavioral Results

We began our analysis by calculating the rate of reported synchrony and reaction time for each of the 6 SOAs used in the experiment. As anticipated, synchrony was reported most frequent for truly synchronous stimuli and decreased with increasing SOAs in either direction (**Fig 3-1D**). While response times in the SJ task are not of direct interest relative to our experimental hypotheses they replicated previous results, in which simultaneity judgement is fastest for synchronous speech stimuli (mean RTs ± SEM for A300V 839 ± 37.6 ms; A150V 792.6 ± 34.7

ms; AV 645.4 ± 35.7 ms; V150A 741.3 ± 35.3 ms; V300A 825.4 ± 37.1 ms; V450A 810.6 ± 37.2 ms. A 1 x 6 SOA repeated measures ANOVA for response time had a significant main effect of SOA ($F_{5,100} = 24.95$, $p = 3.038 \times 10^{-16}$). We also compiled hit and false alarm rates for the SR task, which indicated that participants performed at ceiling for catch trials in both sensory modalities. Participants had a mean hit rate of 99.64% (SD ± 0.96%, all participants > 96%) for auditory catches and a mean hit rate of 99.38% (SD ± 1.35%, all participants > 94%) for visual catches. The false alarm rate (i.e. pressing a response when neither type of catch was present) was extremely low, occurring on 0.06% of trials (SD ± 0.12%, all participants < 0.5%). Participants thus performed both tasks correctly and with a high degree of fidelity.

Effects of Temporal Structure on Theta Oscillations Are Task Dependent

We next analyzed whether oscillations in the lower theta band (3.5 – 5 Hz) were stronger for large asynchronies as we previously demonstrated (D. M. Simon & Wallace, 2017)**,** and whether this effect required attention to temporal configuration. To do so, we analyzed neural responses to the A300V, A150V, and AV conditions only, as the number of trials was strongly weighted towards these conditions (see methods). We averaged time-frequency representations (**Fig 3-2 and Fig 3-3**) from 300 – 500 ms and from 3.5 – 5 Hz at electrode Cz and performed a 3 (SOA) x 2 (task) repeated measures ANOVA. Electrode Cz was selected as we did not have a strong hypothesis regarding lateralization of effects and Cz is well positioned to capture the effects we previously reported.

We found that at this location there was a significant main effect of SOA ($F_{2,40} = 21.167$, $p = 5.368 \times 10^{-7}$), no main effect of task ($F_{1,20} = 3.887 \times 10^{-6}$, $p = 0.9984$), and a significant interaction ($F_{2,40} = 7.887$, $p = 0.0013$) (**Fig 3-4**). We then performed follow-up t-tests which indicated that these effects occurred because low theta power during the SJ task was greater in the A300V condition than in the A150V ($t_{20} = 4.845$, $p = 9.815 \times 10^{-5}$) or AV ($t_{20} = 5.658$, $p = 1.545 \times 10^{-5}$) conditions. Differences in theta power between A150V and AV did not reach significance ($t_{20} = 1.675$, $p = 0.1095$). Theta power did not differ across asynchronies in the SR task (all $t < 1.468$, all $p > 0.1577$). Pairwise comparisons between the two tasks indicated that

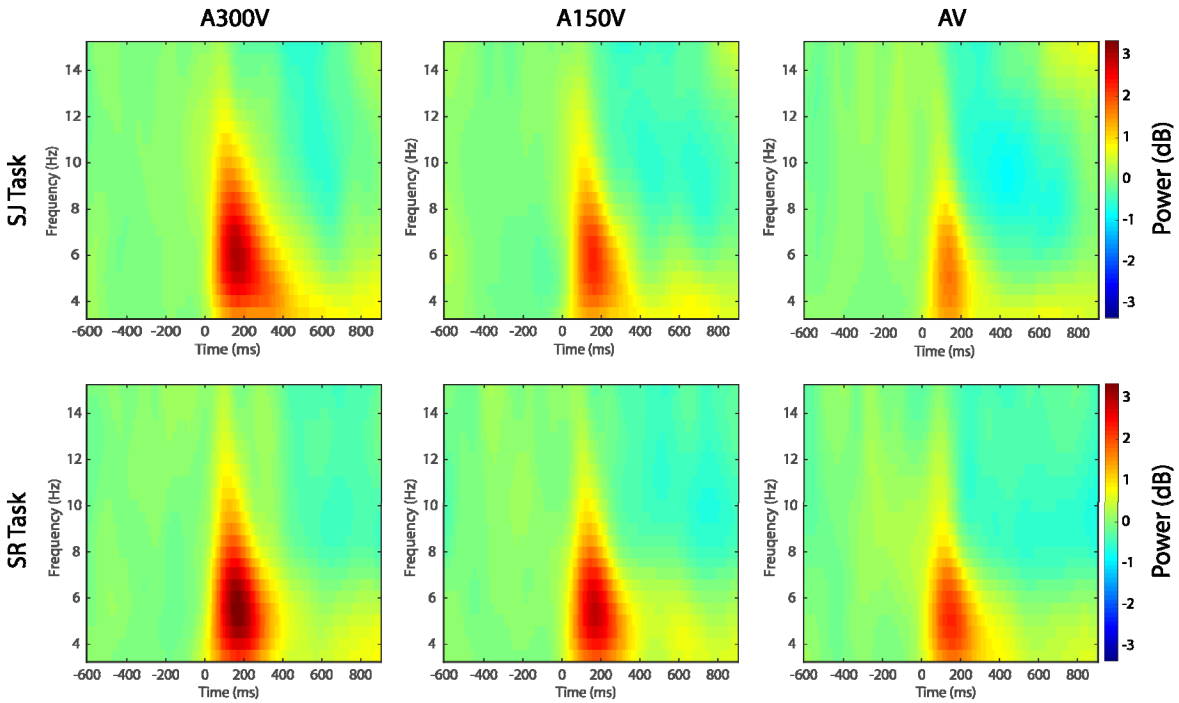

Figure 3-2 Time-Frequency Representations for Electrode Cz
Tasks (SJ and SR) are displayed as rows, while SOAs (A300V, A150V, and AV) are displayed as columns.

- Top Row: Time-frequency representations for the A300V, A150V, and AV conditions during simultaneity judgment. Note the greatest power in the theta band is seen in the most asynchronous condition (A300V).
- Bottom Row: Time-frequency representations for the A300V, A150V, and AV conditions during the speeded response task.

theta power was higher in the SJ task for the A300V condition (t_{20} = 2.4305, p = 0.0246), did not differ in the A150V condition (t_{20} = -0.5435, p = 0.5928) and was lower for SJ than SR in the synchronous (AV) condition (t_{20} = -2.8805, p = 0.0092). These results indicate that actively attending to the temporal structure of the stimuli is required for activity in the low theta band to change with the level of audiovisual asynchrony in the stimulus.

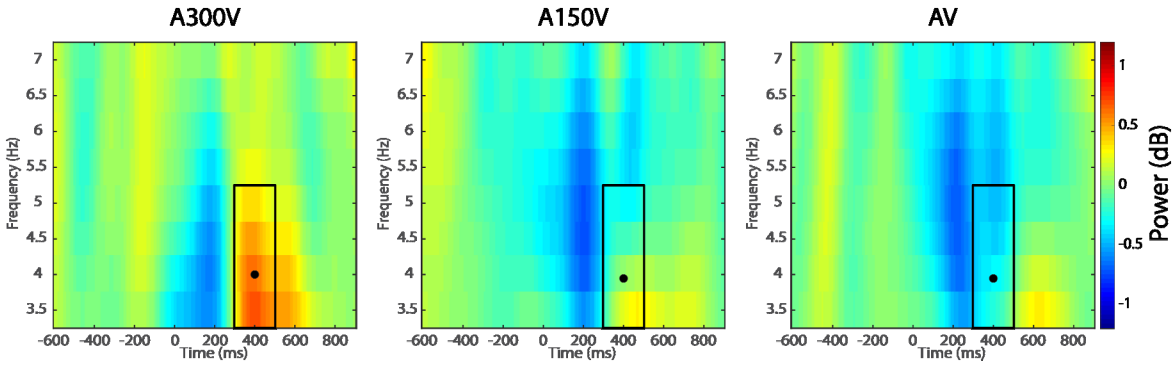

Figure 3-3 Time Frequency Contrasts for Electrode Cz
Expanded (3.5-7 Hz) difference for electrode Cz between time-frequency representations (SJ – SR). Black box indicates the 3.5 – 5 Hz and 300 – 500 ms window selected for further power analysis. Black dot indicates the time-frequency seed used for connectivity analysis. Note the striking difference in power within this frequency band and temporal epoch between tasks, most notably in the A300V condition.

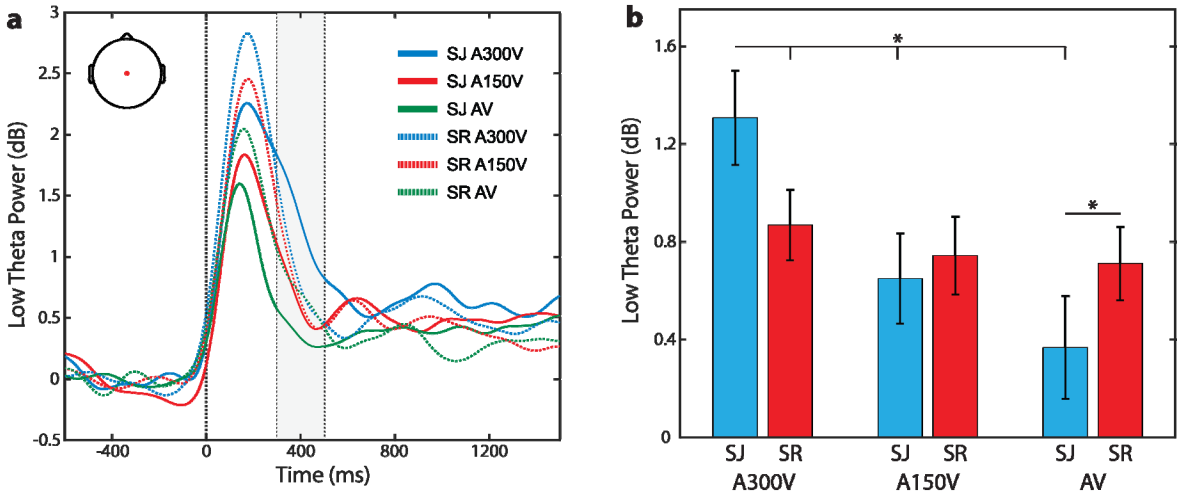

Figure 3-4 Low Theta Power for Electrode Cz
Power in the low (3.5 – 5 Hz) theta band as a function of task and temporal asynchrony

A) Time course of low theta power at electrode Cz (inset depicts electrode location) for the A300V, A150V, and AV conditions during both tasks. The grey shaded region indicates the 300 - 500ms analytical window.
B) Low theta power for the SJ task (blue) and SR task (red) in the 300 – 500 ms window for each of the three levels of asynchrony. Note that power is greatest for the largest asynchrony in the SJ task, and fails to differ based on temporal structure for the SR task.

As a control analysis to determine if these effects could be due to temporal smearing of the initial auditory response, we calculated power in the canonical theta band (4 – 8 Hz) using a time window of 50 – 250 ms. This window captures both the canonical N1 and P2, and which have previously been shown to be modulated by temporal structure (D. M. Simon & Wallace, 2017). A 3 (SOA) x 2 (task) ANOVA on theta power in this window indicated there was a significant main effect of SOA ($F_{2,40} = 45.702$, $p = 4.667 \times 10^{-11}$), a main effect of task ($F_{1,20} = 10.620$, $p = 0.0039$), and no interaction effect ($F_{2,40} = 0.9107$, $p = 0.4104$) (**Fig 3-5**). Consistent with our previous work, follow-up t-tests indicated that for the SJ task the main effect of SOA was attributable to graded reductions in theta power as the visual stimulus approached true synchrony (A300V vs A150V $t_{20} = 4.875$, $p = 9.17 \times 10^{-5}$; A300V vs AV $t_{20} = 7.649$, $p = 2.315 \times 10^{-7}$; A150V vs AV $t_{20} = 3.431$, $p = 0.0026$). Similar results were present in the SR task (SR task

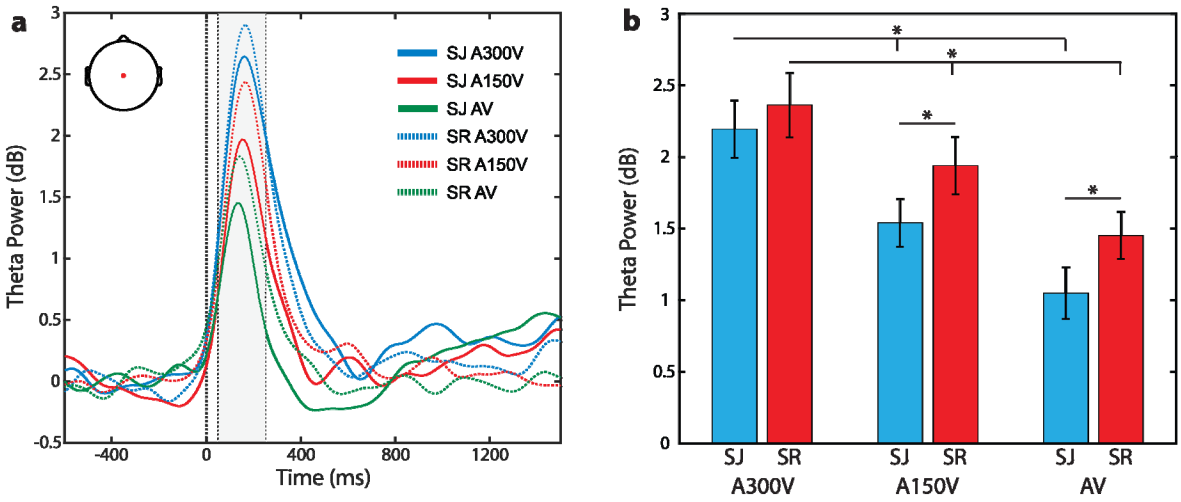

Figure 3-5 Canonical Theta Power for Electrode Cz
Power in the canonical (4 – 8 Hz) theta band as a function of task and temporal asynchrony.

A) Time course of canonical theta power at electrode Cz for the A300V, A150V, and AV conditions during both tasks. The grey shaded region indicates the 50 - 250ms analytical window.
B) Theta power for the SJ task (blue) and SR task (red) in the 50 – 250 ms window for each of the three levels of asynchrony.

A300V vs A150V $t_{20} = 2.832$, $p = 0.0103$; A300V vs AV $t_{20} = 5.8082$, $p = 1.106 \times 10^{-5}$; A150V vs AV $t_{20} = 3.358$, $p = 0.0031$). Pairwise comparisons also indicated that power in the SR task was higher than for the SJ task in the A150V and AV conditions (A300V SJ vs SR $t_{20} = -0.8726$, $p = 0.3928$; A150V SJ vs SR $t_{20} = -3.42$ $p = 0.0027$; AV SJ vs SR $t_{20} = -2.984$, $p = 0.0073$). We repeated this process using only the 3.5 – 5 Hz band as the larger wavelet size at these frequencies might allow temporal blurring unnoticed in the full theta band analysis. This analysis yielded similar results to the full theta band; a main effect of SOA ($F_{2,40} = 15.229$, $p = 1.209 \times 10^{-5}$), a main effect of task ($F_{1,20} = 18.743$, $p = 3.258 \times 10^{-4}$), and no interaction effect ($F_{2,40} = 0.1086$, $p = 0.8974$). Crucially, this early theta power was centrally distributed, which is consistent with auditory cortical generators, and did not demonstrate a SOA x task interaction. The effects we observe in the primary low theta analysis thus cannot be attributed to temporal

blurring of early auditory cortical responses in the frequency domain.

We then extended this procedure to all electrodes by performing an individual 3 (SOA) x 2 (task) repeated measures ANOVA at each electrode for power averaged from 3.5 – 5 Hz and for

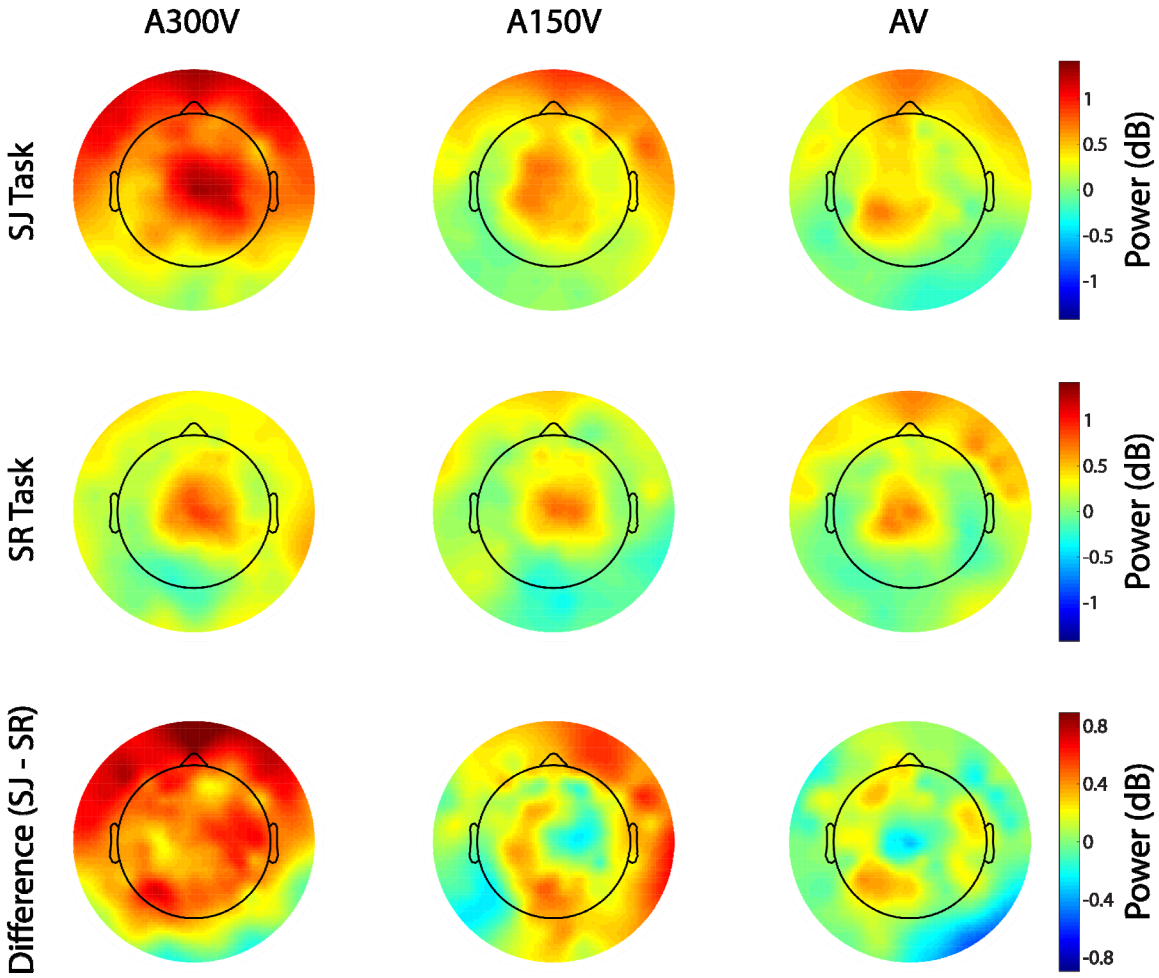

Figure 3-6 Low Theta Power as a Function of Spatial Location
Spatial representations of low theta power (3.5 – 5 Hz) across task and SOA for the interval between 300 – 500 ms. Tasks are displayed in rows, while SOAs are displayed in columns.

- Top Row: Spatial representation of low theta power for the A300V, A150V, and AV conditions during simultaneity judgment.
- Middle Row: Spatial representation of low theta power for the A300V, A150V, and AV conditions during the speeded response task.
- Bottom Row: Spatial representation of the difference in low theta power between conditions (SJ – SR).

the time interval between 300-500 ms. We corrected for multiple comparisons using FDR with q = 0.05 and display topographies for both power (**Fig 3-6**) and F values (**Fig 3-7**). A large number of electrodes with a right central distribution displayed a significant main effect of SOA ($p^{FDR} < 0.05$, 42 contiguous electrodes in the primary cluster). Additionally, a smaller subset (6 contiguous electrodes) with a right central distribution displayed a significant interaction ($p^{FDR} < 0.05$). Both the main effect of SOA and the interaction thus presented at a right lateralized location atypical of initial auditory cortical responses.

Functional Connectivity in the Theta Band Differs Across Tasks

We next examined whether connectivity at low theta frequencies differed between tasks or SOAs. We selected the contiguous right central electrodes showing significant FDR corrected interaction effects and calculated their weighted phase lag index (WPLI) functional connectivity centered at 4 Hz and 400±250 ms after stimulus onset (**Fig 3-8**). These values correspond with our analytical window for power and encompass the frequency bands of interest after accounting for the frequency smoothing of the Hann Window. We then performed spatial randomization

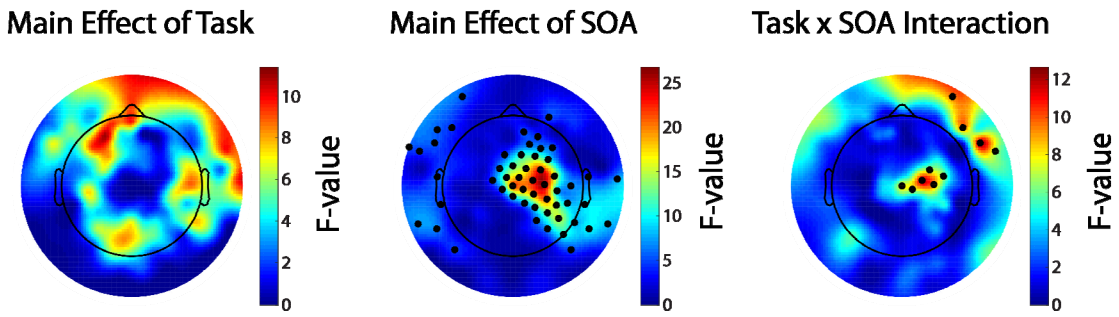

Figure 3-7 Statistical Contrasts as a Function of Spatial Location
Spatial representation of the 3 (SOA) x 2 (task) repeated measures ANOVA for low theta power (3.5 – 5 Hz) averaged over interval of 300 – 500 ms. Colors indicate the F-Value for the statistical contrast at each spatial location. Black dots indicate electrodes with significant effects after correction for multiple comparisons ($p < 0.05$, FDR corrected).

testing for main effects of task and SOA, and for an interaction. For the main effect of task, we found two bilaterally placed significant clusters. The left cluster ($p = 0.0418$) consisted of 10 temporal sensors, while the right cluster ($p = 0.0486$) consisted of 9 temporal sensors. We found no clusters displaying a main effect of SOA. A single cluster was found when testing for the interaction effect ($p = 0.0320$), which consisted of 14 electrodes over right posterior scalp. Cluster locations are depicted as insets in **figure 3-9**. For completeness, we repeated this process for canonical oscillatory frequency bands (see methods) and found no effects for any of these bands (all $p > 0.1$). To better quantify values within the identified clusters we then separately averaged connectivity across electrodes within each cluster and performed 3 (SOA) x 2 (task) repeated measures ANOVA.

As anticipated, these ANOVAs were consistent with the results of the permutation testing (**Table 1**). We then performed pairwise T-tests using each cluster's theta connectivity values

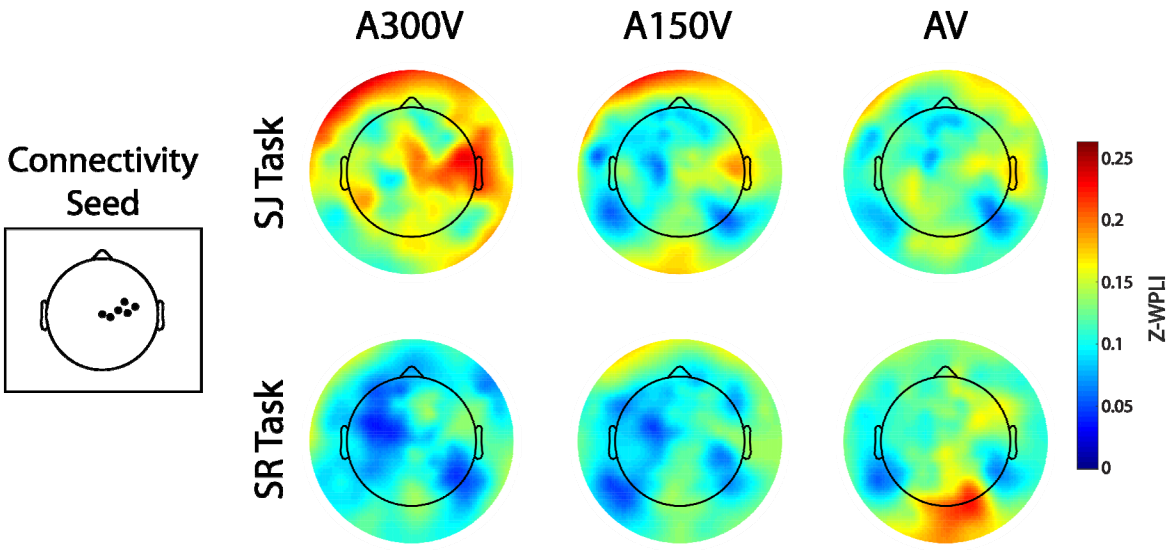

Figure 3-8 Functional Connectivity as a Function of Spatial Location
Topographic representations of 4 Hz synchronization to the right central electrode cluster (inset left; this cluster represents the connectivity seed) for the SJ (top) and SR (bottom) tasks and for the 3 SOAs. Warmer colors indicate increased functional connectivity.

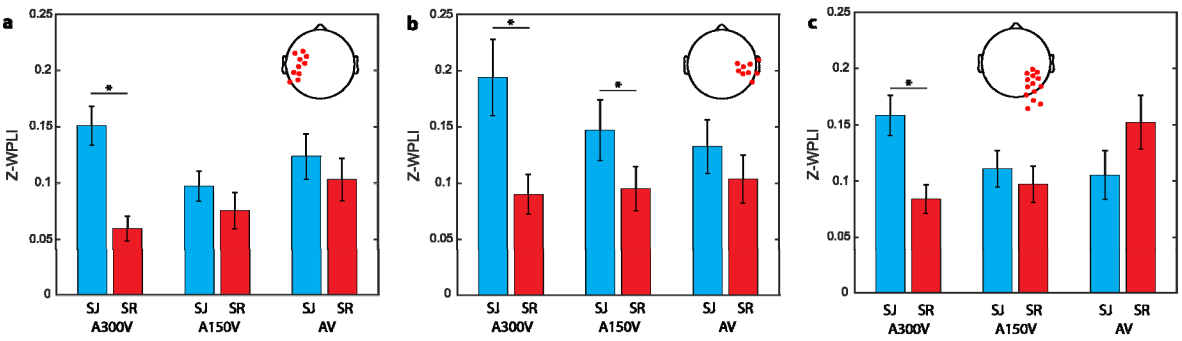

Figure 3-9 Low Theta Connectivity by Task and Cluster

A) Low theta connectivity averaged across the left temporal electrode cluster (inset) for the two tasks and three SOAs.
B) Low theta connectivity averaged across the right temporal electrode cluster (inset) for the two tasks and three SOAs.
C) Low theta connectivity averaged across the right posterior electrode cluster (inset) for the two tasks and three SOAs.

comparing the two tasks (**Fig 3-9**). In the left temporal cluster theta connectivity was higher in the SJ task for A300V ($t_{20} = 4.351$, $p = 3.0996 \times 10^{-4}$), but not in the A150V ($p = 0.304$) or AV conditions ($p = 0.492$). In the right temporal cluster theta connectivity was higher in the SJ task for A300V ($t_{20} = 3.7471$, $p = 0.0013$) and A150V ($t_{20} = 2.168$, $p = 0.0424$), but not in the AV condition ($p = 0.4465$). These main effects represent elevated theta synchronization between temporal sensors and right central electrodes when the stimulus was highly or mildly asynchronous during simultaneity judgment. In the right posterior interaction cluster connectivity was stronger for SJ in the A300V condition ($t_{20} = 3.9565$, $p = 7.7886 \times 10^{-4}$), but was not significantly weaker for SJ in the AV condition ($t_{20} = -1.7548$, $p = 0.0946$). We highlight that theta synchronization between this cluster and right central sensors is elevated for asynchronous stimuli during SJ and somewhat elevated for synchronous stimuli during SR, which yields the interaction effect due to inverted trajectories.

Table 0-1 ANOVAs for Phase Synchrony

Cluster	ANOVA Contrast	F	df	p-value
Left Temporal	Main effect of task	**16.649**	1,20	**5.831×10^{-4}**
	Main effect of SOA	1.858	2,40	0.1691
	SOA x Task Interaction	2.39	2,40	0.1406
Right Temporal	Main effect of task	**9.667**	1,20	**0.0055**
	Main effect of SOA	0.6158	2,40	0.5453
	SOA x Task Interaction	1.9164	2,40	0.1604
Right Occipital	Main effect of task	0.7837	1,20	0.3865
	Main effect of SOA	1.0613	2,40	0.3555
	SOA x Task Interaction	**8.772**	2,40	**6.938×10^{-4}**

To infer directionality, we then computed weighted mean phase lags between the right central cluster and the three identified clusters for each participant. For the SJ task, we restricted this analysis to the A300V and A150V conditions as these conditions presented elevated connectivity using the non-directional WPLI measure (see above ANOVAs and follow up t-tests). We also only depict relationships for which a cluster had more connectivity than in the SR task (A300V all three clusters, A150V only the right cluster). For the SR task we only performed this analysis in the synchronous condition and for the occipital cluster only, as this was the only elevated connectivity in this task as part of the interaction cluster. Phase relationships for each cluster are presented in **Table 2**.

Table 0-2 Cluster Phase Relationships

Task & Condition	Target Cluster	Phase Angle re: central	Time Difference (ms)	MRVL
SJ A300V	L. Temporal	2.5784 rad (147.733°)	22.4 lag (or 102.6 lead)	0.7598
	R. Temporal	0.2437 rad (13.965°)	9.7 lead	0.9764
	R. Occipital	-0.5256 rad (-30.116°)	20.9 lag	0.6997
SJ A150V	R. Temporal	0.2016 rad (11.557°)	8 lead	0.9760
SR AV	R. Occipital	-0.8156 rad (-46.733°)	32.5 lag	0.7111

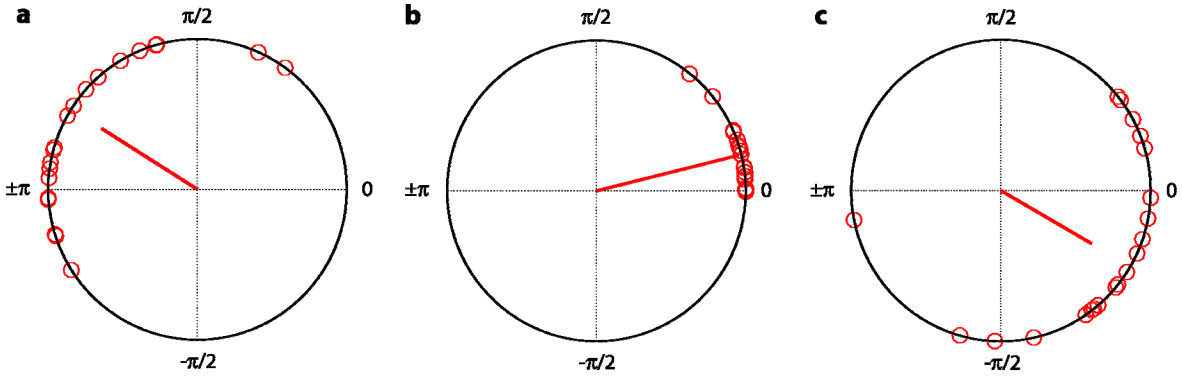

Figure 3-10 Low Theta Phase Lag in the A300V Condition and SJ Task
Red circles indicate data for individual participants, while the red vector indicates the direction and concentration of phases.

A) Weighted phase angle difference between the left temporal cluster and the connectivity seed in the A300V condition and SJ task.
B) Weighted phase angle difference between the right temporal cluster and the connectivity seed in the A300V condition and SJ task.
C) Weighted phase angle difference between the occipital cluster and the connectivity seed in the A300V condition and SJ task.

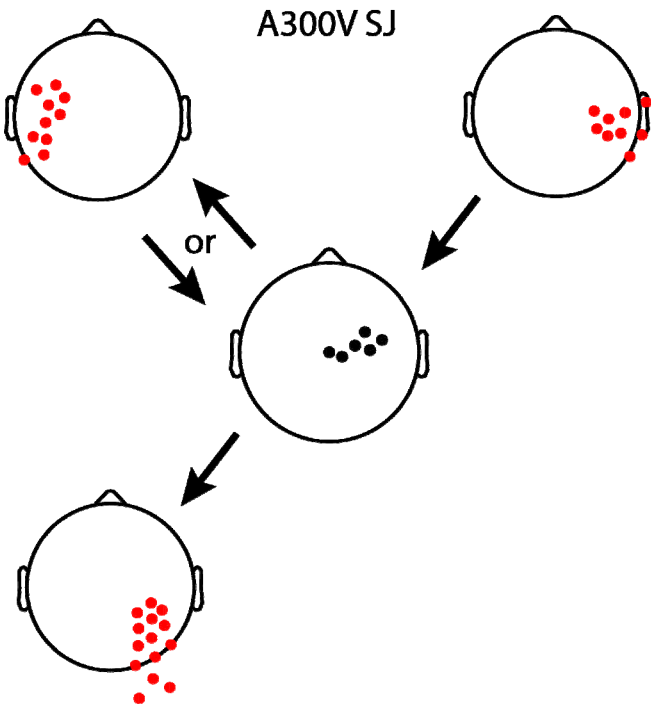

Figure 3-11 Functional Low Theta Network for the A300V Condition and SJ task
Representation of the timing and direction of information flow based on weighted relative 4 Hz phase for the A300V SJ task. Arrows point from the leading to the lagging cluster for each pairing. Phase relationships were found to be similar in the other conditions (see table 3-2).

In the A300V conditions all three of these distributions proved to be highly non-uniform when tested with a Rayleigh test (left main effect cluster Z = 12.123, p = 6.955 x 10^{-7}; right main effect cluster Z = 20.021, p = 8.733 x 10^{-14}; right occipital interaction cluster Z = 10.282, p = 9.074 x 10^{-6}) (**Fig 3-10**). We present a depiction indicating directionality of information transfer in the A300V condition in **figure 3-11** and note that the directionality was identical in the other conditions examined.

In the A150V condition the right temporal cluster was found to be non-uniform (Z = 20.005, p = 9.212 x 10^{-14}) and to lead the central cluster by 0.2016 rad (11.557°), which is equivalent to an 8 ms lead (MRVL = 0.9760). In the AV condition and SR task the phase relationship between the central and occipital clusters were also non-uniform, as expected (Z = 10.62, p = 5.753 x 10^{-}

[6]). A Watson-Williams multi-sample test for equal means did not indicate that this lag was different from the 20.9 ms lag in the A300V SJ task ($F_{1,40} = 1.19$, $p = 0.2819$).

Discussion

We demonstrate that power in the lower theta band varies with the degree of temporal asynchrony between the auditory and visual components of audiovisual speech stimuli. When this temporal disparity is large, theta power is high, and vice versa. Importantly, however, we also demonstrate that these differences occur only when the temporal structure of the stimuli is actively attended during simultaneity judgment. This suggests that this power modulation is reflective of an active temporal processing. Using connectivity analysis, we further demonstrate that neural circuits exhibiting power differences couple to temporal sensors in a task dependent manner. Connectivity to posterior and more putatively visual areas was also modulated differently by asynchrony in the two tasks. These findings serve as evidence that the multisensory error signal we previously reported is only processed during active attending of temporal structure. They also begin to offer tantalizing insights into how timing information might flow between cortical regions during actively attended temporal processing.

Theta Oscillations Index Neural Processing of Temporal Structure

Our primary finding is that low theta (3.5 – 5 Hz) oscillations are enhanced in the presence of temporal mismatch between the auditory and visual components and audiovisual speech. When the streams are substantially out of temporal alignment, there is a relatively robust increase in theta power compared to passive viewing at right central electrodes. Crucially, we further show

that these oscillations are actively suppressed below passive viewing levels when the temporal structure of the paired audiovisual stimuli is synchronous. The right lateralized topography of theta power differences also serves to differentiate them from the initial auditory cortical response despite the limitations in temporal resolution inherent to low frequency wavelet transforms. Taken together we believe that these findings reinforce and extend our previously proposed hypothesis that power in the lower portion of the theta band indexes cognitive processes associated with the identification of temporal mismatch between the auditory and visual speech streams (D. M. Simon & Wallace, 2017).

Previous work in error monitoring has primarily identified cingulate and frontal cortex as the locus of error signals. We note that the right central distribution of this activity in our experiment and our previous work (D. M. Simon & Wallace, 2017) is somewhat incongruent with the expected fronto-central distribution typically associated with these cortical generators (Cavanagh & Frank, 2014). It is, however, consistent with previous work examining perceived simultaneity for audiovisual speech, which demonstrated right lateralization in perception-based changes in the N1 component (Huhn, Szirtes, Lorincz, & Csepe, 2009). This suggests that the lateralization of the observed response may be a marker of hemispheric dominance for this type of processing. We believe this may also occur because other cortical areas may be involved, leading to a complexity in the scalp topographies that mask the precise localization of generators. The posterior temporal sulcus, for example, has been suggested to be an important hub for audiovisual temporal processing (Stevenson, Altieri, Kim, Pisoni, & James, 2010; Wallace & Stevenson, 2014) and projects oscillatory power quite broadly to the scalp. For example, (Schepers, Schneider, Hipp, Engel, & Senkowski, 2013) examined beta oscillations originating from STS in response to audiovisual speech stimuli and found a rather broad scalp projection.

Recent work has also elucidated the importance of low frequency coherence involving motor cortex in audiovisual speech integration (Park, Kayser, Thut, & Gross, 2016), and this might similarly contribute to the observed scalp topography. Previous work has shown that relative temporal information can be found directly in auditory and visual cortex in the form of very low frequency (1 Hz) phase (Kosem et al., 2014). Together these findings suggest that a broad network may be recruited for computation of multisensory temporal error, leading to a complex scalp projection.

Theta Connectivity as a Mechanism for Cognitive Control of Multisensory Processing

Our results further begin to probe how these potential temporal structure-monitoring circuits are selectively synchronized with other brain regions to support temporal processing. During active temporal processing, phase synchronization to bilateral temporal regions was elevated when the stimuli were physically asynchronous relative to what was seen during passive viewing. In contrast, for objectively synchronous stimuli, no differences in phase synchronization were present. Importantly, this change in synchrony is unique to the low theta band and strongly consistent with the explanation that cortical regions registering temporal structure transfer information by coupling at low frequencies when temporal structure is task relevant. We also demonstrate that right lateralized connectivity to posterior, and thus more putatively visual sensors, is regulated in opposite directions by SOA depending on the current task. During the SR task, phase coupling was highest for synchronous stimuli, consistent with the notion that networks processing auditory and visual inputs align their phase relationships when stimuli are being integrated. During the SJ task, however, this pattern is reversed, and phase synchrony is strongest for asynchronous stimuli. We believe this reversal is indicative of the flexible role theta

band phase coupling plays in shaping information transfer to meet task demands (Cavanagh & Frank, 2014; Cohen, 2014). It is also consistent with previous work showing that top down factors can shape phase synchrony between brain areas (Hanslmayr et al., 2012).

Task dependent low frequency synchronization as a mechanism for top down control of multisensory temporal processing has important implications for the study of populations with diminished or dysfunctional multisensory temporal integration. One example of such a population is schizophrenia, in which multisensory temporal deficits are well documented (Hass et al., 2017; Parsons et al., 2013) and diminished theta oscillations have been observed during multisensory processing (Roa Romero, Keil, Balz, Gallinat, & Senkowski, 2016). Disrupted neural synchrony is present over a broad range of contexts in schizophrenia (Uhlhaas & Singer, 2010), suggesting that temporal processing dysfunctions may arise from dysfunctional long range theta synchronization. Another example of a clinical population with altered multisensory temporal processing is individuals with autism spectrum disorder (ASD), who have a diminished ability to distinguish asynchronous speech from synchronous speech (Stevenson et al., 2014). In both of these populations, flexibility in neural processing may be impaired (Akar, Kara, Latifoglu, & Bilgic, 2016; Catarino, Churches, Baron-Cohen, Andrade, & Ring, 2011; D. M. Simon, Damiano, C.R., Woynaroski, T.G., Ibañez, L.V., Murias, M., Stone, W.L., Wallace, M.T., Cascio, C.J., 2017). We suggest that a component of inflexibility in these populations may be a diminished ability to shape task specific neural networks through low frequency neural coherence. Such a deficit would lead to difficulty integrating and segregating timing representations appropriately using the theta band network we demonstrate. Future research would be well served to examine potential synchronization deficits during multisensory temporal processing in these clinical populations.

Multisensory Integration as an Active Feature Selection Process

Our results also serve to reinforce the amenability of the neural circuits mediating multisensory integration to top down task demands. Specifically, they indicate that task specific phase coupling is present to either support or inhibit the transfer of timing information across neural circuits. Neural oscillations as a mechanism for top down control of sensory processing within (Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008) and across (Lakatos et al., 2009; Mazaheri, Nieuwenhuis, van Dijk, & Jensen, 2009) circuits is well documented, and our study highlights that these same processes can be leveraged by the brain to selectively attend or ignore timing differences between the individual components of multisensory stimuli. This selective synchronization to support transfer of timing information is also interesting given the well-documented phenomena of temporal recalibration (Fujisaki et al., 2004; Vroomen et al., 2004), in which the brain adapts to prolonged exposure to asynchrony. We suggest that reductions or alterations in low frequency phase coupling might support this phenomenon by controlling or ablating connectivity between networks containing timing representations.

Limitations of the study include the possibility that participants might have covertly performed simultaneity judgment during the SR task. Importantly, however, such covert task performance would serve to equalize neural activity across tasks. This effect would run counter to the hypothesis that low frequency activity should differ between tasks. The presence of task related differences and task x stimulus interactions thus indicates that either covert simultaneity judgement did not occur or was sufficiently infrequent that it did not interfere with task related differences. Additionally, the temporal resolution of our findings is limited by the spectral transforms used. While we provide control analyses by comparing neighboring activity, temporal

interference in our results is still a possibility. This is particularly true for our connectivity analysis, as the proximity of the ERP may obscure distributed effects in more fronto-centralized regions. Specifically, the WPLI is very conservative compared to other connectivity approaches (in terms of aggressively discounting potentially volume-conducted synchronization) and may discount some true synchronization effects (Cohen, 2015). While we believe this conservative approach is a strength of our study as it grants confidence in our significant connectivity results, it does raise the possibility that a nontrivial portion of true neural synchronization was missed.

Conclusion

Our study demonstrates that top down task demands modulate neural responses to asynchronous audiovisual speech. For early neural activity, active attending to multisensory temporal concordance acts to dampen cortical responses likely generated by primary and secondary auditory cortex. For later cortical responses, active attending of stimulus properties is required for temporal structure to modulate activity. Phase based connectivity analysis further indicated that low frequency functional networks are differentially engaged during active attending and passive viewing of temporally asynchronous stimuli. Our results thus indicate that low frequency neural activity in both local and long-range circuits is mediated by top down task demands during multisensory temporal processing. The task dependencies we highlight here may also contribute to the variability seen in across experiments using ERP component based approaches to multisensory integration of speech signals (Baart, 2016). Future work utilizing a more continuous design with fewer sharp stimulus transients may be able to elucidate that exact temporal dynamics and cortical localization of task dependent activity. Such work might make

substantial contribution to elucidating top down control of multisensory information flow in the human brain.

References

Akar, S. A., Kara, S., Latifoglu, F., & Bilgic, V. (2016). Analysis of the Complexity Measures in the EEG of Schizophrenia Patients. *Int J Neural Syst, 26*(2), 1650008.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn Sci, 16*(7), 390-398.

Baart, M. (2016). Quantifying lip-read-induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays. *Psychophysiology, 53*(9), 1295-1306.

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software, 31*(10), 1-21.

Catarino, A., Churches, O., Baron-Cohen, S., Andrade, A., & Ring, H. (2011). Atypical EEG complexity in autism spectrum conditions: a multiscale entropy analysis. *Clin Neurophysiol, 122*(12), 2375-2383.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences, 18*(8), 414-421.

Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends Neurosci, 37*(9), 480-490.

Cohen, M. X. (2015). Effects of time lag and frequency matching on phase-based connectivity. *J Neurosci Methods, 250*, 137-146.

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci, 16*(9), 1348-1355.

Cooper, P. S., Wong, A. S., Fulham, W. R., Thienel, R., Mansfield, E., Michie, P. T., & Karayanidis, F. (2015). Theta frontoparietal connectivity associated with proactive and reactive cognitive control processes. *Neuroimage, 108*, 354-363.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial

EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21.

Engel, A. K., Gerloff, C., Hilgetag, C. C., & Nolte, G. (2013). Intrinsic Coupling Modes: Multiscale Interactions in Ongoing Brain Activity. *Neuron, 80*(4), 867-886.

Ewald, A., Aristei, S., Nolte, G., & Abdel Rahman, R. (2012). Brain Oscillations and Functional Connectivity during Overt Language Production. *Front Psychol, 3*, 166.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9*(10), 474-480.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci, 7*(7), 773-778.

Hanslmayr, S., Volberg, G., Wimber, M., Oehler, N., Staudigl, T., Hartmann, T., . . . Bauml, K. H. (2012). Prefrontally driven downregulation of neural synchrony mediates goal-directed forgetting. *J Neurosci, 32*(42), 14742-14751.

Hass, K., Sinke, C., Reese, T., Roy, M., Wiswede, D., Dillo, W., . . . Szycik, G. R. (2017). Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion. *Cogn Neuropsychiatry, 22*(2), 145-158.

Haufe, S., Nikulin, V. V., Muller, K. R., & Nolte, G. (2013). A critical assessment of connectivity measures for EEG data: a simulation study. *Neuroimage, 64*, 120-133.

Hipp, J. F., Engel, A. K., & Siegel, M. (2011). Oscillatory Synchronization in Large-Scale Cortical Networks Predicts Perception. *Neuron, 69*(2), 387-396.

Huhn, Z., Szirtes, G., Lorincz, A., & Csepe, V. (2009). Perception based method for the investigation of audiovisual integration of speech. *Neurosci Lett, 465*(3), 204-209.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol, 111*(10), 1745-1758.

Klimesch, W., Hanslmayr, S., Sauseng, P., Gruber, W., Brozinsky, C. J., Kroll, N. E., . . . Doppelmayr, M. (2006). Oscillatory EEG correlates of episodic trace decay. *Cereb Cortex, 16*(2), 280-290.

Kosem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage, 92*, 274-284.

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., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron, 64*(3), 419-430.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods, 164*(1), 177-190.

Mazaheri, A., Nieuwenhuis, I. L., van Dijk, H., & Jensen, O. (2009). Prestimulus alpha and mu activity predicts failure to inhibit motor responses. *Hum Brain Mapp, 30*(6), 1791-1800.

Murray, M. M., & Wallace, M.T. (2012). *The Neural Bases of Multisensory Processes* (M. M. Murray, Wallace, M.T. Ed.). Boca Raton, FL: CRC Press.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin Neurophysiol, 115*(10), 2292-2307.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci, 2011*, 156869.

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

Parsons, B. D., Gandhi, S., Aurbach, E. L., Williams, N., Williams, M., Wassef, A., & Eagleman, D. M. (2013). Lengthened temporal integration in schizophrenia. *Neuropsychologia, 51*(2), 372-376.

Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr Clin Neurophysiol, 66*(1), 75-81.

Phillips, J. M., Vinck, M., Everling, S., & Womelsdorf, T. (2014). A long-range fronto-parietal 5- to 10-Hz network predicts "top-down" controlled guidance in a task-switch paradigm. *Cereb Cortex, 24*(8), 1996-2008.

Roa Romero, Y., Keil, J., Balz, J., Gallinat, J., & Senkowski, D. (2016). Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. *J Neurophysiol, 116*(3), 1396-1407.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. *Cerebral Cortex, 17*(5), 1147-1153.

Sauseng, P., Klimesch, W., Freunberger, R., Pecherstorfer, T., Hanslmayr, S., & Doppelmayr, M. (2006). Relevance of EEG alpha and theta oscillations during task switching. *Exp Brain Res, 170*(3), 295-301.

Schepers, I. M., Schneider, T. R., Hipp, J. F., Engel, A. K., & Senkowski, D. (2013). Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. *Neuroimage, 70*, 101-112.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci, 31*(8), 401-409.

Simon, D.M., Damiano, C.R., Woynaroski, T.G., Ibañez, L.V., Murias, M., Stone, W.L., Wallace, M.T., Cascio, C.J. (2017). Neural Correlates of Sensory Hyporesponsiveness in Toddlers at High Risk for Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*.

Simon, D.M., & Wallace, M. T. (2017). Integration and Temporal Processing of Asynchronous Audiovisual Speech. *Journal of Cognitive Neuroscience*.

Spielberg, J. M., Miller, G. A., Heller, W., & Banich, M. T. (2015). Flexible brain network reconfiguration supporting inhibitory control. *Proc Natl Acad Sci U S A, 112*(32), 10020-10025.

Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., & James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *Neuroimage, 49*(4), 3308-3318.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *J Neurosci, 34*(3), 691-697.

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus

dependencies. *Exp Brain Res, 227*(2), 249-261.

Uhlhaas, P. J., & Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nat Rev Neurosci, 11*(2), 100-113.

van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron, 81*(6), 1240-1253.

van Driel, J., Knapen, T., van Es, D. M., & Cohen, M. X. (2014). Interregional alpha-band synchrony supports temporal cross-modal integration. *Neuroimage, 101*, 404-415.

van Eijk, R. L., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. *Percept Psychophys, 70*(6), 955-968.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat Rev Neurosci, 2*(4), 229-239.

Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., & Pennartz, C. M. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage, 55*(4), 1548-1565.

von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *Int J Psychophysiol, 38*(3), 301-313.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys, 72*(4), 871-884.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res, 22*(1), 32-35.

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64C*, 105-123.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science, 316*(5831), 1609-1612.

Womelsdorf, T., Vinck, M., Leung, L. S., & Everling, S. (2010). Selective theta-synchronization of choice-relevant information subserves goal-directed behavior. *Front Hum Neurosci, 4*, 210.

CHAPTER IV

SINGLE TRIAL PLASTICITY IN EVIDENCE ACCUMULATION RATE UNDERLIES RAPID RECALIBRATION TO ASYNCHRONOUS AUDIOVISUAL SPEECH

The contents of this chapter are drawn from a manuscript in preparation:

Simon, D.M., Nidiffer A.R., Wallace M.T., Single Trial Plasticity in Evidence Accumulation Rate Underlies Rapid Recalibration to Asynchronous Audiovisual Speech

Abstract

Asynchronous arrival of audiovisual information at the peripheral sensory organs is a ubiquitous property of signals in the natural environment due to differences in the propagation time of light and sound. As these audiovisual cues are constantly changing their distance from the observer, rapid adaptation to these asynchronies is thus crucial for their appropriate integration, and consequently for the creation of a coherent perceptual representation of our dynamic world. We investigated the neural basis of rapid recalibration to asynchronous audiovisual speech in humans using a combination of psychophysics, drift diffusion modeling, and electroencephalography (EEG). Consistent with previous reports, we found that participant's perception of audiovisual temporal synchrony/asynchrony depends on the temporal ordering of the previous trial. Drift diffusion modelling indicated that this temporal recalibration effect was well accounted for by

trial order dependency in the rate of evidence accumulation (i.e. drift rate). Neural responses as indexed via evoked potentials were similarly found to vary based on the temporal ordering of the previous trial. Furthermore, these neural signals displayed both response locking and a build-to-threshold structure that have previously been established as neural correlates of evidence accumulation. Within and across subject correlations indicated that the observed changes in drift rate and the modulation of evoked potential magnitude were related. These results indicate that the rate and direction of evidence accumulation are affected by immediate sensory history and that these changes contribute to single trial recalibration to audiovisual temporal asynchrony.

Introduction

Adaptation to Temporal Asynchrony as a Contributor to Multisensory Integration

Objects and events in the natural environment frequently generate informative signals in multiple sensory modalities. Combining these multisensory signals into a unified percept has previously been shown to offer substantial behavioral and perceptual benefits (Murray & Wallace, 2012). These advantages have been particularly well described for speech signals, in which visual speech can dramatically facilitate speech comprehension in noisy environments (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby, 1954). An important cue for correctly integrating signals emanating from a common source is their temporal relationship, and studies have confirmed that the temporal relationship between auditory and visual speech cues directly affects whether these signals are perceptually bound (van Wassenhove, Grant, & Poeppel, 2007). This reliance on temporal concordance, however, presents a unique challenge to the nervous system, as the temporal relationship for audiovisual speech signals continually changes due to differences

in the propagation speed of light and sound across space.

Due to these differences in propagation time, it makes substantial ecological sense for the brain to adapt to temporal asynchrony in an attempt to take differences in arrival time into account. For audiovisual signals, this adaptation has been shown to occur in response to sustained exposure to asynchronous signals (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004), and more recently adaptation has been shown to occur at the level of the single trial (Van der Burg, Alais, & Cass, 2013; Van der Burg & Goodbourn, 2015). By recalibrating in this manner, the brain is able to appropriately bind signals with a common source, even if their arrival times at the sensory periphery are somewhat misaligned. Recently, we investigated the neural basis of single trial temporal recalibration and demonstrated that neural responses to simple audiovisual stimuli (i.e., flashes and beeps) differ in magnitude depending on the temporal ordering of the stimulus on the previous trial (D. M. Simon, Noel, & Wallace, 2017). Specifically, when the temporal order of the previous stimulus was the opposite of the temporal order of the current stimulus (e.g., visual preceding auditory followed by auditory preceding visual), voltage at centro-parietal electrode sites was substantially larger. The centro-parietal location and relatively late timing of the differences observed (>325 ms after the first stimulus, >125 ms after the second stimulus) strongly suggested that, rather than indexing changes in low-level sensory processing, these differences in neural activity were indexing supramodal decisional processes (O'Connell, Dockree, & Kelly, 2012).

Motivations for the Current Study

The current study aimed to extend these findings in three important ways. First, we sought to test

whether similar effects are present for audiovisual speech stimuli, which are more ecologically relevant than flashes and beeps. Second, the reduced temporal precision afforded to simultaneity judgements based on speech stimuli (i.e., speech stimuli seem to be "bound" over larger temporal intervals) allowed us to directly test whether these neural effects are restricted to occurring after the second stimulus of asynchronous pairings. Most importantly, given that we previously hypothesized these physiological and behavioral effects are related to changes in evidence accumulation rate, we examined whether diffusion modelling of the decisional process could account for the observed neural effects. Specifically, we sought to compare these to drift rate, which indexes the quality of sensory evidence (Gold & Shadlen, 2007; Voss, Rothermund, & Voss, 2004). To address these questions, we employed a speeded simultaneity judgment task in response to audiovisual speech stimuli and concurrently recorded electroencephalography (EEG). Our findings indicate that participant's perceptual judgements and the magnitude of neural responses are mediated by trial-to-trial differences in audiovisual temporal ordering across speech events. Secondly, we found that given a large enough temporal delay, these neural effects occur before onset of the second stimulus. Further, we then demonstrate that the magnitude of changes in neural responses correspond well with modeled differences in evidence accumulation rate. These results indicate that the rate and direction of evidence accumulation are affected by immediate sensory history and that these changes contribute to single trial recalibration to audiovisual temporal asynchrony.

Methods and Materials

Participants

Data was drawn from the study of adult audiovisual speech processing already reported in (D.M. Simon & Wallace, 2017) and inclusion criteria were identical across studies. All participants reported that they were right handed, had normal or corrected-to-normal vision, and had normal hearing. Data from 28 participants were initially collected. Two participants were excluded from analysis due to behavioral performance indicating they did not correctly perform the task and 1 participant did not complete the task, leaving a total of 25 analyzed participants (16 female) with a mean age of 22.08 years (± 4.21). The study was conducted in accordance with the declaration of Helsinki, and informed written consent was obtained from all participants. All procedures were approved by the Vanderbilt University institutional review board.

Psychophysical Task

The speeded simultaneity judgement task utilized is described in detail in (D.M. Simon & Wallace, 2017). Briefly, participants performed a speeded two alternative forced choice simultaneity judgment task for audiovisual speech stimuli presented on a computer monitor and bilateral speakers. The experimental stimuli consisted of an audiovisual movie of a female saying the syllable 'BA'. Each trial began with a still face presented for 1700-2000ms with a uniform distribution. This was followed by the audiovisual movie, with a duration of 2000 ms. Following the movie, a still face consisting of the last video frame was presented for 750 ms. If no response was given by the end of the still face period a response screen appeared until a response was given or for a maximum of 2500 ms. Participants were instructed to use their right hand to indicate whether the stimuli were perceived to occur at the same time (i.e., synchronously) or at different times (i.e., asynchronously) via keyboard button press. Participants were also explicitly told to respond as quickly and accurately as possible, and that the appearance of the response

screen was an indicator that their responses were too slow. Participants completed a practice block before the main experiment.

To create the experimental temporal asynchronies, we manipulated the audiovisual stimulus by delaying either the visual stimulus (to create an AV trial) or the auditory stimulus (to create a VA trial). We created six total asynchronies ranging from audition leading vision by 450 ms (A450V) to vision leading audition by 450 ms (V450A) in steps of 150 ms, resulting in 7 conditions including the original movie featuring synchronized stimuli. Blocks consisted of 105 stimuli presented in a random order and participants completed a total of 13 or 14 blocks, and thus 1365 or 1470 trials. Stimulus onset for all stimuli was considered relative to the leading stimulus. That is, for auditory leads stimulus onset was at the time of auditory onset, while for visual leads stimulus onset was the onset of the video frame associated with auditory onset in the original video (as to not include pre-articulatory motion). These events occurred simultaneously in the synchronous video. In other words, time 0 corresponded with the first point at which task relevant information was present.

Behavioral Data Analysis

Behavioral data was analyzed by compiling rate of reported synchrony and response times (RTs). This was performed separately for each of the seven SOAs, and then repeated for each of the seven SOAs sorting trials based on whether the previous trial was auditory leading or visual leading. Synchronous trials were omitted from analysis when sorting by lead type. RTs were analyzed directly via a 2 lead type x 7 SOA repeated measures ANOVA with follow-up paired sample t-tests. Reports of synchrony were fit with single-term Gaussian psychometric functions with free parameters of amplitude, mean, and standard deviation (MATLAB fit.m). The mean of

the best fitting distribution is taken as the point of subjective simultaneity (PSS) while the standard deviation is taken as a measure of temporal binding window (TBW) size. We then compared amplitude and PSS of these distributions using paired sample t-tests to determine if they changed based on the previous trial. This process was then extended by compiling separate distributions for each of the possible seven SOAs on the previous trial, including synchrony. These seven data distributions were then once again fit with Gaussian functions and the PSS was compared across them using repeated measures ANOVA. Additionally, based on previous reports that plasticity in the PSS (i.e., ΔPSS) and TBW size are related (D. M. Simon, Noel, et al., 2017; Van der Burg et al., 2013) we investigated whether this relationship was true using linear regression (Pearson correlation). We also investigated whether differences between RTs for each lead type were related to individual perceptual thresholds using linear regression between change in response time and overall rate of reported synchrony separately for each condition.

For calculation and display of differences between auditory and visual leads we primarily performed subtraction of visual leading values from auditory leading values (A lead – V lead), but used the opposite subtraction (V lead – A lead) for the PSS to TBW correlation to maintain comparability with the existing literature (i.e. (Van der Burg et al., 2013)). We note that the direction of this subtraction is arbitrary and chosen for consistency with previous publications, and thus that the direction (i.e. sign) of effects is not directly interpretable.

Drift Diffusion Modelling

Choice and RT data were fit to a drift diffusion model (DDM) using the Diffusion Model Analysis Toolbox (DMAT; (Vandekerckhove & Tuerlinckx, 2007, 2008)). The DDM was used

due to its wide prevalence, well-validated parameters (Ratcliff & McKoon, 2008; Voss, Nagler, & Lerche, 2013; Voss et al., 2004) and ability to explain a wide variety of phenomena in choice and reaction time data (Ratcliff & Rouder, 1998). We allowed three DDM parameters to vary: non-decision time, which quantifies the amount of time related to sensory encoding processes and generation of motor responses, drift rate, which corresponds with the strength of the sensory evidence and the corresponding trajectory (slope) of the decision process, and drift-rate variability, which quantifies the consistency of drift rate across trials. The remaining variability parameters (starting point variability and non-decision time variability) were fixed at zero, while boundary separation and starting evidence were constrained across conditions but allowed to vary across subjects. Constraint of the boundary and starting point parameters across conditions was based on a-priori knowledge that, because direction of the rapid recalibration effect depends on a relationship between the current *and* previous stimulus, it cannot be explained by either bias towards a particular perceptual choice or change in decisional boundary distance. Fit values (Bayesian Information Criterion, Chi Square, and Log Likelihood Ratio) confirmed that this model provided the best fit of the data. This was the best fitting model for 13 of 25 participants. The best fitting models for the remaining participants were evenly split between including just drift rate (3 participants); drift rate and starting point (4 participants); and drift rate and boundary separation (5 participants).

We excluded drift rates for one participant in the A450V condition based on an implausible drift rate delta between the auditory and visual leads (A lead – V lead = -0.6317), which was >4 standard deviations from the mean drift rate delta across all conditions and participants. We similarly excluded a single non-decision time for one participant in the A300V condition based on an implausible non-decision time fit (1.51 seconds) which was >6 standard deviations from

the mean. The effects of removing these values are noted in the results where appropriate.

EEG Recording and Processing

Continuous EEG was recorded from 128 electrodes referenced to the vertex (Cz) using a Net Amps 400 amplifier and Hydrocel GSN 128 EEG cap (EGI systems Inc.). Data were acquired with NetStation 5.3 with a sampling rate of 1000 Hz and were further processed using MATLAB and EEGLAB (Delorme & Makeig, 2004). Continuous EEG data were band-pass filtered from 0.15 to 50 Hz with a 6 dB roll-off of 0.075 to 50.075 Hz. Epochs 3s long from 1000 ms before to 2000 ms after onset of the first stimulus were then extracted. Artifact contaminated trials and bad channels were identified and removed, and data were then recalculated to the average reference. Data were then submitted to ICA using the Infomax algorithm (Jung, Makeig, Humphries, et al., 2000; Jung, Makeig, Westerfield, et al., 2000), and artifact related components were removed. Lastly, bad channels were reconstructed using spherical spline interpolation (Perrin, Pernier, Bertrand, Giard, & Echallier, 1987) and data were re-inspected for residual artifacts. Overall a mean of 1081 (79% ± 9.5%) of trials were retained, while 4.17 (SD ± 2.42) channels and 10.56 (SD ± 4.14) Independent components were removed per participant. For a more thorough description of EEG procedures see (D.M. Simon & Wallace, 2017).

Stimulus Locked ERP analysis

To determine if the temporal ordering of the previous trial influenced neural responses we first separated data into two sets based on whether the previous trial was auditory leading or visual leading. The first trial of each block, which has no previous trial, and trials in which the previous trial was synchronous were excluded from this analysis. We performed this binning process

separately for each SOA on the current trial. We then averaged data in the time domain resulting in 14 ERPs for each (2 lead types x 7 SOAs). These ERPs were then statistically compared using non-parametric randomization testing with cluster-based correction for multiple comparisons (Maris & Oostenveld, 2007) as implemented in FieldTrip (http://www.fieldtriptoolbox.org/) (Oostenveld, Fries, Maris, & Schoffelen, 2011). The statistical test used for cluster inclusion was the dependent samples T test, cluster alpha was set to $a = 0.05$ and we used a permutation significance threshold of $a = 0.025$, which is equivalent to a two tailed test. Trend level ($0.025 < p < 0.05$) permutation results are also reported where appropriate. Given the potential for stimulus timing related differences in neural activity, we restricted testing to paired comparisons in which the physical stimulus is identical. We further employed Bonferroni correction on the permutation thresholds to rule out the possibility of type one errors and note Bonferonni corrected results when appropriate. For each significant spatiotemporal cluster, we then selected the electrodes participating when the cluster reached its maximum size, in terms of number of electrodes, and averaged them into a single ERP. We tested these ERPs again using non-parametric randomization testing. For completeness, this process was also performed in conditions which did not produce a significant spatiotemporal cluster. As there was no cluster to select from in these conditions we utilized a group of six electrodes (E54, E56, E61, E62, E78, and E79) common to all significant clusters and positioned over parietal scalp.

Response Locked ERP Analysis

To analyze decisional signals, we first low pass filtered single trials at 25 Hz to improve signal to noise ratio and then re-aligned single trials to participant's RTs. Trials were then time averaged into response locked ERPs to examine the Centro-parietal positivity (CPP) (O'Connell et al.,

2012). For CPP analysis we pooled values from the same six parietal electrodes (E54, E56, E61, E62, E78, and E79) selected for stimulus locked ERP analysis when no spatiotemporal cluster was present. To determine the slope of the CPP we fit a line to the CPP from -500 to -50 ms relative to participant's neural response for each condition separately (Twomey, Murphy, Kelly, & O'Connell, 2015) and then compared these slopes with a one way repeated measures ANOVA. Raw voltage of the CPP was examined using randomization testing with cluster alpha set to 0.05 and a permutation threshold of 0.025. For defining CPP onset in each condition, we performed a point-by-point paired sample t-test using the response locked ERP against 0 and defined onset as the first point for which $p < 0.01$ for at least ten consecutive samples, starting at that point. For display purposes, single trials were smoothed with a Gaussian moving average with a standard deviation of 49.9 trials, and thus a full window size of 300 trials.

Correlational Analysis

Correlational analysis between behavioral results, in the form of drift diffusion model parameters, and ERPs was performed using rank correlation and linear regression. ERPs were reduced to a single value for each lead type by pooling over the electrodes common to all significant spatiotemporal clusters found in the A450V, A300V, V300A, and V450 conditions (20 electrodes total) and time points similarly significant in all four spatiotemporal clusters (368 - 567ms). Subtracting these values (A lead – V lead) yielded a per participant voltage difference. We performed a similar drift rate subtraction and then correlated these values, pooled across conditions, using both Spearman rank correlation and linear regression (Pearson correlation). We also performed a similar Δdrift x ΔERP correlation within each participant using linear regression, and formally tested if these correlations were significantly different from zero using a

one-sample t-test. Finally, we tested whether the number of participants with a significant within subject correlation exceeded the null probability using a one-sample proportion test.

Statistical Tests

Time series data were tested using two-tailed spatiotemporal randomization testing with cluster-based correction for multiple comparisons (Maris & Oostenveld, 2007). Behavioral data and model parameter estimates were analyzed using repeated measures ANOVA and follow-up two-tailed t-tests. Significance thresholds for permutation tests were corrected for multiple comparisons using Bonferroni correction, while behavioral and model comparisons are presented uncorrected. All analysis was performed in MATLAB.

Results

Perception on Individual Trials Depends on the Temporal Ordering of the Previous Trial

We first tested whether our participants demonstrated single trial recalibration to audiovisual asynchrony as shown previously (D. M. Simon, Noel, et al., 2017; Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015). We fit Gaussian distributions to participant's behavioral responses on the speech SJ task (see methods), and these Gaussians were found to describe the distribution of perceptual reports well for all participants (mean $r^2 = 0.9611$, SD +/- 0.0333, r^2 for all participants > 0.874). Across participants, the mean point of subjective simultaneity (PSS) was 75.24 ms +/- 33.51 (t-test vs 0; $t_{24} = 11.226$, $p = 4.902 \times 10^{-11}$), and the mean temporal binding window (TBW) size was 296.10 ms +/- 97.74 (**Fig 4-1A**). We also calculated mean response time separately for each SOA (**Fig 4-1B**).

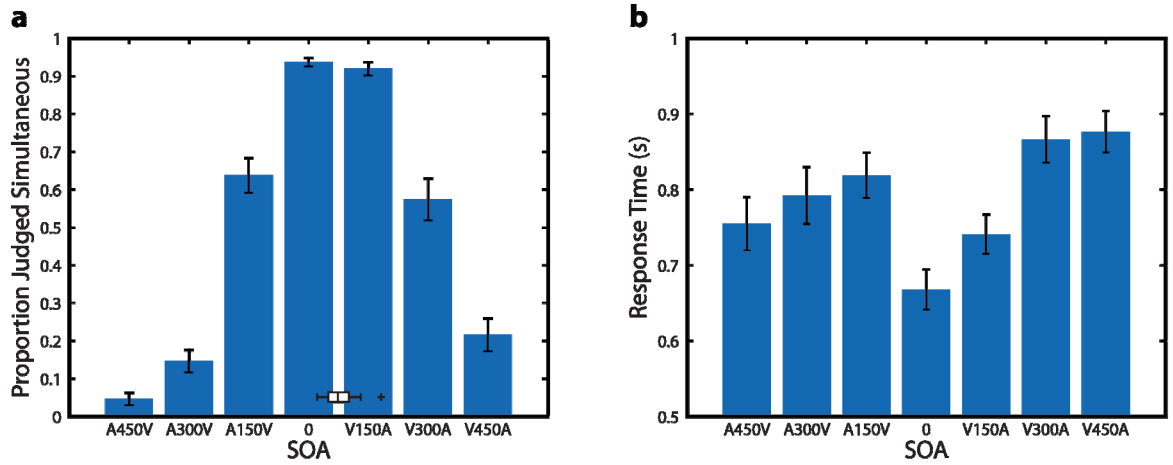

Figure 4-1 Behavioral Results for All Trials

A) Proportion of trials reported as synchronous for all 7 SOAs. Error bars indicate standard error of the mean. Inset box plot indicates PSS values.
B) Response times for all 7 SOAs. Error bars indicate standard error of the mean.

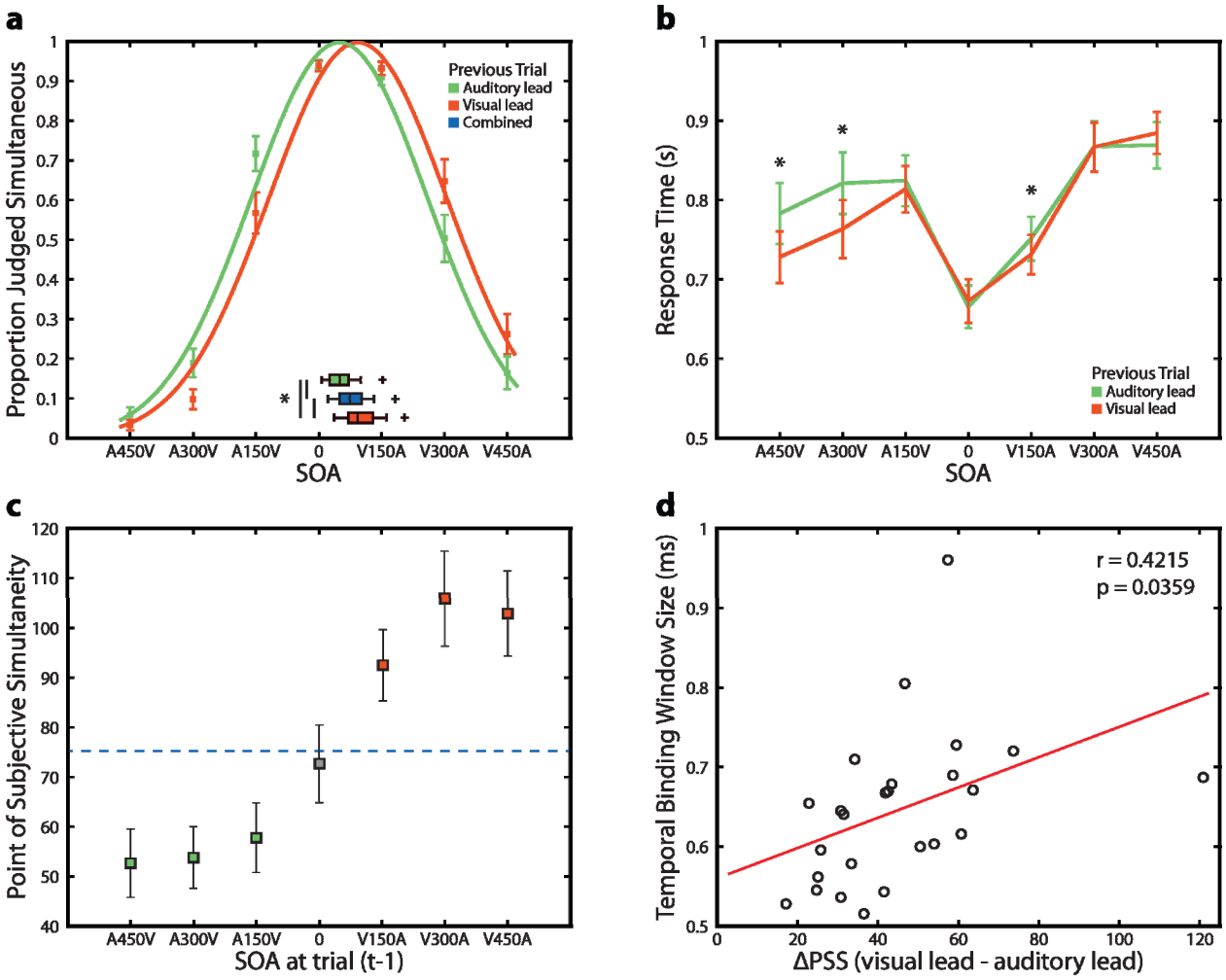

Figure 4-2 Behavioral Consequences of the Previous Trial

A) Proportion of trials reported as synchronous for all 7 SOAs separated by whether the previous trial was auditory leading (green) or visual leading (orange) and fit to a Gaussian distribution. Error bars indicate standard error of the mean. Gaussian fits are scaled to an amplitude of 1. The inset box plots indicate PSS values for auditory leading trials (green), visual leading trials (orange) or the combination of both (blue). Asterisk marked bars indicate a significant paired sample t-test between PSS values.

B) Response time for all 7 SOAs separated by whether the previous trial was auditory leading or visual leading. Error bars indicate standard error of the mean. Asterisks indicate a significant difference between lead types.

C) PSS for psychometric functions fit separately to data sorted into 7 bins based on the SOA of the previous trial. Error bars indicate standard error of the mean. Solid line indicates the PSS when all data is pooled.

D) Relationship between TBW size and change in PSS. Change in PSS is calculated as [(PSS V lead) – (PSS A lead)] for consistency with previous reports. Red line indicates the linear regression fit.

We then split trials into two categories; those in which the previous trial was auditory leading,

and those in which the previous trial was visual leading (**Fig 4-2A**). Trials in which the previous stimuli were synchronously presented were omitted from this analysis. We then re-fit data for the two lead types to determine if the psychometric functions differed. The Gaussian distribution fit to trials in which the previous trial was auditory leading had a mean PSS of 53.93 SD +/- 31.06 ms, whereas trials in which the preceding trial was visually leading had a mean PSS of 99.02 SD +/- 38.45 ms. We tested whether recalibration occurred by comparing the PSS values for these distributions to both the overall PSS and to each other. Both of these PSS values were found to be significantly different from the PSS of unsorted trials (auditory lead vs all trials $t_{24} = -11.117$, $p = 5.98 \times 10^{-11}$; visual lead vs all trials $t_{24} = 9.499$, $p = 1.330 \times 10^{-9}$). A paired sample T-test for PSS between these distributions was also significant ($t_{24} = -10.43$, $p = 2.143 \times 10^{-10}$). Goodness of fit between these distributions was found to be equally good (Auditory lead mean $r^2 = 0.9498$ +/- 0.0381, Visual lead mean $r^2 = 0.9539$ +/- 0.0360, paired sample t test $t_{24} = -0.6375$, $p = 0.5298$). The amplitude of the Gaussian fits, which serves as a measure of response bias, was also not different between the two lead types (Auditory lead mean = 1.048, SD +/- 0.0773, Visual lead mean = 1.0533 +/- 0.0769, $t_{24} = -0.6033$, $p = 0.5520$). These results thus strongly support the conclusion that the temporal ordering of the preceding trial directly affects participant's perceptual report on the current trial.

In addition to analyzing perceptual judgments (i.e., synchronous vs. asynchronous) on this task, response times (RTs) were also analyzed using a 2 (lead type) x 7 (SOA) repeated measures ANOVA. There was a significant main effect of SOA ($F_{6,144} = 25.425$, $p = 1.89 \times 10^{-20}$), a main effect of lead type ($F_{1,24} = 22.455$, $p = 8.07 \times 10^{-5}$), and a significant interaction ($F_{6,144} = 7.046$, $p = 1.365 \times 10^{-6}$). We then used follow-up t-tests which indicated that RTs were faster in three conditions when the previous trial had been visually leading (A450V $t_{24} = -5.3022$, $p = 1.937$ x

10^{-5}; A300V $t_{24} = -6.454$, $p = 1.128 \times 10^{-6}$; and V150A $t_{24} = -2.548$, $p = 0.0176$) (**Fig 4-2B**). No conditions were found in which a previous trial being auditory leading resulted in faster RTs (all other $p > 0.21$). While the RT difference for the V150A condition was quite modest (16.5 ms), the RT advantage in the A450V and A300V conditions was substantial (A450V: 48.55 ms, A300V: 54.63 ms). This indicates that the temporal ordering of the preceding trial (i.e., auditory lead versus visual lead) can substantially speed RTs, and that this effect is particularly pronounced when the current trial is a large auditory lead.

We then extended the Gaussian fitting procedure to individual SOAs by subdividing all trials into 7 bins each corresponding with a single SOA on trial t-1. We compared PSS across these 7 distributions and found a significant difference ($F_{6,144} = 41.35$, $p = 5.2290 \times 10^{-29}$), indicating that the PSS shifted to more positive values when the previous trial was visually leading (**Fig 4-2C**). Goodness of fit across these seven distributions did not differ ($F_{6,144} = 1.0897$, $p = 0.3714$). We further tested whether shifts in the PSS for visual and auditory leads were differentially modulated by the magnitude of the leads. To do this, we isolated conditions of these two types and performed two separate 1 x 3 repeated measures ANOVAs (factor of SOA). Differences in the magnitude of auditory leads were found to not contribute to the magnitude of the PSS shift ($F_{2,48} = 1.04$, $p = 0.3614$), but a trend was found for the magnitude of the visual lead contributing to the magnitude of the PSS shift ($F_{2,48} = 2.98$, $p = 0.060$). Lastly, we investigated whether the magnitude of the PSS shift depended on individual binding window size as has been previously reported (D. M. Simon, Noel, et al., 2017; Van der Burg et al., 2013). We found that this relationship was once again present, with larger PSS shifts being associated with larger TBWs ($r = 0.4215$, $p = 0.0359$) (**Fig 4-2D**).

Changes in Response Time Due to Recalibration Are Individualized

No differences in overall RT were found between auditory leads and visual leads at many SOAs, despite relatively large changes in the rate of reported synchrony. We next investigated if this occurred because changes in RT were heterogeneous across participants and might correspond with the substantial individual variability found in audiovisual temporal acuity (i.e., individual TBW sizes - see (Powers, Hillock, & Wallace, 2009; Stevenson & Wallace, 2013). Specifically, a given lead type might drive some participants towards greater perceptual ambiguity, while driving others toward greater perceptual clarity, depending on individual thresholds. To investigate this possibility, we subtracted visual leading RTs from auditory leading RTs (A lead – V lead) and then performed a multi-sample test for equal variances (**Fig 4-3A**). This indicated that the variance of RT changes between lead types differed across conditions (Bartlett statistic = 19.3863, df = 6, p = 0.0036). Notably, the highest variability (standard deviation) in RT change was found in the A150V and V300A conditions, which were the most perceptually ambiguous at the group level (A450V = 45.78 ms, A300V = 42.32 ms, A150V = 65.42 ms, AV = 35.53 ms, V150A = 32.28 ms, V300A = 60.96 ms, V450A = 43.31 ms).

For these two conditions, we then used linear regression to determine if perceptual thresholds corresponded with the direction and change in RT. For the A150V condition there was a significant negative correlation between overall reported rate of synchrony and RT change (r = -

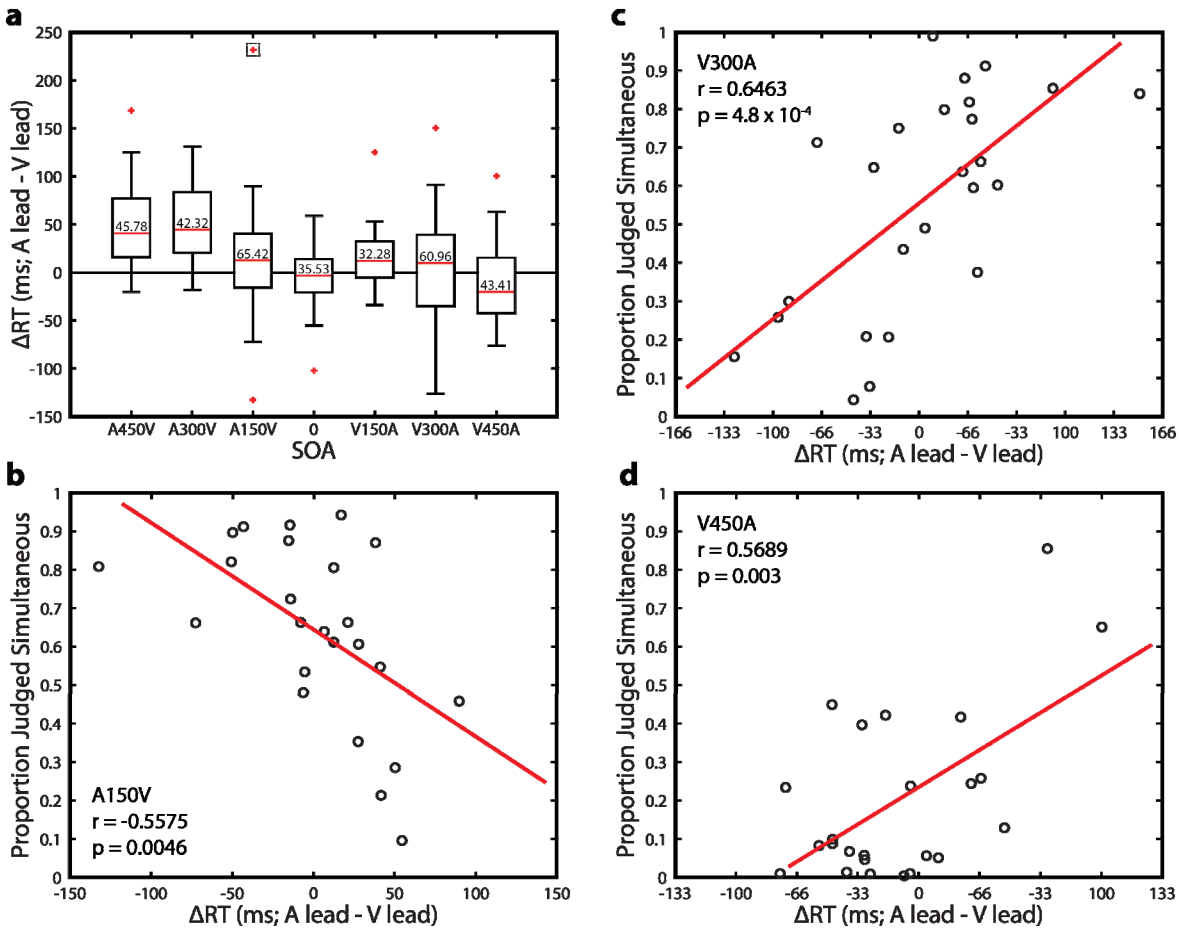

Figure 4-3 Effects of Rapid Recalibration on Response Time are Individualized

A) Change in response time (A lead RT – V lead RT) for all SOAs. Inset numbers are the standard deviation of the change in response time. The dotted box indicates an outlier value not shown in panel B.
B) Correlation between change in response time (A lead RT – V lead RT) and participant mean perception rate pooled across lead types for the A150V condition. A single outlier was omitted (see main text for statistics with the outlier included).
C) Correlation between change in response time (A lead RT – V lead RT) and participant mean perception rate pooled across lead types for the V300A condition.
D) Correlation between change in response time (A lead RT – V lead RT) and participant mean perception rate pooled across lead types for the V450A condition.

0.5127, p = 0.0088), which remained significant even with the removal of a large outlier (r = -0.5575, p = 0.0046) (**Fig 4-3B**). The V300A condition demonstrated a similarly significant correlation in the opposite direction (r = 0.6463, p = 4.8 x 10^{-4}) (**Fig 4-3C**). We then expanded this approach to the V450A condition, as it was the sole remaining condition without a response time effect and found a similar positive correlation (r = 0.5689, p = 0.003) (**Fig 4-3D**). In all the other conditions (A450V, A300V, AV, V150A) no such correlations were found (all p > 0.374). These correlations are thus present *only* at SOAs without a significant RT difference between lead types, and in the three conditions which are most ambiguous at the group level (i.e. the closest to a 50% rate of reported synchrony). Changes in RT due to rapid recalibration thus depend on whether rapid recalibration pushes a participant towards higher perceptual ambiguity. When recalibration pushes a participant towards ambiguity, RTs are slower due to a protracted

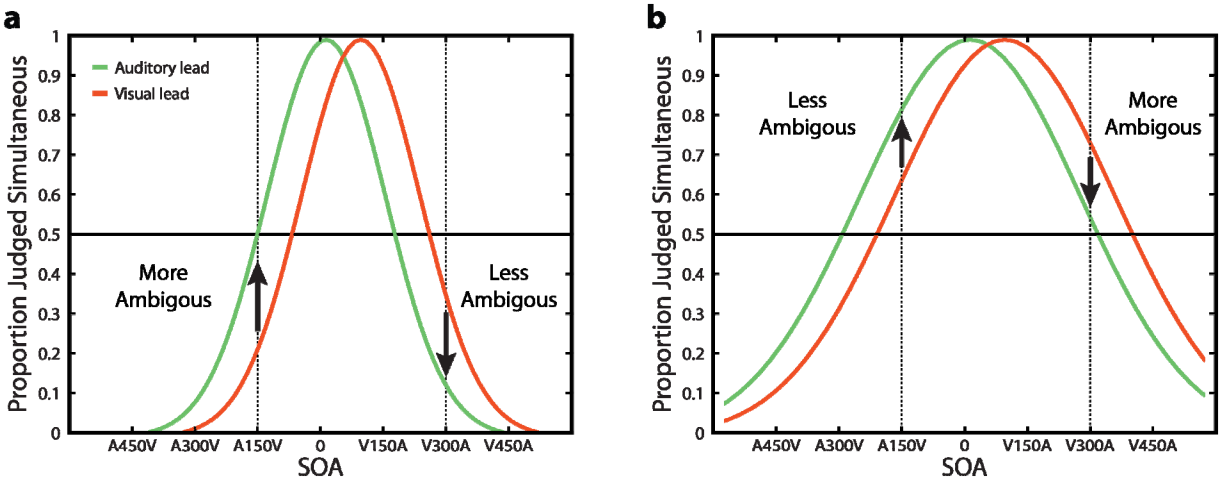

Figure 4-4 Proposed Origin for Individualized Changes in Response Time

A) Illustration of the effect of lead type on perceptual ambiguity for a participant with a narrow binding window. Relative to a visual lead, auditory leading trials are more perceptually ambiguous in the A150V condition, but less ambiguous in the V300A condition. The PSS for each distribution is identical to panel B.

B) Illustration of the effect of lead type on perceptual ambiguity for a participant with a wide binding window. Relative to a visual lead, auditory leading trials are less perceptually ambiguous in the A150V condition, but more ambiguous in the V300A condition. The PSS for each distribution is identical to panel A.

decisional process, while shifts away from ambiguity lead to faster RTs. We Illustrate in **Fig 4-4A** and **Fig 4-4B** how an identical change in PSS can have opposite effects on perceptual ambiguity that depend on individual audiovisual temporal acuity (TBW width). These individualized differences in RT changes thus explain why differences in mean RT only occur at certain SOAs, despite the ubiquity of changes in perceptual report across conditions.

Single Trial Recalibration Changes the Rate of Sensory Evidence Accumulation

We hypothesized that the changes in perceptual report occur because the previous trial affects the rate and direction of evidence accumulation on the current trial. To test this hypothesis we employed drift diffusion modelling (see methods for parameters) (Vandekerckhove & Tuerlinckx, 2007, 2008). The drift diffusion model posits that decisions are the result of a stochastic diffusion process that is absorbed by one of two decision boundaries and offers robust quantification of evidence accumulation in two alternative forced choice designs such as that employed in the current study. After fitting the drift diffusion model, we utilized a 2 (lead type) x 7 (SOA) repeated measures ANOVAs for each free model parameter to determine whether they were affected by SOA, the nature of the previous trial, and any possible SOA x lead type interactions.

We first examined drift rate, which is the average rate and direction of the decision variable during the decision process. Drift rate demonstrated a significant main effect of SOA ($F_{6,144}$ = 108.1982, p = 8.0746 x 10^{-51}), while the main effect of lead type was not significant ($F_{1,24}$ = 0.1142, p = 0.7383). There was also a significant interaction effect ($F_{6,144}$ = 9.7869, p = 4.8950x10^{-9}) (**Fig 4-5A-C).** This interaction indicated that the effect of lead type on drift rate reversed depending on the SOA, and we thus performed follow-up paired sample t-tests for all conditions contrasting lead types. These tests indicated that significant differences were present in 5 conditions with large SOAs (**Fig 4-5C**) (A450V, t_{23} = 4.9988, p = 4.6778 x 10^{-5} uncorrected; A300V, t_{24} = 5.8133, p = 5.4018 x 10^{-6} uncorrected; A150V, t_{24} = 4.0112, p = 5.122 x 10^{-4} uncorrected; V300A, t_{24} = 4.5270, p = 1.3828 x 10^{-4} uncorrected; V450A, t_{24} = -2.3983, p = 0.0246 uncorrected). Drift rate thus not only varied significantly across SOAs, but was also directly affected in a directional manner by the temporal order of the previous trial. When the temporal order of the current trial and the previous trial were mismatched, sensory evidence for

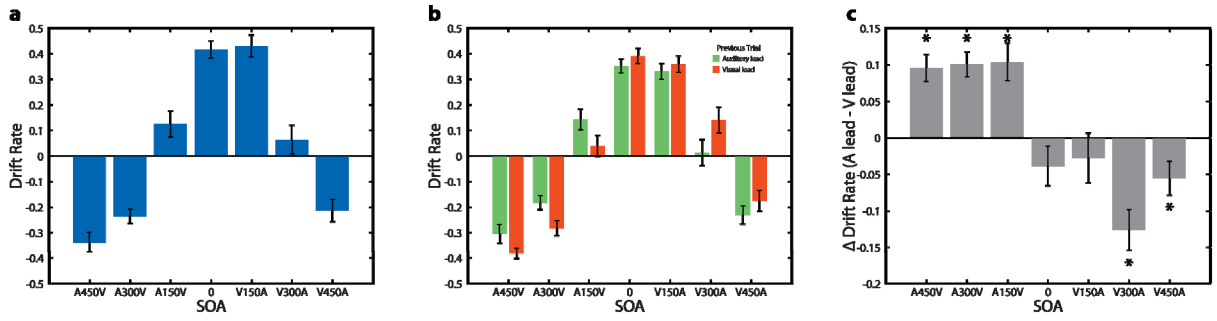

Figure 4-5 Diffusion Model Results: Drift Rate

A) Drift rate for each of the seven SOAs for a model pooled across auditory leading and visual leading trials. Error bars indicate standard error of the mean.
B) Drift rate for each of the seven SOAs for two models fit separately based on whether the previous trial was auditory leading or visual leading. Error bars indicate standard error of the mean.
C) Difference in drift rate (auditory lead – visual lead) between the two lead types in panel B. Error bars indicate standard error of the mean.

asynchrony was stronger and the decision variable thus accumulated more strongly towards an asynchronous choice (i.e. a more negative drift rate). In other words, despite the stimulus on the current trial being physically identical for both lead types the strength of the sensory evidence differed.

We next examined non-decision time, which accounts for processes that take place both before the decisional process begins as well as other factors such as motor preparation. Non-decision time demonstrated a significant main effect of SOA ($F_{6,144} = 10.2024$, $p = 2.1472 \times 10^{-9}$), no significant main effect of lead type ($F_{1,24} = 0.0824$, $p = 0.7765$), and a trend towards an interaction effect ($F_{6,144} = 2.0831$, $p = 0.0587$) (**Fig 4-6A-C**). The interaction trend was not robust to removal of a single large non-decision time value (see methods; outlier removed $F_{6,138} = 1.69$, $p = 0.1272$) and was thus not pursued further. In sum, non-decision time was shorter for synchronously presented stimuli, but appeared to be unaffected by the temporal order of the preceding trial. As non-decision time encompasses low-level sensory encoding, this serves as

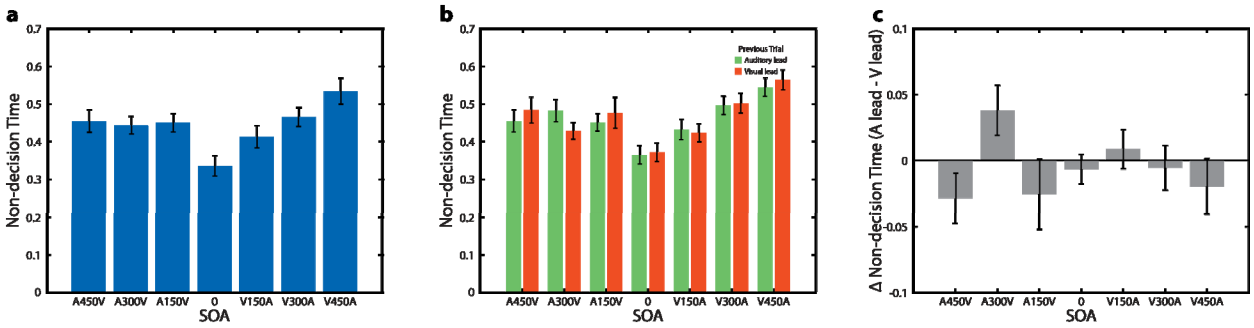

Figure 4-6 Diffusion Model Results: Non-Decision Time

A) Non-decision time for each of the seven SOAs for a model pooled across auditory lead and visual lead trials. Error bars indicate standard error of the mean.
B) Non-decision time for each of the seven SOAs for two models fit separately based on whether the previous trial was auditory leading or visual leading. Error bars indicate standard error of the mean.
C) Difference in non-decision time (auditory lead – visual lead) between the two lead types in panel B. Error bars indicate standard error of the mean.

evidence that the overall speed of sensory encoding is unaffected by the nature of the preceding trial.

Last, we examined drift rate variability, which is the magnitude of trial-to-trial variability in drift rates for a particular physical stimulus. Drift rate variability demonstrated a main effect of SOA ($F_{6,144} = 3.5442$, $p = 0.0027$), no main effect of lead type ($F_{1,24} = 0.7572$, $p = 0.3928$), and no significant interaction ($F_{6,144} = 0.9045$, $p = 0.4935$)

Neural Responses to Audiovisual Speech Stimuli Vary Based on the Temporal Ordering of the Previous Stimulus

We next employed an analytical strategy for our EEG data similar to the behavioral data analysis by binning trials depending on whether the preceding trial was auditory leading or visual leading. Trials in which the previous stimulus was synchronous were once again excluded from this analysis. We time averaged these trials into 14 total ERPs (2 lead types x 7 SOAs) and compared each pairing at a given SOA utilizing spatiotemporal randomization testing with cluster based correction for multiple comparisons. In this analytic approach, comparisons were thus between neural responses to physically identical stimuli, with the only difference being the temporal ordering of the previous trial. These spatiotemporal tests indicated that significant differences were present in the four conditions in which stimuli were most offset in time (A450V, $p = 0.001$; A300V, $p = 0.0042$; V300A, $p = 0.0053$ and V450A, $p = 0.0001$) (**Fig 4-7A-D)**. Note that all four of these conditions remain at least marginally significant after Bonferonni correction for 7 total comparisons (all $p < 0.0371$, Bonferroni corrected). Additionally, p-values of 0.0001 represent the floor of a randomization test with 10,000 randomizations and indicate that the maximum cluster in the real data was larger than the maximum cluster observed in all

permutations. No significant spatiotemporal clusters were identified in the A150V, AV, or V150A conditions (all $p > 0.09$, note that permutation tests are one tailed and thus this does not constitute a robust trend). We also note that sufficiently strong centro-parietal clusters (A450V and V450A) are associated with a significant dipolar effect at electrodes near the edge of the montage due to the average reference. This secondary cluster was smaller and ring shaped when significant, and we thus focused our analysis on the centro-parietal locations.

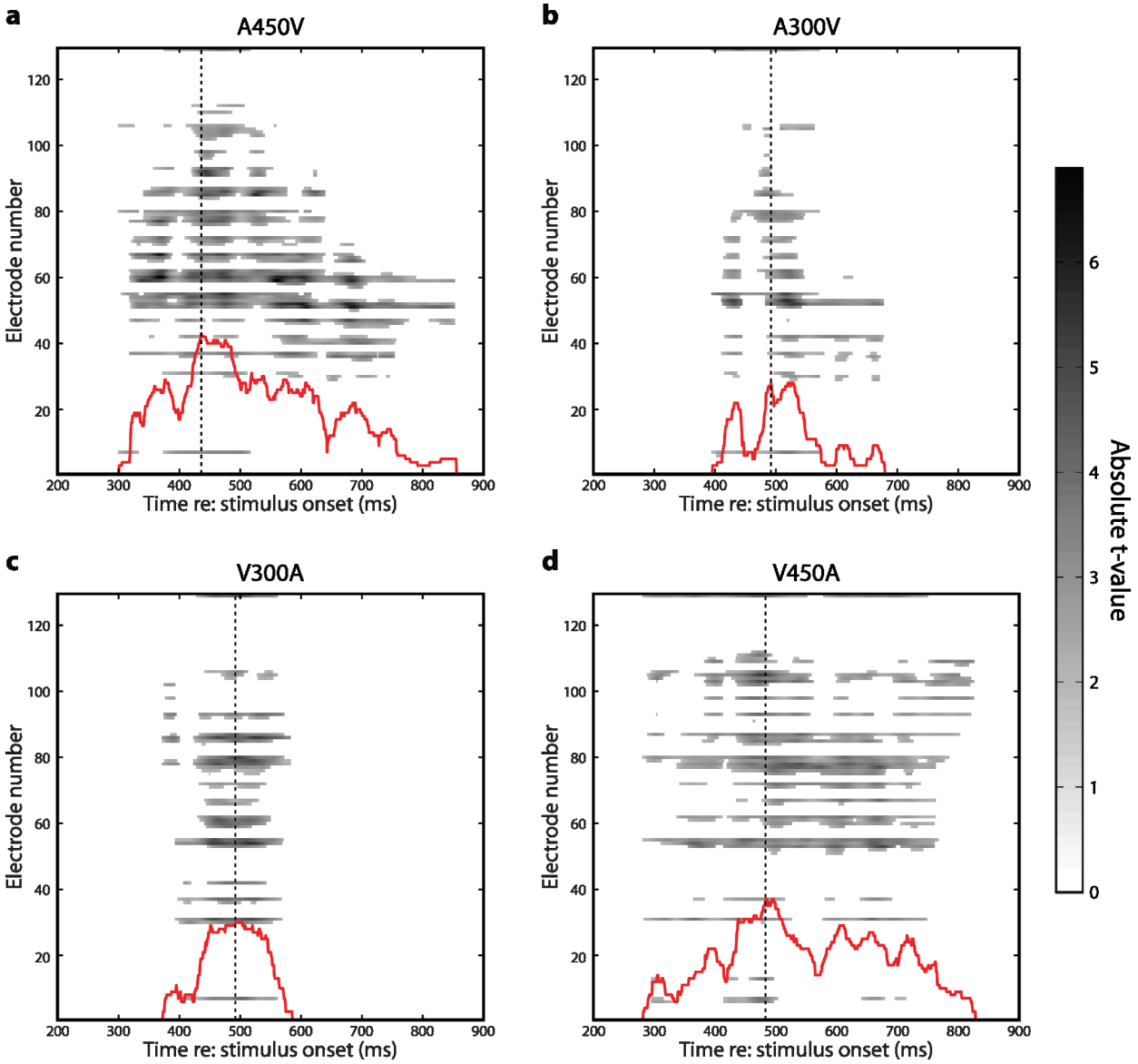

Figure 4-7 Spatiotemporal Clustering Results

A) Raster plot representation of the spatiotemporal cluster in the A450V condition. The absolute value of the t-statistic is indicated for each time point for comparisons which are both individually significant ($p < 0.05$) and part of the significant spatiotemporal cluster. The inset red line indicates the number of significant electrodes at each time point (same scale as electrode number). The dashed line indicates the time point when the cluster reached maximum spatial size for the first time.

B) Raster plot representation of the A300V condition. Inset lines as in panel A.

C) Raster plot representation of the V300A condition. Inset lines as in panel A.

D) Raster plot representation of the V450A condition. Inset lines as in panel A.

While these spatiotemporal clusters clearly illustrate the presence of effects, they fail to best capture the temporal nature of the differences in evoked activity. To better elucidate this, we identified the time point where each spatiotemporal cluster first reached its maximum size in terms of number of electrodes and selected those electrodes (A450V – 436ms, 43 electrodes; A300V – 492ms, 28 electrodes; V300A 492ms, 30 electrodes; V450A 483ms, 37 electrodes) (**Fig 4-8**). Event related potentials for data averaged across these electrodes was then tested for significance using randomization testing with cluster-based correction for multiple comparisons. In all four conditions, we found at least one significant temporal cluster (all $p < 0.0224$, Bonferroni corrected for 7 comparisons). We depict the spatial cluster selected and ERPs averaged across those electrodes in **figure 4-9**. Notably, in the A450V and V450A conditions, the onset times of significant differences substantially precedes the onset of the second stimulus, indicating that the neural processes influenced by the temporal ordering of the previous stimulus do not rely on both stimuli having occurred.

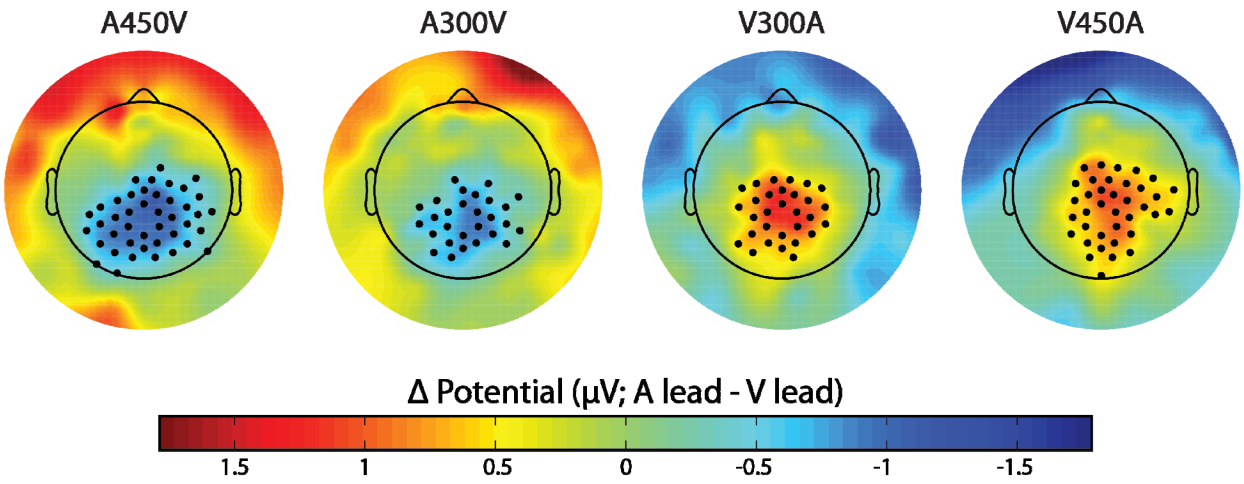

Figure 4-8 Topographic Representation of Spatiotemporal Clusters
Topographic representation of voltage differences between auditory and visual leading trials (A lead – V lead) for each SOA. Black dots indicate electrodes which are both individually significant ($p < 0.05$) and part of the significant spatiotemporal cluster. Time points for each SOA correspond with the dashed line in figure 7.

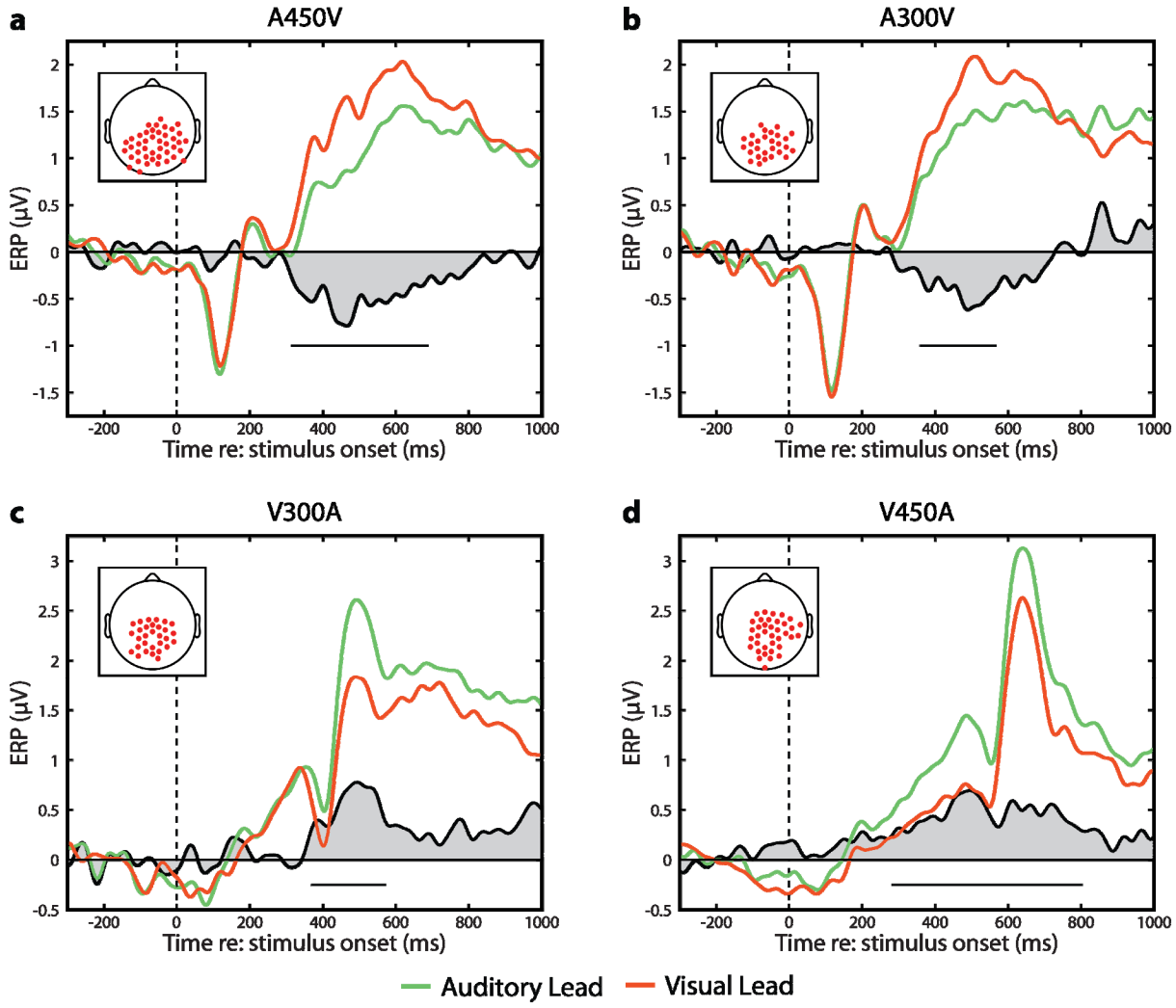

Figure 4-9 ERPs for large SOAs divided by the lead type of the previous trial.
Insets depict the cluster of electrodes selected in each condition.

A) ERPs in the A450V condition binned depending on whether the previous trial was an auditory lead (green) or a visual lead (orange). Significant differences between A lead and V lead trials are present from 313-689 ms ($p = 0.0003$, randomization test), and are depicted by the black line. The grey shaded area highlights the difference between conditions.

B) ERPs in the A300V condition binned depending on whether the previous trial was an auditory lead (green) or a visual lead (orange). Significant differences are present from 358-567 ms ($p = 0.0013$, randomization test), and are depicted by the black line. The grey shaded area highlights the difference between conditions.

C) ERPs in the V300A condition binned depending on whether the previous trial was an auditory lead (green) or a visual lead (orange). Significant differences are present from 368-573 ms ($p = 0.0032$, randomization test), and are depicted by the black line. The grey shaded area highlights the difference between conditions.

D) ERPs in the V450 condition binned depending on whether the previous trial was an auditory lead (green) or a visual lead (orange). Significant differences are present from 280-804 ms ($p = 0.0002$, randomization test), and are depicted by the black line. The grey shaded area highlights the difference between conditions. A marginally significant effect is present from 183-261 ms ($p = 0.0428$, randomization test).

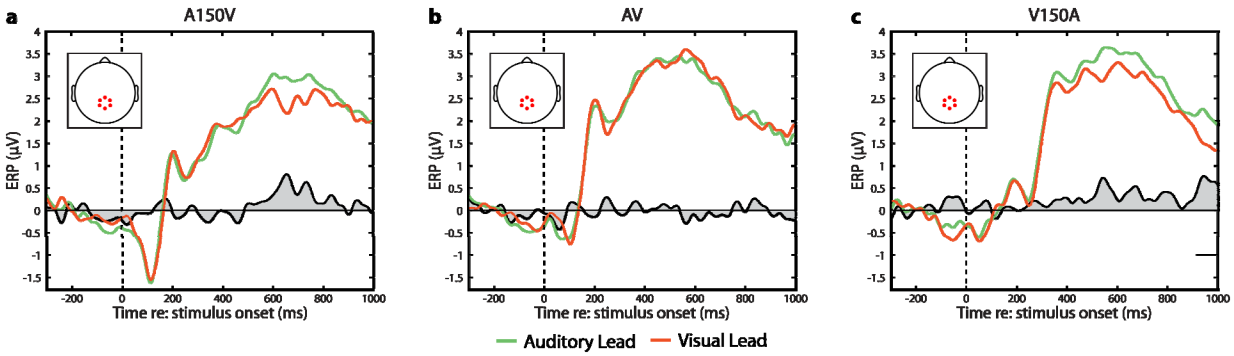

Figure 4-10 ERPs for small SOAs divided by the lead type of the previous trial.
Data were averaged across the six electrodes indicated.

A) ERPs in the A150V condition binned depending on whether the previous trial was an auditory lead (green) or a visual lead (orange). No significant differences are present.
B) ERPs in the AV condition binned depending on whether the previous trial was an auditory lead (green) or a visual lead (orange). No significant differences are present.
C) ERPs in the V150 condition binned depending on whether the previous trial was an auditory lead (green) or a visual lead (orange). Significant differences are present from 914-1000ms ($p = 0.013$, randomization test), and are indicated by the black underline. A marginally significant effect is present from 517-573 ms ($p = 0.0443$, randomization test).

For completeness, we also depict the neural responses in the conditions in which no significant spatiotemporal differences were found. As there were no significant spatiotemporal clusters selected in these conditions, we utilized a centro-parietal electrode cluster selected to optimally capture decisional components of ERPs (E54, E55, E61, E62, E78, and E79). This cluster was also composed of electrodes found to be significant in all four large SOAs as described above. We tested these time series using randomization testing with cluster based correction for multiple comparisons and found no significant differences in the A150V and AV conditions. In the V150A condition we found a small and very late effect from 914-1000 ms after stimulus onset ($p = 0.013$), which does not survive Bonferonni correction (**Fig 4-10**).

Neural Activity Exhibits Decisional Dynamics

Recent work examining build-to-threshold decision variables has focused on removing stimulus transients (i.e. sharp stimulus onsets and offsets) to allow clearer examination of decisional activity (Kelly & O'Connell, 2013; Loughnane et al., 2016; O'Connell et al., 2012). We thus started our examination of decisional signals by confirming that clear decisional dynamics were present despite the presence of onset and offset transients in our speech stimulus. We first sorted activity at centro-parietal electrodes, pooled across all conditions and participants, by single trial RT. A clear relationship between positive voltage buildup and participant RT is visible when trials are sorted in this manner **(Fig 4-11A)**. Re-aligning trials to participant response yielded a robust centro-parietal positivity (CPP) response in all conditions, which built to a consistent threshold (**Fig 4-11B**). As expected, the duration of the buildup was longer in conditions in which visual stimuli preceded auditory stimuli, consistent with the hypothesis that auditory leading temporal precision (i.e. strength of temporal sensory evidence) is greater than visual leading temporal precision (i.e. at moderate offset a V-A ordering is more ambiguous than an A-V ordering). This is highlighted by the significant difference in voltage across conditions resulting from visually leading conditions beginning the CPP buildup substantially earlier relative to the response (-704 to -289 ms, repeated measures multivariate F test at each time point; cluster $p = 0.0003$, randomization test). This was confirmed when we determined the point of CPP onset relative to response for each condition (A450V = -350ms; A300V = -396ms; A150V = -508ms; AV = -487ms; V150A = -501ms; V300A = -838 ms; V450A = -640ms).

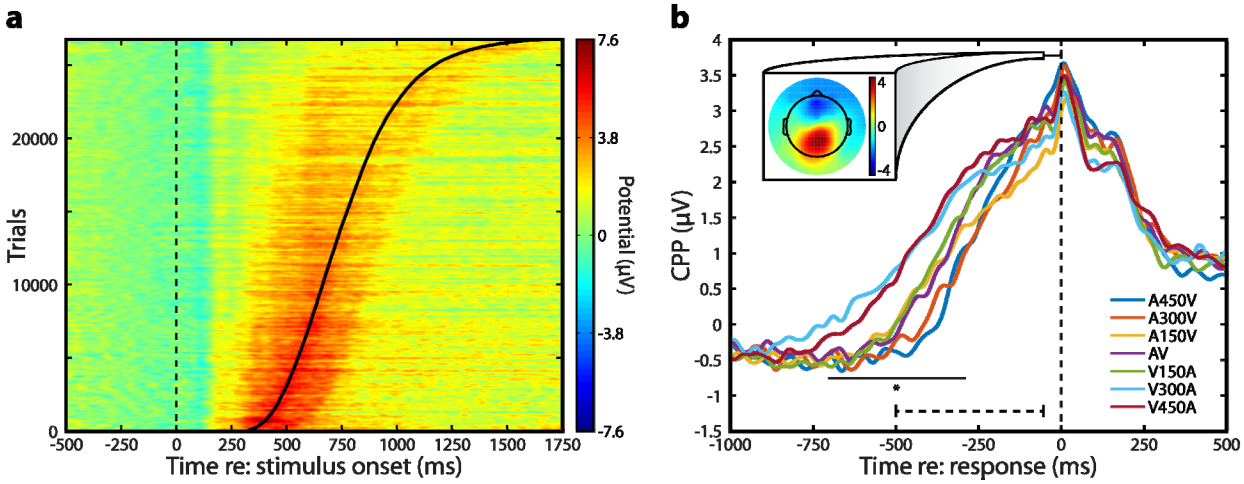

Figure 4-11 Decisional signal dynamics during simultaneity judgment.

A) Amplitude of the centro-parietal positivity (CPP) derived from single trial EEG and pooled across all 7 SOAs and all participants. The dotted line indicates stimulus onset, while the sigmoidal solid line indicates response times for individual trials. Data were sorted by response time and vertically smoothed with a 300 trial Gaussian moving average.

B) CPP calculated separately for each of the 7 SOAs. The dashed line at bottom indicates the time window used for slope analysis (-500 to -50 ms relative to response; see figure 12). Solid black line indicates time points with a significant difference in voltage (repeated measures multivariate F test at each time point; cluster p = 0.0003, randomization test). Top bar indicates the time region averaged for CPP topography (-50 to 0 ms relative to response), which is shown in the inset and averaged across all 7 SOAs.

We next examined the rate of evidence accumulation, which is linked to the strength of the available sensory evidence. To do so we calculated the linear slope of the CPP from -500 to -50ms for each condition (see (Twomey et al., 2015) for a similar approach). The CPP slope was found to differ significantly across conditions ($F_{6,144} = 13.17$, $p = 7.47 \times 10^{-12}$) (**Fig 4-12**). Additionally, we then split stimuli into visual first (A450V, A300V, A150V) and auditory first (V150A, V300A, V450A) categories (omitting synchronous trials). A 2 (sensory order [i.e. AV or VA current stimulus]) x 3 (SOA) repeated measures ANOVA indicated that there was still a main effect of SOA ($F_{2,48} = 21.27$, $p = 2.4325 \times 10^{-7}$), no main effect of leading modality ($F_{1,24} = 1.33$, $p = 0.2599$), but that a significant interaction effect was present ($F_{2,48} = 9.525$, $p = 3.2828$ x

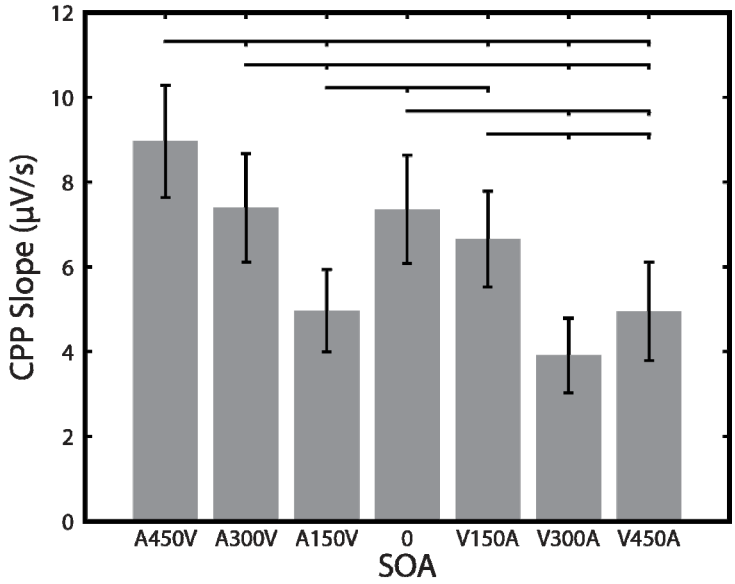

Figure 4-12 CPP Build Rate by SOA
Linear CPP slope for all 7 SOAs. Error bars indicate standard error of the mean. Slope was calculated from -500 to -50 ms relative to participant response (dashed region in figure 11). Bars indicate significant differences in slope (t-test, $p < 0.05$, uncorrected)

10^{-4}). This indicates that sensory evidence accumulation rates are asymmetric, in line with larger delays being closer to perceptual threshold for stimuli in which vision leads. Furthermore, the CPP slope was lowest in conditions that participants found perceptual ambiguous (A150V and V300A). This is consistent with these conditions being assigned drift rates closest to zero by the diffusion model (see **Fig 4-5** earlier), indicating that the available sensory evidence was relatively weak.

Neural Response Modulations Index Changes in Information Accumulation Rate

Lastly, based on previous work hypothesizing that the neural instantiation of rapid recalibration relates to differences in information accumulation rate (D. M. Simon, Noel, et al., 2017), we investigated whether the observed neural effects correlated with differences in drift rate across conditions and within participants. To do so, we pooled the voltage differences between lead

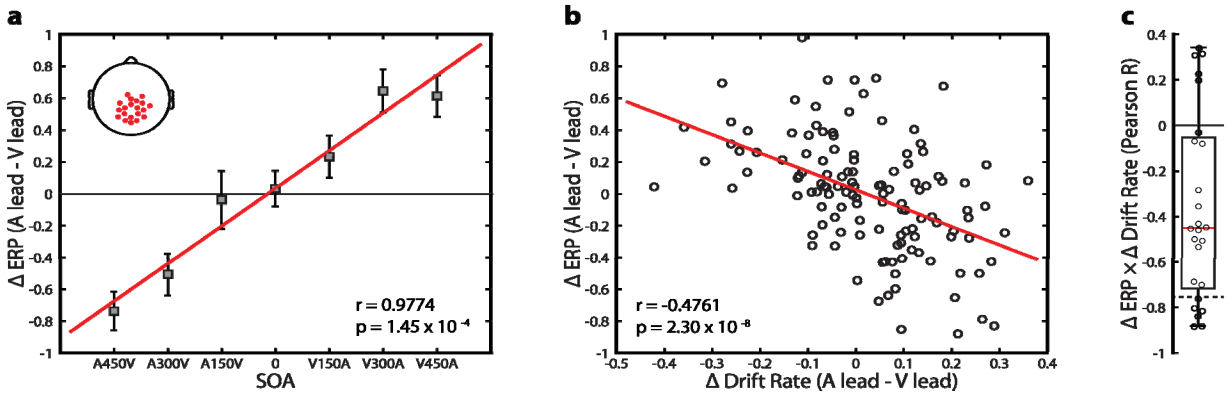

Figure 4-13 Relationship between neural activity and changes in drift rate.

A) Change in voltage between the auditory lead and visual lead conditions for each of the seven SOAs on the current trial. Data was pooled across the 20 electrodes and 200 time points (368-567 ms) found to be significant in all spatiotemporal clusters. Error bars indicate standard error of the mean. Inset: The 20 selected electrodes. Red line indicates the linear regression fit.

B) Relationship between change in drift rate and change in evoked amplitude pooled across the five SOAs which demonstrated significant drift rate modulation based on lead type. Values were pooled as in A. Red line indicates the linear regression fit.

C) Distribution of within subject correlations between change in drift rate and change in evoked potential amplitude. Each dot is one participant. The dashed line indicates the critical value ($r < -0.7545$) for a significant within subject negative correlation ($p < 0.05$, two-tailed) between change in drift rate and change in voltage.

types over a total of 20 electrodes and the 200 time points (368 – 567 ms) found to be significant in all four large SOA conditions. These pooled values differed strongly across conditions ($F_{6,144} = 15.20$, $p = 1.9301 \times 10^{-13}$) and demonstrated an extremely strong linear relationship between SOA and the difference between A leading and V leading ERPs (mean values Pearson $r = 0.9774$, $p = 1.4516 \times 10^{-4}$) (within subject mean Pearson $r = 0.6285$ +/- 0.203, all individual subject $r > 0.157$) (**Fig 4-13A**). We then correlated these voltage differences with individual differences in drift rate, pooling all conditions in which the previous trial was shown to significantly affect drift rate (A450V, A300V, A150V, V300A, and V450A). The relationship between these values was significant using Spearman rank correlation ($r = -0.4930$, $p = 9.4703 \times 10^{-9}$). For display purposes we also computed Pearson correlation, which was similarly

significant ($r = -0.4761$, $p = 2.2952 \times 10^{-8}$) (**Fig 4-13B**). These correlations indicate that individual changes in evoked potential amplitude and individual changes in drift rate attributable to the temporal ordering of the previous stimulus relate to one another across multiple experimental conditions.

We then determined whether changes in drift rate within participants correlated with neural response modulations. We performed linear regression on individual participant's change in evoked potential amplitude and change in drift rate between the two lead types. These correlations were found to be predominantly negative and significantly different from zero (paired sample t-test against zero, $t_{24} = -4.4538$, $p = 1.6659 \times 10^{-4}$) (**Fig 4-13C**). Additionally, of the twenty-five participants, six (24%) presented individually significant negative correlations ($r < -0.7545$), which is a significant proportion when tested with a one sample proportion test against the null probability of 2.5% ($Z = 6.886$, $p = 5.7383 \times 10^{-12}$). Individual participants thus show changes in neural response based on lead type that appear to correspond modestly with their changes in evidence accumulation rate, as might be expected given the noisy nature of EEG signals.

Discussion

We investigated the neural processes underpinning single trial adaptation to temporal asynchrony in audiovisual speech signals. Our behavioral and neural results converge to indicate that this rapid adaptation is engendered by changes in the rate and direction of sensory evidence accumulation. Specifically, when the temporal order of the current stimulus is inverted relative to the previous trial (e.g., visual precedes auditory followed by auditory precedes visual), evidence

accumulates more strongly for an asynchronous choice. These results indicate that rapid temporal recalibration is based in the ability of neural systems to dynamically evaluate feed forward sensory evidence based on immediate sensory history. This adaptive reweighting process, particularly between the senses, is likely to contribute to appropriate integration of sensory inputs and ecologically advantageous behavioral adaptation to the statistics of a dynamic natural world.

Single Trial Recalibration as a Manifestation of Adaptive Decisional Processes

Our primary finding is that the perceptual phenomenon of rapid audiovisual temporal recalibration is rooted in single trial changes in the evidence accumulation process. This is most evident in our modelling results, in which the previous trial strongly affects drift rate, which indexes the strength of available sensory evidence (Gold & Shadlen, 2007). It is also evident in the individualized relationship between RT changes and perceptual threshold. Decisional evidence is often relative, in that a contrast between sensory statistics is what determines decisional outcomes. A salient example of such a relative process can be found in sensory oddball tasks, in which targets typically deviate from standard stimuli along a single stimulus dimension (e.g. frequency, luminance, etc.). Recent work has shown that for both auditory and visual oddballs, evidence accumulation towards a 'target' choice is more rapid when the relative distance between the target and standard, in terms of frequency or luminance, is larger (Twomey et al., 2015). Our diffusion modelling demonstrates a similar finding, in which drift rate is stronger towards an asynchronous choice when the previous stimulus had a highly dissimilar temporal relationship. Importantly, however, our study demonstrates this principle for second order sensory evidence (i.e. a difference of differences).

Physiologically, our results further support the notion that sensory history affects evidence accumulation. We strongly replicate prior work demonstrating that parietal ERPs consistent in topography and timing with the P3B are larger when temporal ordering switches (D. M. Simon, Noel, et al., 2017). We also demonstrate that response locked ERPs strongly exhibit the expected features of decisional signals (Kelly & O'Connell, 2015; O'Connell et al., 2012). Under the aforementioned relative distance framework, in which 'evidence' is the gap between the 'target' (current stimulus) and immediate sensory history (previous stimulus), these neural response modulations would be expected to vary linearly in magnitude with the difference between the current and prior stimulus, as observed. Our physiological and behavioral results thus converge to indicate that second order sensory evidence, in the form of the temporal relationship between the current and former trial, follows the same computational principles as primary sensory evidence.

The Past as a Contributor to Current Sensory Evidence

Our results highlight that the immediate sensory past is a major contributor to feed forward processing of sensory inputs. The nature of the previous stimulus results in substantial changes in behavior on the current trial, highlighting both the flexibility of temporal decision-making and the underlying neuronal processes. While such flexibility has long been recognized in Bayesian (Stocker, 2006) and predictive coding (Rao & Ballard, 1999) accounts of sensory processing, specific instantiations of the neural adaptation process have remained somewhat elusive. Our results strongly indicate that decisional processes are the primary driving force behind rapid audiovisual temporal adaptation. We note that no adaptation was observed in early sensory evoked potentials (i.e. auditory N1 or P2), which is in agreement with both our previous work

(D. M. Simon, Noel, et al., 2017) and sustained audiovisual recalibration experiments (Kosem, Gramfort, & van Wassenhove, 2014). While the EEG signal is limited to indexing phase aligned population signals and thus is blind to more subtle network dynamics (Nunez & Srinivasan, 2006), we nonetheless believe this indicates that processing changes within the earliest cortical sensory circuits are not primary contributors to single trial recalibration. We do not rule out, however, that changes within primary sensory networks could function to store information about the temporal environment, as has been shown during sustained adaptation (Kosem et al., 2014). Features such as selective synchronization of feed forward and feedback connections (Bosman et al., 2012) or internal spike correlation (Carnevale, de Lafuente, Romo, & Parga, 2012) are hidden within the EEG population signal and may be highly distributed across cortical regions (Siegel, Buschman, & Miller, 2015). As drift rate indexes the strength of sensory evidence feeding into the decision process, such 'hidden' mechanisms likely make major contributions to changes in feed forward evaluation of sensory information.

We previously hypothesized that the lack of neural modulations in early evoked responses may represent a dissociation between rapid and sustained audiovisual temporal recalibration (D. M. Simon, Noel, et al., 2017). These differing types of adaptation have been shown to be behaviorally separable (Van der Burg, Alais, & Cass, 2015), but to our knowledge have not yet been fully dissociated in terms of neural mechanism. Previous work on sensorimotor recalibration over longer time scales suggests that changes in more putatively sensory ERPs occurs during sustained recalibration (Stekelenburg, Sugano, & Vroomen, 2011). In light of this work, our current findings indicate that flexible circuits feeding into decisional networks adapt within a single trial, while a larger number of trials is likely required to transfer adaptation to sensory processing regions. Alternatively, the computed nature of the evidence being probed in

our experiment might play an important role. Computing a relational comparison across trials may require a computational process in which relative timing of individual trials (i.e. stimulus pairings) is already stored in a supramodal form. Such representations are ill suited to primary visual or auditory areas and are likely to be mediated through activity changes in distributed or fronto-parietal networks (Bizley, Jones, & Town, 2016; Heekeren, Marrett, & Ungerleider, 2008).

Perceptual Recalibration in Decisional Networks

Our study places audiovisual rapid recalibration into a generalized decisional framework known to operate supramodally (O'Connell et al., 2012; Twomey et al., 2015). We also show that this decisional framework applies despite the second order nature of the perturbations in task relevant sensory information. These findings suggest rapid recalibration may be a specific manifestation of a more generalized process by which flexible decisional networks evaluate incoming information differently based on sensory history. Previous approaches have highlighted the flexibility of the decisional processes subtending perception, including their ability to dynamically adapt to perturbations of the available evidence (Huk & Shadlen, 2005; O'Connell et al., 2012). We believe that dynamic weighting of incoming sensory inputs based on immediate history is a similarly important form of cognitive flexibility. Consideration of such flexibility is only further emphasized by our demonstration that this adaptation process makes particularly strong contributions to behavioral variability when the available sensory evidence is ambiguous.

The notion that flexibility in feed forward neural networks underlies rapid recalibration is also of particular interest due to recent demonstrations that this process is atypical in individuals with autism spectrum disorder (Noel, De Niear, Stevenson, Alais, & Wallace, 2016; Turi,

Karaminis, Pellicano, & Burr, 2016). ASD is often characterized by a loss of behavioral (D'Cruz et al., 2013) and neural (Catarino, Churches, Baron-Cohen, Andrade, & Ring, 2011; D. M. Simon, Damiano, et al., 2017) flexibility, and has well documented differences in P3B amplitude (Cui, Wang, Liu, & Zhang, 2017). Furthermore, differences in perceptual performance in ASD have been hypothesized to relate to differences in neural instantiations of Bayesian inference (Rosenberg, Patterson, & Angelaki, 2015; Turi et al., 2016). Impaired flexibility in the neural mechanisms subtending transformation of sensory evidence into decisions forms a ready explanation for this constellation of findings and clearly warrants further investigation.

Despite our convergent behavioral, modelling, and physiological results, our study is not without limitations. In particular, the impact of the decision itself (synchronous or asynchronous) on neural responses remains unexplored, as our current experimental design is unable to disentangle differences in choice from differences in the preceding stimulus. Future work would be well served to examine the importance of this factor. Elucidating the location and nature of neural activity which encodes the temporal order of the previous stimulus also remains an important goal for understanding the mechanistic basis of rapid recalibration.

Conclusion

Our results indicate that changes in decisional processes substantially contribute to single trial adaptation to audiovisual temporal asynchrony. Specifically, the rate and direction of evidence accumulation is affected by the similarity between current sensory evidence and the immediate sensory past. Despite the audiovisual speech content of the stimuli in our experiment, the neural recalibration processes we report might also underpin numerous other single trial audiovisual

adaptation phenomena. Such rapid adaptation may make substantial contributions to the flexibility and adaptability of human behavior in the face of a dynamic sensory environment.

References

Bizley, J. K., Jones, G. P., & Town, S. M. (2016). Where are multisensory signals combined for perceptual decision-making? *Curr Opin Neurobiol, 40*, 31-37.

Bosman, C. A., Schoffelen, J. M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., . . . Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron, 75*(5), 875-888.

Carnevale, F., de Lafuente, V., Romo, R., & Parga, N. (2012). Internal signal correlates neural populations and biases perceptual decision reports. *Proc Natl Acad Sci U S A, 109*(46), 18938-18943.

Catarino, A., Churches, O., Baron-Cohen, S., Andrade, A., & Ring, H. (2011). Atypical EEG complexity in autism spectrum conditions: a multiscale entropy analysis. *Clin Neurophysiol, 122*(12), 2375-2383.

Cui, T., Wang, P. P., Liu, S., & Zhang, X. (2017). P300 amplitude and latency in autism spectrum disorder: a meta-analysis. *Eur Child Adolesc Psychiatry, 26*(2), 177-190.

D'Cruz, A. M., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., & Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology, 27*(2), 152-160.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci, 7*(7), 773-778.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu Rev Neurosci, 30*, 535-574.

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate

human perceptual decision making. *Nature Reviews Neuroscience, 9*(6), 467-479.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience, 25*(45), 10420-10436.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol, 111*(10), 1745-1758.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *J Neurosci, 33*(50), 19434-19441.

Kelly, S. P., & O'Connell, R. G. (2015). The neural processes underlying perceptual decision making in humans: recent progress and future directions. *J Physiol Paris, 109*(1-3), 27-37.

Kosem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage, 92*, 274-284.

Loughnane, G. M., Newman, D. P., Bellgrove, M. A., Lalor, E. C., Kelly, S. P., & O'Connell, R. G. (2016). Target Selection Signals Influence Perceptual Decisions by Modulating the Onset and Rate of Evidence Accumulation. *Curr Biol, 26*(4), 496-502.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods, 164*(1), 177-190.

Murray, M. M., & Wallace, M.T. (2012). *The Neural Bases of Multisensory Processes* (M. M. Murray, Wallace, M.T. Ed.). Boca Raton, FL: CRC Press.

Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2016). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Res*.

Nunez, Paul L., & Srinivasan, Ramesh. (2006). *Electric fields of the brain : the neurophysics of EEG* (2nd ed.). Oxford ; New York: Oxford University Press.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat Neurosci, 15*(12), 1729-1735.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software

for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci, 2011*, 156869.

Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr Clin Neurophysiol, 66*(1), 75-81.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci, 2*(1), 79-87.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput, 20*(4), 873-922.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347-356.

Rosenberg, A., Patterson, J. S., & Angelaki, D. E. (2015). A computational perspective on autism. *Proc Natl Acad Sci U S A, 112*(30), 9158-9165.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. *Cerebral Cortex, 17*(5), 1147-1153.

Siegel, M., Buschman, T. J., & Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science, 348*(6241), 1352-1355.

Simon, D. M., Damiano, C. R., Woynaroski, T. G., Ibanez, L. V., Murias, M., Stone, W. L., . . . Cascio, C. J. (2017). Neural Correlates of Sensory Hyporesponsiveness in Toddlers at High Risk for Autism Spectrum Disorder. *J Autism Dev Disord*.

Simon, D. M., Noel, J. P., & Wallace, M. T. (2017). Event Related Potentials Index Rapid Recalibration to Audiovisual Temporal Asynchrony. *Front Integr Neurosci, 11*, 8.

Simon, D.M., & Wallace, M. T. (2017). Integration and Temporal Processing of Asynchronous Audiovisual Speech. *Journal of Cognitive Neuroscience*.

Stekelenburg, J. J., Sugano, Y., & Vroomen, J. (2011). Neural correlates of motor-sensory temporal recalibration. *Brain Res, 1397*, 46-54.

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Exp Brain Res, 227*(2), 249-261.

Stocker, A. A., Simoncelli, E. P. (2006). *Sensory adaptation within a Bayesian framework for perception*.

Sumby, W.H., Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *J. Acoust. Soc. Am., 26*(2), 212-215.

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports, 6*.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *Eur J Neurosci, 42*(1), 1636-1643.

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *J Neurosci, 33*(37), 14633-14637.

Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Sci Rep, 5*, 14526.

Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proc Biol Sci, 282*(1804), 20143083.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychon Bull Rev, 14*(6), 1011-1026.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: a DMAT primer. *Behav Res Methods, 40*(1), 61-72.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion Models in Experimental Psychology A Practical Introduction. *Experimental Psychology, 60*(6), 385-402.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: an empirical validation. *Mem Cognit, 32*(7), 1206-1220.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res, 22*(1), 32-35.

CHAPTER V

DISCUSSION

Summary of Results

Multisensory temporal processing and multisensory temporal integration are fundamental contributors to the transformation of sensory inputs into a unified and coherent perceptual representation of the world. These processes play a critical role in everyday functions such as audiovisual speech perception, and thus make substantial contributions to ecologically important behavior (Crosse, Butler, & Lalor, 2015; Crosse, Di Liberto, & Lalor, 2016; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). The demonstration that these integrative processes are disrupted in a host of developmental disabilities motivates basic research into the role temporal structure on the processing and integration of sensory inputs (Francisco, Jesse, Groen, & McQueen, 2017; Hass et al., 2017; Martin, Giersch, Huron, & van Wassenhove, 2013; Stevenson et al., 2017; Stevenson, Siemann, Schneider, et al., 2014). In this document, the individual stages of multisensory temporal processing in the human brain were investigated through a combination of EEG and concurrent psychophysical tasks. By taking an admittedly reductionist view focused on individual processing stages this work succeeded in offering a unique characterization of how multisensory temporal information is processed.

Chapter 2 thoroughly explored how temporal concordance between the auditory and visual streams modulates the degree of multisensory integration for simple audiovisual speech. This

study revealed a striking resemblance between the neural tuning functions for alpha and theta power and participant's behavioral responses. This serves as comprehensive evidence that cross modal temporal concordance contributes to reductions in the amplitude of auditory evoked potentials, which has been identified as an important neural correlate of audiovisual speech integration (Baart, 2016; Besle, Fort, Delpuech, & Giard, 2004; van Wassenhove, Grant, & Poeppel, 2005). Importantly, the surprising width of this tuning function (~500ms) is in strong agreement with complimentary approaches such as stimulus reconstruction (a form of system identification), which have similarly found an extended window for audiovisual speech integration (Crosse et al., 2015; Crosse et al., 2016). This extended temporal window also strongly contrasts with experiments examining temporal integration of auditory only speech using similar reconstruction methods and comparable speech stimuli. These studies typically find that even under extremely noisy conditions, which extend integration times, temporal integration of auditory speech occurs over a substantially shorter timescale of approximately 200-300ms (N. Ding & Simon, 2013a, 2013b). Chapter 2 thus adds new evidence to the notion extended by (Crosse et al., 2016), that audiovisual temporal integration of speech signals might be a privileged operation occurring over larger temporal scales than unisensory temporal integration. This notion of an extended audiovisual integration period is particularly compelling in light of the 2-5 Hz frequencies (200-500ms cycle time) at which auditory and visual speech streams correlate (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). These extended cycle times necessitate a 500ms window of integration to fully capture auditory and visual temporal correspondence. Previous work has also shown that audiovisual speech integration is in fact strongest at the lower end of this frequency range (Crosse et al., 2015), which further emphasizes the need for a relatively extended audiovisual speech integration window.

This study also isolated for the first time a potential oscillatory correlate of temporal information processing. This signal manifested in the lower portion of the theta band and not only changed across conditions, but also correlated with individual participant's performance on the task. This signal also shared both spatial and frequency characteristics with previous investigations of stimulus congruence processing for both visual stimuli (Hanslmayr et al., 2008) and audiovisual speech stimuli (Roa Romero, Keil, Balz, Gallinat, & Senkowski, 2016). These characteristics strongly suggest that this low theta signal is related to processing of temporal structure, and is associated with the brain computing a mismatch between the timing of sensory components.

The isolation of a neural correlate of temporal processing has exciting implications as it allows examinations focused on this signal's properties. We took just such an approach in chapter 3, where we interrogated whether this signal requires deliberate processing of temporal structure and then characterized the nature and direction of information flow during temporal processing. We found that this potential congruence signal only appeared during deliberate processing of temporal structure, indicating that it is indeed associated with top-down volitional processing of audiovisual timing. We also found that phase coupling in a larger network involving this signal's local circuit was modulated by the presence or absence of the temporal processing task. This finding is consistent with the hypothesis that coherence between brain regions forms the backbone of directed information flow in the brain and interacts with localized processing (Fries, 2005; Siegel, Donner, & Engel, 2012; Womelsdorf et al., 2007; Womelsdorf, Vinck, Leung, & Everling, 2010). For asynchronous stimuli in particular, coupling was substantially strengthened during active temporal processing, and the phase lags suggested that information was moving from right temporal regions to central areas and then to putatively

visual occipital areas. This suggests that a functional network, constructed via coherence in the lower theta band, supports audiovisual temporal processing. The bilateral nature and temporal location of the connected electrodes also suggests that the STS might constitute a contributor to this network, consistent with its role as a hub for audiovisual temporal processing and audiovisual speech processing (Powers, Hevey, & Wallace, 2012; Schepers, Schneider, Hipp, Engel, & Senkowski, 2013). A role for the STS in coherence based networks has also previously been shown in the form of preferential synchronization with auditory cortex when useful visual inputs are present (Maier, Chandrasekaran, & Ghazanfar, 2008). The data thus suggests that connectivity between STS and other brain areas such as auditory cortex is substantially enhanced when transfer of temporal information is needed. This study thus offers a glimpse of the neurophysiological substrates involved in top down control of audiovisual temporal information flow.

Lastly, in chapter 4 we established the physiological basis of an important form of neural plasticity known as rapid recalibration. Previous reports have indicated that for audiovisual simultaneity judgment the reported percept is not based solely on the sensory stimulus presented, but also incorporates information from both the immediate (Van der Burg, Alais, & Cass, 2013; Van der Burg & Goodbourn, 2015; Van der Burg, Orchard-Mills, & Alais, 2014) and long term (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Van der Burg, Alais, & Cass, 2015; Vroomen, Keetels, de Gelder, & Bertelson, 2004) sensory history. This previous work further hypothesized that the mechanistic basis of this rapid adaptation would be rooted in functional changes in sensory processing regions. Judging whether a stimulus is synchronous or asynchronous is a form of perceptual decision-making, however, and performance of the SJ task thus necessarily involves evidence accumulators located in frontal and parietal cortex (Heekeren, Marrett, &

Ungerleider, 2008). This accumulation process utilizes the sensory evidence available, which consists of relative timing computations by the neural network examined in chapter 3, and could just as easily be the locus of rapid adaptation.

Importantly, this decisional process has well described theoretical characteristics in which sensory evidence integrates over time until it is absorbed by a decisional boundary (Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004). This accumulation process is also directly observable as ramping activity both in single neurons (Bollimunta, Totten, & Ditterich, 2012; L. Ding & Gold, 2012; Shadlen & Newsome, 2001) (for reviews see: (Gold & Shadlen, 2007; Smith & Ratcliff, 2004)) and in the neural population signals indexed by EEG and MEG (O'Connell, Dockree, & Kelly, 2012; Schurger, Sitt, & Dehaene, 2012; Smyrnis et al., 2012). This ramping form of temporal integration, which is distinct from multisensory perceptual integration, is observable as a build-to-action threshold that rises at a rate equivalent to the integral of the sensory evidence (O'Connell et al., 2012). Such temporal integration is believed to be an optimal form of decision making when individual decisional units (i.e. neurons) have sparse or noisy representations of the available evidence (Gold & Shadlen, 2007). This accumulation process is also known to underlie other canonical action computations such as volitional initiation of motor action (Hanes & Schall, 1996), suggesting that temporal integration and boundary absorption are ubiquitous properties of action and perception. This strong conceptual framework and its corresponding biologically informed models such as drift diffusion (Vandekerckhove & Tuerlinckx, 2007) allow for robust interrogation of the decisional contribution to multisensory temporal perceptual plasticity.

Using a combination of electroencephalography and drift diffusion modelling we demonstrated conclusively that physiological indices of rapid adaptation to asynchronous

audiovisual speech occurs seemingly exclusively in decisional circuits. We found no evidence of adaptation in earlier sensory ERP components or sensory encoding time and found that the accumulation process itself conformed to canonical descriptions. While many signal dynamics are hidden in the EEG population signal (Nunez & Srinivasan, 2006), the strong correspondence between changes in physiology and modelling of the underlying decision variable serves as evidence for a major decisional component to this phenomenon. This does not exclude the contribution of sensory circuits altogether, but constrains their possible contributions to be either non-phase locked (i.e. induced phase changes, see: (Kosem, Gramfort, & van Wassenhove, 2014) for an example), or to take the form of differences in forward information flow to the decisional circuit.

Together, the results of these studies offers a unique characterization of temporal processing for audiovisual speech as information ascends the cortical hierarchy. Processing begins in low-level sensory cortical regions, in which suppressive multisensory interactions are already present. More distributed processing circuits are then recruited through neural coherence, which compare or combine temporal representations from these areas. Finally, this sensory evidence accumulates in parietal decision circuits until reaching a decisional threshold and triggering the behavioral response.

Implications of Main Findings

Basic Science Implications

The current work indicates that the precise timing of auditory and visual stimuli plays a crucial role in the degree to which visual inputs dampen cortical responses in early auditory areas.

Importantly, this effect occurs over a notably prolonged temporal window congruent with both behavioral (Munhall, Gribble, Sacco, & Ward, 1996; van Wassenhove, Grant, & Poeppel, 2007), imaging (Stevenson, Altieri, Kim, Pisoni, & James, 2010), and physiological (Crosse et al., 2016) studies, all of which point to substantial temporal tolerance in audiovisual speech processing. Similar sub-additive multisensory interactions have previously been identified to be a potential hallmark of improvement in information content in neural signals (Angelaki, Gu, & DeAngelis, 2009; Kayser, Logothetis, & Panzeri, 2010; Kayser, Petkov, & Logothetis, 2008) and a form of anticipatory predictive coding (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). The current work particularly demonstrates the sensitivity to temporal alignment of early alpha band activity during audiovisual speech processing. This finding is strongly consistent with previous invasive physiological approaches demonstrating ~10 Hz (alpha) inhibition in the auditory cortex during audiovisual processing (Kayser et al., 2008). Alpha oscillations have previously been proposed as an important marker of cortical inhibition crucial for the control of information flow within and between sensory regions (Klimesch, Sauseng, & Hanslmayr, 2007) and have been specifically highlighted as a mechanism of information refinement and resource allocation for auditory sensory representations (Strauss, Wostmann, & Obleser, 2014) (for a speech perception example see (Strauss, Henry, Scharinger, & Obleser, 2015)). Importantly, there is also strong congruence between the generally inhibitory nature of alpha oscillations (Jensen & Mazaheri, 2010; Klimesch et al., 2007) and the generally sub-additive nature of multisensory integration found in experiments using audiovisual speech signals (Baart, 2016). Combined with the more rapid onset of alpha band effects relative to theta band effects, this suggests an important role for rapid inhibition in refinement of the primary auditory cortical response and thus a reduction in overall response magnitude. In the presence of

informative (i.e. concurrent or leading) visual speech, the need for such cortical dampening to occur would be substantially reduced, and thus serves as a marker of the improved processing efficiency proposed by (van Wassenhove et al., 2005). The current work thus suggests that inhibitory processes and information sharpening correspond with audiovisual temporal congruence in a manner consistent with the low frequency correspondence found in the statistics of natural audiovisual speech (Chandrasekaran et al., 2009; Schwartz & Savariaux, 2014).

Representational sharpening could also be characterized as reduction in the neural uncertainty regarding the sensory signal, and thus form the basis of visual influences on auditory Bayesian inference in the brain (Knill & Pouget, 2004). Taken in this context, we demonstrate that cortical representations of simple acoustic signal features sharpen to varying degrees depending on the temporal concordance of the visual input. This sharpening likely makes important contributions to perception, but empirically establishing this relationship remains an important future step. Such a question could be addressed by determining whether temporal tuning and response sharpening, both at the level of the individual neurons (i.e. single unit spikes or multiunit activity) and at the level of the circuit (i.e. LFP) are preferentially present in perceptually relevant neurons. At the level of the individual neuron this could be interrogated using metrics such as choice probability (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996) and determining whether neurons exhibiting response suppression consistent with the observed scalp signal have categorically higher choice probability than neurons which lack response suppression or exhibit disorganized response suppression. This approach has previously been utilized in non-human primates to indicate that visual-vestibular neurons with expected tuning functions likely make larger perceptual contributions than neurons with suboptimal tuning (Gu, Angelaki, & DeAngelis, 2008). Such an approach is further motivated by recent work

demonstrating causal links between auditory cortical activity and perceptual decisions (Tsunada, Liu, Gold, & Cohen, 2016). The current findings thus create an avenue for directly addressing the perceptual contributions of multisensory temporal tuning in auditory cortical neurons by determining if visual speech selectively suppresses or enhances activity in the neurons that most strongly contribute to perception.

The protracted tuning function demonstrated also offers constraints on the neural mechanisms responsible for reductions in auditory cortical response magnitude. A major mechanism proposed for multisensory interactions within the auditory cortex is phase reset (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Lakatos et al., 2009; Thorne, De Vos, Viola, & Debener, 2011), which is proposed to occurring through direct anatomical inputs from other sensory regions (Falchier, Clavagnier, Barone, & Kennedy, 2002; Falchier et al., 2010; Hackett et al., 2007; Smiley & Falchier, 2009). This reset mechanism offers a way to rapidly and flexibly optimize the neural state in the auditory cortex for input based on visual activity or other sensory inputs. In light of the tuning function presented, however, several questions arise. First, generation of such a slow tuning function would require primarily (~2Hz) delta phase reset, but the phase resets observed in auditory cortex do not fall exclusively within this band (Lakatos et al., 2007; Lakatos et al., 2009). In particular, low theta (~4-5Hz) oscillations likely also reset, and given the width of the tuning function have enough time to completely cycle. Itis possible that the lower frequency reset is preferentially considered (i.e. each oscillatory frequency has a distinct integrative process; a mechanism proposed in (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008) and supported by (Chandrasekaran & Ghazanfar, 2009)). Alternatively, the slow tuning function might result from broadband phase resets with distributed timing (i.e. temporal jittering), which would cancel at the population level for high frequencies while maintaining low

frequencies that are more resistant to temporal jitter. Such temporally imprecise reset might be present due to the low temporal precision and relatively smooth rate of change found in visual speech signals, and has been demonstrated to serve adaptive functions in cross-modal attention tasks (Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008). Disambiguating these mechanisms at the level of the EEG signal is not feasible, as the number of neurons active and the phase alignment of their activity both manifest as increased signal amplitude and ITC at the scalp (Nunez & Srinivasan, 2006). Approaching this question with invasive physiological techniques is thus highly appropriate and necessary for answering the question of whether phase reset by the visual input is heterogeneous or whether specific frequency bands are preferentially considered.

Recent work has also revealed the precise nature of timing modulated audiovisual integration in the auditory cortex of the macaque for ecologically valid temporal offsets, which are strongly consistent with the effects we observed (Kayser et al., 2008; Perrodin, Kayser, Logothetis, & Petkov, 2015). We note, however, that the signatures of integration in our study extend well beyond the < 200 ms asynchronies which demonstrated integrative effects in the aforementioned work. The neural correlates of processing during large asynchronies (>200 ms) at the level of individual neurons and the local circuit are thus currently unexplored. Such work is also particularly interesting given the ubiquity of perceptual binding (i.e. reports of 'same time' or illusory perception) at offsets of this size (i.e. the current work or (Stevenson et al., 2010; van Wassenhove et al., 2007)), despite the seeming falloff of integrative effects observed at larger offsets in (Kayser et al., 2008). Examining neural function for these large offsets may thus yield an important opportunity to dissociate low-level integration and high-level perceptual binding (i.e. formation of a multisensory "object") within auditory cortical circuits.

Determining the invasive neural correlates of the observed auditory response suppression at

large temporal offsets might also yield information about these neuron's potential contribution to statistically optimal integration (Ernst & Banks, 2002). A widely observed phenomenon in the integration of multisensory inputs is that integration is optimal relative to the reliability of the sensory inputs. In other words, a sensory input with a high reliability (i.e. spatial localization with vision) carries substantially more integrative weight than one with low reliability (i.e. spatial localization with audition) (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002; Kersten, Mamassian, & Yuille, 2004; Maloney & Mamassian, 2009). Examples of this weighted integration include the dominance of the visual input for spatial localization of the classical ventriloquist's illusion (Slutsky & Recanzone, 2001) and the dominance of audition for temporal ventriloquism (Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Kuling, Kohlrausch, & Juola, 2013). Theoretically, refinement of the auditory neural representations of sensory inputs should lead to increased representational reliability, and thus increased weight during tasks of this type. This could be examined behaviorally by using the temporal offset based sharpening we demonstrate, while probing the optimality of an orthogonal paradigm such as integration of auditory and tactile cues. Such an approach could also be combined with the choice probability method suggested above to determine if choice probability and optimality are linked via this neural tuning function.

Bridging Low-Level integration with decision making

The current work also suggests that changes in decisional processes account for trial-by-trial variability in perceptual thresholds. Previously this temporal adaptation process, rapid recalibration, was implied to depend on adaptation within sensory systems (Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015) based on its specificity to audiovisual stimuli (Van der

Burg et al., 2014). The results in this work indicate that this is not the case, as rapid recalibration causes no observable changes in evoked responses which occur early in time and are readily attributable to feed forward encoding in primary auditory cortices (i.e. the N1 or P2). At the same time, we observed very large differences in neural activity attributable to modality independent decisional processes. This finding of decisional specificity in the rapid recalibration process is further reinforced by the fact that non-decision time, which includes both sensory processing and generation of motor responses (Ratcliff & McKoon, 2008), is unaffected by the nature of the sensory past. Combined with the lack of auditory or visual ERP adaptation found in a nearly identical paradigm using flashes and beeps (Simon, Noel, & Wallace, 2017), adaptation in initial sensory representations is thus unlikely to be a major contributor to the rapid recalibration effect. This finding motivates careful consideration in future work regarding how single trial perceptual recalibration occurs or fails to occur at each stage of the transformation from a sensory input to perception.

For example, temporal adaptation in low-level sensory cortices clearly occurs for low-level stimuli given sufficient adaptation time consisting of either multiple consecutive trials (Stekelenburg, Sugano, & Vroomen, 2011) or long stimulus trains (Kosem et al., 2014). In the first of these studies, changes in early sensory ERPs were observed, while in the latter a clear neural correlate of temporal recalibration was observed in the form of phase shifts in auditory and visual cortex. This indicates that primary sensory regions can indeed adapt to distortions in the temporal statistical structure of the environment under certain circumstances, and furthermore, that this adaptation in auditory cortex has perceptual relevance (Kosem et al., 2014). What parameters of the stimuli or experimental design are necessary to engender this low-level plasticity, however, is currently unknown. In particular, given the differences in temporal

processing observed between simple and complex stimuli (reviewed extensively in chapter 1), it is unclear if stimulus experience, which could also be viewed as the strength of Bayesian priors, plays any role in these low-level adaptation effects. Similarly, both of these studies utilized sustained adaptation to varying degrees, and it is possible that putatively 'low level' neural adaptation only emerges after repeated stimulus presentations. Empirical examination of the role stimulus complexity and experimental structure play in early adaptation effects is still needed, yielding a number of opportunities for future research.

Given the lack of evidence of changes in early sensory encoding strength in the current work, the next stage to consider is transfer of information between the networks encoding stimulus features and the networks making perceptual decisions. By nature, the formation of a perceptual decision based on sensory evidence requires feed forward transmission of information from sensory circuits to fronto-parietal circuits governing the decision process (Heekeren et al., 2008). In other words, circuits processing low-level sensory information *must* transfer relevant information to circuits that contain the evidence accumulators and form the decision. For a practical example of this process, when making a decision regarding motion coherence, neurons in the middle temporal area (MT/V5) strongly represent motion (i.e. sensory evidence), but only weakly represent choice (Britten et al., 1996; Britten, Shadlen, Newsome, & Movshon, 1992; Shadlen, Britten, Newsome, & Movshon, 1996). These neurons then and project to neurons in the lateral intraparietal area (LIP), which very strongly represent motor choice and demonstrate accumulation characteristics (Huk & Shadlen, 2005; Shadlen & Newsome, 2001) (for a more recent and nuanced view of the representation of decision and choice in these areas see: (Huk, Katz, & Yates, 2017; Yates, Park, Katz, Pillow, & Huk, 2017)). It has previously been demonstrated that in motion coherence tasks, motion adaptation causes changes in the directional

tuning of the neurons representing the motion (Kohn & Movshon, 2004). In other words, the gross population activity present in a scalp ERP is unchanged, but the fundamental nature of the message (i.e. sensory evidence) being read by the decisional circuit is altered. The current work suggests a similar phenomenon, except that rather than area MT the 'evidence' likely resides in the distributed network elucidated in chapter 3.

The final decisional stage, accumulation, was observed to follow the canonical model of decisional evidence accumulation. This included a build to threshold decision variable, which was found to be consistent across all conditions and a slope (i.e. information accumulation rate) that corresponded with the perceptual ambiguity of the sensory stimulus. These findings are novel in the current multisensory context, but are well explored for other types of perceptual decision-making. For example, the accumulation process for both audition and vision was well characterized in (O'Connell et al., 2012), which found that the trajectory of the decision signal tracked the integral of the sensory evidence present in primary cortices. This is consistent with the theoretical underpinnings of an optimal accumulator model, which integrates noisy evidence over an interval until a decision threshold is reached (Gold & Shadlen, 2007; Smith & Ratcliff, 2004). By integrating information in this manner, neural systems are able to quickly reach decisions which are robust to sensory and biological noise.

The current work thus establishes a strong impetus for understanding how adaptation occurs in the intermediate messaging step, as this step is likely responsible for the behavioral plasticity observed as well as the trial-by-trial changes in accumulation rate. Additionally, trial-by-trial variability in this process likely utilizes specialized circuitry involving auditory and visual brain regions, as rapid temporal recalibration has previously been shown to be specific to audiovisual stimuli (Van der Burg et al., 2014). Work addressing the underlying coupling process might

require invasive techniques, given the fine spatial and neural population specificity required to carry out such a specific computational operation.

More generally, the finding that the audiovisual sensory past shapes the trajectory of the decision process is important, because it highlights the need to conceive of perceptual processes as interconnected with decisional and motor circuitry in the brain. This is reinforced by recent examinations noting that motor cortical regions actively participate in audiovisual speech integration (Park, Kayser, Thut, & Gross, 2016). Consideration of such mechanisms is important in psychophysical experiments such as SJ, in which binary behavioral reports (and consequently, binary decisions) are extracted from neural representations which are almost certainly initially continuous. This is also particularly true for multisensory timing experiments, in which comparison across potentially disparate sensory timing computations may be required to extract the pertinent decisional information. Recent calls to examine multisensory decision-making have emphasized that multisensory interactions occur at many levels of the decisional process, but that the *decisional relevance* of integration at each step is rather poorly understood (Bizley, Jones, & Town, 2016). The current work expands this call by strongly indicating that, for audiovisual multisensory temporal decisions, it is crucially important to consider adaptation within sensory circuits and mechanisms of selective feed forward transfer to decisional circuits.

The role of coupling between sensory systems and decisional accumulators also remain relatively unexplored in multisensory contexts. Important mechanisms of neural information transmission such as selective synchronization (Bosman et al., 2012) and neural coherence (Engel, Gerloff, Hilgetag, & Nolte, 2013) almost certainly play a crucial role in this processing stage. Coupling between brain circuits must play an important role in multisensory decision making, as neurons in primary sensory cortices seem to encode decision relevant information

only for the primary sense despite responding to other modalities (Lemus, Hernandez, Luna, Zainos, & Romo, 2010). This indicates that the information for a multisensory perceptual decision is likely not present within a single sensory circuit, necessitating some form of interregional transfer. In chapter 3, a task switch paradigm was used to isolate low frequency phase coupling which supports the SJ task, but additional clarity is clearly needed in regards to the directionality and content of information transfer. The coupling observed in the SJ task also did not target frontal or parietal scalp locations where decisional accumulation signals are typically found, indicating that it may not be an indicator of sensory to decisional coupling (Kelly & O'Connell, 2013, 2015; Loughnane et al., 2016; O'Connell et al., 2012; Twomey, Murphy, Kelly, & O'Connell, 2015). This discrepancy may occur because of methodological differences between established physiological approaches to connectivity and the nature of the accumulation process. In particular, the temporal integration process found in decisional accumulators is generally considered optimal due to the noisy nature of evidence present for individual trials. Non-invasive investigations of cortical connectivity, however, are generally based on consistency in phase relationships across trials (for discussions of analytical methods for brain synchrony see: (Lachaux, Rodriguez, Martinerie, & Varela, 1999; Nolte et al., 2004; Stam, Nolte, & Daffertshofer, 2007; Vinck, Oostenveld, van Wingerden, Battaglia, & Pennartz, 2011)). In other words, increased consistency (i.e. increased functional connectivity) *across* trials may not correspond well with changes in information flow *within* trials. Additional work is clearly needed to clarify the role of neural coherence in decision making, and how differences in coupling strength affect the evidence accumulation process.

Future Directions

Clinical Applications

The current work suggests a number of productive directions for further research. One of these is to extend these approaches to clinical populations which demonstrate multisensory temporal dysfunction such as ASD, schizophrenia, and dyslexia. Multisensory dysfunction in these populations has been hypothesized to cascade into higher order social and cognitive deficits that in many cases primarily characterize these disorders (Wallace & Stevenson, 2014). Currently, the physiological basis of multisensory perceptual disruptions in all of these clinical populations is poorly understood. This is particularly relevant for ASD, in which integrative deficits have long been theorized to underpin differences in perceptual performance (Frith & Happe, 1994; Happe & Frith, 2006). These integrative processes are believed to rely on both high and low frequency coherence (Engel et al., 2013), which has recently been demonstrated to be disrupted during visual perceptual processing in ASD (Peiker et al., 2015). Multisensory interactions necessarily involve information transfer between sensory systems, a process reliant on neural coherence (Senkowski, Schneider, Foxe, & Engel, 2008). Multisensory interactions thus may be particularly vulnerable to oscillatory disruptions observed during sensory processing in ASD (Simon & Wallace, 2016). Similarly, differences in the P3B ERP component, which is the stimulus locked manifestation of evidence accumulation (Kelly & O'Connell, 2015; O'Connell et al., 2012; Twomey et al., 2015) have been widely reported in ASD (Cui, Wang, Liu, & Zhang, 2017). Both theta coherence and the P3B ERP component demonstrated important multisensory interactions in the current work, and extension of these approaches to ASD might yield substantial insight regarding the neural bases of disruptions noted in simultaneity judgment (Noel, De Niear, Stevenson, Alais, & Wallace, 2016; Stevenson, Siemann, Schneider, et al.,

2014; Turi, Karaminis, Pellicano, & Burr, 2016).

Developmental Applications

A second important extension of the current work is to determine the developmental trajectory of these physiological processes. Behavioral measures of audiovisual temporal integration are known to have a strong developmental trajectory in which sensitivity matures in early adulthood (Hillock-Dunn, Grantham, & Wallace, 2016; Hillock-Dunn & Wallace, 2012; Hillock, Powers, & Wallace, 2011; Kaganovich, 2016; Noel, De Niear, Van der Burg, & Wallace, 2016). The neural mechanisms of this relatively slow maturational process are currently unexplored, and the physiological measures established by the current work could be readily generalized to pediatric populations. This is particularly important for temporal acuity in regards to speech stimuli, as audiovisual facilitation of speech comprehension is known to share a similarly extended developmental trajectory which (Ross et al., 2011). Importantly, audiovisual speech processing also has a well-described developmental trajectory in terms of neural correlates (Kaganovich & Schumaker, 2014; Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014), offering a potentially robust extension to the temporal asynchrony manipulations used in the current work. Such an approach is only further motivated by the documented developmental trajectory for audiovisual speech illusions (Sekiyama & Burnham, 2008), which reinforces the notion of strong developmental influences on audiovisual integration and perception. Given the aforementioned focus on extending the current work to clinical populations, it also bears mention that developmental and clinical factors clearly interact with one another in terms of audiovisual speech perception (Foxe et al., 2015; Stevenson, Siemann, Woynaroski, et al., 2014). Understanding the aforementioned neural bases of dysfunction in clinical populations thus also

requires a comprehensive characterization of multisensory temporal function in developing individuals.

Identifying Mechanisms of Audiovisual Perceptual Learning

Lastly, by characterizing the neural mechanisms involved in multisensory temporal processing at multiple levels, the current work creates a unique opportunity to determine how perceptual learning paradigms shape neural activity. Perceptual learning has been previously shown to improve multisensory temporal acuity for visually leading impulse stimuli (Cecere, Gross, & Thut, 2016; Powers, Hillock, & Wallace, 2009). In an fMRI experiment, these behavioral improvements were shown to occur due to strengthening of a distributed multisensory network involving the SC, STS, and primary sensory cortices (Powers et al., 2012). The physiological basis of training based improvements in temporal acuity is currently unexplored, but has previously been proposed to involve changes in neural coherence. The current work establishes that local and long-range theta synchronization is involved in temporal processing, and determining how perceptual training shapes synchronization is an important future step. Recent physiological evidence also indicates the presence order specific network recruitment during the SJ task (Cecere, Gross, Willis, & Thut, 2017), which mirrors the asymmetry of trainability. This suggests that coherence based approaches might disambiguate how activity in the visually leading brain network is shaped by training relative to the training resistant auditory leading brain network. Combined with the clinical and developmental characterizations suggested above, such work could serve as a crucial step in determining whether perceptual training might serve as a method of remediation for multisensory temporal dysfunction in clinical populations.

Conclusion

Multisensory temporal processing is fundamentally important for the appropriate integration of ecologically important audiovisual signals such as speech and makes substantial contributions to human perception and behavior. The current work offers a uniquely comprehensive physiological characterization of the neural processes involved in this important function in the mature human brain. Physiological indicators of multisensory temporal interactions were observed at every level of the cortical hierarchy, and spanned from initial sensory processing in localized cortical regions to high-level decisional circuits which accumulate sensory evidence to form decisions. Taken as a whole, the current work offers substantial insights into how the brain aggregates and integrates information from an environment in which sensory signals occur with varying degrees of temporal correspondence. These fundamental neural computations underlie the flexibility and adaptability of human behavior in naturalistic environments and offer numerous forward paths for productive future research.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol, 14*(3), 257-262.

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Curr Opin Neurobiol, 19*(4), 452-458.

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension - 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology, 50*(1-2), 157-163.

Baart, M. (2016). Quantifying lip-read-induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays. *Psychophysiology, 53*(9), 1295-1306.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am A Opt Image Sci Vis, 20*(7), 1391-1397.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension - 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology, 50*(1-2), 147-155.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci, 20*(8), 2225-2234.

Bizley, J. K., Jones, G. P., & Town, S. M. (2016). Where are multisensory signals combined for perceptual decision-making? *Curr Opin Neurobiol, 40*, 31-37.

Bollimunta, A., Totten, D., & Ditterich, J. (2012). Neural Dynamics of Choice: Single-Trial Analysis of Decision-Related Activity in Parietal Cortex. *Journal of Neuroscience, 32*(37), 12684-12701.

Bosman, C. A., Schoffelen, J. M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., . . . Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron, 75*(5), 875-888.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience, 13*(1), 87-100.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci, 12*(12), 4745-4765.

Cecere, R., Gross, J., & Thut, G. (2016). Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *Eur J Neurosci, 43*(12), 1561-1568.

Cecere, R., Gross, J., Willis, A., & Thut, G. (2017). Being First Matters: Topographical Representational Similarity Analysis of ERP Signals Reveals Separate Networks for Audiovisual Temporal Binding Depending on the Leading Sense. *J Neurosci, 37*(21), 5274-5287.

Chandrasekaran, C., & Ghazanfar, A. A. (2009). Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. *J Neurophysiol, 101*(2), 773-788.

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

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *J Neurosci, 35*(42), 14195-14204.

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye Can Hear Clearly Now: Inverse Effectiveness in Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal Integration. *J Neurosci, 36*(38), 9888-9895.

Cui, T., Wang, P. P., Liu, S., & Zhang, X. (2017). P300 amplitude and latency in autism spectrum disorder: a meta-analysis. *Eur Child Adolesc Psychiatry, 26*(2), 177-190.

Ding, L., & Gold, J. I. (2012). Neural Correlates of Perceptual Decision Making before, during, and after Decision Commitment in Monkey Frontal Eye Field. *Cerebral Cortex, 22*(5), 1052-1067.

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

Ding, N., & Simon, J. Z. (2013b). Robust cortical encoding of slow temporal modulations of speech. *Adv Exp Med Biol, 787*, 373-381.

Engel, A. K., Gerloff, C., Hilgetag, C. C., & Nolte, G. (2013). Intrinsic Coupling Modes: Multiscale Interactions in Ongoing Brain Activity. *Neuron, 80*(4), 867-886.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci, 22*(13), 5749-5759.

Falchier, A., Schroeder, C. E., Hackett, T. A., Lakatos, P., Nascimento-Silva, S., Ulbert, I., . . . Smiley, J. F. (2010). Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cereb Cortex, 20*(7), 1529-1538.

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H. P., Russo, N. N., Blanco, D., . . . Ross, L. A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with Autism Spectrum Disorder (ASD) and their resolution during early

adolescence. *Cereb Cortex, 25*(2), 298-312.

Francisco, A. A., Jesse, A., Groen, M. A., & McQueen, J. M. (2017). A General Audiovisual Temporal Processing Deficit in Adult Readers With Dyslexia. *J Speech Lang Hear Res, 60*(1), 144-158.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9*(10), 474-480.

Frith, U., & Happe, F. (1994). Autism: beyond "theory of mind". *Cognition, 50*(1-3), 115-132.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci, 7*(7), 773-778.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu Rev Neurosci, 30*, 535-574.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience, 11*(10), 1201-1210.

Hackett, T. A., Smiley, J. F., Ulbert, I., Karmos, G., Lakatos, P., de la Mothe, L. A., & Schroeder, C. E. (2007). Sources of somatosensory input to the caudal belt areas of auditory cortex. *Perception, 36*(10), 1419-1430.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274*(5286), 427-430.

Hanslmayr, S., Pastotter, B., Bauml, K. H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *J Cogn Neurosci, 20*(2), 215-225.

Happe, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J Autism Dev Disord, 36*(1), 5-25.

Hass, K., Sinke, C., Reese, T., Roy, M., Wiswede, D., Dillo, W., . . . Szycik, G. R. (2017). Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion. *Cogn Neuropsychiatry, 22*(2), 145-158.

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience, 9*(6), 467-479.

Hillock-Dunn, A., Grantham, D. W., & Wallace, M. T. (2016). The temporal binding window for audiovisual speech: Children are like little adults. *Neuropsychologia, 88*, 74-82.

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory

temporal binding window persist into adolescence. *Dev Sci, 15*(5), 688-696.

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia, 49*(3), 461-467.

Huk, A. C., Katz, L. N., & Yates, J. L. (2017). The Role of the Lateral Intraparietal Area in (the Study of) Decision Making. *Annu Rev Neurosci, 40*, 349-372.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience, 25*(45), 10420-10436.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front Hum Neurosci, 4*, 186.

Kaganovich, N. (2016). Development of sensitivity to audiovisual temporal asynchrony during midchildhood. *Dev Psychol, 52*(2), 232-241.

Kaganovich, N., & Schumaker, J. (2014). Audiovisual integration for speech during mid-childhood: Electrophysiological evidence. *Brain Lang, 139C*, 36-48.

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Visual enhancement of the information representation in auditory cortex. *Curr Biol, 20*(1), 19-24.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cereb Cortex, 18*(7), 1560-1574.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *J Neurosci, 33*(50), 19434-19441.

Kelly, S. P., & O'Connell, R. G. (2015). The neural processes underlying perceptual decision making in humans: recent progress and future directions. *J Physiol Paris, 109*(1-3), 27-37.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271-304.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res Rev, 53*(1), 63-88.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci, 27*(12), 712-719.

Knowland, V. C., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. (2014). Audio-visual speech perception: a developmental ERP investigation. *Dev Sci, 17*(1), 110-124.

Kohn, A., & Movshon, J. A. (2004). Adaptation changes the direction tuning of macaque MT neurons. *Nature Neuroscience, 7*(7), 764-772.

Kosem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage, 92*, 274-284.

Kuling, I. A., Kohlrausch, A., & Juola, J. F. (2013). Quantifying temporal ventriloquism in audiovisual synchrony perception. *Attention Perception & Psychophysics, 75*(7), 1583-1599.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum Brain Mapp, 8*(4), 194-208.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

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., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron, 64*(3), 419-430.

Lemus, L., Hernandez, A., Luna, R., Zainos, A., & Romo, R. (2010). Do sensory cortices process more than one sensory modality during perceptual judgments? *Neuron, 67*(2), 335-348.

Loughnane, G. M., Newman, D. P., Bellgrove, M. A., Lalor, E. C., Kelly, S. P., & O'Connell, R. G. (2016). Target Selection Signals Influence Perceptual Decisions by Modulating the Onset and Rate of Evidence Accumulation. *Curr Biol, 26*(4), 496-502.

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology, 18*(13), 963-968.

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience, 26*(1), 147-155.

Martin, B., Giersch, A., Huron, C., & van Wassenhove, V. (2013). Temporal event structure and timing in schizophrenia: preserved binding in a longer "now". *Neuropsychologia, 51*(2), 358-371.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys, 58*(3), 351-362.

Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2016). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Res*.

Noel, J. P., De Niear, M., Van der Burg, E., & Wallace, M. T. (2016). Audiovisual Simultaneity Judgment and Rapid Recalibration throughout the Lifespan. *PLoS One, 11*(8), e0161698.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin Neurophysiol, 115*(10), 2292-2307.

Nunez, Paul L., & Srinivasan, Ramesh. (2006). *Electric fields of the brain : the neurophysics of EEG* (2nd ed.). Oxford ; New York: Oxford University Press.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat Neurosci, 15*(12), 1729-1735.

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

Peiker, I., David, N., Schneider, T. R., Nolte, G., Schottle, D., & Engel, A. K. (2015). Perceptual Integration Deficits in Autism Spectrum Disorders Are Associated with Reduced Interhemispheric Gamma-Band Coherence. *J Neurosci, 35*(50), 16352-16361.

Perrodin, C., Kayser, C., Logothetis, N. K., & Petkov, C. I. (2015). Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. *Proc Natl Acad Sci U S A, 112*(1), 273-278.

Powers, A. R., 3rd, Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *J Neurosci, 32*(18), 6263-6274.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput, 20*(4), 873-922.

Roa Romero, Y., Keil, J., Balz, J., Gallinat, J., & Senkowski, D. (2016). Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. *J Neurophysiol, 116*(3), 1396-1407.

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J.

(2011). The development of multisensory speech perception continues into the late childhood years. *Eur J Neurosci, 33*(12), 2329-2337.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. *Cerebral Cortex, 17*(5), 1147-1153.

Schepers, I. M., Schneider, T. R., Hipp, J. F., Engel, A. K., & Senkowski, D. (2013). Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. *Neuroimage, 70*, 101-112.

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

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc Natl Acad Sci U S A, 109*(42), E2904-2913.

Schwartz, J. L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Comput Biol, 10*(7), e1003743.

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Dev Sci, 11*(2), 306-320.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci, 31*(8), 401-409.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience, 16*(4), 1486-1510.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology, 86*(4), 1916-1936.

Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience, 13*(2), 121-134.

Simon, D. M., Noel, J. P., & Wallace, M. T. (2017). Event Related Potentials Index Rapid Recalibration to Audiovisual Temporal Asynchrony. *Front Integr Neurosci, 11*, 8.

Simon, D. M., & Wallace, M. T. (2016). Dysfunction of sensory oscillations in Autism Spectrum

Disorder. *Neurosci Biobehav Rev, 68*, 848-861.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12*(1), 7-10.

Smiley, J. F., & Falchier, A. (2009). Multisensory connections of monkey auditory cerebral cortex. *Hear Res, 258*(1-2), 37-46.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci, 27*(3), 161-168.

Smyrnis, N., Mylonas, D. S., Rezaie, R., Siettos, C. I., Ventouras, E., Ktonas, P. Y., . . . Papanicolaou, A. C. (2012). Single-trial magnetoencephalography signals encoded as an unfolding decision process. *Neuroimage, 59*(4), 3604-3610.

Stam, C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp, 28*(11), 1178-1193.

Stekelenburg, J. J., Sugano, Y., & Vroomen, J. (2011). Neural correlates of motor-sensory temporal recalibration. *Brain Res, 1397*, 46-54.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci, 19*(12), 1964-1973.

Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., & James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *Neuroimage, 49*(4), 3308-3318.

Stevenson, R. A., Park, S., Cochran, C., McIntosh, L. G., Noel, J. P., Barense, M. D., . . . Wallace, M. T. (2017). The associations between multisensory temporal processing and symptoms of schizophrenia. *Schizophrenia Research, 179*, 97-103.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *J Neurosci, 34*(3), 691-697.

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014). Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. *J Autism Dev Disord, 44*(6), 1470-1477.

Strauss, A., Henry, M. J., Scharinger, M., & Obleser, J. (2015). Alpha phase determines successful lexical decision in noise. *J Neurosci, 35*(7), 3256-3262.

Strauss, A., Wostmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Front Hum Neurosci, 8*, 350.

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *J Neurosci, 31*(10), 3853-3861.

Tsunada, J., Liu, A. S. K., Gold, J. I., & Cohen, Y. E. (2016). Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nature Neuroscience, 19*(1), 135-+.

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports, 6*.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *Eur J Neurosci, 42*(1), 1636-1643.

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *J Neurosci, 33*(37), 14633-14637.

Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Sci Rep, 5*, 14526.

Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proc Biol Sci, 282*(1804), 20143083.

Van der Burg, E., Orchard-Mills, E., & Alais, D. (2014). Rapid temporal recalibration is unique to audiovisual stimuli. *Exp Brain Res*.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A, 102*(4), 1181-1186.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychon Bull Rev, 14*(6), 1011-1026.

Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., & Pennartz, C. M. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage, 55*(4), 1548-1565.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res, 22*(1), 32-35.

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64C*, 105-123.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science, 316*(5831), 1609-1612.

Womelsdorf, T., Vinck, M., Leung, L. S., & Everling, S. (2010). Selective theta-synchronization of choice-relevant information subserves goal-directed behavior. *Front Hum Neurosci, 4*, 210.

Yates, J. L., Park, I. M., Katz, L. N., Pillow, J. W., & Huk, A. C. (2017). Functional dissection of signal and noise in MT and LIP during decision-making. *Nat Neurosci, 20*(9), 1285-1292.