Large-scale Neural Dynamics of Cross-modal Speech Perception

G Vinodh Kumar

Thesis submitted to
National Brain Research Centre
for the award of
Doctor of Philosophy
in Neuroscience

National Brain Research Centre
(Deemed University)
Manesar, Haryana, India - 122052

DECLARATION

I, **G. Vinodh Kumar,** hereby declare that the work presented in the thesis entitled "**Large-scale neural dynamics of cross-modal speech perception**" was carried out by me under the guidance of Dr. Arpan Banerjee, Additional Professor/ Scientist V, National Brain Research Centre, Manesar, Haryana-122052, India.

I also declare that no part of this thesis has been previously submitted for the award of any degree or diploma of the National Brain Research Centre (Deemed University) or any other university.

I also affirm that all the experiments involving human volunteers described in the thesis were approved by Human Ethics Committee of the National Brain Research Centre and were in accordance with the Declaration of Helsinki.

Vinodh

(G. Vinodh Kumar)

Date: 31.01.2019

Place: Manesar

ACKNOWLEDGEMENTS

Pursuing Ph.D. in NBRC has been an enriching experience. Many people have contributed directly and indirectly in shaping me into a better researcher. I wish to pen down my immense gratitude to those people here.

Firstly, I am deeply indebted to my mentor, Dr. Arpan Banerjee for his relentless support and guidance. He has been hugely patient with me, providing the freedom and time to pursue my ideas while extending the required facilities. Coming from a completely different academic background, I was overwhelmed by the neuroimaging tools and analyses. But, his constant encouragement to pay more emphasis to the question rather than the experimental methodology is the only reason I have been able to pursue my research in cognitive-neurophysiology. The perfect blend of his scientific guidance and encouragement has transformed me from a science admirer to a researcher. I hope I have been able to imbibe some of his excellent qualities. Ph.D. under his tutelage will be one of my cherished memories.

I extend my sincere gratitude to Dr. Dipanjan Roy for his constant encouragement and guidance. His immense literature knowledge has always motivated me to look at problems from a different perspective. Discussions on both academic and non-academic topics with him has been invaluable. His guidance, especially in road-blocks during my Ph.D. has always helped me better my spirits.

I am thankful to my doctoral committee members - Prof. Nandini Chatterjee Singh, Prof. Neeraj Jain and Dr. Soumya Iyengar - for giving their time to understand and critique my work. I also thank all the faculty at NBRC for being amazing teachers and nurturing my interest in neuroscience. Dr. Iyengar and Prof. Chatterjee Singh deserve a special mention for their guidance in neuroanatomy and tips on scientific writing respectively. Their guidance has helped me to understand and propagate neuroscience effectively.

My experiment relied primarily on human participants. I extend my gratitude to all the participants in my studies for their cooperation and enthusiasm. Also, I thank all current members of my lab for their support and constructive criticisms.

I sincerely acknowledge the financial support from NBRC core for my doctoral work and the support from Department of Biotechnology (DBT) for traveling to conferences that has enriched me in my research.

I thank all my wonderful friends at NBRC, they have been more like a family and were with me during my ups and downs. Dida, Guruji, Abhishek, Tamesh and Siddharth, being the initial ones to join the lab have significantly helped me setting up my experiments. Get-together and discussions of everything under the sun with Shanah, Kautuk, Guncha, Apoorv, Bala, Indra, Deborah, Shankamala, Raghav, Priyanka and Sujata have always maintained my morale at high. Shrey deserve a special mention for being a great critique towards my approach to the research question at hand. My acquaintance with all these people have transformed me positively in academic and personal spheres.

Lastly, but most importantly, there is norm to measure my gratitude to Appa, Amma, Sumi and my life-partner Divya. Without their support, I would have not even imagined of pursuing science. My father has been a source of abundant inspiration personally. His guidance on patience and perseverance has helped me tackle various road-blocks both academic and non academic. I would always be highly indebted to my mother and sister for having great hope on whatever I pursue. Divya's encouragement to think rationally even in the toughest of time has helped me push myself only higher.

Contents

List of figures

List of tables

Chapter 1

Introduction

Spoken language, more than anything else, is what makes us human. It appears that no other communication system in the animal kingdom can more precisely shape events in each others' brains than spoken language. It has played the central role in propelling our immediate ancestors, *hominids* from east-African ecological community to the most dominant species on the planet. Late British neurologist Oliver Sacks broadly elucidates the importance of language in his book *Seeing Voices* as -

> *And language, (...) is not just another faculty or skill, it is what makes thought possible, what separates thought from nonthought, what separates the human from the non human.*

The evolution of spoken language is thus the most significant event in the history of life on earth. Unarguably, it is so pivotal to humanity that it permeates all aspects of human cognition, behavior and culture, making it an avenue of immense investigation for linguists, anthropologists, geneticists, evolutionary biologists, computer scientists and neuroscientists.

The emergence of human languages however, is firmly tied to the evolution of human speech which is the vocal medium to convey language. The evolution of a complex vocal apparatus that could produce vocalizations adequate to serve the linguistic needs of the modern man is, however, inconceivable without a concurrent perceptual specialization machinery. Therefore, conceptualization of the aforementioned perceptual specialization machinery holds an essential niche particularly in neuroscience known as ′neurobiology of speech perception′. In addition to providing inferences on evolution, speech perception research has had paramount clinical, social, technological and cultural implications. As we discuss in the subsequent sections, over the decades, the exploration of the neurobiology of speech perception has increasingly proven its complexity and vital role in human cognition.

1.1 Speech as a multisensory phenomenon

Face-to-face conversation, video-conferencing and watching a television are some of the activities that we perform routinely. Perception of speech in such scenarios involves the participation of more than one sensory modality (primarily vision and audition) wherein watching the speaker provides concurrent visual cues. The fact that, we can communicate effectively without the visual cue in specific scenarios (e.g. talking over a telephone), perhaps undermine the significance of visual cues during speech perception. Nevertheless, the research over the past few decades have incontrovertible evidences in support of the conclusion that visual speech cues when accessible, supplement and modulate speech perception.

Visual cues from the lips of the speaker carry relevant linguistic information that aid speech perception in noisy acoustic environment [126] and in deaf individuals (*lipreading*) [7]. Moreover, studies by Weikum and collegues [140] that demonstrate infants as young as 4-6 months old can discriminate languages from just viewing silently presented articulations, leads us to infer that sensitivity to visual speech arises as a part of normal development rather than compensatory strategy to cope with deafness or noisy acoustic conditions. With evidences pinpointing speech perception as a multisensory phenomena, theories and scientific investigations were directed to address the question - how does the visual source of information about speech get integrated with the auditory speech.

1.1.1 Audio-visual (AV) integration of speech : Theories and evidences

AV integration in infants

Unlike theories of speech perception in adults, two major school of thoughts exist that divide the research on the early development of AV speech perception. The first is the *integration* framework that refers AV speech perception as 'cross-modal ' or 'intermodal'. This is to indicate that that two independently functioning sensory systems need to act coherently for the perception to emerge. Nevertheless, discrepancy within this approach exist as to whether the integration occurs through general principles of associative learning [140] or more active hypothesis testing [60]. Contary to the *integration* framework is the *differentiation* framework that refer early AV integration as *amodal* to indicate the notion that senses are not seperate from birth and respond simultaneously [1] to external input. In differentiation framework, experiences are also considered to play a role in attuning our senses to integrate only the conforming inputs.

Support for the *integration* framework primarily come from the studies looking, for example, at infants' AV matching abilities. In a typical AV matching paradigm, the infants are presented with side-by-side talking faces articulating different utterances. These video are presented along with a centrally presented speech sound that matches one of the utterances. Behavioral indication of the AV matching is obtained when the infants spend longer duration looking at the matching face than a the mismatching face. Following such paradigm, Dodd

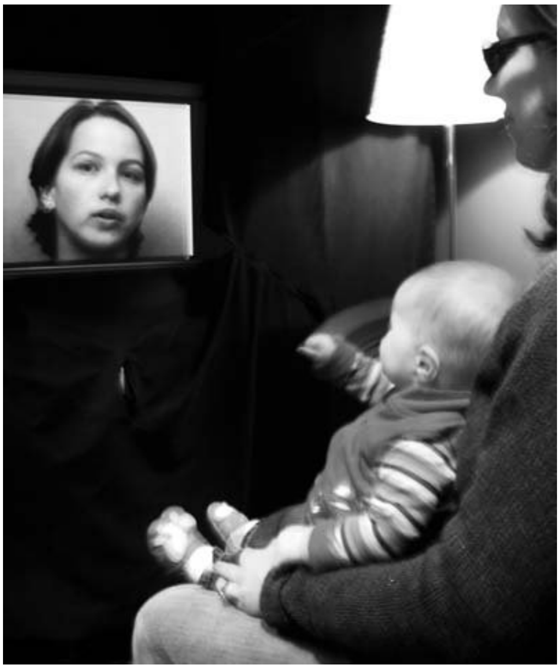

Figure 1.1: Experimental setup used by Weikum et al., 2007 to test visual language discrimination in infants aged 4–6 months. Two groups of infants were habituated by presenting a series of videoclips showing faces silently articulating sentences in French and English respectively. Subsequently, the infants were tested on their ability to discriminate visual speech of the language they were not habituated. Infants discriminated English from French speech just from viewing silent articulations

and others have show that infants as young as 2.5 -4 months are able to match the affective content of faces and voices [33]. Such ability of infants AV matching as young as 2 months of age has been attributed to the experience gathered from substantial parent-child face-to face interactions wherein they learn to associate conforming inputs,

Evidences in support of the *differentiation* view come especially from the studies that demonstrate that AV matching may also be possible without experience. Aldridge and colleagues have shown that even neonates can match articulatory movements with the heard sound [4]. The most convincing evidence in support of the *differentiation* comes from a recent study by Pons et al., wherein they presented AV syllables /*ba*/ and /*va*/ to American and Spanish infants for AV matching. Notably, the phonetic distinction of syllables /*ba*/ and /*va*/ are not contrastive in Spanish. Both American and Spanish infants at 4 months succeeded whereas by 10 months only the American infants showed the ability of matching the sound of /*ba*/ and /*va*/ to their respective visual speech [107]. These results suggests the existence of AV integration prior to specific experiences with AV speech.

Combining the evidences accumulated over the years in support of the both views, it is clear that neural systems supporting AV integration emerge very early in life. Furthermore, one can infer that although capability of AV speech perception exist from birth, experience also plays a prominent role in sharpening and attuning these capabilities.

AV integration in adults

Substantial amount of research has been pursued to characterize the mechanisms of AV speech perception- the set of mental operations that facilitate the integration of information from multiple sensory modalities. However, there exists a theoretical divide among the researchers about the underlying processes, and representation of the speech inputs and the output of the integrative process. The three most influential models proposed till date and neuroscientific evidences in support of the respective models are discussed in turns.

Convergence views

The *convergence* theory that gained the maximum attention in the past few decades was the 'motor theory of speech perception' [81]. It proposes that speech inputs are represented as specific speech gestures. In other words, the theory hypothesizes that spoken words are perceived by the articulatory gestures rather than identifying the sound patterns of the incoming speech. A major claim of the theory is the participation of motor representations during speech perception. Studies looking at motor evoked potentials (MEP) of the lip and tougue related areas induced by transcranial magnetic stimulations (TMS) have shown modulation in the MEP during the perception of speech. However, the theory also faces criticisms from the studies that have shown the intactness of the speech perception abilities in patients with damages in the motor brain areas.

Associationist views

The *associationist* view hypothesizes that the speech information undergo extensive sensory processing in the respective sensory-specific cortices before being integrated to elicit a percept. Also according to this view, multisensory speech are represented in the incoming speech gestures, but instead in abstract features stored as prototype in the memory. These prototypes are constantly assessed during the coarse of integration. The fuzzy-logical model of perception (FLMP) claims a niche in the associationist view of speech perception [89]. Although the FLMP offers to explain a substantial amount of empirical data [35], it face criticisms for its excessive impetus it confers to the memory. Also, the previously discussed AV integration in infants negates the view and its dependence on the memory.

Analysis-by-Synthesis views

This relatively recent theory, assumes that the role of salient and distinguishable visual information is to modulate the processing of the conforming auditory input. According to van-Wassenhove and others, a salient visual information (for example, the closure of the mouth during the articulation of $/b/$) that precedes the onset of the corresponding acoustic information by several tenths of a second, offers predictions for the incoming acoustic input, thereby speeding up the processong of speech [131]. However, extension of the model is restricted by the ambiguity that exists in the visual speech signal itself (for example the articulatory gestures of $/p/$, $/b/$ and $/m/$ are indifferent).

Each of the aforementioned theories although explain many empirical findings, they equally face criticisms. To address the discrepancy, exploiting a percept arising from the combination of the features specified by heard and seen speech would serve as the apt model to study the neuronal operation during multisensory speech perception. And, McGurk effect precisely fulfills those requirements. We eleborate on McGurk effect in the following sections.

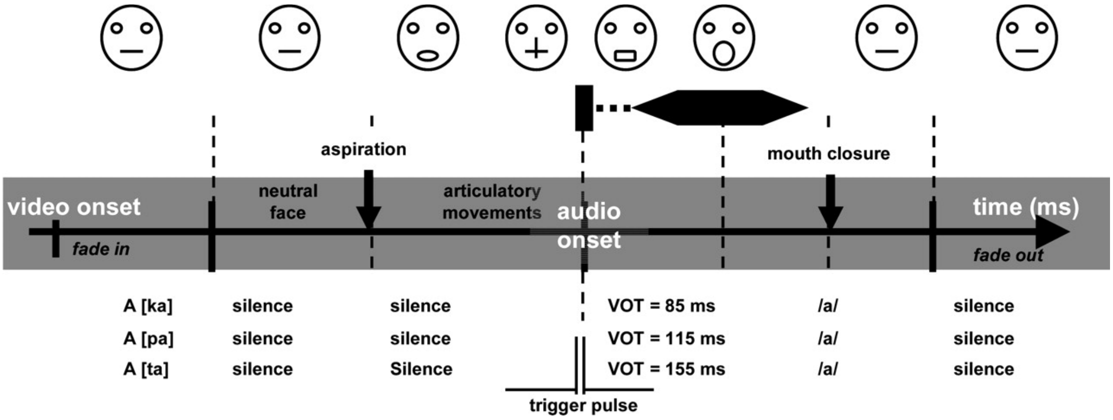

Figure 1.2: Speech articulatory gestures often precede their corresponding sounds. This early visual information has an impact on the way speech sounds are processed [131].

1.2 McGurk effect: An entry point to understand AV integration

Since its discovery nearly 40 years ago, McGurk effect has been predominantly employed as the prototypical paradigm to understand multisensory speech perception. Harry McGurk and John McDonald published *Hearing lips and Seeing Voices* [90], a study in which they illustrate a remarkable audiovisual speech phenomena known as McGurk effect. McGurk effect occurs when a acoustic signal of a phoneme is dubbed onto specific *semantically-incongruent* visual signal of a phoneme. Observers of such incongruent audio-visual pairs (audo /*ba*/ + visual /*ga*/) do not recognize the inter-modal differences and perceive (i.e. hear) a phoneme (/*da*/) different from the auditory or visual signal.

McGurk effect effectively illustrates that speech perception is not only an auditory process. Besides, a stronger evidence of AV integration is provided when a unitary percept arises from the combination of the features specified by heard and seen speech. For these reasons, McGurk effect has been used a extensively used a s proxy measure to understand AV speech integration [6, 88, 100]. Furthermore, the technical advances of neuroimaging tools has enabled contemporary cognitive-neuroscientists to exploit various psychophysical parameters of McGurk effect to understand different aspects of AV speech perception.

1.2.1 Neurobiology of AV speech perception as revealed by McGurk effect

Employing McGurk effect, substantial amount of studies have employed different neuroimaging tools such as functional magnetic resonance imaging (fMRI), magnetoncephalography (MEG), positron emission tomography (PET), electroencephalography (EEG), transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS) to explore the neural workings during AV speech perception. Although each of the aforementioned tools come with some peculiarities, It is worth noting the nature of information that each tool provides. fMRI and PET studies offer information regarding the hemodynamic state of the brain structures involved, thereby providing the best spatial resolution amongst the non-invasive tools. EEG and MEG provide electrophysiological information, with much higher temporal resolution. Finally, neuromodulation techniques TMS and tDCS allows us to understand the causal links between brain activity and behavioral responses. Converging evidences employing the McGurk effect demonstrates activation of specific cortical modules like the pSTS (posterior Superior Temporal Sulcus) [61, 98, 99, 117], frontal and parietal areas [61, 122] being responsible for the McGurk perception. Although,

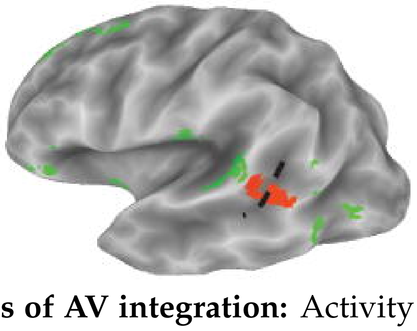

Figure 1.3: **Cortical locus of AV integration:** Activity in the left superior temporal sulcus (orange colored regions) during the McGurk perception [12].

Nath and colleagues have shown existence of a positive correlation only between activity levels in STS levels and the propensity of McGurk susceptibility [99]. Furthermore, decreasing STS activity with 1 Hz TMS have been shown to result in decreased perception of illusion and not the perception the congruent AV stimulus [12]. This finding demonstrates the important role of the cortical loci STS in AV integration, proved by the lack of influence on unisensory processes since the congruent stimuli presentation can be perceived accurately by each of the isolated sensory modalities. However, to better understand the specific role of different brain regions in this process it is important to explore the temporal dynamics of the recruitment of each area.

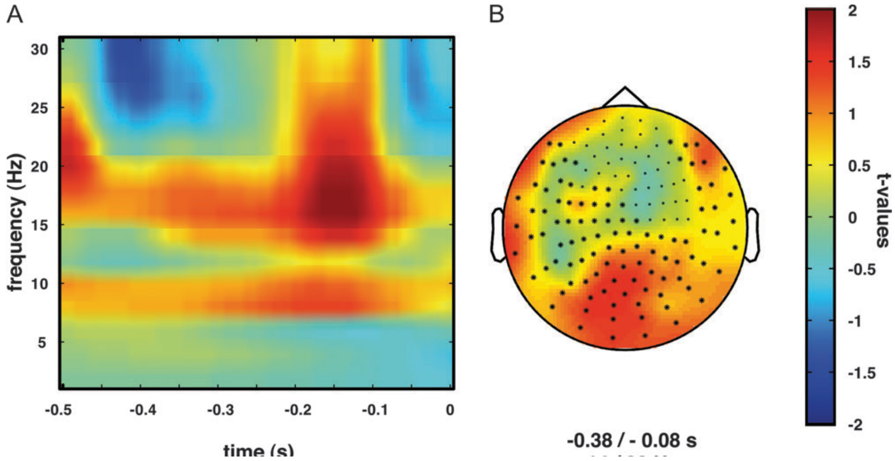

Figure 1.4: **Prestimulus activity:** (A) Time–frequency representation of the prestimulus interval at sensor level for the comparison between "fusion" and "unimodal" trials. Time 0 ms indicates the onset of mouth movement and audio stream. (B) Topography (14-30 Hz, -380 to -80 ms) of the positive beta- band cluster found in the prestimulus interval at sensor level for the comparison between fusion and unimodal trials. [70].

Addressing the question on temporal dynamics, electrophysiological evidences have explored signatures in the event-related brain potentials (ERPs) and oscillations in specific frequency ranges. Seminal studies by Keil and others have shown that activity in the beta band can be predict the perception of McGurk effect of the observers [70]. Also, evidences further show the significance of beta [110] and gamma band activity [63] toward illusory (cross-modal) perceptual experience. Studies employing connectivity measures on functional imaging and electrophysiological data primarily demonstrate the functional connetness of left Superior temporal gyrus to the frontal and parietal regions crucial for ilusory speech perception [70].

Put together, McGurk effect employed as a proxy to understand multisensory speech perception accentuate the role of brain structures especially in the temporal lobe to be pivotal for cross-modal perception. Also, the electro-physiological evidences pinpoint the implications of specific brain oscillations during speech perception. Importantly, these evidences have enabled scientists in the conceptualization and design of effective diagnostic markers for speech related disorders.

McGurk effect in clinical populations

Impaired multisensory integration have been reported in population with Schizophrenia, Dyslexia and Autism Spectrum Disorders. And, more and more research is being done using the McGurk illusion to better understand multisensory integration in these patients.

One of the most studied clinical group in the field of sensory integration is the Autism Spectrum Disorder (ASD) - is characterized by deficits in social interaction, language development and motor impairments with neurological causes. It has been observed that propensity of McGurk perception is much less than in the neurotypical population [128]. The lesser incidence of McGurk perception has been attributed by some to peculiar gaze pattern for face stimuli known to take place in the ASD patients. However, contrasting evidences show no significant difference in the eye movement in ASD patients and the controls indicating lack of integration. In addition to studies with individuals where specific multisensory integration processes are compromised, McGurk effect has been employed to assess development and semantic representations in dyslexic children [11] and also understand social interaction in Schizophrenic patients [11].

The McGurk illusion is the most predominantly used multisensory integration illusion paradigms. As reviewed above, several studies applied this illusion as a tool to better comprehend how audiovisual speech information is processed. Over the years, with availability of larger datasets and advanced neuroimaging tools, the robustness of experimental results have increased. However, a prominent approach still followed in the field is to localize the cortical loci pivotal for a perception or looking at connectivity between key cortical regions. Understanding the large-scale network organization of the brain is a crucial for formulating a comprehensive theory on the information processing of multisensory speech perception.

1.3 Towards understanding the large-scale networks of AV integration

Our understanding of the workings of the brain and cognition have primarily come from modular paradigm. The modular paradigm postulates that our cognitive abilities emerge as result of activations in brain areas working as independent processes [36]. However, converging evidences over the years elucidate limitations of the approach [44]. Even the sensory cortices, considered to be highly modular structures functionally has been shown to possess cross-modal interactions [45]. A more emerging view posits that information processing associated with the functioning of higher order brain functions (action, perception, learning, language, and cognition) is carried out by large scale neural networks [19]. Although the structural architecture of the brain has been extensively studied, the complex dynamics elicited by the neural networks as brain oscillations and synchronizations during any cognitive task remains still at a nascent stage.

1.4 Scope of the thesis

Speech perception is a quintessential human trait that inextricably involves multisensory integration. Incidence of visual cues (talker's mouth movements) aids speech perception especially in noisy surrounding. Understanding the neural processes that engender the integration of audio-visual (AV) cues in speech perception has been a topic of intense research for more than six decades. A predominantly employed paradigm in AV speech perception studies is the McGurk effect. During McGurk effect, a listener perceives a completely different syllable (illusory percept) when presented with an auditory phoneme dubbed onto certain semantically-incongruent visual phoneme (e.g audio/*ba*/ dubbed onto visual/*ga*/ is perceived as /*da*/). McGurk effect has been used as a proxy measure for AV speech perception as the frequency with which an observer perceives the illusion offers an index to AV integration.

We focused on understanding the dynamics of large-scale cortical network that foster AV speech perception. We analyzed the properties of large-scale oscillatory network with the following goals that we elaborate in the subsequent chapters in turn :

(1) Identify the markers in the large-scale network during the temporal integration of AV stimuli.

(2) Understand segregation and integration of cortical information processing during cross-modal perception.

(3) Determine the markers underlying inter-individual and inter-trial variability during cross-modal perception.

(4) Employ a neural-mass model to provide a neurodynamic explanation for the empirical signatures of inter-individual and inter-trial variability.

Chapter 2

Neuro-cognitive networks during temporal integration of audio-visual speech

2.1 Introduction

Perception of the external world involves the efficient integration of information over multiple sensory systems [137]. During speech perception, visual cues from the speaker's face enhances the intelligibility of auditory signal [21, 53, 126]. Also, the incidence of specific semantically-incongruent visual information modulates auditory perception, for example, an auditory speech sound /ba/ superimposed with a speaker's lip movement of /ga/, gives rise to a perception of /da/ [90]. Similarly, an incongruent AV combination of /pa/-/ka/ elicits an 'illusory' (cross-modal) percept /ta/ [84,90,132]. However, such multisensory-mediated effects are influenced by the relative timing of the auditory and visual inputs [96,124,130,132]. Consequently, the temporal processing of the incoming multiple sensory (auditory and visual) information and their integration to yield a crossmodal percept is pivotal for speech perception [32]. Where and how the underlying information processing takes place is subject of several research studies which we review in the following paragraph. Cortical and sub-cortical regions and functional brain networks with specific patterns of connectivity becomes the prime target for these investigations. In a nutshell, characterization of the multi-scale representational space of temporal processing underlying multisensory stimuli is an open question to the community.

As we discuss in the following paragraph, a dominant strategy in multisensory research is the search for loci comprising of brain areas that are responsible for triggering the multisensory experience [12, 61, 99]. However, from the perspective of functional integration [17,19] understanding the large-scale network organization underlying the temporal processes is a critical component of formulating a comprehensive theory of multisensory speech perception. Numerous neuroimaging and electrophysiological studies have explored the neural mechanism that underpins audio-visual integration employing McGurk effect [52, 61, 63, 70, 99, 114, 117, 122, 125, 131, 137]. A majority of these studies accentuate the role of primary auditory and visual cortices, multisensory areas such as posterior superior temporal sulcus (pSTS) [61,98,99,117] and other brain regions including frontal and parietal areas [61,122] in the perception of the illusion. In particular, the electrophysiological evidences primarily emphasizes the significance of beta [70,110] and gamma band activity [63] toward illusory (cross-modal) perceptual experience. Source-level functional connectivity among brain areas employing phase synchrony measures, reveal interactions among cortical regions of interest (left Superior Temporal Gyrus) and the whole brain that correlates with cross-modal perception [70]. These studies either reveal the activations in the cortical loci or the functional connectedness to particular cortical regions of interest that are elemental for the illusory percept. On the other hand, the role of timing between auditory and visual components in AV speech stimuli has been studied from the perspective of the main modules in multisensory processing [61]. Recently, we have addressed this issue using a dynamical systems model to study the interactive effects between AV lags and underlying neural connectivity onto perception [129]. Interestingly, how these network are functionally connected in the context of behavioral performance or perceptual experience are increasingly being revealed [70,98]. Nonetheless, the identification and systematic characterization of these networks under cross-modal and unimodal perception is an open question.

A traditional measure of large-scale functional connectivity in EEG is the sensor-level global coherence [2, 9, 24, 25, 42]. Global coherence can be described as either the normalized vector sum of all pairwise coherences between sensor combinations, the frequency domain representation of cross-correlation between two time-series [24,80] or the ratio of the largest eigenvalue of the cross-spectral matrix to the sum of its eigenvalues [94]. An increased global coherence confirms the presence of a spatially extended network that spans over several EEG sensors, since local pairwise coherence would not survive statistical threshold after averaging. To the best of our knowledge, global coherence has not been used in the domain of audio-visual (AV) speech perception to evaluate the presence of whole brain networks. Furthermore, characterization of the differences in whole brain network organization underlying cross-modal vs. unimodal perceptual experience vis-à-vis the timing of sensory signals will be critical to understanding the neurobiology of multisensory perception.

Here, we used the incongruent McGurk pair (audio /pa/ superimposed on the video of the face articulating /ka/) to induce the illusory percept /ta/. Further, we generated a temporal asynchrony in the onset of audio and visual events of the McGurk pair to diminish the rate of cross-modal responses. Subsequently, we exploited the inter-trial perceptual variability to study integration both at behavioral levels by accounting perceptual response and eye-tracking as well as neural levels using EEG. We considered subjects' /pa/ responses as unimodal perception since it represents only one sensory stream and /ta/ responses as cross-modal perception since it represents an experience resulting from integrating features from two modalities [32]. We studied the spectral landscape of perceptual categorization as function of AV timing and found patterns that matched with previous reports. Finally, we evaluate the large-scale brain network organization dynamics using time-frequency global coherence analysis for studying perceptual categorization underlying different temporal processing scenarios at various AV lags. In the process, we reveal the complex spectro-temporal organization of networks underlying multisensory perception.

2.2 Materials and Methods

2.2.1 Participants

Nineteen [10 males and 9 females, ranging from 22–29, (mean age 25; SD = 2)] healthy volunteers participated in the study. No participant had neurological or audiological problems. They all had normal or corrected-to-normal vision and were right handed. The study was carried out following the ethical guidelines and prior approval of Institutional Review Board of National Brain Research Centre, India.

2.2.2 Stimuli and Trials

The experiment consisted of 360 trials overall in which we showed the videos of a male actor pronouncing the syllables /*ta*/ and /*ka*/ (**Figure 2.1**). One-fourth of the trials consisted of congruent video (visual /*ta*/ auditory /*ta*/) and the remaining trials comprised incongruent videos (visual /*ka*/ auditory /*pa*/) presented in three audio-visual lags: 450 ms (audio lead), 0 ms (synchronous), +450 ms (audio lag), each comprising one-fourth of the overall trials. The stimuli were rendered into a 800 600 pixels movie with a digitization rate of 29.97 frames per second. Stereo soundtracks were digitized at 48 kHz with 32 bit resolution. The stimuli were presented via Presentation software (Neurobehavioral System Inc.). The video was presented using a 17 LED monitor. Sounds were delivered at an overall intensity of 60 dB through sound tubes.

The experiment was carried out in three blocks each block consisting of 120

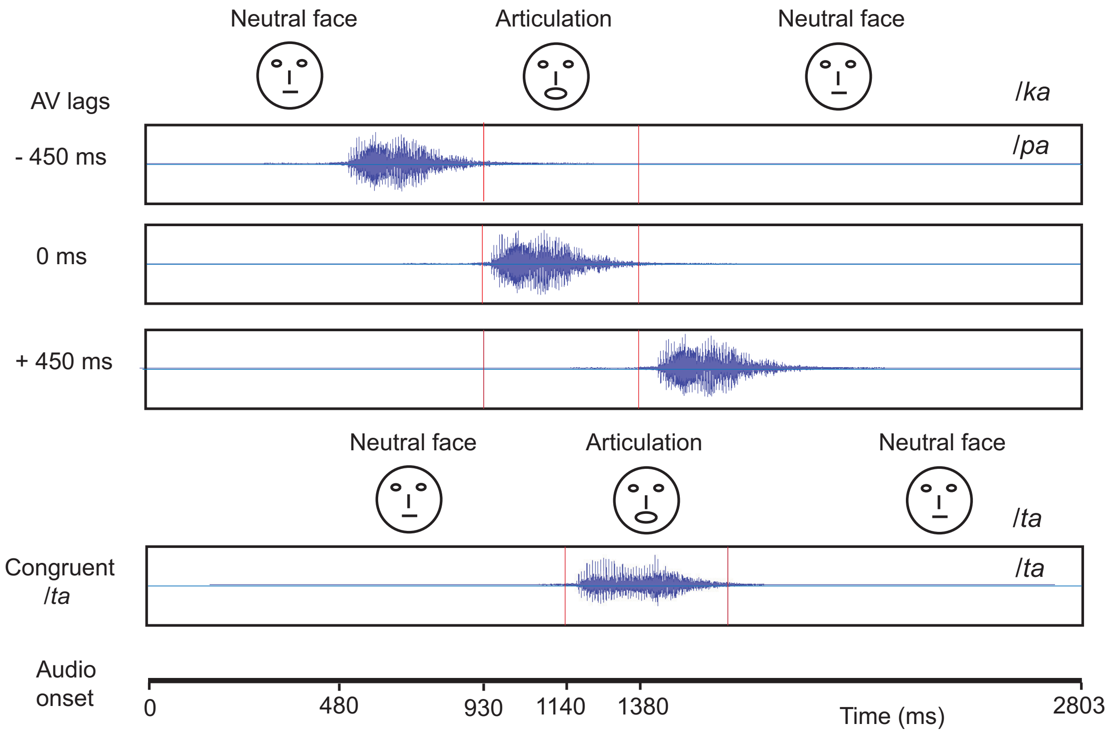

Figure 2.1: **Stimuli** : Each block represents a video. **(A)** The McGurk stimuli: Audio */pa/* superimposed on visual (lip movement) */ka/* was presented under different audio-visual (AV) lag scenarios. The location of onset of audio is varied with respect to a person's initiation of lip-movement */ka/* at -450, 0, and 450 ms. **(B)** In congruent */ta/* condition, audio */ta/* is presented synchronously with onset of lip movement */ta/*.

trials. Inter-trial intervals were pseudo-randomly varied between 1.2 and 2.8s. Each block comprised the four stimuli types (30 trials of each): Congruent video and three incongruent videos with the AV lags. The subjects were instructed to report what they heard while watching the articulator using a set of three keys. The three choices were /pa/, /ta/ and "anything else" (Other).

Post EEG scan, the participants further performed a behavioral task. The task comprised of 60 trials, comprising 30 trials each of auditory syllables */pa/* and */ta/*. Participants were instructed to report their perception using a set of two keys while listening to syllables. The choices were */pa/* and */ta/*.

2.2.3 Data Acquisition and Analysis

EEG

EEG recordings were obtained using a Neuroscan system (Compumedics NeuroScan, SynAmps2) with 64 Ag/AgCl sintered electrodes mounted on an elastic cap of Neuroscan in a 10–20 montage. Data were acquired continuously in AC mode (sampling rate, 1 kHz). Reference electrodes were linked mastoids, grounded to AFz. Channel impedances were kept at < 5 kΩ. All subsequent analysis was performed in adherence to guidelines set by [69].

Eye Tracking

Gaze fixations of participants on the computer screen were recorded by EyeTribe eye tracking camera with resolution 30 Hz (https://theeyetribe.com). The gaze data were analyzed using customized MATLAB codes. The image frame of the speaker video was divided into 3 parts, the head, the nose and the mouth (**Figure 2.2A**). The gaze locations at these quadrants over the duration of stimulus presentation were converted into percentage measures for further statistical analysis.

Pre-Processing of EEG Signals

The collected EEG data were subsequently filtered using a bandpass of 0.2–45 Hz. Epochs of 400 and 900 ms before and after the onset of first stimuli (sound or articulation) were extracted and sorted based on the responses, */ta/*, */pa/*, and "other" respectively. Epochs were baseline corrected by removing the temporal mean of the EEG signal on an epoch-by-epoch basis. Epochs with maximum signal amplitude above $100\mu V$ or minimum below $-100\mu V$ were removed from all the electrodes to eliminate the response contamination from ocular and muscle-related activities. Approximately 70–75 % (250 trials) trials of each subject were preserved after artifact rejection. In the final data analysis, a mean of 24 (SD = 9), 18 (SD = 9), and 25 (SD = 13) incongruent trials at 450, 0, +450 ms AV lags respectively in which the participants responded /pa/ were included. Similarly, a mean of 32 (SD = 15), 42 (SD = 13), and 32 (SD = 14) incongruent trials at 450, 0, +450 ms AV lags respectively in which the participants responded */ta/* were included in the final analyses. Approximately 2–6% of trials were excluded from each of the aforementioned trial categories. The response category with lowest number of occurrences was */pa/* at 0 ms AV lag with 270 hits from a total of 1350 trials across all volunteers (15 90). Subsequently, we randomly resampled 270 trials from */ta/* responses at 0 ms AV lag, and */pa/* and */ta/* responses at other AV lags. Thus, for each AV lag condition, 270 trials chosen randomly from the respective sorted response epochs (*/pa/* or */ta/*) entered the final analyses.

Spectral Analysis

Power spectra of the preprocessed EEG signals at each electrode were computed on a single trial basis. We computed the spectral power at different frequencies using customized MATLAB (www.mathworks.com) codes and the Chronux toolbox (www.chronux.org). Time bandwidth product and number of tapers were set at 3 and 5 respectively while using the Chronux function mtspecgramc.m to compute the power spectrum of the sorted time series in EEG data. Subsequently, the differences in the power during */ta/* and */pa/* responses at each AV lag were statistically compared by means of a cluster-based permutation test [87] using the fieldtrip toolbox (www.fieldtriptoolbox.org). The fieldtrip function ft_freqstatistics.m was used to perform the cluster computation. During the statistical comparison, an observed test statistic value below the threshold of 0.05 in at least 2 of the neighborhood channels were set for being considered in the cluster computation. Furthermore, 1000 iterations of trial randomization were carried out for generating the permutation distribution at a frequency band. Subsequently, a two tailed test with a threshold of

0.025 was used for evaluating the sensors that exhibit significant difference in power. Statistical analysis was carried out separately for alpha (8–12 Hz), beta (13–30 Hz), and gamma (30–45 Hz) frequency ranges.

Large-Scale Network Analysis

For deciphering the coordinated oscillatory brain network underlying the AV integration, we employed global coherence analyses [16, 24, 80, 87] on the perceptual categories (*/ta/* and */pa/*). A higher value of this measure will indicate the presence of strong large-scale functional networks. We computed the global coherence by decomposing information from the cross-spectral matrix employing the eigenvalue method [94]. The cross-spectrum value at a frequency f between sensor pair i and j was computed as:

$$C_{\mathrm{ij}}^{\mathrm{X}} = 1/K \cdot \textstyle\sum_{k=1}^{k} X_{\mathrm{i}}^{\mathrm{k}}(f).X_{\mathrm{j}}^{\mathrm{k}}(f) \tag{2.1}$$

where $X_{\mathrm{i}}^{\mathrm{k}}$ and $X_{\mathrm{i}}^{\mathrm{k}}$ are tapered Fourier transforms of the time series from the sensors i and j respectively, at the frequency f. A 6262 matrix of cross spectra, that represents all pairwise sensor combination, was computed in our case. Conversely, to characterize the dynamics of coordinated activity over time, we evaluated the time-frequency global coherogram. We employed the Chronux function cohgramc.m to obtain the time-frequency cross-spectral matrix for all the sensor combinations. Subsequently, for each trial we obtained the global coherence at each time point and frequency bin by computing the ratio of the largest eigenvalue of the cross-spectral matrix to the sum of the eigenvalues employing the following equation:

$$C_{\mathrm{Global}}(f) = \frac{S_1^{\mathrm{Y}}(f)}{\sum_{i=1}^{n} S_{\mathrm{i}}^{\mathrm{Y}}(f)} \tag{2.2}$$

where $C_{\text{Global}}(f)$ is the global coherence, $S_1^{\text{Y}}(f)$ is the largest eigenvalue and the denominator $\sum_{i=1}^{n} S_{\text{i}}^{\text{Y}}(f)$ represents the sum of eigenvalues of the cross-spectral matrix [24]. Time-frequency global coherogram computed for */ta/* and */pa/* responses were further compared at each time point for significant difference in different frequency bands (alpha, beta, and gamma) by means of cluster-based permutation test [87].

For every frequency bin at each time point, the coherence difference between /ta/ and /pa/ was evaluated using the Fisher's Z transformation

$$Z(f) = \frac{\tanh^{-1}(C_1(f)) - \tanh^{-1}(C_2(f)) - (\frac{1}{2m_1-2} - \frac{1}{2m_2-2)!})}{\sqrt{\frac{1}{2m_1-2} + \frac{1}{2m_2-2}}} \tag{2.3}$$

where $2m_1$, $2m_2$ = degrees of freedom; $Z(f) \approx N(0,1)$ a unit normal distribution; and C_1 and C_2 are the coherences at frequency$f$.

The coherence Z-statistic matrix obtained from the above computation formed the observed Z-statistics. Subsequently, from the distribution of observed Z-statistics, 5^{th} and the 95^{th} quantile values were chosen as upper and lower threshold i.e., the values below and above the threshold values respectively were considered in the cluster computation. Based on spectral adjacency (4–7 Hz, theta; 8–12 Hz, alpha; 13–30 Hz, beta; 30–45 Hz, gamma), clusters were selected at each time point. Consequently, cluster-level statistics were computed by taking the sum of positive and negative values within a cluster separately. Following the computation of the cluster-level statistics of the observed Z-statistics, 1000 iterations of trial randomization were carried out. For every iteration, cluster-level statistic was computed on the randomized trials to generate the permutation distribution. Subsequently, the values of observed cluster-level statistics were compared with the 2.5^{th} and the 97.5^{th} quantile values of the respective permutation distribution. The observed cluster-level statistics value that were below 2.5^{th} and above 97.5^{th} quantile consequently for two time points formed the negative and positive clusters respectively.

2.3 Results

2.3.1 Behavior

Behavioral responses corresponding to McGurk stimuli with the AV lags were converted to percentage measures for each perceptual category (*/pa/*, */ta/*, or "other") from all subjects. We set a minimum threshold of 60% of */ta/* response in any AV lag, 450, 0, and +450 ms to qualify a participant as an illusory perceiver. 15 participants passed this threshold and 4 participants failed to perceive above the set threshold (**Figure 2.2B**). Data from only 15 perceivers were used for further group level analysis. We observed that maximum percentage of illusory (*/ta/*) responses occurred at 0 ms AV lag when the lip movement of the speaker was synchronous with the onset of auditory stimulus (**Figure 2.2C**). Also, the percentage of */pa/* responses was minimum at 0 ms AV lag. We ran one-way ANOVAs on the percentage responses for */pa/*, */ta/*, and "other" with AV lags as the variable. We observed that AV lags influenced the percentage of */ta/* [$F_{(2,44)} = 27.68, p < 0.0001$] and */pa/* [$F_{(2,44)} = 5.89, p = 0.0056$] responses. However, there was no influence of AV lags on "other" responses [$F_{(2,44)} = 0.36, p = 0.700$]. We also performed paired Student's *t*-test on the percentage of responses (*/ta/* and */pa/*) at each AV lag. Insignificant differences of 10.20–11.40 % were observed between /ta/ and /pa/ responses at 450 ms AV lag [$t_{(14)} = 0.63, p = 0.27$] and +450 ms AV lag [$t_{(14)} = 0.45, p = 0.67$] respectively. However, at 0 ms AV lag we observed the percentage of */ta/* responses were significantly higher by 36.58 % than the percentage of /pa/ responses,$t_{(14)} = 10.20, p < 0.0001$. Furthermore, the hit rate of */ta/* responses during congruent */ta/* was observed to be 0.97. Also, the hit rate of */ta/* and */pa/* during auditory alone conditions were observed to be 0.96 and 0.98 re-

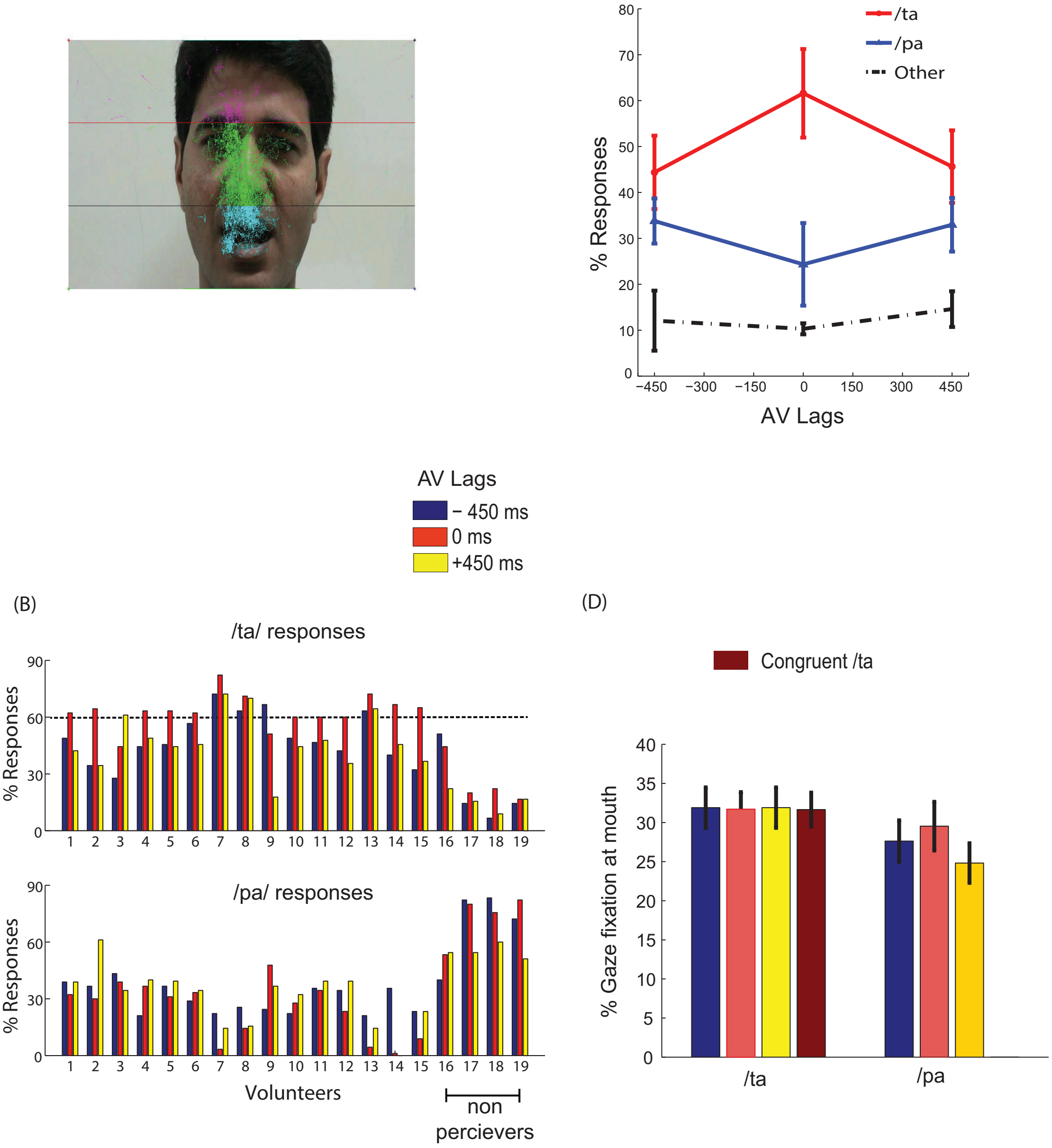

Figure 2.2: **Behavior** : **(A)** overall eye gaze fixation overlaid over a single frame of the stimuli **(B)** the bar graphs show the percentage of */ta/* and */pa/* responses for each subject at the AV lags:450,0, +450 ms as indicated by the colors guide **(C)** shows the number of normalized group responses in each of the three perceptual categories: "*/pa/*", "*/ta/*", and "other" for each AV lag. The error bars represent the 95% confidence interval **(D)** Mean gaze fixation percentages at mouth for each perceptual category at the respective stimuli (incongruent AV lags 450, 0, +450 ms, and congruent */ta/*) across trials and participants. The error bars represents 95% confidence interval.

spectively.

Gaze fixations at different locations on the speaker's, head, nose and mouth areas were converted into percentage measures trial-by-trial for each subject and stimuli conditions. **Figure 2.2A** indicates that most of the gaze fixations were around head, nose, and mouth areas only. We ran a repeated measures 2-way ANOVA on mean gaze fixation percentages across trials at mouth areas with lags and perceived objects (*/pa/* or */ta/*) as variables. No significant differences were found for gaze fixations across lags [$F_{(2,89)} = 0, p = 0.95$] and perceptual categorization [$F_{(2,89)} = 1.33, p = 0.27$] as well as their interactions [$F_{(2,89)} = 0.01, p = 0.85$]. Number of */pa/* responses for congruent */ta/* stimulus was negligible (<1%), to do meaningful statistical comparisons. We also performed paired Student's *t*-tests on the mean gaze fixation percentages for */pa/* and */ta/* responses at each lag. Increases in gaze fixation at mouth during */ta/* perception by 15.5 % at 450 ms AV lag [$t_{(14)} = 0.90, p = 0.38$], 7.2 % at 0 ms AV lag [$t_{(14)} = 0.90, p = 0.38$] and 28.54% at +450 ms AV lag [$t_{(14)} = 0.32, p = 0.74$] (see **Figure 2.2D** for the mean values) were not statistically significant.

2.3.2 Oscillatory Activity

Subsequent to replicating the perceptual [96,132]and the eye gaze behavior [50] results as reported earlier, the focus of interest was what differentiates the two perceptual states (*/ta/* and */pa/*) in terms of brain oscillations and large-scale functional brain networks. Therefore, spectral power at different frequency bands during */ta/* and */pa/* perception were compared at different AV lags. Power spectra at each sensor computed in the time window before (**Figure 2.3A**) and after (**Figure 2.3B**) the onset of first stimuli showed distinct changes in power for the two states. Cluster-based permutation tests employed for comparing the spectral power between the perceptual states show that */ta/* perception is associated with an overall suppression in power for all AV lags (**Figure 2.3**). The magenta "*" on the topoplots highlight the position of the negative clusters showing a significant suppression at 95% confidence levels in power.

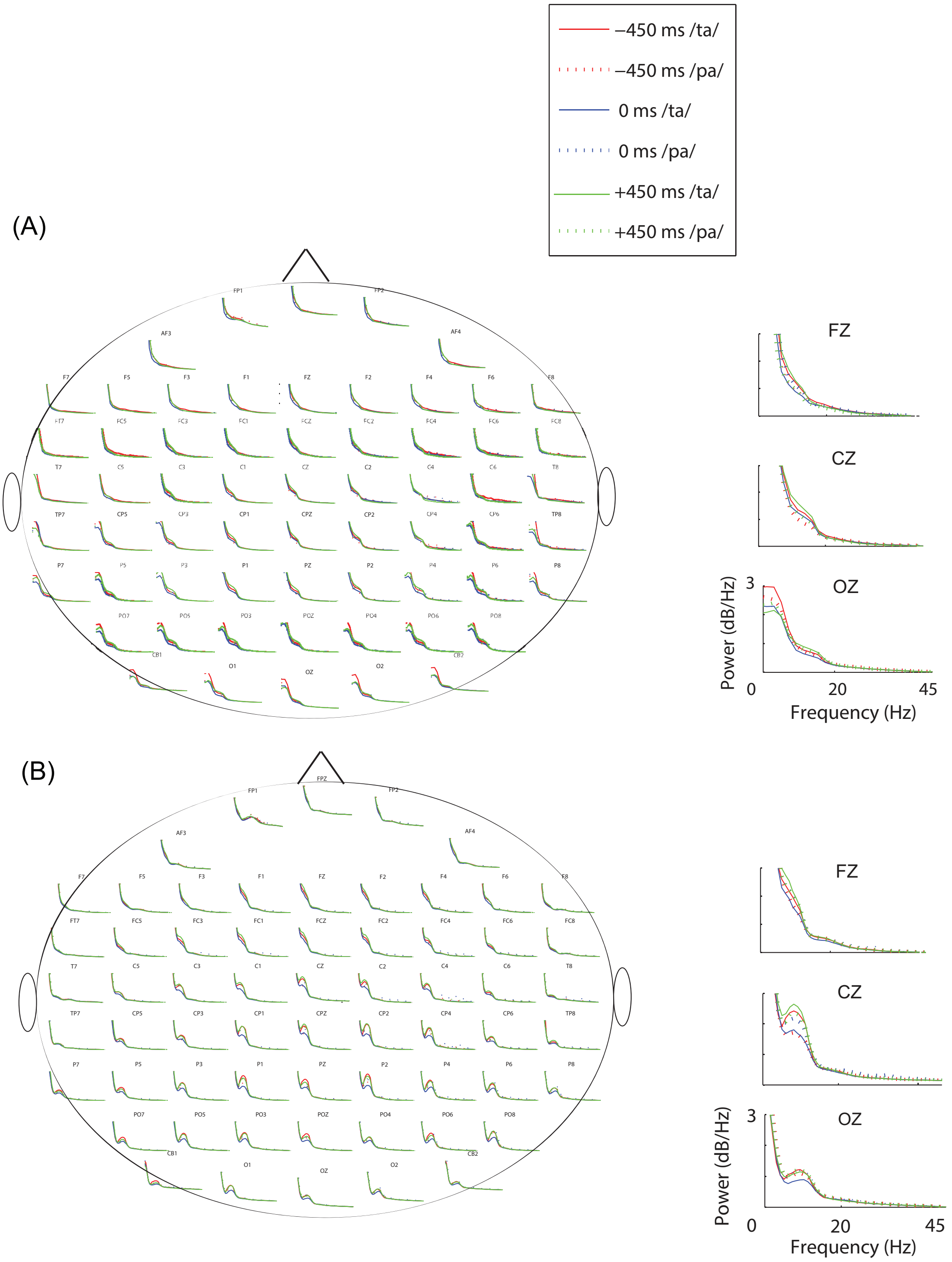

Figure 2.3: **Power Spectrum**: Spectral-power at each condition and perceptual category during **(A)** Pre-stimulus onset. **(B)** Post-stimulus onset periods. The plots adjacent to the scalp maps show the enlarged plots of the power spectrum at the sensors: Fz, Cz, and Oz.

The blue areas on the scalp map highlight the regions that show decrease in the spectral power and the orange and red regions highlight the regions that show an increase in the spectral power. During the pre-stimulus period, one significant negative cluster [$t_{(269)} = 2.04, p = 0.02$] over temporo-occipital sensors, two over frontal and occipital sensors [$t_{(269)} = 3.57, p = 0.002$ and $t_{(269)} = 3.14, p = 0.0002$] and one over occipital sensors [$t_{(269)} = 2.18, p = 0.01$] were observed for alpha, beta, and gamma bands respectively in 0 ms AV lag (**Figure 2.3A**). Also, one significant negative cluster over fronto-temporal and occipital sensors [$t_{(269)} = 2.65, p = 0.004$], one over frontal and occipital sensors [$t_{(269)} = 2.31, p = 0.01$] were observed at alpha and beta bands respectively during +450 ms AV lag (see **Figure 2.3C**). However, no significant difference was found during 450 ms AV lag.

Furthermore, during post-stimulus onset period, the */ta/-/pa/* comparison revealed one significant negative cluster over all sensors [$t_{(269)} = 1.93, p = 0.02$], one over frontal, parietal, and occipital sensors [$t_{(269)} = 2.70, p = 0.004$] and one over occipital sensors [$t_{(269)} = 2.54, p = 0.006$] at alpha, beta, and gamma bands respectively during 450 ms AV lag (see **Figure 2.4C**). During 0 ms AV lag, one significant negative cluster [$t_{(269)} = 2.22, p = 0.01$] spanning over all sensors and one over occipital sensors [$t_{(269)} = 2.10, p = 0.02$] was observed at alpha and beta bands respectively (see **Figure 2.4D**). However, no significant difference in power between */ta/-/pa/* trials was observed during the post-stimulus period at +450 ms AV lag. Overall, significant spectral power was lower during */ta/* than */pa/* as reflected from cluster-based analysis during pre- and post-stimulus periods.

2.3.3 Time-Frequency Global Coherogram

Eigenvalue based time-frequency global coherogram [24] was computed for the epochs of 1.3 s duration (0.4 s pre-stimulus, and 0.9 s post-stimulus segments). The time locking was done to the first sensory component, audio or visual, for

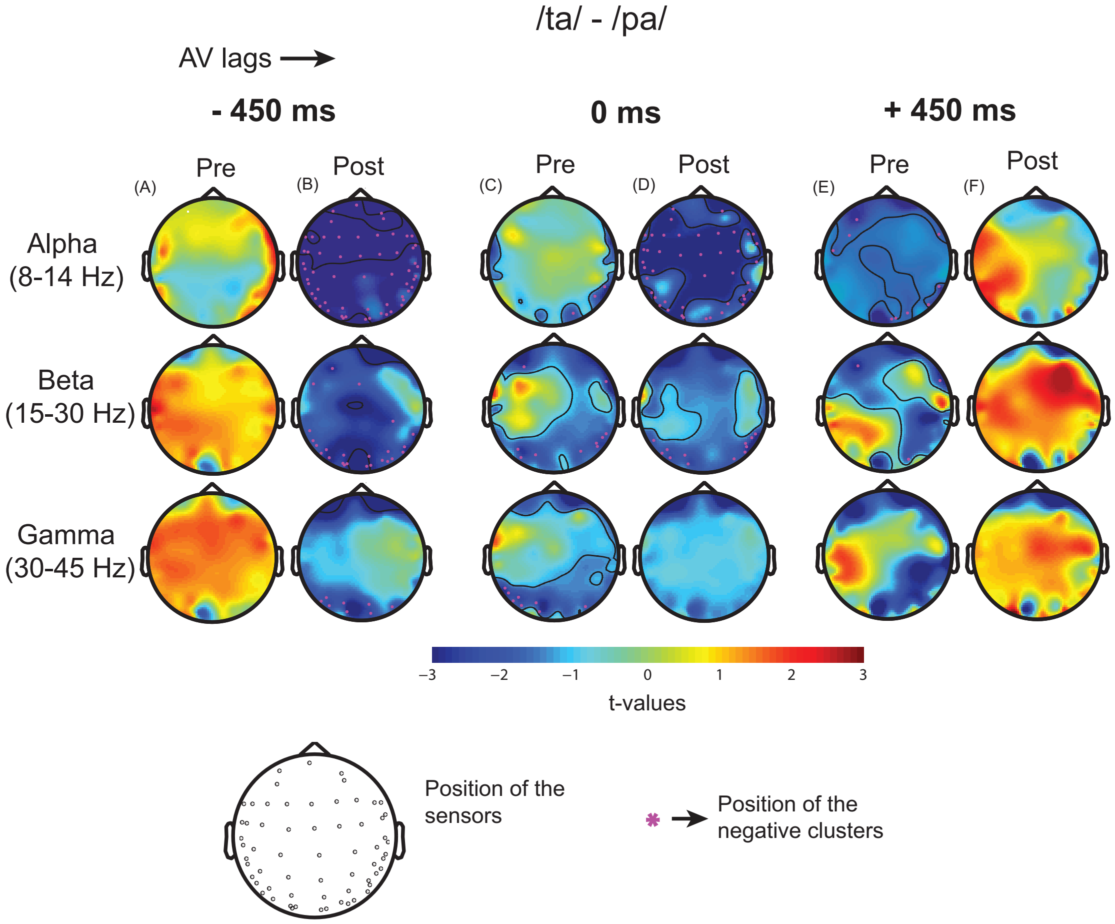

Figure 2.4: **Spectral Difference**: The topoplots and the magenta "*" highlight the clusters that show significant difference between the perceptual categories /*ta*/-/*pa*/ during the three stimulus conditions: 450 ms AV lag at **(A)** pre-stimulus onset **(B)** post-stimulus onset, 0 ms AV lag **(C)** pre-stimulus onset **(D)** post-stimulus onset, +450 ms AV lag **(E)** pre-stimulus onset **(F)** post-stimulus onset.

450 and +450 ms AV lag and the onset of AV stimulus for 0 ms AV lag. The mean coherogram plots for the perceptual categories /*ta*/ and /*pa*/ and their difference at AV lags: 450 ms (see **Figures 2.5A–C**), 0 ms (see **Figures 2.5D–F**), +450 ms (see **Figures 2.5G–I**) showed relatively heightened global coherence in the theta band (4–8 Hz) throughout the entire epoch duration. Cluster-based permutation tests employed to compare the mean coherogram for /*ta*/ and /*pa*/ at the respective AV lags revealed both positive and negative clusters (see

Figures 2.5C,F,I). Positive clusters highlighted in black dashed rectangles signify time-frequency islands of increased synchrony and the negative clusters in red dashed boxes signify islands of decreased synchrony in the global neuronal network.

In the pre-stimulus period, we observed two positive and one negative cluster each during 450 and +450 ms AV lag. The first and second positive clusters during 450 ms AV lag were observed in the frequency bands beta (16–30 Hz) ($z_{97.5} = 0.29$) and gamma (>30 Hz) ($z_{97.5} = 0.78$) respectively and the negative cluster was found in theta band (4–7 Hz) ($z_{0.025} = 0.29$). Here, $z_{97.5}$ and $z_{0.025}$ represent the two-tailed thresholds at $p = 0.05$ set by permutation tests to compute the significantly different cluster (for details, see Methods section). Similarly during +450 ms AV lag the first and second positive clusters were observed in the frequency bands beta ($z_{97.5} = 0.26$) and gamma ($z_{97.5} = 0.34$) respectively and the negative cluster was found in the alpha band (8–12 Hz) ($z_{0.025} = 0.78$). However, during 0 ms AV lag, only a significant positive cluster was observed in the alpha frequency band ($z_{97.5} = 0.58$).

In the post-stimulus onset period, during 450 ms AV lag (see **Figure 2.5C**), three positive clusters were observed, (1) in alpha band with temporal range between 200 and 560 ms ($z_{97.5} = 0.50$), (2) in beta band with temporal range between 50 and 500 ms ($z_{97.5} = 0.29$), and (3) in gamma band between 50 and 400 ms ($z_{97.5} = 0.78$). Also, a negative cluster ($z_{0.025} = 1.02$) was observed in the theta band between 800 and 900 ms. During +450 ms AV lag (see **Figure 2.5I**), two positive clusters were observed, one in the theta band ($z_{97.5} = 0.73$) between 0 and 500 ms and the other one in gamma band ($z_{97.5} = 0.34$) between 0 and 200 ms. A negative cluster was also observed in the theta band ($z_{0.025} = 0.68$) between 700 and 850 ms. Interestingly, during 0 ms AV lag (see **Figure 2.5F**) we observed a positive cluster ($z_{97.5} = 0.26$) precisely in the gamma band (300 and 700 ms) and three negative clusters ($p \leq 0.05$). Two of the negative clusters ($z_{0.025} = 0.31$) were observed in the theta band around 300 and 600 ms and

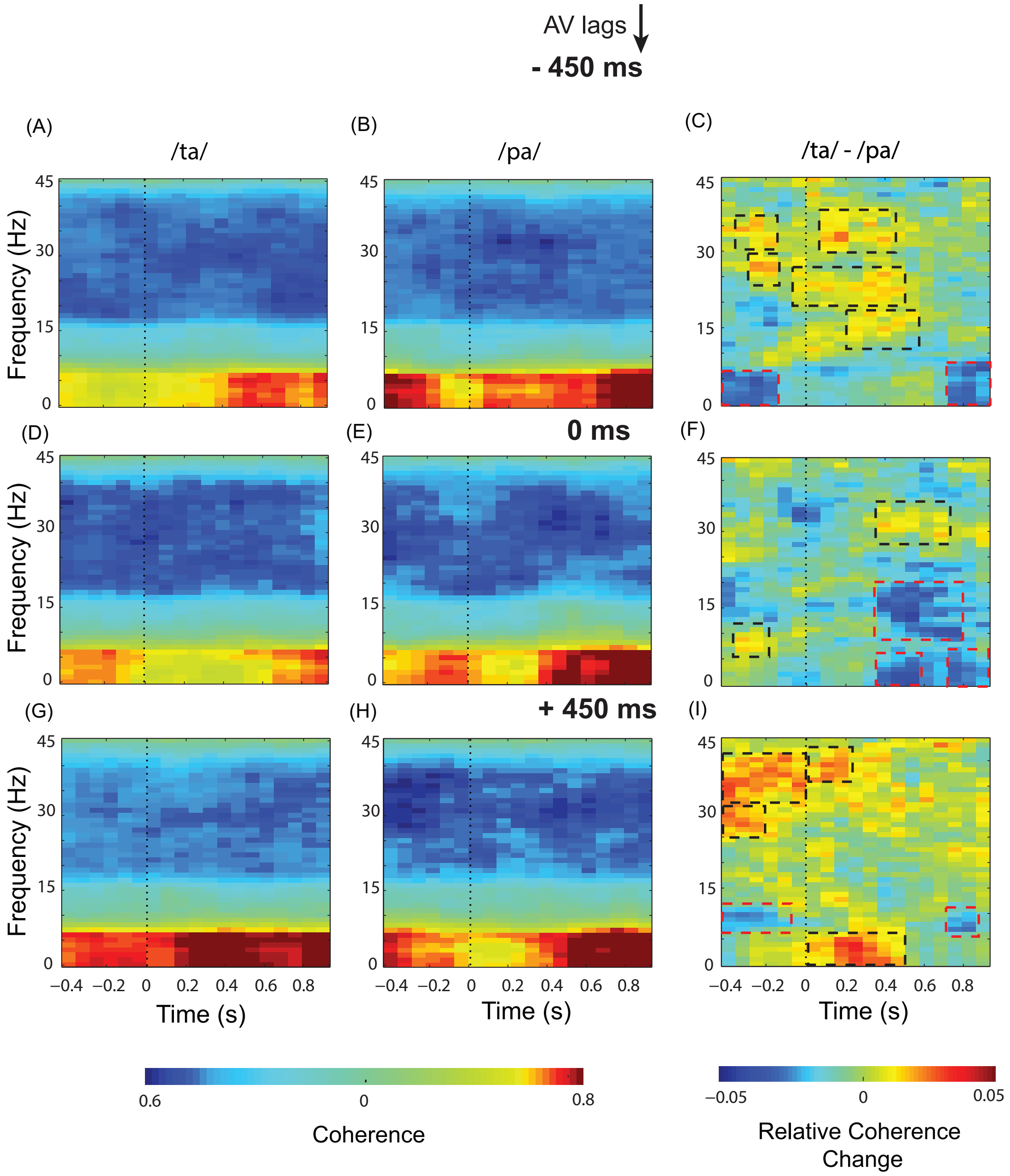

Figure 2.5: **Time-frequency representations of large-scale functional brain networks**: Mean time frequency coherogram for different perceptual categories time locked to the onset of the first sensory component (A or V) during the three conditions and the mean coherence difference between */ta/* and */pa/* responses at different AV lags: for 450 ms **(A)** */ta/* **(B)** */pa/* **(C)** */ta/-//pa/*; for 0 ms **(D)** */ta/* **(E)** */pa/* **(F)** */ta/-//pa/*; for 450 ms **(G)** */ta/* **(H)** */pa/* **(I)** */ta/-//pa/*..

700 and 900 ms and the third negative cluster incorporated both alpha and beta bands (9–21 Hz) ($z_{0.025} = 0.25$) and appeared between 300 and 800 ms.

2.4 Discussion

Characterizing the dynamics of the whole brain network is essential for understanding the neurophysiology of multisensory speech perception. We have shown that the spatiotemporal dynamics of the brain during speech perception can be represented in terms of brain oscillations and large-scale functional brain networks. We explicitly focused on investigating the characteristics of the brain networks that facilitate perception of the McGurk illusion. We exploited the perceptual variability of McGurk stimuli by comparing the oscillatory responses and network characteristics within identical trials. The main findings of the study are: (1) heightened global coherence in the gamma band along with decreased global coherence in the alpha and theta bands facilitates multisensory perception (2) a broadband enhancement in the global coherence at theta, alpha, beta, and gamma bands aids multisensory perception for asynchronous AV stimuli, as brain engages more energy for multisensory integration. We discuss the behavioral and neural-level findings in following sub-sections.

Variability of Perceptual Experience

A vast body of literature has reported that under controlled settings one can induce illusory perceptual experience in human participants [70,84,90,98,132]. Here, we constructed incongruent AV stimuli (auditory */pa/* superimposed onto video of face articulating */ka/*) using three different AV lags: 450 ms (audio precede articulatory movements), 0 ms (synchronous onsets of audio and articulatory movements), and +450 ms (articulatory movements precede audio) (see **Figure 2.1**). We identified that a categorical perceptual difference appeared with variation in AV lags. Synchronous AV stimuli resulted in higher percentage response of crossmodal (*/ta/*) perception (**Figure 2.2C**) whereas AV lags of 450 and +450 ms resulted in lowering of the percentage of crossmodal percept and higher occurrence of the unimodal percept */pa/*. Furthermore, we observed high hit rate of */ta/* responses both during congruent /ta/ stimuli (>90%) and during our post-hoc "auditory alone" behavioral experiment (>95%). Behavioral studies by van Wassenhove et al. (2007) [132] demonstrate 200 ms of asynchrony as the temporal window of bimodal integration. However, electrophysiological studies especially in the domain of preparatory processes demonstrate the elicitation of ERP components up to 600–800 ms in response to a cue followed by a target stimulus [121]. Extending this line of reasoning to our experimental paradigm, we believe an existence of temporal integration mechanisms beyond 200 ms does not allow the percentage of */pa/* perception to reach the level for congruent multisensory or purely auditory perception. In the current study we focused on the boundaries of stable illusory perception but the temporal boundaries of multisensory integration needs to be tested by future studies.

Interestingly percentage of gaze fixation at the mouth of the speaker for crossmodal response trials did not vary significantly at any AV lags based on t-test. Also, the interaction between lags and perceptual categorization was not significant when analyzed with 2-way ANOVA. Even though not statistically significant, the mean gaze fixation percentages at mouth for crossmodal perception were slightly higher than unimodal perception at all AV lags. Therefore, we cannot completely rule out the findings of an earlier study that show that frequent perceivers of McGurk effect fixate more at the mouth of the speaker [50] as well as we were limited by the number of participants to evaluate correlations between the behavioral results and the percentage of gaze fixation at 0 ms AV lag. On the other hand the subjective behavioral response for perceptual categorization clearly showed an interaction effect between AV lags and perceived objects. It is important to note that the identical multisensory stimuli generated varying responses for different trials. All stimuli being multisensory, differential perception served as an efficient handle to tap into the perceptual processing underlying speech perception. Our behavioral response results are consistent with previous studies on McGurk stimuli [96, 132] that demonstrate the influence of AV lags on perceptual experience. Hence, we expected to identify the neurophysiological processes underlying different multisensory perceptual scenarios.

Spectral Landscape of the Cortical Activity

Non-parametric statistical comparison between the perceptual categories (*/ta/–/pa/*) showed suppression of the spectral power in alpha, beta, and gamma frequency bands (see **Figure 2.4**). Suppression of alpha-band power has been associated with attention and language comprehension processes by enabling controlled access to knowledge [10,51,76,105]. Accordingly, the suppression of alpha-band power observed in our study can be attributed to the attention related network aiding access to stored knowledge and filter redundant information.

Beta-band power was observed to be suppressed at frontoparietal to occipital sensors during 450 ms AV lag and at occipital scalp regions during 0 ms AV lag but no such suppression was observed during +450 ms AV lag. Beta band power has been linked with various cognitive facets including top-down control of attention and cognitive processing [39]. Besides, in the domain of multisensory integration and language processing, suppression of beta-band power has been associated with the occurrence of unexpected stimuli [10, 141]. Furthermore, recent studies also show suppression of beta power during the perception of the McGurk illusion [110]. Extending the line of reasoning from the aforementioned studies, suppression of beta-band power might be associated to the occurrence of an unexpected stimulus and its processing. Visual-lead condition, wherein we observed no significant difference in the beta power, is possibly the most predictable situation and hence significant beta power modulation was not detected. Behaviorally, Munhall et al. (1996) [96], report McGurk illusion is most dominant between an AV lag of 0–200 ms and there is a slight asymmetry toward positive AV lags (visual lead). In fact, our data from a different experiment also replicated this result.

Gamma-band power was observed to be significantly suppressed only during

450 ms AV lag at the occipital scalp regions. Also, in the pre-stimulus period significant reduction in gamma band power was observed at occipital scalp regions during 0 ms AV lag. Existing studies have demonstrated the role of gamma-band oscillations in cognitive functions like visual perception, attention and in the processing of auditory spatial and pattern information [62,65]. Also, gamma band activity over sensory areas has been attributed to the detection of changes in AV speech [64]. However, we observed a suppression in gamma band activity which may be linked with preparatory processes over wider network that waits for the expected visual information to arrive. Although, the brain oscillatory responses to multisensory perception have been extensively studied, a consensus on the mechanisms associated with these oscillations remains elusive. Our study contributes to this vast body of work in conveying that multisensory speech perception requires complex signal processing mechanisms that involves the participation of several brain regions. Therefore, understanding the process requires analyzing the whole brain operating as large scale neurocognitive network. In the subsequent section we discuss the network analysis results.

Neurocognitive-Network Level Processing Underlying Illusory Perception

Global time-frequency coherogram (see **Figure 2.5**) computed for the perceptual categories quantifies the extent of coordinated neuronal activity over the whole brain. Global coherence reflects the presence of neuro-cognitive networks in physiological signals [17]. Previous studies posits that neuronal coherence could provide a label that binds those neuronal assemblies that represent same perceptual object [37,38,136]. Besides, going by the communication-through-coherence (CTC) hypothesis, only coherently oscillating neuronal groups communicate effectively as their communication window for spike output and synaptic input are open at the same time [43,118]. Hence, coherent transmission poses a flexible mechanism that facilitates the integration of converging streams in time windows of varying duration. In our analysis we observed a relatively

heightened theta-band coherence for both the perceptual categories at all the AV lags (see **Figures 2.5 A,B,D,E,G,H**). Theta band coherence has been associated to cognitive control processes [26]. Accordingly, the enhanced theta-band coherence might reflect the control processes preparing for upcoming stimuli.
Non-parametric statistical analysis employed to test the global coherence differences between */ta/* and */pa/* during 0 ms AV lag, revealed a positive cluster, signifying enhanced synchrony specifically at the gamma band (between 300 ms and 700 ms). Also, we observed negative clusters (between 300 and 900 ms) in the theta, alpha and beta bands that signify decreased synchrony among the underlying brain regions. Overall temporal congruence of AV stimuli results in a narrow-band coherence whereas lagged AV stimuli seemed to engage a broadband coherence (see **Figure 2.5 C,F,I**). However, we had one limitation because of the nature of our stimuli. A direct statistical comparison across lagged conditions was not meaningful since each lagged condition had a different temporal sequence of audio-visual components.

Inter-areal coherence of oscillatory activity in the beta frequency range (15–30 Hz) has been associated with top-down processing [139]. Moreover, top-down processing involves the modulation of the hierarchical sensory and motor systems by pre-frontal and frontal brain areas [92]. The dense anatomical interconnectivity among these association areas give rise to self-organized large scale neuronal assemblies defined as neuro-cognitive networks (NCNs), with respect to the cognitive demands [20]. In this context, our finding of increased coherence in the beta band during 450 and +450 ms AV lag is especially relevant as it enables us to hypothesize that synchronization of the beta oscillations provides long range inter-areal linkage of distributed cortical areas in NCNs. Such networks can readily process the retrieval of well learnt audio-visual associations suggested by Albright (2012) [3].

Gamma band coherence are shown to be associated with voluntary eye movements, saccades [9]. Besides, stimulus selection by attention also induces local gamma band synchronization [55]. Our results show enhanced gamma coherence (positive cluster) at all AV lags. Considering the increased gaze fixation at mouth during /ta/ perception, heightened gamma coherence reflects the recruitment of the visual attention areas. A recent review proposes that gamma band (30–90 Hz) coherence activates postsynaptic neurons effectively by modulating the excitation such that it escapes the following inhibition [43]. Besides rendering effective communication, gamma coherence has also been proposed to render communication that are precise and selective [22, 43]. Importantly, gamma band coherence has also been demonstrated to be implicated in associative learning [93]. Thus, our observation of enhanced coherence exclusively at gamma and desynchronization at alpha and beta-bands during 0 ms AV lag portrays an attention network working in harmony with the NCNs most likely linked to associative memory retrieval. This conjecture is also supported by the secondary evidence in case of 450 and 450 ms AV lags, where an additional working memory process is competing for processing and integration of the multisensory stimuli and leading to a broadband enhancement in global coherence. A more detailed delineation of working memory processing and associative memory recall needs to be carried out with other kinds of multisensory stimuli and will be a major focus of our future endeavors.

Chapter 3

Multi-scale cortical representation of cross-modal perception

3.1 Introduction

Combination of information from different senses enhances our perceptual and response ability. For example, although speech perception is based on the processing of the auditory signals, speech intelligibility can be influenced when it is accompanied by the visual articulatory gestures of the speaker. This can either result in enhancement of the auditory perception [53, 126] or modulate it when accompanied with semantically-incongruent lip movements [90]. Numerous research papers have explored the cortical correlates of multisensory perception, and demonstrated the involvement of specific modules and distributed cortical networks. However, it remains unclear at what scales these networks are engaged and what the most pertinent substrate is for representing the mechanism of multisensory perception.

The conventional view of sensory processing is that convergence and integration of information across different modalities occurs in specific cortical modules post extensive processing within sensory-specific subcortical and cortical regions. However, evidence from recent studies shows that multisensory integration extends beyond modularity and suggests that multisensory convergence is considerably widespread in the brain [15, 23, 91]. Furthermore, even the primary sensory areas have been claimed as a part of the emerging network of multisensory regions [5, 15]. From the perspective of localization of function via integration hypothesis [18, 56, 83, 91], it is fundamentally important to understand the network-level mechanisms at various spatiotemporal scales over which multisensory information processing is represented.

Behavioral and neuroimaging studies in the domain of speech perception have extensively used McGurk effect to gain insights on mechanism of audio–visual (AV) integration and multisensory perception [47, 61, 63, 70, 78, 125, 131, 137]. During the McGurk effect, an auditory speech sound */ba/* superimposed onto the visual lip movement of */ga/* gives rise to an illusory (cross-modal) percept of */da/* [90]. A substantial amount of evidence employing the McGurk effect demonstrates activation of specific cortical modules like the pSTS (posterior Superior Temporal Sulcus) [61,98,99,117], frontal and parietal areas [61,122] being responsible for the cross-modal perception. On the other hand, studies employing connectivity measures on functional imaging and electrophysiological data primarily reveal interactions among cortical regions of interest [70] or characterize the properties of the global network [78] endorsing the mechanism of functional integration. However, to our knowledge no study has reported that both mechanisms are operational on a putative data set along with their variability across trials. Therefore, investigating the interplay between the modular components of an extended cortical network of multisensory regions concomitantly with dynamic changes within the components would help us develop a comprehensive account of underlying mechanisms involved in multisensory perception. In the present study, we used an incongruent McGurk pair (audio */pa/* superimposed on a video of the face articulating */ka/*) to induce the cross-modal percept */ta/*. Further, we introduced a temporal asynchrony in the onset of audio and visual events of the McGurk stimuli to diminish the rate of cross-modal responses */ta/*, in comparison to the unimodal response of */pa/*, thus creating two perceptual categories which can be further studied from the perspective of integration and segregation of information processing in the brain at different spatial scales. We observed the representation of dynamical information processing at each spatial scale, the individual sensor level in EEG data (using time series and spectro-temporal representations of sensor-level power) and large-scale brain networks (using the imaginary coherence to extract the between-sensor interactions), indicating that multi-scale representation of the AV integration is pertinent for a comprehensive understanding of multisensory speech processing.

3.2 Materials and Methods

3.2.1 Participants

Nineteen healthy volunteers (10 males and 9 females, in the range of 22–29 years of age; mean age 25, SD = 2) participated in the study. All participants gave written informed consent, and they had no neurological or audiological problems. They all had normal or corrected-to-normal vision and were right-handed. The study was carried out following the ethical guidelines and prior approval of the Institutional Review Board of the National Brain Research Centre, India. The data from four volunteers were not included in the study because they reported to hear only the auditory stimuli and did not perceive the McGurk effect when audio–visual stimuli were incongruent.

3.2.2 Stimuli and Trials

Stimuli

Each participant responded to 360 trials which consisted of videos of a native Hindi-speaking male articulating the syllables /*ka*/ and /*ta*/ (see **Figure 3.1**). One-fourth (90 trials) of the trials consisted of congruent video (visual /*ta*/ auditory /*ta*/). The remaining three-fourths of the trials comprised incongruent videos (visual /*ka*/ auditory /*pa*/) presented with AV lags: 450 ms (audio leads the articulation), 0 ms (synchronous) and +450 ms (articulation leads the audio), each encompassing one-fourth of the overall trials. The audio syllable was extracted from a video of the speaker articulating /*pa*/ using the software Audacity (https://www.audacityteam.org). Subsequently, the extracted audio syllable was superimposed onto the muted video of the speaker articulating the syllable /ka/ using the software Videopad Editor (https://www.nchsoftware.com). The stimuli were rendered into a 800 600 pixels movie with a digitization rate of 29.97 frames per second. Stereo soundtracks were digitized at 48 kHz with 32 bit resolution. Presentation software (Neurobehavioral System Inc.) was used to present the stimuli using a 17 LED monitor. Sound was delivered using sound tubes at an overall intensity of 60 dB.

Experimental Design

The experiment was divided into three blocks. Each block consisted of 120 trials comprising four kinds of videos (30 trials of each): a congruent video and the three incongruent McGurk pair videos with AV lags. Inter-stimulus intervals were pseudo-randomly varied between 1200 ms and 2800 ms to minimize expectancy effects. The subjects were instructed to report what they heard while watching the speaker using a set of three keys: /*pa*/, /*ta*/ or 'anything else'. The subjects also performed a behavioral task post EEG scan. The task consisted of 60 trials, comprising 30 trials of the auditory syllables /*ta*/ and /*pa*/ each. The subjects were instructed to report what they heard using the choices /*ta*/ and /*pa*/.

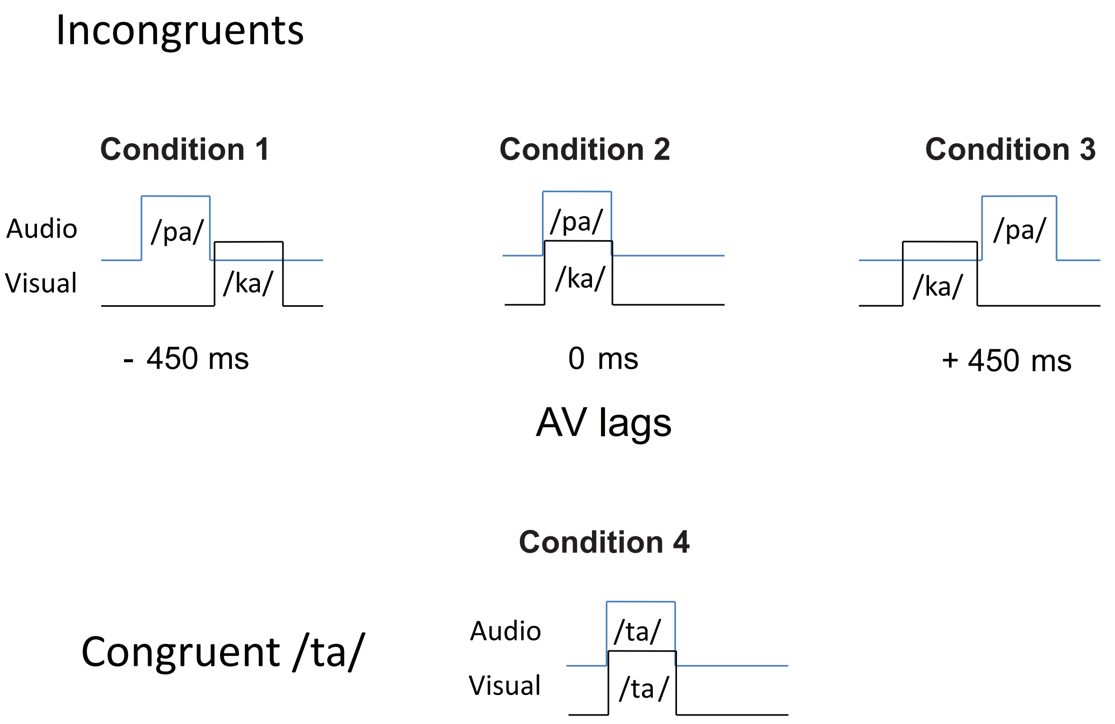

Figure 3.1: **Stimuli**: Each condition represents a video of speaker articulating a speech sound. AV lags show the temporally incongruent placement of the audio /*pa*/ with respect to the articulation (lip movement) of /*ka*/. The congruent /*ta*/ represents a video with audio /*ta*/ dubbed onto a video of a person articulating /ta/.

3.2.3 Data Acquisition and Analysis

EEG

Continuous EEG scans were acquired using a Neuroscan system (Synamps2, Compumedics, Inc.) with 64 Ag/AgCl scalp electrodes sintered on an elastic cap in a 10–20 montage. Recordings were made against a centroid (Cz) reference and digitized at a sampling rate of 1000 Hz. Channel impedances were kept at values < 5 kΩ.

Preprocessing of EEG Signals

The EEG data acquired was initially re-referenced to linked mastoids and filtered using a bandpass of 0.2–45 Hz. Subsequently, the continuous EEG was divided into epochs (400 ms to 900 ms surrounding the onset of the first stimulus, i.e., the sound or articulation) and sorted based on the responses, */ta/*, */pa/* and 'other', respectively. Epochs were baseline-corrected by removing the temporal mean of the EEG signal on an epoch-by-epoch basis. Subsequently, we performed artifact rejection to eliminate the response contamination from ocular and muscle-related activities. However, depending on the analysis, we used two different thresholds. For statistical analysis of the event-related potentials, to minimize false positives arising from high amplitude in the low-frequency waveforms, epochs with a maximum signal amplitude above 50μV or a minimum below 50μV were removed from all electrodes. For spectral and network analysis, we used a signal amplitude threshold of $\pm 100\mu$V for artifact rejection as amplitude differences in waveforms will have no relevance in the spectral domain.

Event-Related Potential (ERP) Analysis

The preprocessed EEG data were further sorted according to the responses using customized MATLAB codes. After pooling across all subjects, the ERPs for each condition contained a minimum of 128 trials, were averaged and plotted across all electrodes. As we specifically focused on the difference in the ERP pattern between the */ta/* and */pa/* responses, the sorted epochs for each stimulus condition were compared statistically. Ms-by-ms paired t-tests were performed between the */ta/* and */pa/* responses across all electrodes to evaluate the spatio-temporal properties of AV integration. For each scalp electrode, the first time point where the t-test yielded a p-value < 0.05 and continued to do so for at least 20 consecutive data points (20 ms) was considered significantly different. The method serves as an alternative to Bonferroni correction for multiple comparisons, which would increase the possibility of false negatives [97].

Spectral Analysis

A time–frequency spectrogram of EEG signals at each electrode was computed on a single-trial basis and sorted based on the responses, /*ta*/, /*pa*/ and 'other', respectively. We computed the spectral power at different frequencies over time using customized MATLAB (https://www.mathworks.com) codes and the Chronux toolbox (https://www.chronux.org). The time bandwidth product and the number of tapers were set at 3 and 5, respectively, and a fixed time window of 0.3 s was applied while using the Chronux function mtspecgramc.m to compute the time–frequency spectrogram of the sorted time series in EEG data. The time–frequency spectrogram computed for the perceptual categories /*ta*/ and /*pa*/ were compared channel by channel employing cluster-based permutation tests [87]. During the cluster-based permutation tests, 1000 iterations of trial randomization were carried out to generate the permutation distribution at a frequency band at a time point. Subsequently, a two-tailed test with a threshold of 0.025 was used to evaluate the positive (increased spectral power) and negative (decreased spectral power) clusters at the respective sensors.

Network Analysis

To comprehend the cortico-cortical interactions underlying AV integration, we assessed the imaginary component of pairwise sensor-level coherence introduced by Nolte and colleagues [101]. This functional connectivity estimate captures the 'true' brain interactions that occur with a certain time lag, neglecting the spurious interactions arising from common references, volume conduction and crosstalk. Imaginary coherence refers to the complex part of the coherency C_{ij} that quantifies the phase relationship between two time series $\hat{x}_i(t)$ and $\hat{x}_j(t)$ at a specific frequency f . Coherency $C_{ij}(f)$ is the normalized cross-spectrum between two signal pairs, which in the current study are the EEG signals from different sensor pairs i and j .

$$C_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}S_{jj}}} \tag{3.1}$$

where S_{ij} is the cross-spectrum obtained by performing the complex conjugate of the Fourier transforms of $\hat{x}_i(t)$ and $\hat{x}_j(t)$. Imaginary coherence was evaluated in the time window of 0.9 s post the onset of the first stimulus (audio or visual) for each perceptual category (*/ta/* and */pa/*) at all the AV lags. We employed the Chronux function crossSpecMatc.m to obtain the normalized cross-spectral matrix for all sensor combinations. Subsequently, we extracted the imaginary part of the cross-spectral values that constitute the imaginary coherence. The values of the imaginary coherence in the three frequency bands (alpha, beta and gamma) were further averaged using the Circular statistics function `circ_mean.m`. Imaginary coherence computed for */ta/* and */pa/* responses were further compared between each channel pair for significant difference at different frequency bands (alpha, beta and gamma) explicitly by means of the cluster-based permutation test [87]. For each channel pair, the imaginary coherence difference between */ta/* and */pa/* was evaluated using the Fisher's Z transformation

$$Z(f) = \frac{\tanh^{-1}(C_1(f)) - \tanh^{-1}(C_2(f)) - (\frac{1}{2m_1-2} - \frac{1}{2m_2-2)!})}{\sqrt{\frac{1}{2m_1-2} + \frac{1}{2m_2-2}}} \tag{3.2}$$

where $2m_1$, $2m_2$ = degrees of freedom; $Z(f) \approx N(0,1)$ a unit normal distribution; and C_1 and C_2 are the coherences at frequency$f$.

The coherence Z-statistic matrix obtained from the above computation formed the observed Z-statistics. Consequently, 1000 iterations of trial randomization were carried out to generate the permutation distribution at a frequency band for each channel pair. Subsequently, a two-tailed test with a threshold of 0.001 was used to evaluate the channel pairs that showed significantly different interactions between the two perceptual categories. The same statistical tests were carried out to test the differences at different AV lags.

3.3 Results

3.3.1 Behavior

We converted the behavioral responses corresponding to McGurk stimuli with the AV lags to percentage measures for each perceptual category (*/pa/*, */ta/* or 'other') from all subjects using customized Matlab codes. To qualify a participant as an illusory (cross-modal) perceiver, we set a minimum threshold of 60% of */ta/* response in any AV lag, 450, 0 and +450 ms. Fifteen participants qualified and four participants failed to perceive above the set threshold. Data from only 15 perceivers were used for further group-level analysis. We observed that a maximum percentage of illusory (*/ta/*) responses occurred at 0 ms AV lag (**Figure 3.2**). The percentage of */pa/* responses was also at minimum at 0 ms AV lag. We ran a repeated-measures two-way ANOVA on the percentage responses with AV lags and the perceptual categories (*/ta/* and */pa/*) as the variables. We observed that there was no influence of AV lags in the percentage of responses of */ta/* and */pa/* [$F_{(2,89)} = 0.84, p = 0.44$]. However, we found a significant difference in the percentage responses between the two perceptual categories [$F_{(1,89)} = 19.46, p < 0.0001$]. Also, the interaction between perceptual categorization and AV lags was significant [$F_{(2,89)} = 23.83, p < 0.0001$]. Furthermore, we performed a post-hoc test using the Scheffe method on the perceptual categories. We observed a significant difference in the percentage responses between the two perceptual categories at the 95% confidence level. We also performed a paired Student's t-test on the percentage of responses (*/ta/* and */pa/*)

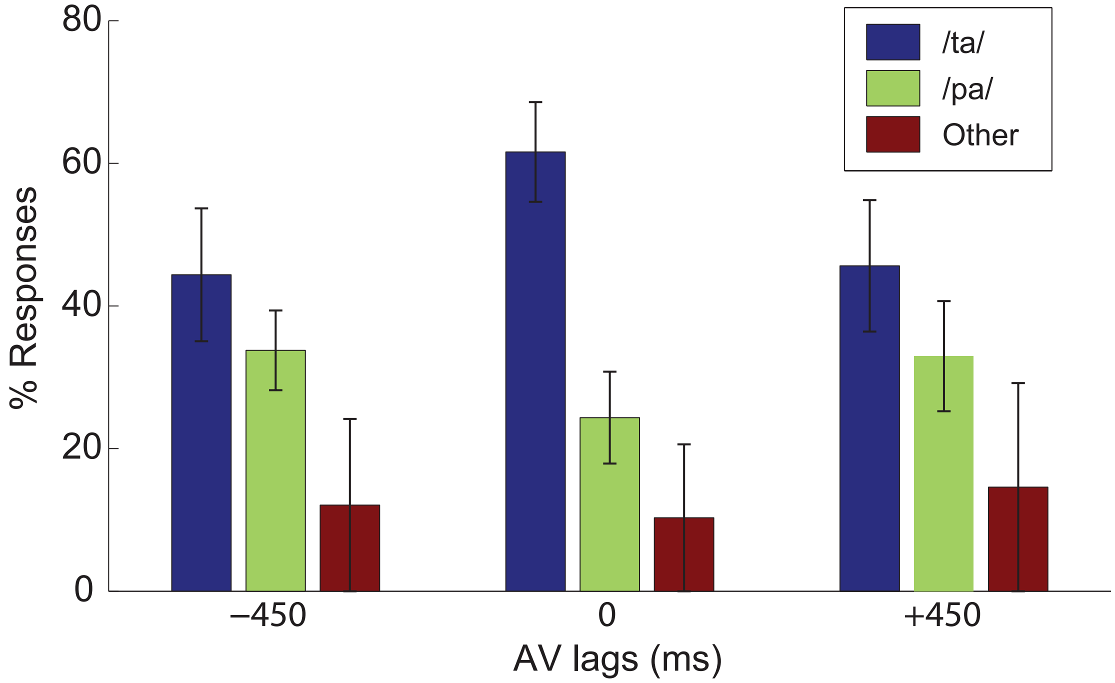

Figure 3.2: **Behavior**: Percentage of perceptual categorization for */pa/*, */ta/* and 'other' percepts as a function of AV lags, normalized and grouped over all 15 perceivers. The error bars represent 95% confidence interval.

at each AV lag. Insignificant differences of 10.20% and 11.40% were observed between */ta/* and */pa/*responses at 450 ms AV lag [$t_{(14)} = 0.63, p = 0.27$] and +450 ms AV lag [$t_{(14)} = 0.45, p = 0.67$], respectively. However, at 0 ms AV lag we observed that the percentage of */ta/* responses was significantly higher by 36.58% than the percentage of */pa/* responses [$t_{(14)} = 10.20, p < 0.0001$]. The hit rate of */ta/* responses during congruent /ta/ was observed to be 0.97. Also, the hit rate of */ta/* and */pa/* during auditory-alone conditions was observed to be 0.96 and 0.98, respectively.

3.3.2 Event-Related Activity

The difference wave obtained by subtracting the event-related responses of /pa/ from the responses of */ta/* (*/ta/* − */pa/*) for the AV lags 450 ms, 0 ms

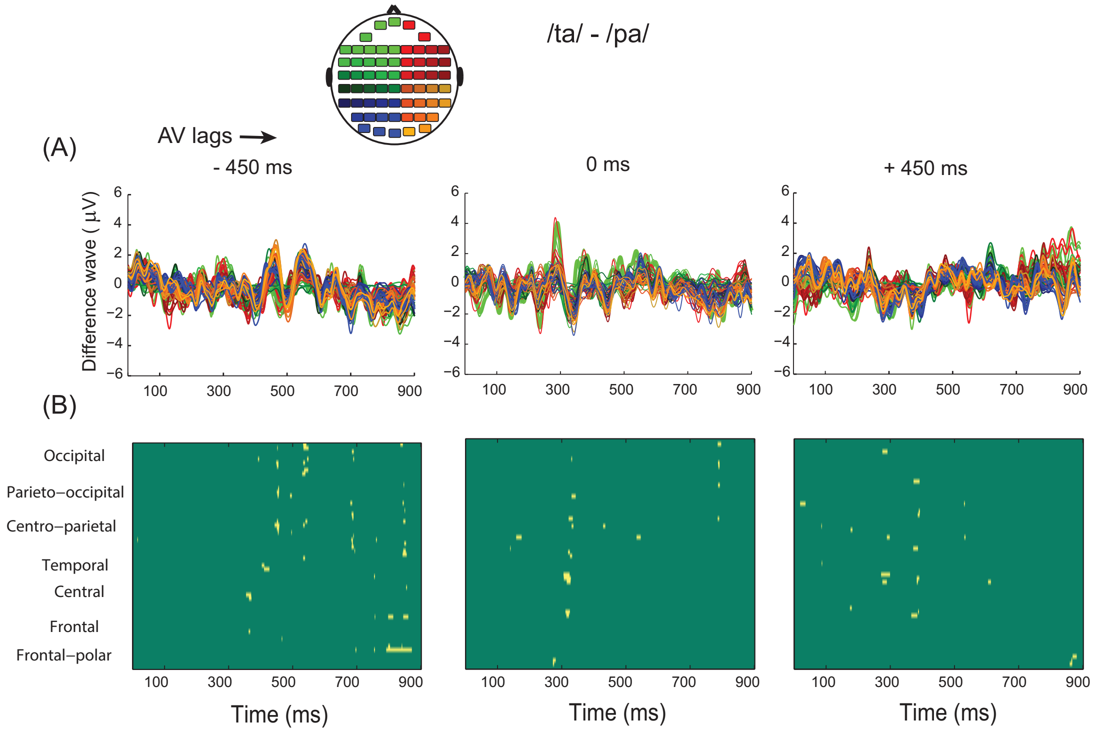

Figure 3.3: **Event-Related Potential**:**(A)** The difference wave between ERPs sorted out from /*ta*/ and /*pa*/ response trials at 450 ms, 0 ms and +450 ms AV lag. The topoplot at the top left displays the color code used for plotting ERPs assigned to respective scalp channel locations. For example, the green and red positive peaks around 300 ms represent the peak of activity in the left frontal and right frontal sensors. **(B)** Statistical cluster plots of the difference between the perceptual categories (/*ta*/ and /*pa*/) for each stimulus. The clusters indicate the time points where the p-values were < 0.05 for more than 20 ms. General sensor positions are arranged from frontal to posterior regions (bottom to top).

and +450 ms at all scalp electrodes are shown in Fig. 3A. In the difference wave, we observed a positive peak between 300–380 ms in frontal-polar, frontal and centro-parietal sensors at 450 ms AV lag and in frontal-polar, central, temporal, centro-parietal and parieto-occipital sensors at 0 ms AV lag, respectively. However, we did not observe any such peaks in the difference wave at +450 ms AV lag.

To compute the sensors eliciting significantly different amplitude during /*ta*/ responses than /*pa*/ responses, we performed millisecond-by-millisecond t-tests between the two conditions. To ignore transient responses, the criteria for significance were chosen such that at the onset latency the first point in the time series was where the p-value was less than 0.05 and remained so for at least 20 ms consecutively. The cluster plots in **Figure 3B** exhibit such temporal windows. At 450 ms AV lag we observed a difference in the frontal and central sensors at 370 ms followed by which we observed a difference in the temporal, centro-parietal, parieto-occipital and occipital sensors ranging from

3.3.3 Power of Oscillatory Activity

The relative difference in the time–frequency spectrogram between */ta/* and */pa/* responses (*/ta//pa/*) at each sensor obtained after the cluster-based permutation test is shown in **Figure 3.4**. Figures **3.4A, B and C** plot the differences in the spectral power between */ta/* and */pa/* responses at 450 ms, 0 ms and +450 AV lag, respectively. At 450 ms AV lag, we observed negative clusters predominantly in the theta and alpha bands denoting a decrease in the spectral power in the left frontal, left temporal and bilateral occipital sensors. However, at 0 ms AV lag we observed positive clusters, denoting an increase in the spectral power in the theta and alpha frequency bands predominantly in the occipital sensors and left temporo-parietal and right centro-parietal sensors in addition to suppression of alpha and theta power in left frontal areas. At +450 ms AV lag, we observed a bilateral decrease in the spectral power in the frontal and temporal sensors. On the left frontal and temporal sensors, negative clusters were observed in the theta, alpha, beta and the gamma bands. However, in the right temporal sensors negative clusters were observed in the theta bands.

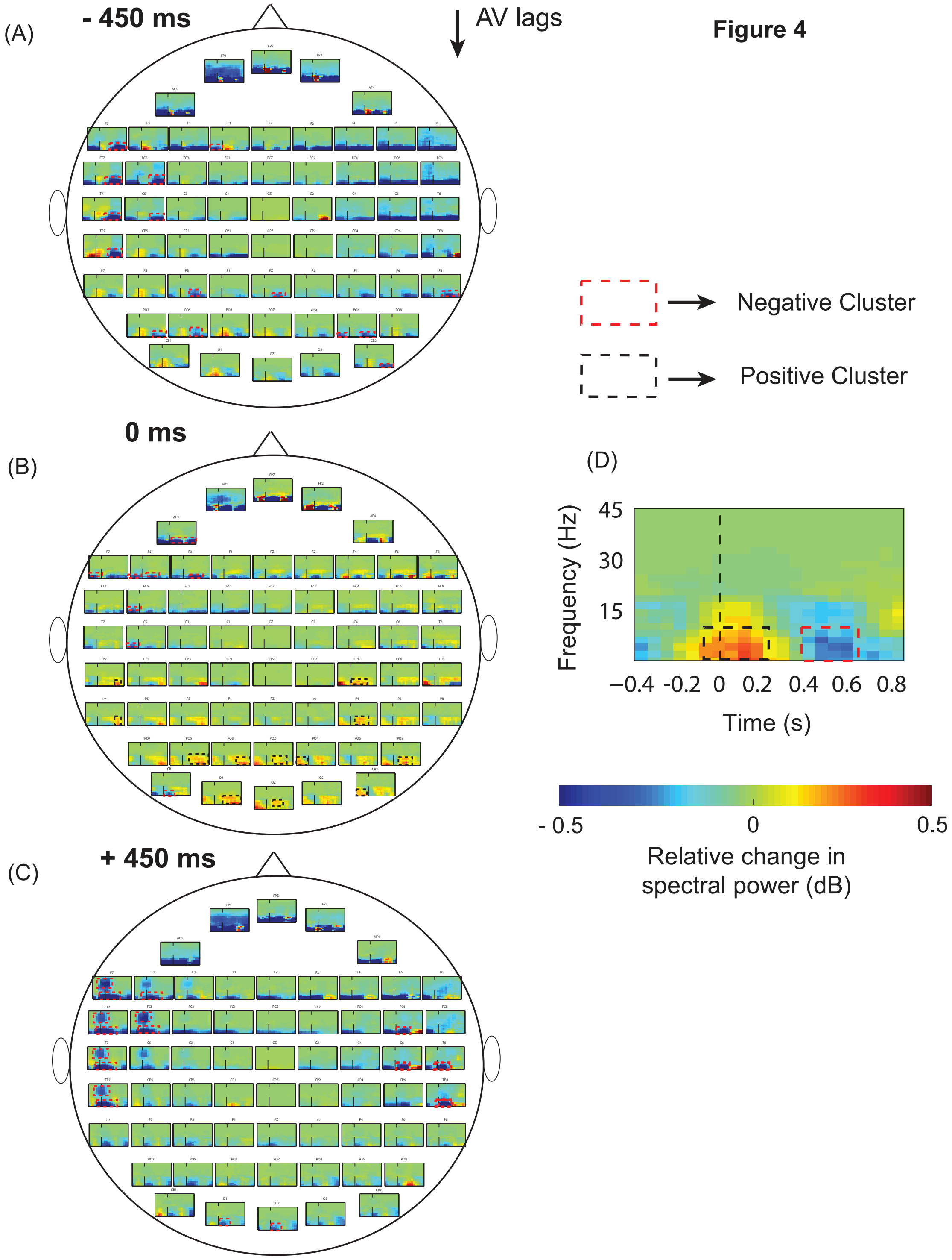

Figure 3.4: **Power spectral analysis**: Time–frequency spectrogram difference at each sensor time locked to the onset of the first stimulus (400 ms pre-stimulus and 900 ms post stimulus): **(A)** 450 ms AV lag, **(B)** 0 ms AV lag and **(C)** +450 ms AV lag. The red and black dotted boxes represent the areas in the respective sensors that exhibit a significant difference between the perceptual categories (/*ta*/ and /*pa*/). Panel **(D)** represents an enlargement of the spectrogram at each sensor showing islands of increased and decreased power.

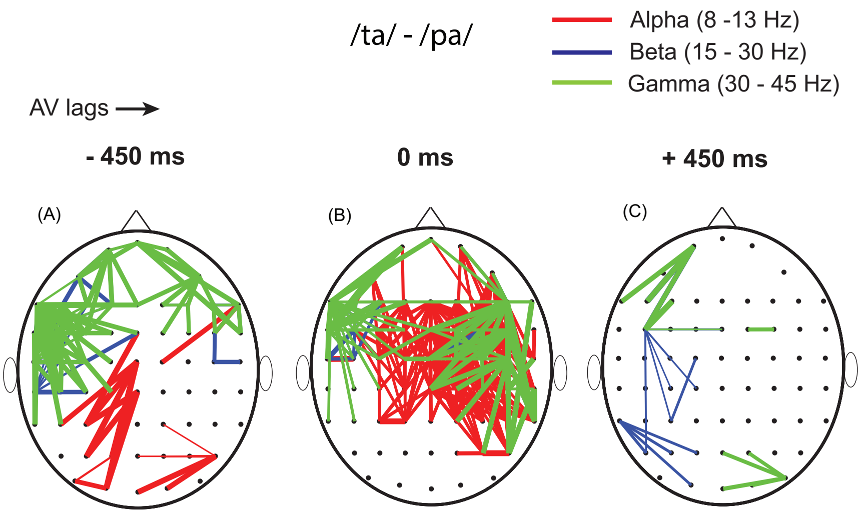

Figure 3.5: **Functional connectivity changes:** Imaginary coherence difference between /*ta*/ and /*pa*/ response trials. Dots indicate the channel location and the lines indicate channel pairs with statistically significant (p < 0.001) imaginary coherence changes at different frequency bands as indicated by the color codes in the top right at **(A)** 450 ms, **(B)** 0 ms, and **(C)** +450 ms AV lags.

.

3.3.4 Functional Connectivity

To assess the functional connectivity underlying AV integration, we non- parametrically compared the imaginary coherence between (/*ta*/) and unisensory (/*pa*/) responses from all the pairwise sensor combinations. At 0 ms AV lag (**Figure 5B**), significant differences in the connectivity were observed in the alpha band bilaterally between frontal-parietal sensors, unilateral right frontal-temporal, frontal-occipital temporal and temporal-occipital sensors; in the beta band between left frontal-temporal and right frontal-parietal sensors; in the gamma band between bilateral frontal-parietal and frontal-temporal sensors, right frontal-occipital, temporal-parietal and parietal-occipetal sensors. At +450 ms AV lag (**Figure 5C**), significant differences in interaction were observed in the beta band between left temporal-parietal, temporal-occipital and among occipital sensors; in the gamma band among left frontal sensors and among right occipital sensors.

3.4 Discussion

In the present study, we used EEG to investigate the spatiotemporal structure of cortical activity underlying multisensory speech perception. We exploited the trial-by-trial variability in the perception of McGurk stimuli to identify the neural representation of multisensory speech perception at different scales. We compared the neural correlates of unisensory and cross-modal perception using identical stimuli at the ERP, spectral and large-scale functional network level. Thus, we could capture the trial-by-trial variability of a participant as well as the segregation-based information-processing mechanisms at the individual sensor level (from ERP, spectral methods) and integration-based information-processing mechanisms (using imaginary coherence) in one single study. The main findings of the study are: (1) A positive peak in the latency range of 300–400 ms serves as a temporal marker of AV integration; (2) decreased post stimulus theta (4–7 Hz), alpha (8–12 Hz) and beta (13–30 Hz)band activity across frontal, temporal sensors and enhanced theta and alpha band activity across occipital sensors act as a spectral signature for cross-modal perception; (3) enhanced functional connectedness at the gamma band with the frontal sensors is pivotal for cross-modal perception.

Previous studies have shown that by presenting certain *semantically-incongruent* AV stimuli, one can induce an illusory (cross-modal) perceptual experience in the participants [84, 90, 132]. In the current study, we constructed incongruent AV stimuli by superimposing auditory /*pa*/ onto video of the speaker articulating /*ka*/ to induce an illusory percept of /*ta*/. Furthermore, studies have also demonstrated that the illusory experience can be modulated by the introduction of AV lags [96, 132]. Therefore we introduced an AV lag of 450 ms to our incongruent AV stimuli to generate three conditions overall: 450 ms (audio preceding video), 0 ms (synchronous onsets of audio and video) and +450 ms (video preceding audio) AV lag. We observed that the stability of the illusory percept varied with the introduction of the AV lags. Synchronous AV stimuli resulted in a response of illusory perception that was stable and at a significantly higher frequency of occurrence than the unisensory percept /*pa*/, whereas AV lags of 450 ms and +450 ms resulted in lowering of the illusory percept and a higher occurrence of the unisensory percept /*pa*/. Additionally, we observed a hit rate of /*ta*/ responses above 90% for congruent /*ta*/ stimuli and above 95% during our post-hoc 'auditory alone' behavioral experiment. Our behavioral response results corroborate existing studies of the McGurk effect [96,132] that demonstrate the effect of AV lags on perception. Furthermore, variability in the perception of identical incongruent stimuli served as an efficient handle to compare and understand the processing of multisensory speech stimuli [129].

Segregation of Information Processing Underlying Illusory Perception

Timing of Neural Information Processing

Converging evidence suggests that conscious perception is marked by a higher P300 component [106,109,113]. Our results demonstrate a robust positive peak in the temporal window of 300–400 ms as seen in the ERP difference plot (**Figure 3A**) at 450 ms and 0 ms AV lags. The results are further validated by cluster plots of ERPs obtained from millisecond-by-millisecond paired t-tests (**Figure 3B**) between /*ta*/ and /*pa*/ at all the AV lags. Although no robust peak around 300 ms was observed during +450 ms AV lag, cluster plots demonstrate a difference across central, temporal, centro-parietal and occipital sensors around 300 ms post stimulus onset. Importantly, significant differences in the ERP start only post 300 ms stimulus onset at 450 ms and 0 ms. In addition, interestingly the difference persists longer at 450 AV lag than at 0 ms AV lag, where the difference was observed in most sensors closely around the 300 ms window. We attribute the persistence of difference beyond 300 ms at 450 AV lag to the neurophysiological processes involved in binding the information across the two modalities. Considering the asynchronous AV stimuli, one can hypothesize that the neurophysiological process is the working memory that holds the first incoming stimulus (audio or visual) before integrating with the upcoming stimulus. Behavioral studies by Van Wassenhove and colleagues [132] demonstrate 200 ms of asynchrony as the temporal window of AV integration. However, electrophysiological studies understanding preparatory processes show the elicitation of ERP components upto 600–800 ms in response to a cue followed by a target stimulus [121]. In light of this finding we can endorse our speculation of the persistent difference post 300 ms at 450 AV lag arising from the underlying binding processes. The smaller difference window observed at 0 ms AV lag indicates an integration mechanism that is distinct from the processing when the AV stimuli are time-lagged. These mechanisms can be understood further by inspecting the signals at different scales. Furthermore, we also observe a difference before 300 ms, primarily at the central and parietal electrodes at +450 ms AV lag. These might arise from the anticipatory processes trying to predict the auditory representation following articulatory cues. Our findings here primarily point to the P300 component as the temporal marker of cross-modal perception.

Spectro-Temporal Structure of Brain Rhythms at Each Sensor

Oscillatory cortical activity modulates and drives perception [133]. Non-parametric statistical comparison of the time–frequency spectrogram between the perceptual categories (*/ta/* - */pa/*) (**Figure 3.4**) highlights the durations and frequencies at each sensor that have significantly different signal power change. The patterns of spectral difference allow us to speculate on the mechanism of AV integration which we discuss in the following paragraph. At 450 AV lag, we observed a suppression in the theta and alpha bands primarily in the frontal, left-temporal and occipital sensors. Similarly, at +450 ms AV lag, we observed a bilateral suppression of spectral power in the frontal and temporal sensors in the theta, alpha, beta and gamma bands. Theta band activity has been implicated in the encoding of new information and retrieval of episodic memories [75,102]. Furthermore, suppression of alpha band power has been implicated in attention and language comprehension processes by enabling controlled access to knowledge [10,51,54,76,105,120]. From an information processing perspective, event-related desynchronization in a local area indicates the onset of preparatory processes [54]. Also, differences across the sensors might reflect the activity in the underlying sensory-specific and working memory areas endorsing the fuzzy logical model of perception, in which each input is first independently evaluated with prototypes stored in memory followed by its integration and perception [89]. Our claim arises in the first place from the nature of the stimuli (450 ms and +450 ms AV lag) as in both cases either the audio */pa/* precedes the articulation or vice-versa. Furthermore, suppression of beta band power has been implicated in top-down control of attention [39]. Additionally, gamma band oscillations have been associated with visual perception, attention and the processing of auditory and spatial information [62,65]. Therefore, the suppression in the beta and gamma bands observed in the left temporal sensors at +450 ms AV lag might be associated with the attention network guiding the perceptual processing. Interestingly, at 0 ms AV lag, we observed a difference in the spectral power predominantly in the occipital sensors followed by the frontal and temporal sensors. We observed enhanced theta and alpha band activity in the occipital sensors; however, we observed suppression in those bands in the frontal and left temporal sensors. Here, a plausible hypothesis behind the emergence of cross-modal perception is the engagement of associative memory networks aided by the synchronous presentation of visual stimuli that integrate the well-learnt audio–visual cues [3].

Integration of Information Underlying Multisensory Perception

To gain insight into the integration of information that occurs in the functional network that disambiguates the two perceptual states, we evaluated the variation in the coherence of ongoing oscillatory activity. In an earlier study [78], we showed evidence of a global network being operational during multisensory perception. However, the local sub-networks giving rise to such large-scale interactions were unknown as the real part of coherency is susceptible to volume conduction effects. In the current article, we focus our analysis on the complex part of the coherency, i.e., the imaginary coherence, because this measure is sensitive only to synchronization of two processes that occur with a time lag and are minimally affected by volume conduction [101]. Also, it reduces the false positive estimates of interactions existent in functional connectivity measures such as absolute coherence and phase synchrony [49].

Upon non-parametric comparison of the imaginary coherence of */ta/* and*/pa/* responses, we observed an enhanced functional connectivity in the alpha band at 450 ms AV lag among the parietal-temporal-occipital sensors and at 0 ms AV lag among bilateral frontal-parietal-temporal and occipital sensors. However, at +450 ms AV lag we did not observe any significant difference in the functional connectivity at the alpha band between the */ta/* and */pa/* responses. Thalamo-cortical and cortico-cortical interactions are thought to be the generators of the human alpha rhythms, with the magnitude of the alpha coherence dependent on the frequency selectivity of the underlying network and the similarity of the inputs. Besides, alpha band synchronization has been associated with short term attentional processes [72]. Therefore, in the light of the aforementioned studies, the enhanced functional connectivity observed at 0 ms AV lag can be attributed to the attentional network. In addition, at 0 ms AV lag, the AV inputs being synchronous, the enhanced connectivity also reflects the processes involved in scrutinizing the congruency of the AV inputs. At 450 ms AV lag the difference in the connectivity alerts the short-term attentional network operating to integrate the auditory information to the upcoming visual information. However, as auditory processing is faster than the visual [58, 119], at +450 ms AV lag, the temporal lag makes time available for the visual processing of the lip movement and therefore we do not observe an enhanced connectivity emerging from the short-term attentional processes.

Inter-areal coherence of oscillatory activity in the beta frequency range (15– 30 Hz) has been has been implicated in top-down processing [139]. Furthermore, promoted by the dense anatomical connectivity, the neurons self-organize themselves into large-scale neuronal assemblies called neuro-cognitive networks (NCN), in reaction to the cognitive demands [20]. In this context, the increased interaction we observed between the temporal-parietal, bilateral temporal-parietal and temporal-occipital sensors at 450 ms (**Figure 3.5A**), 0 ms (**Figure 3.5B**) and +450 ms AV lag (**Figure 3.5C**), respectively, provides a long-range inter-areal linkage of distributed cortical areas in NCNs. These also enable the processing of the retrieval of well learnt audio–visual associations as suggested by Albright and colleagues [3].

Enhanced functional connectivity, primarily between the frontal and parietal sensors in the gamma band, was observed at all AV lags. Ther fronto-parietal network has been shown to selectively bias the processing of lower-order sensory systems [27]. Besides, gamma band coherence has been shown to be implicated in voluntary eye movements, saccades and linguistic processing [9, 108]. Stimulus selection by attention also induces local gamma band synchronization [55]. Furthermore, our gaze fixation results on the current data reported in Kumar et al. (2016) [78] show enhanced gaze fixation on the mouth during /ta/ perception. Combining these data, we hypothesize that selective attention paid to the mouth is the result of a top-down interaction that governs the perceptual processing. Most interestingly, the enhanced functional connectivity (slightly more extensive in right hemisphere) between fronto-temporal, fronto-parietal and fronto-occipital sensors signifies an increase in crosstalk between visual association areas and multisensory and integrative centers of the brain when AV information is synchronous. On the other hand, during the presentation of asynchronous AV stimuli at 450 ms, a more left-hemisphere-dominant network is operational, presumably due to the presence of pseudo-linguistic stimuli (*/pa/-/ka/-/ta/*). From the perspective of predictive coding [115, 127], one can infer that the prediction error and the internal representation of the brain can be updated within a small temporal window to process the incoming incongruent AV stimulus. Future studies can explore the boundaries of the temporal windows over which predictive coding is possible.

Overall, we present a multi-scale representation of multisensory speech processing. Although we observe markers at the individual sensor level, our results indicate that a comprehensive account of underlying neural processes emerges only when one analyzes the physiological signals at multiple scales. In the current study, due to the nature of the stimuli we were not able compare between temporal lags. However, future studies can explore such lags at the source level to build a complete picture of multisensory speech processing.

Chapter 4

Oscillatory cortical networks underlying inter-individual and inter-trial heterogeneity during cross-modal speech perception

4.1 Introduction

Speech perception during face-to-face conversation inextricably involves multisensory integration of auditory and visual cues. This is nicely demonstrated in laboratory settings by the McGurk effect [90], in which the video stimulus of a human speaker with the sound of /*ba*/ superimposed on the lip movements /*ga*/ is perceived by the listener as a completely different syllable /*da*/ (illusory/ cross-modal percept). Subsequently, several studies have identified the psychophysical parameters that play a dominant role in generating cross-modal effects [96, 129, 132] and their underlying neural correlates [12, 61, 63, 70, 110, 114, 131]. Nonetheless, the distribution of responses to McGurk stimulus is heterogeneous and some individuals rarely perceive the illusion [99]. While the neural correlates underlying illusory/ cross-modal perception has been extensively studied in a group of McGurk perceivers, the electrophsiological evidences in terms of large-scale brain networks underlying the perceptual heterogeneity is unknown.

Recent evidences show that subject-wise variability in the illusory perception is contingent on the McGurk stimulus and the response choice employed in the experimental paradigm [86]. Concurrently, neuroimaging evidences attribute the heterogeneity across individuals to the extent of activation at the superior temporal sulcus (STS) [12, 99]. Neurophysiological studies highlight the pre-stimulus activity in STS and its functional connectedness to front-parietal regions as a viable neuromarker of illusory perception within a group of individuals [70]. Converging evidences have indicated that beyond a specific region of interest, networks of brain regions facilitate perceptual processing. In this perspective, a recent review suggests neuronal oscillations as a key substrate of neuronal information processing that needs to be fully explored to answer the individual's perceptual experience. A recent study has also indicated that a large-scale network of oscillatory brain networks are involved in effectuating cross-modal perception [79]. A key question emerges, how robust is this network across a group of individuals and whether the organization of such networks is contingent on the stimulus configurations or the perceptual outcome, specifically in the case of McGurk incongruent stimulus.

It is well known that robust oscillations observed from macroscopic recordings such as EEG/MEG are an outcome of network interactions among local subpopulations of excitatory and inhibitory neurons. Empirically such interactions result in global coherence dynamics observed by earlier studies such as Kumar et al. [78]. In the current study, exploiting the perceptual variability within and across the individuals, we demonstrate how distinct coherence patterns become the hallmark of category-specific (inter-trial) and individual-specific (inter-individual) perceptual experience.

4.2 Materials and Methods

4.2.1 Participants

Eighteen healthy right handed volunteers (10 males, 8 females, mean age = 24.9, SD = 2.8) gave written informed consent under an experimental protocol approved by the Institutional Human Ethics Committee of the National Brain Research Centre, Gurgaon which is in agreement with the Declaration of Helsinki.

4.2.2 Stimuli and Trials

A digital video of a native Hindi speaking male articulating the syllables /*pa*/, /*ka*/ and /*ta*/ was recorded and edited using – the audio editing software Audacity (https://www.audacityteam.org) and the video editing software Videopad Editor (https://www.nchsoftware.com). The duration of each video clip ranged from 1.5 to 1.7 seconds to include the neutral, mouth closed position and all the mouth movements of articulation to closing. The duration of auditory syllables in the videos ranged from 0.4 to 0.5 seconds. The stimuli (Fig. 1) consisted of four kinds of videos: three congruent (auditory and visual matching) syllables (/*pa*/, /*ka*/, /*ta*/) and one McGurk/incongruent (auditory and visual mismatch) syllable (auditory /*pa*/ + visual /*ka*/) producing the McGurk percept of /*ta*/ or /*tha*/.

The experiment contained five blocks – each block consisting 120 trials (30 trials of each video presented in random). Inter-stimulus intervals were pseudorandomly varied between 1200 ms and 2800 ms to minimize expectancy effects. Using a forced choice task, on every trial, participants reported their percept by pressing a specified key on the keyboard corresponding to /*pa*/, /*ka*/, /*ta*/ or something else (others) while watching the videos.

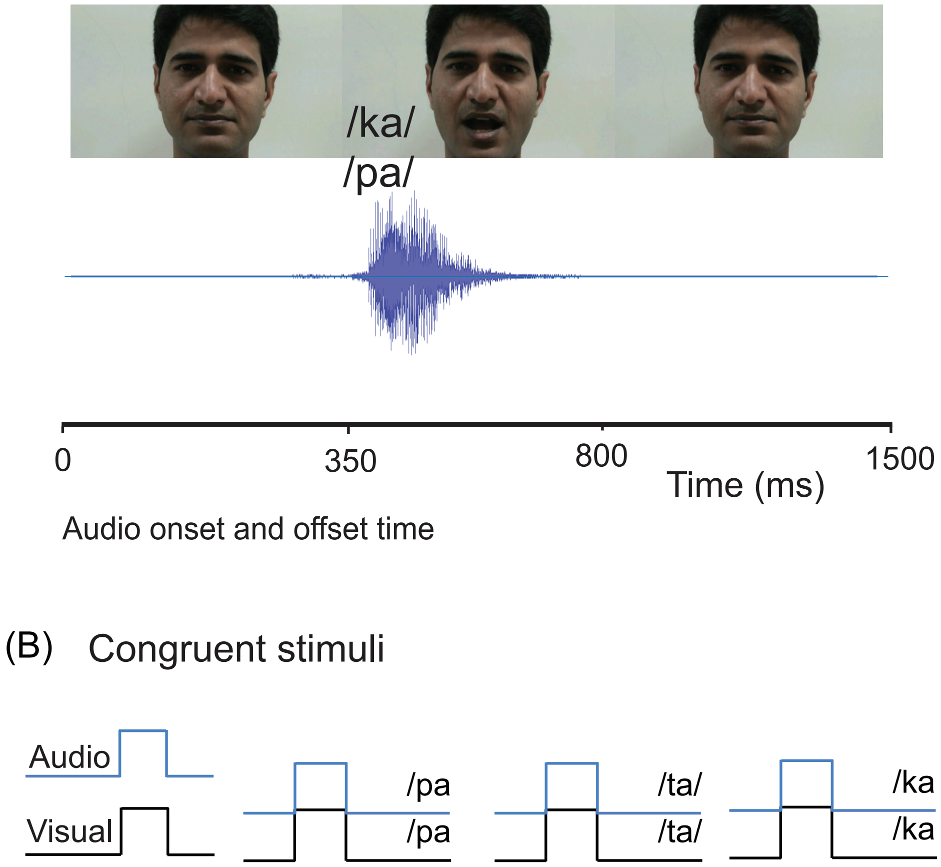

Figure 4.1: **Stimuli:** (A) Example trial with 3 video frames from McGurk stimulus (audio */pa/* + video */ka/*) used in this experiment (top row), the audio trace of the syllable */pa/* presented simultaneously to the video (middle row) and the onset and offset time of the audio. (B)The congruent AV stimuli : each block represents a video with audio */pa/*, */ta/* and */ka/* dubbed onto a video of a person articulating */pa/*, */ta/* and */ka/* respectively.

.

4.2.3 Data acquisition and Analysis

EEG

Continuous EEG scans were acquired using a Neuroscan system (Synamps2, Compumedics, Inc.) with 64 Ag/AgCl scalp electrodes sintered on an elastic cap in a 10-20 montage. Individual electrode locations were registered using the Fastrak 3D digitizing system (Polhemus Inc.). Recordings were made against the centre (near Cz) reference electrode on the Neuroscan cap and digitized at a sampling rate of 1000 Hz. Channel impedances were monitored to be at values < 10 kΩ.

Preprocessing

Preprocessing and off-line data analysis were performed using EEGLAB [31], Fieldtrip [104], and custom made MATLAB scripts (The MathWorks, Natick, MA). Continuous data were high-pass [0.1 Hz, finite impulse response (FIR)], low-pass (80 Hz, FIR), and notch filtered (45–55 Hz, 9th-order 2-pass Butterworth filter). The noisy channels were removed and the data was re-referenced. For the data analysis, epochs of 0.8s were post the onset of the sound were extracted and sorted based on the stimuli (congruent AV stimuli: */ta/*, */pa/*, */ka/* and incongruent McGurk stimulus) and responses (*/ta/*, */pa/* and 'others'). The sorted epochs were then baseline corrected by removing the temporal mean of the EEG signal 200 ms before the onset of sound on an epoch-by-epoch basis. Finally, in order to remove the response contamination from ocular and muscle-related artifacts, epochs with amplitudes above or below$\pm 75\mu$V were removed from all electrodes.

Network analysis

To investigate frequency specific functional connectivity (FC) that sub-serves cross-modal perception and characterizes inter-trial and inter-individual differences, we computed global coherence. The techniques allows to capture and quantify the strength of the covariation of neural oscillations at the global scale [24, 78]. We employed the Chronux [94] function CrossSpecMatc.m to obtain trial-wise global coherence of the epochs sorted based on stimuli and responses. We considered 3 orthogonal discrete prolate spheroidal sequences (dpss), also known as Slepian tapers, to avoid leakage of the spectral estimates into nearby frequency bands. The time-bandwidth of 5 was taken that resulted in a frequency bandwith of 0.25Hz. The output variable 'Ctot' of the function yields the global coherence value at frequency f. The function initially multiplies the epochs with the set specified number of Slepian tapers before performing fast-fourier transform (FFT). The resulting FFT values are averaged and the cross-spectrum for all sensor combinations at frequency f are computed. The cross-spectral density between between two sensors was computed from the tapered fourier transforms using the following equation

$$C_{ij} = conj(X_i(f).X_j(f)) \tag{4.1}$$

where, C_{ij} represents the cross spectrum, X_i and X_i represent the tapered fourier transforms from the sensors i and j. Subsequently, singular value decomposition (SVD) was applied on the cross-spectral density matrix for every frequency f which yields the following

$$C(f) = VCV^T \tag{4.2}$$

The diagonal matrix C comprises the values proportional to the variance explained by the orthogonal set of eigenvectors (V, V^T). Finally, global coherence $C_{\text{Global}}(f)$ at frequency bin f was computed by normalizing largest eigenvalue (first entry of the $C(f)$ at frequency f) on a trial-by-trial basis for each participant employing the following equation:

$$C_{\mathrm{Global}}(f) = \frac{C_1(f)}{\sum_{i=1}^{n} C_i(f)} \tag{4.3}$$

The global coherence computed on a trail-by-trial basis was further sorted based on the perceptual categories (*/ta/* and */pa/*) and averaged over all the participants and statistically compared.
We further analyzed if changes in global coherence values at specific frequency bands (alpha: 8-13Hz, beta: 14-30Hz, gamma: 31-45Hz) correlated with the participants' susceptibility of McGurk perception. Participant-wise mean of the global coherence in specific frequency bands were computed and were statistically analyzed using Spearman rank correlation and t-tests.

Source analysis

We employed dynamic imaging of coherent sources (DICS) [48]— a frequency-domain adaptive spatial filtering algorithm - to identify the sources of the effects found in the frequency specific global coherence dynamics. This algorithm has proven to be particularly powerful in localizing oscillatory sources. We used Fieldtrip toolbox for localizing the sources. Firstly, ft_prepare_leadfeild.m was used employing the Boundary Element Method (BEM) to generate the leadfield matrix for each participant from their respective magnetic resonance imaging (MRI) scans. The leadfield matrix corresponds to the tissue and geometrical properties of the brain represented as discrete grids or voxels. Subsequently, we employed ft_freqanalysis.m to evaluate the participant-wise cross-spectral matrices of the epochs as the adaptive spatial filters are constrained by the covariance and leadfield matrices. These spatial filters regulate the amplitude of brain electrical activity passing from a specific location while attenuating activity originating at other locations. The distribution of the output amplitude of the spatial filters provides the metric for source localization. We employed ft_sourceanalysis.m to compute the source activity in the grids specified by the spatial filters. However, in order to find the contrast activity against the baseline conditon, we performed simaliar source analysis with pre-stimulus time series and eventually computed the contrast activity using the following formula

$$C_{Contrast} = \frac{P_{stim} - P_{prestim}}{P_{prestim}} \tag{4.4}$$

where $C_{Contrast}$ represent the source power at the respective grids, $P_{prestim}$ and P_{stim} represent the power at discrete grids in the leadfeild matrix during pre-stimulus and stimulus. These source activity was interpolated onto individual anatomical magnetic resonance imaging images using ft_sourceinterpolate.m and subsequently normalized onto a standard Montreal Neurological Institute (MNI) brain using ft_volumenormalize.m in order to calculate group statistics and for illustrative purposes.

4.3 Results

4.3.1 Behavior

We employed the incongruent McGurk stimulus, visual /ka/ paired with auditory /pa/ to induce the illusory response /ta/. We used four kinds of AV stimuli: McGurk incongruent pair and congruent AV stimuli (*/ta/*, */pa/* and */ka/*). As the participants observed the four stimuli presented randomly, they reported if they heard */ta/*, */pa/*, */ka/* or 'something else. We observed a high degree of inter-individual variability in McGurk susceptibility (**Figure 4.2A**). We classified the participants based on their McGurk susceptibility into two groups - rare perceivers (<50% */ta/* percept) and frequent perceivers (>50% */ta/* percept) for characterizing any attribute in the subsequent analysis that is

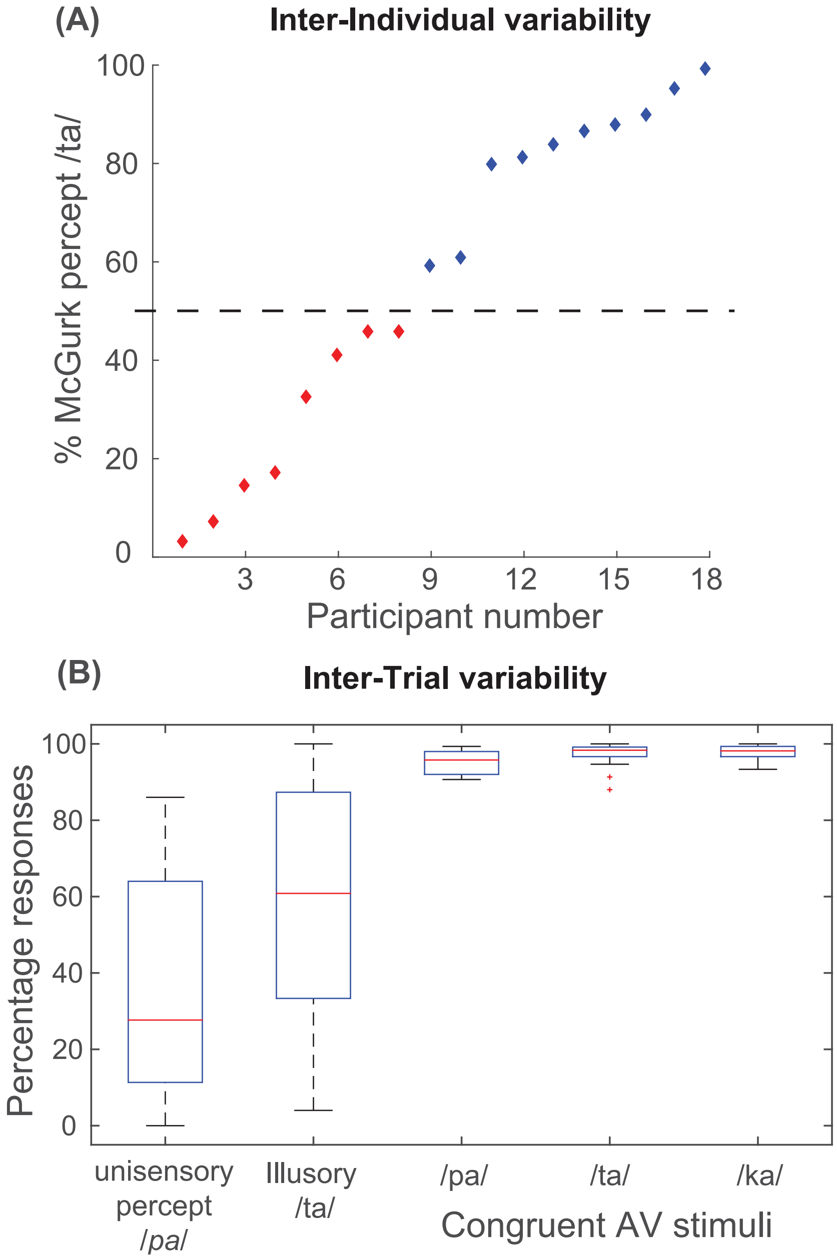

Figure 4.2: **Behavior:** (A) Inter-individual variability - Propensity of McGurk effect for each of the 18 participants expressed as the percentage of */ta/* percept during the presentation of McGurk stimulus. The participants were categorized as frequent perceivers (blue diamonds, > 50%) and rare perceivers (red diamonds, <=50%). (B) Inter-trial variability - Percentage of */ta/* (illusory) and */pa/* (unisensory)percept during the presentation of McGurk stimulus and the hit rate during the presentation of congruent AV stimuli(*/pa/*, */ta/* and */ka/*) averaged over all the participants.

associated with these two specific groups.

In the analysis of inter-trail variability(response tendency), which was computed as the relative proportion of illusory (*/ta/*) responses in all McGurk trials across all participants, we found that participants reported a */ta/* percept 58.19% trials, whereas a unisensory */pa/* percept was reported in 37.56% trails. Congruent AV stimuli (*/pa/*, */ta/* and */ka/*) were correctly identified in 96.56% trails. The difference between between the percentage of */ta/* and */pa/* percept was not significant (t=1.92, degrees of freedom = 34, p = 0.06, for details, see Figure 4.2B).

4.3.2 Large-scale functional connectivity dynamics

Inter-Individual variability

We were interested in the influence of the dynamics of the oscillatory large-scale functional connectivity on inter-individual differences in the perception of McGurk effect. For this purpose, we divided our participants into two groups based on their susceptibility of McGurk effect: rare perceivers (<50% */ta/* percept) and frequent perceivers (>50% */ta/* percept). It specifically allowed us to interogate if inter-individual heterogeneity stems from the differences in the inherent processing of multisensory stimuli in the two groups of perceivers.

We computed the time-averaged global coherence on the epochs during McGurk stimulus. We observed that rare perceivers elicited an enhanced global coherence than frequent perceivers in theta, alpha and beta bands. Frequent perceivers, however, were characterized by enhanced global coherence in the gamma band (**Figure 4.3A**). Interestingly, across all participants, a significant negative correlation was observed only between each participant's alpha band global coherence and and their propensity of experiencing the McGurk percept (r=0.14, p=0.04) (**Figure 4.3B**).

Furthermore, to overrule the possibility of that negative correlation a result of stimulus specific sensory processing and not necessarily from cross-modal as-

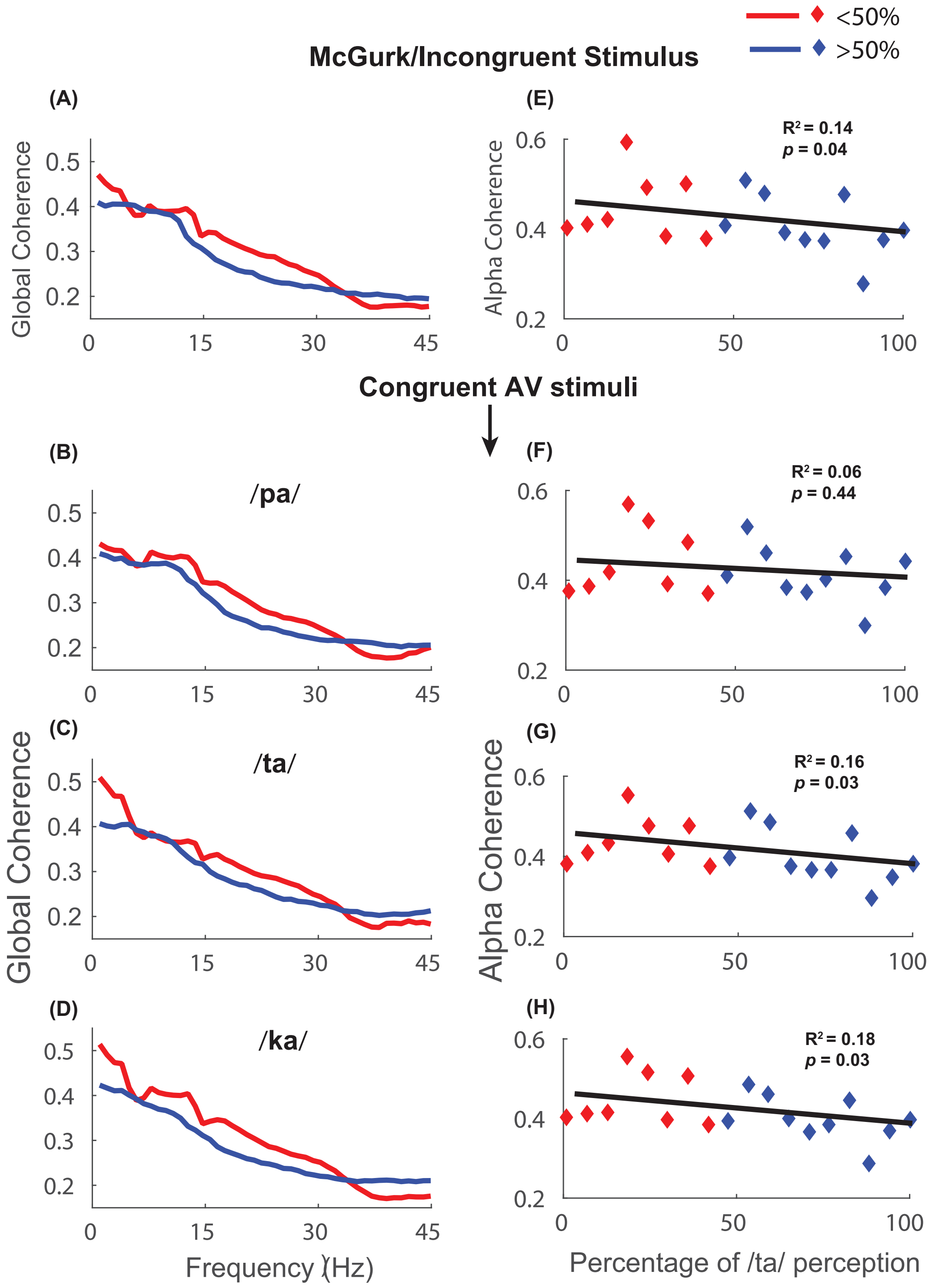

Figure 4.3: **Large scale functional connectivity dynamics :** Inter-individual variability - Time averaged global coherence during stimulus : (A) McGurk (B) Congruent */pa/* (C) Congruent */ta/* (D) Congruent */ka/*. Mean alpha band coherence plotted against participants susceptibility of McGurk effect during stimulus: (E) McGurk (F) Congruent */pa/* (G) Congruent */ta/* (H) Congruent */ka/*.

pects, we investigated the global coherence dynamics for congruent AV stimuli (/*pa*/, /*ta*/ and /*ka*/). We observed again that rare perceivers elicited an enhancement of global coherence in the theta, alpha and beta band. And, frequent perceivers were characterized by an enhanced gamma band global coherence (**Figure 4.3C, E, G**). Evidently, a significant negative correlation was observed only between each participant's alpha band global coherence and and their propensity of experiencing the McGurk percept during congruent /*ta*/ AV stimulus (r=0.16, *p*=0.03) (**Figure 4.3F**) and congruent /*ka*/ AV stimulus (r=0.18, *p*=0.03) (**Figure 4.3H**). No such significant correlation was not observed during congruent /*pa*/ AV stimulus (**Figure 4.3D**).

Inter-Trial variability

The time-averaged global coherence was computed on trials sorted based on the perceptual categories (/*ta*/ and /*pa*/) over all the participants and compared. We observed that /*ta*/ perception was characterized by a significant decrease in global coherence in alpha (*t(8)*=-4.72, p = 0.002) and beta band (*t(30)*=-2.88, p = 0.007). However, we observed a significant increase in global coherence in the gamma band between 40-45Hz (*t(14)*=11.47, $p < 0.001$) (**Figure 4.4A**). Furthermore, we computed global coherence on the trials during congruent AV stimuli (/*pa*/, /*ta*/ and /*ka*/) to understand if the differences observed between the perceptual categories (/*ta*/ and /*pa*/) elicit large-scale neural substrate of AV integration. No significant difference was observed in the global coherence dynamics between the congruent AV stimuli (**Figure 4.4B**).

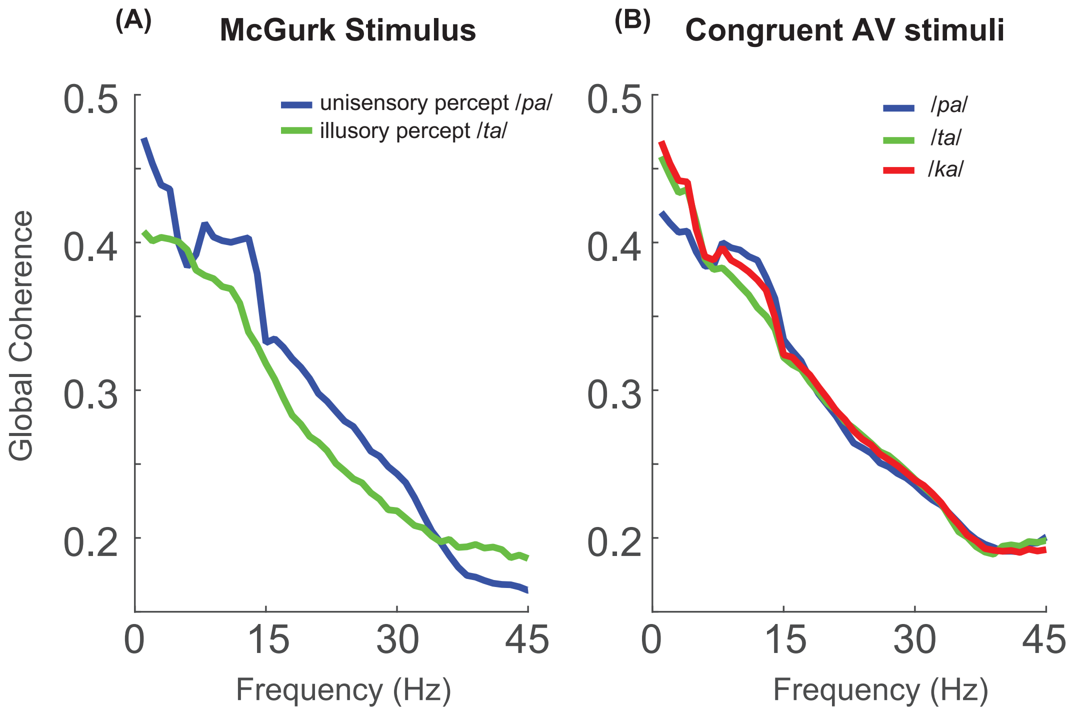

Figure 4.4: **Large scale functional connectivity dynamics :** Inter-trial variability (A) Time-averaged global coherence of trials during /*ta*/ (illusory) and /*pa*/ percept averaged across all the participants (B) Time-averaged global coherence during congruent stimuli (/*ta*/, /*pa*/ and /*ka*/) averaged across all the participants.

4.3.3 Source analysis reveals cortical areas participating in functional connectivity dynamics

We employed source analysis to identify the the possible cortical generators of the global coherence dynamics that characterized inter-individual variability in the perception of McGurk effect. Therefore, we primarily focused on identifying the cortical sources of alpha band activity. Beamformer source analysis (DICS, Gross et al. 2001) suggested the areas that contribute significantly are illustrated in **figure 4.5**. Interestingly, we observed that the source locations were consistent for McGurk stimulus and congruent AV stimuli. The cortical locations that showed significant activations listed in **table 4.1**.

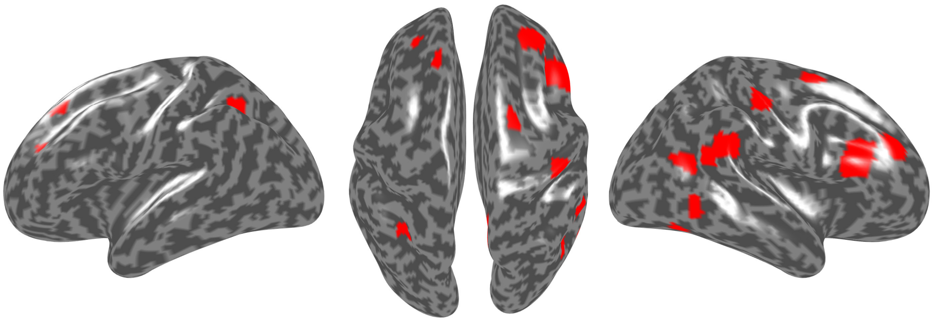

Figure 4.5: **Cortical sources :** Cortical sources underlying alpha band oscillations identified using DICS beamformer algorithm from the sensor time series. The sources eliciting power higher than the set threshold (>99.5 percentile) are highlighted.

Table 4.1: Cortical loci eliciting power higher than the set threshold (> 99.5 percentile) in the source analysis

	Left hemisphere	**Right hemisphere**
Frontal lobe	Middle frontal gyrus Superior frontal gyrus	Middle frontal gyrus Superior frontal gyrus
Temporal lobe	Fusiform gyrus	Fusiform gyrus Middle temporal gyrus posterior Superior temporal gyrus
Parietal lobe		Precuneus

4.4 Discussion

An ongoing challenge in multisensory speech perception is to accurately identify and characterize the possible neural mechanisms that govern the perceptual variability across individuals and trials. Traditionally studies have focused on identifying the neural correlates in terms of candidate brain regions or region of interest specific interactions that are responsible for the observed heterogeneity [61]. A more emerging understanding suggest the existence of networks of brain regions facilitating perceptual processing [13]. Neuronal oscillations has been identified as a key substrate of neuronal information processing that needs to be fully explored to answer the individual's perceptual experience [70,71,78]. Robust oscillations observed from macroscopic recordings such as EEG/ MEG are an outcome of network interactions among local subpopulations of excitatory and inhibitory neurons [14, 30, 143]. Empirically such interactions result in global coherence dynamics [78, 79] . The present study demonstrates how distinct coherence patterns correlate with inter-trial and inter-individual heterogeneity.

The key empirical observations in our study are: 1) The susceptibility of an individual's McGurk effect is negatively correlated to their respective alpha band global coherence, indicating desynchronization of large-scale neural assemblies in the alpha band with increasing propensity to perceive the effect 2) McGurk effect (cross-modal perception) is characterized by decreased alpha, beta and enhanced gamma band coherence.

The most significant achievement of our study was to capture the network correlates of inter-individual and inter-trial variability with the same measure of global coherence. Notably, the presence of alpha band coherence was consistent in the rare perceivers in McGurk and congruent AV stimuli (/*pa*/, /*ta*/ and /*ka*/). This emphasizes the role of synchronization of large-scale neural assemblies in alpha band in effectuating variability in the inherent processing of multisensory speech. Previous evidences accentuate the modulations in alpha band coherence to central executive processes [75, 116] that are postulated to be involved in allocating working memory storage to phonological loop that maintains verbal information, and the visuo-spatial sketchpad that maintains transient visuo-spatial information [8]. Another possible role of the alpha coherence in the inhibitory mechanisms involves processing of spatially and temporally extended visual objects [74, 76]. Both of these processes are involved in the cross-modal perception and communication via alpha coherence may be the most plausible neuromarker of inter-group variabilty at the network level.

Recent study by Fernández and colleagues demonstrates an increase in the power of theta oscillations in response to an incongruent McGurk stimulus, thereby accentuating its role in the prediction of the conflict [95]. Noticeably, we observed an enhanced global coherence in the theta band in frequent and rare perceivers which indicates even if theta band communication is present in both group of perceivers, it is a not necessarily a specific marker of inter-individual differences or trial specific perception. In general it is quite possible that different neuro-cognitive processes can be operating simultaneously involving communication at various frequencies via coherence [118]. Hence, it is important to identify which of these are meaningful to the ongoing task and the subtle differences that vary with the context in which the task evolves.

Inter-trial variability at network level was explored in our previous study [78] where we have shown global coherogram differences: desynchronization in the alpha, beta bands and enhancement in gamma band. Here, we replicate those results in time-averaged global coherence. However, in the current study our global coherence results of the congruent AV stimuli (*/pa/*, */ta/* and */ka/*), wherein we see no significant differences (**Figure 4.4B**), highlight the robustness of these oscillations as a reliable marker of AV integration. An obvious question emerges, do cross-frequency couplings among theta, alpha, beta and gamma band exist in a context specific way? Questions of such nature become a prime candidate to answer for future studies. A detailed account of cross-frequency coupling via coherence, phase-amplitude or phase-frequency coupling is currently out of scope of the present study.

Our source analysis shows activations in posterior STS, fusiform gyrus, left inferior frontal gyrus and bilateral superior frontal gyrus (**Figure 4.5**). Activations in these areas corroborate with the earlier findings. Notably, the consistency of the source activations during McGurk stimulus and congruent AV stimuli further emphasize the relevance of understanding the underlying information processing in terms of whole brain network rather than isolated brain modules.

Overall, global coherence acts as a robust global functional connectivity marker as it is affected to a lesser degree by volume conduction, simply because the functional connections that can spuriously affect a distinct pattern of coherence are unlikely to survive the normalized vector summation procedure that is undertaken. However, the neural mechanisms that give rise to the network level correlates require a combination of electro physiology and neurobiologically inspired large-scale computational model.

Chapter 5

Neurodynamic explanation of perceptual variability in the perception of cross modal speech

5.1 Introduction

Our understanding of the workings of the brain and cognition have primarily come from modular paradigm. The modular paradigm postulates that our cognitive abilities emerge as result of activations in brain areas working as independent processes [36]. However, converging evidences over the years elucidate limitations of the approach [44]. Even the sensory cortices, considered to be highly modular structures functionally has been shown to possess cross-modal interactions [45]. A more emerging view posits that information processing associated with the functioning of higher order brain functions (action, perception, learning, language, and cognition) is carried out by large scale neural networks [19]. Although the structural architecture of the brain has been extensively studied, the mechanisms of the complex dynamics elicited by the neural networks as brain oscillations and synchronizations during any cognitive task remains elusive.

In the context of multisensory speech perception, a recent study has indicated that beyond a specific region of interest, a large-scale network of oscillatory brain networks are involved in effectuating cross-modal perception [78]. Furthermore, our findings from the previous chapter indicate that synchronization of the neural networks in specific oscillatory bands (e.g alpha and gamma especially) are the most pertinent representation of inter-inter-trial and inter-individual variability in cross-modal perception. However, a systems-level insight on the workings of the neuronal assemblies requies a neuro-biologically inspired computational model.

The existing models of multisensory integration are either motivated from the context of response choices and probabilistic distribution of stimulus cues in the environment [77] or explanation of behavior from a purely phenomenological models [28, 129]. Typically these models attempt to explain the firing rate dynamics of single neurons or the local population using a combination of synaptic and stimuli inspired parameters. Thus, the neurodynamical explanation of large-scale functional connectivity patterns observed during cross-modal percept at the macroscopic scale of observation in EEG and MEG remains elusive because of the dearth of a network model that captures the large-scale brain network dynamics.

We employed a neural mass model approach to investigate the alpha and gamma coherence dynamics associated with inter-individual and inter-trial variability respectively. Since EEG data does not necessarily reflect the local synaptic activity, neural mass model which operates to phenomenologically explain mesoscopic and macroscopic features in EEG/MEG data offers an attractive tool to understand the underlying neural mechanisms. A neural mass is essentially an abstraction of summed synapto-dendritic activity of several thousand neurons in an area which can be in a cooperative dynamical state such as synchronous firing that gives rise to low-frequency oscillations. Such shared dynamical states allow us to reduce the population dynamics in terms of coupled ordinary differential equations where explicit spatial effects can be ignored. We considered broadly a network of three neural masses (each comprising a population of excitatory and inhibitory Hindmarsh-Rose(HR) neurons) as the underlying neuro-cognitive network comprising of auditory, visual and cross-modal masses(nodes)(**Figure 5.1**). Subsequently, by varying the coupling between these three nodes, we capture the neural mechanisms through which coherence dynamics evolve in the brain. Overall, we present an attractive mechanistic proposal that underlie the observed inter-individual and inter-trial variability in multisensory speech perception.

5.2 Materials and Methods

5.2.1 Large-scale dynamical model of three neural masses

Our objective was to construct a large-scale dynamical model which is biologically realistic to explain the generative mechanisms underlying observed coherence spectra and frequency specific functional connectivity (**Chapter 4, Figure 4.3 and 4.5**) emerging as a results of inter-trial and inter-individual variability. Our proposed model is a network of three neural masses, each comprising of excitatory and inhibitory neurons representing auditory (A), visual (V) and higher order multisensory (M) cortical regions (**Figure 5.1**). We follow a previously established practice and convention in computational modelling by treating each cortical region as an individual node as suggested by Stefenascu and Jirsa [123].

Furthermore, we incorporated the following biophysically realistic factors in our model construction-

- Auditory node is assumed to be most sensitive to ambient temporal fluctuations hence operating with a fast time-scale [112], visual node the slowest in terms of sensitivity [142] and somewhat intermediate time-scale for multisensory node.
- Two of the ways visual inputs are directed to the auditory cortex are: 1)

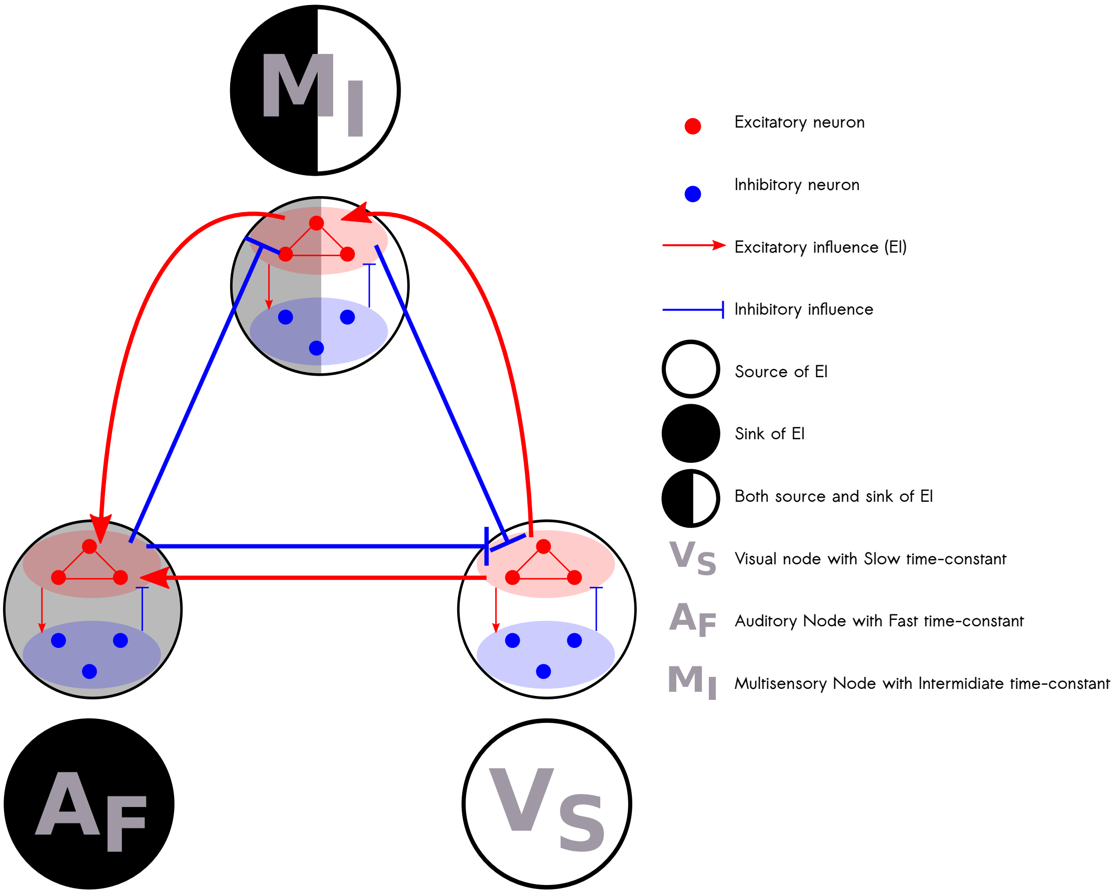
M_I
A_F
V_S
Excitatory neuron
Inhibitory neuron
Excitatory influence (EI)
Inhibitory influence
Source of EI
Sink of EI
Both source and sink of EI
V_S Visual node with Slow time-constant
A_F Auditory Node with Fast time-constant
M_I Multisensory Node with Intermidiate time-constant

Figure 5.1: **Large scale dynamical model consisting of a network three neural masses with different time-constants**: The model comprises three nodes representing auditory(fast time-constant), visual(slow time-constant) and higher order multisensory regions(intermediate time-constant). Each node consists of network of 100 Hindmarsh-Rose excitatory and 50 inhibitory neurons. Each neuron can exhibit isolated spiking, periodic spiking and bursting behavior. Excitatory influences between the nodes are balanced by their inhibitory counterpart. The source and sink represent the flow of excitatory influence.

visual cortex could directly influence the auditory cortex in a feedforward manner due to direct projections [41, 111, 138] and 2) feedback from the higher multisensory association areas [15]. Hence, in our proposed model visual node influences the auditory node in both manners: directly and indirectly via multisensory node.

- As post-synaptic potentials of pyramidal cells, which are excitatory, are shaped by their connections with other excitatory cells and inhibitory cells [73]. We use a population of excitatory and inhibitory neurons in each node where the number of excitatory neurons are considerably higher [103]. Thus, 150 excitatory neurons and 50 inhibitory neurons are selected to have a 3:1 ratio between them, an approach previously followed by Stefanescu and Jirsa [123]. Inhibitory neurons in one neural area do not directly influence inhibitory neurons within the same area since such connections are sparse in nature [123, 143].

5.2.2 Dynamics of Hindmarsh-Rose neurons

The neurons in the human brain have different types of dynamics. A neuron can fire regular spikes, fast spikes or it can fire in bursts. Hindmarsh-Rose model is a three dimensional model (see (5.1)) that capture most of these dynamics as shown in **Figure 5.2**. Therefore, we used Hindmarsh Rose neurons in our model. The different dynamics depend on the parameter I (external current). The state variables x and y vary on a faster time scale while the state variable z evolves on a slower time scale as per the following equations:

$$\begin{aligned}\dot{x} &= y - ax^3 + bx^2 - z + I \\ \dot{y} &= c - dx^2 - y \\ \dot{z} &= r(s(x - x0) - z)\end{aligned} \tag{5.1}$$

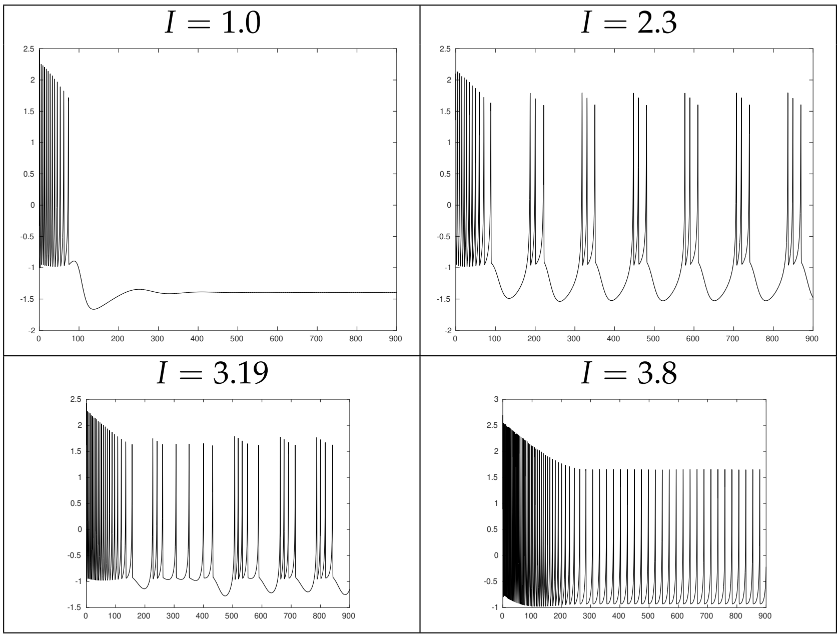

Figure 5.2: **Dynamics of Hindmarsh Rose neurons:** Depending on the parameter I, the system shows a wide range of behaviors, from regular spiking to bursting to chaotic regimes and fixed point behavior

5.2.3 Dynamic Field Model

Incorporating the aforementioned biophysically realistic factors, we define a dynamic mean field model that comprises of three equations for an excitatory Hindmarsh Rose (HR) neuron (number of excitatory neurons are 150 within an area, $N_E = 150$) and three equations for an inhibitory HR neuron (number of inhibitory neurons are 50 within an area, $N_I = 50$). The three variables account for the membrane dynamics and two kinds of gating currents, one fast and one slow respectively. Thus, the entire network can be represented as a network of coupled non-linear differential equations comprising of -

$$\begin{aligned} &\qquad\qquad \textit{Excitatory Subpopulation}\\ \tau_L \dot{x}^L_{n_E} =& y^L_{n_E} - a{x^L_{n_E}}^3 + b{x^L_{n_E}}^2 - z^L_{n_E} + K_{EE}(x^L_{n_E} - x^L_{n_E}) \\ & - K_{IE}(x^L_{n_I} - x^L_{n_E}) + I^L_{n_E} + \sum_{M=1}^{3} w_{ML} x^M_{n_E} + \epsilon \\ \tau_L \dot{y}^L_{n_E} =& c - d{x^L_{n_E}}^2 - y^L_{n_E} \\ \tau_L \dot{z}^L_{n_E} =& r[s(x^L_{n_E} - x_0) - z^L_{n_E}] \ ; \ n_E = 1, ..., N_E \ ; \ L = 1(A), 2(M) \& 3(V) \end{aligned} \tag{5.2}$$

$$\begin{aligned} &\qquad\qquad \textit{Inhibitory Subpopulation}\\ \tau_L \dot{x}^L_{n_I} =& y^L_{n_I} - a{x^L_{n_I}}^3 + b{x^L_{n_I}}^2 - z^L_{n_I} + K_{EI}(x^L_{n_E} - x^L_{n_I}) + I^L_{n_I} \\ \tau_L \dot{y}^L_{n_I} =& c - d{x^L_{n_I}}^2 - y^L_{n_I} \\ \tau_L \dot{z}^L_{n_I} =& r[s(x^L_{n_I} - x_0) - z^L_{n_I}] \ ; \ n_I = 1, ..., N_I \ ; \ L = 1(A), 2(M) \& 3(V) \end{aligned} \tag{5.3}$$

where L: A, V and AV for auditory, visual and audio-visual areas that are driven by a common noise distribution (ϵ). In our model auditory node has the fastest time-constant ($\tau_A \sim$ 0.05ms), visual node has the slowest time-constant ($\tau_V \sim$ 2.5ms) and the time constant value for the multisensory node is chosen to be between the two nodes ($\tau_M \sim$ 1ms) as it integrates information from both the modalities. The mean activity of excitatory neurons in a node ($E(x_{n_E} = \frac{1}{N_E} \sum_{n_E=1}^{N_E} x_{n_E})$) influences neuronal activities of other nodes that is governed by coupling parameters: W_{AV} (auditory-visual coupling), W_{AM} (auditory-multisensory coupling) and W_{VM} (visual-multisensory coupling). Positive value of coupling parameters reflects excitatory influence and negative value of coupling reflects inhibitory coupling. Inhibitory influennces are chosen to maintain a balance with the excitation. For example, visual node's excitatory influence of $+W_{AV}$ on the auditory node is balanced with inhibitory influence of of the same strength: $-W_{AV}$ from the auditory node.

In this configuration, visual node is referred as source node as it is the source of excitatory coupling whereas auditory node is referred to as sink node as all excitatory couplings are directed towards auditory node and multisensory node behaves as both source and sink. We place each individual neuron in a dynamical regime where both spiking and bursting behavior is possible depending on the external input current (I) that enters the neuron when other parameters are held constant at the following values: $a = 1; b = 3; c = 1; d = 5; s = 4; r = 0.006; x_0 = 1.6$, following the values according to Stefenascu and Jirsa [123]. The coupling between neurons within a node is linear and its strength is governed by the following parameters: K_{EE} for excitatory coupling, K_{EI} for excitatory-inhibitory coupling and K_{IE} for inhibitory-excitatory coupling. As excitatory and inhibitory synapses are not independent processes, their relation is captured by the ratio $n = \frac{K_{IE}}{K_{EE}}$. As alpha (8-12 Hz) and delta (1-4 Hz) rhythms are observed during resting state [46], the inhibition to excitation ratio ($n = 3.39$) is chosen when the average activity of nodes in a disconnected network has higher power at alpha and delta frequencies in the absence of stimulus ($\mu(I_{A,V,M} = 0.1)$; baseline). The external currents to both the excitatory and inhibitory subpopulations are drawn from a Gaussian distribution where μ and σ are the mean and standard deviation. As the input stimulus relays to auditory, visual and multisensory regions via thalamus, we interpret lateral geniculate nucleus (LGN) and medial geniculate nucleus (MGN) to be the source of external current (I_A, I_V, I_M). pulse of 450 ms in the nodes when the model was simulated for 1 sec. In rhesus monkey, the projections of MGN to pSTS were found to be sparse [144]. Therefore, we choose lower mean value of external current to multisensory node ($\mu(I_M = 0.85)$) in comparison to the visual node ($\mu(I_V = 2.8)$) and the auditory node ($\mu(I_M = 2.8)$) while keeping the standard deviation of the external current at 0.4 for all nodes.

Table 5.1: Description of model parameters and their corresponding values used in the model

Parameter	Value	Description of parameters
a, b, c, d	$1, 3, 1, 5$	Constants for faster variables x and y
r	0.006	Lower value of r is responsible for slower time-scale of z
s	4	Constant affecting slower variable z
x_0	-1.6	$(x_0, y_0, 0)$is a stable equilibrium point of Hindmarsh-Rose model
N_E, N_I	150, 50	Number of excitatory and inhibitory neurons in each node
μ, σ	2.8, 0.4	Mean and dispersion of input current in each node
$I_{n_E, I_{n_I}}$	Computed from μ, σ	Input current for excitatory and inhibitory neurons in each node
τ_A, τ_V, τ_M	0.05, 2.5, 1	Time-scales for auditory, visual and multisensory nodes
ϵ	0.1	Constant affecting the dispersion of noise
Within nodes connectivity		
K_{EE}	0.5	Coupling between excitatory neurons
K_{EI}	0.5	Excitatory to inhibitory coupling
η	3.39	Constant affecting slower variable z
K_{IE}	Computed from η and K_{EE}	Inhibitory to excitatory coupling
Between nodes connectivity		
W_{AV}, W_{AM}, W_{VM}		Explored in the range 0 to 1 0r -1 to 0
State variables		
x_{n_E}, x_{n_I}		Membrane potentials for excitatory and inhibitory subpopulation
y_{n_E}, y_{n_I}		Spiking variables for excitatory and inhibitory subpopulation
z_{n_E}, z_{n_I}		Bursting variables for excitatory and inhibitory subpopulation

5.3 Results

5.3.1 Inter-individual variability

Empirical findings from Chapter 4 highlight a significant negative correlation between each participant's alpha coherence and McGurk susceptibility. We hypothesized that mediation of interaction between auditory(A) and visual (V) node via the multisensory node (M) (i.e. stronger A-M and V-M coupling than A-V coupling) to be associated with participants with higher susceptibity of McGurk effect. To test this hypothesis, we started with a balanced network coupling state, $W_{AV} = W_{AM} = W_{VM} = 0.35$ where alpha and gamma band coherences co-exist (**Figure 5.3A**) and studied the change in the coherence peaks with decreasing direct A-V coupling (W_{AV}). In **Figure 5.5B** , we observe a suppression of alpha coherence peak as A-V coupling decreases; however gamma coherence peak remains more or less intact. Further, when A-V coupling becomes negligible ($W_{AV} < 0.05$) we observe disappearance of alpha coherence peak. This suggests that decrease in alpha coherence can stem from stronger A-M and V-M coupling than direct A-V coupling in participants with higher susceptibility of McGurk effect.

5.3.2 Inter-trial variability

McGurk percept (/*ta*/) was characterized by enhanced beta and gamma coherence along with decreased alpha coherence. Here, by systematically varying the coupling parameters between the three nodes, we try find the optimal model configuration that could elicit the patterns observed empirically. Our results from the previous section show that frequent perceivers are characterized by decreased alpha band coherence elicited by negligible A-V coupling. Furthermore, a decreased A-V coupling cannot directly explain the illusory perception frequent perceivers. However, the emergence of gamma band (observed empir-

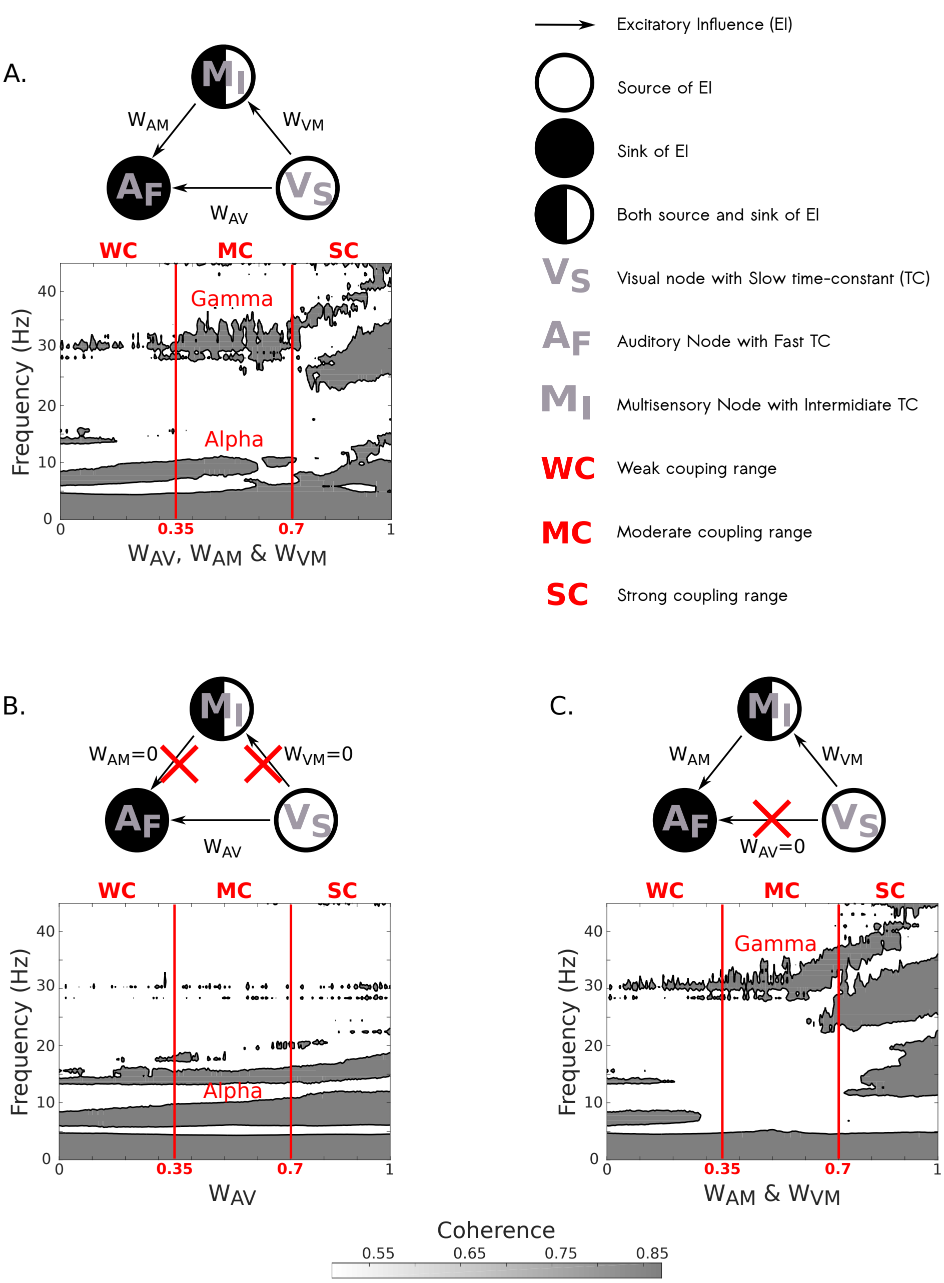

Figure 5.3: **Prediction of alpha and gamma coherences from neural mass model:** A) Alpha and gamma band coherences co-exist in moderate coupling range. B) Only direct A-V coupling generates alpha coherence independently. C) Indirect A-V coupling via multisensory node generates gamma coherence at the limit case scenario of weak direct coupling.

ically during /*ta*/ perception) coherence on incorporation of indirect A-V interactions via multisensory node (**Figure 5.3A, C**) allow us to hypothesize that the indirect communication via multisensory node is crucial for cross-modal per-

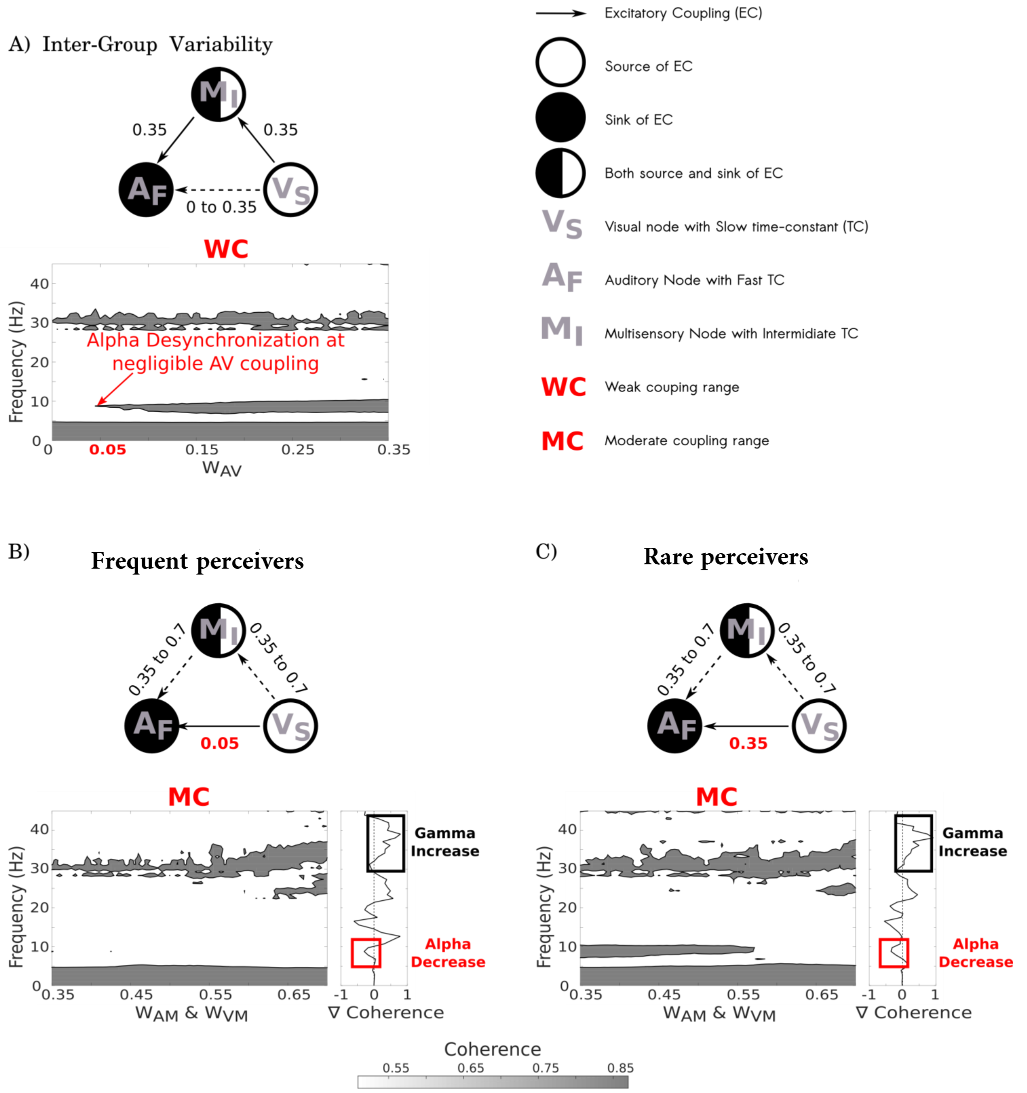

Figure 5.4: **Mechanistic understanding of Inter-individual and inter-trial variability:** A) Alpha de-synchronization characteristic of frequent perceivers resulted due to negligible A-V coupling. B) C) Enhanced gamma coherence and reduced alpha coherence observed in illusory perception are due to an increase in indirect coupling involving multisensory node irrespective of the influence of direct A-V coupling.

ception.

To test this hypothesis, for frequent perceivers we start with a network configuration that generates peak only around gamma band ($W_{AV} = 0.05, W_{AM} =$

$W_{VM} = 0.35$, **Figure 5.4A, 5.3C**). And, for rare perceivers we start with a balanced network configuration that generates co-existing alpha band and gamma band coherences ($W_{AV} = W_{AM} = W_{VM} = 0.35$). Then, we track the change in gamma coherence in the medium coupling (MC) range. In line with our hypothesis, we observe an increase in gamma coherence in network configurations for both frequent and rare perceivers. Interestingly, increasing indirect A-V interactions not only increases gamma band coherence but also display a decrease around alpha band coherence in network configurations of frequent as well as rare perceivers even though frequent perceivers exhibit overall weaker alpha band coherence (**Figure 5.4B, C**). Thus, our model highlights that an increase in indirect A-V interaction via multisensory node leads to an increase in gamma band coherence and decrease in alpha band coherence thereby facilitating illusory perception.

5.4 Discussion

Robust oscillations observed from macroscopic recordings such as EEG/ MEG are an outcome of network interactions among local subpopulations of excitatory and inhibitory neurons. A neural mass model that operates to neurodynamically explain mesoscopic and macroscopic features in EEG/ MEG data offers an attractive tool to understand the underlying neural mechanisms [29, 59, 82]. Therefore, using a computational model of interacting large-scale brain networks, we tried to explain neural mechanisms that generate the coherence dynamics associated with inter-individual and inter-trial variability. The key findings of the study are: 1) Inter-individual variability stems from the differences in the coupling between auditory- and visual node 2) Cross-modal/illusory perception is mediated by the stronger coupling between auditory-multisensory and visual-multisensory than coupling between auditory-visual.

The time scale of processing in our computational model is most disparate for the auditory and visual system, with auditory the fastest and visual the slowest [112, 142]. Without the presence of an intermediate time-scale, one "mode of communication" (alpha coherence) is sustained by the neural mass model within biologically relevant parameter regimes. Once there is another neural mass operating at intermediate time-scale participating in processing of information, the higher dimensionality of the resultant dynamical system allows creation of another mode of communication (gamma coherence). Hence, our model suggests that gamma coherence could emerge due to the communication between primary auditory and visual areas but routed indirectly via areas such as pSTS or inferior parietal or frontal areas. Our suggestion is in line with earlier observations of visual stimuli modulating auditory perception either directly resulting in alpha coherence [67] or indirectly via higher order regions (STS) resulting primarily in gamma coherence [68, 85].

Drawing from the model, inter-individual differences primarily emerge from the differences in the A-V coupling. A stronger A-V coupling implies a lesser mediation via the multisensory node (stronger A-M and V-M) leading to weaker integration which is akin to rare perceivers. Conversely, lesser A-V coupling implies a larger involvement of the multisensory node (stronger A-M and V-M) in mediating the integration of the sensory inputs characteristic of frequent perceivers. Our findings corroborates with the earlier findings suggesting the activity in multisensory node - superior temporal sulcus (STS) as the neural basis for inter-individual differences [99]. However, our findings unfolds the underlying mechanism that places STS as the prime locus responsible for heterogeneity.

Phenomenological model of communication via coherence: A key principle of multisensory integration

Alpha and/or gamma coherences have been observed in other audio-visual perception studies involving speech phrases [34], natural scenes [67] and also artificially generated A-V looming signals [85]. Strong A-V interactions that distinguish the two kinds of perceiver groups (**Figure 5.5A**) also explain the increase in alpha phase consistency observed during natural A-V scenes in rhesus monkeys [67]. Increase in gamma coherence and reduction in alpha and beta coherences were observed during the perception of incongruent (lagged) A-V speech phrases [34]. Similarly, in rhesus monkeys communication through coherence between auditory cortex and STS was high in gamma band during congruent A-V looming signals [85]. We have established that increase in the interaction between fast and slow nodes via intermediate node increases the gamma coherence and decreases coherences in alpha and beta bands (**Figure 5.5B**). Extending to studies beyond audio-visual perception, direct interactions between fast and slow nodes can explain the observed high alpha coherence during good performance while matching tactile Braille stimulus with its visual counterpart [57] and the fast-slow indirect interactions via intermediate node can explain high gamma band coherence during rubber-hand illusion when visuo-tactile stimuli were congruent [66]. Therefore, our proposed model is capable of explaining wider range of observations which are similar in neuronal network mechanisms.

Although our model of three interacting neural masses with different time-constants that generate band specific coherences can come across as a canonical model of neuronal network dynamics, certain limitations exist. Multi-parametric and unbounded nature of the parameter space results in myriads of dynamics including chaos which is non-biological [123]. Therefore, such models should not be used to directly fit the data using optimization techniques. Nonetheless, our model will be useful as a phenomenological or minimalistic model in providing mechanistic insights into many findings to the ones we have discussed as well as several others [40, 43] including pathological.

Chapter 6

Conclusions

Speech is complex by nature because of its transient acoustic properties. Retrieval of as much information as possible from the visual cues becomes crucial for efficient communication. Therefore, perception of speech, most of the time is a multisensory phenomenon involving atleast two sensory modalities: vision and audition. We discussed several examples in this thesis where incidence of conforming information from different modalities leads to robust perception. Nevertheless, the temporal alignment of the conforming modalities are more effective in bringing reliable perception. The neural basis of this robust human ability has been under intense investigation. However, extant data are far from conclusion with regard to the pertinent representational space of multisensory speech perception and its heterogeneity across individuals and contexts.

In the current thesis, we tried to address these issues by performing psychophysical experiments employing the prototypical paradigm - McGurk effect in tandem with EEG recordings. We specifically focused on identifying the dynamics of large-scale oscillatory cortical network because the neural machinery that effectuate multisensory perception would require dynamic interaction of distributed brain regions operating as large-scale network.

We initially sought on identifying the markers in the large-scale oscillatory network underlying temporal integration of AV speech. The study was primarily motivated by the fact that binding of AV speech streams is less sensitive to AV asynchrony during speech perception than perception of other stimuli [96,132,134,135]. While replicating our behavioral results using synchronous and asynchronous McGurk stimuli, we demonstrated that cross-modal percept during synchronous McGurk stimuli involves increase in gamma and decrease in alpha band global coherence indicating the existence of the states of synchronization and de-synchronization of large-scale network in the respective frequency bands. Conversely, cross-modal percept during asynchronous McGurk stimuli involves a broad-band increase in the global coherence indicating the engagement of the cortical network in more frequency bands to achieve the cross-modal percept [78].

We further investigated if the markers underlying the temporal integration of AV speech span across various spatial scales of neuronal organization that can be measured from EEG data. Specifically, we were interested to know how the local information processing manifested in the individual sensors are orchestrated into the global integrative network that facilitate AV speech perception. To this end, employing the dataset from the previous study, we identified the neural representation of subjective cross-modal perception at different organizational levels - at specific locations in sensor space and at the level of the large-scale brain network estimated from between-sensor interactions. We demonstrated that an enhanced positivity in the event-related potential peak around 300 ms following stimulus onset associated with cross-modal perception. At the spectral level, we show that cross-modal perception involved an overall decrease in power at the frontal and temporal regions at multiple frequency bands during synchronous and asynchronous AV stimuli, along with an increased power at the occipital scalp region for synchronous AV stimuli. Finally, at the level of large-scale neuronal networks, our analysis show that enhanced functional connectivity at the gamma band involving frontal regions serves as a marker of AV integration. Put together, we report that segregation of information processing at individual brain locations and integration of information over candidate brain networks underlie multisensory speech perception [79]. The large-scale network properties elicited in the dynamics of global coherence as shown in the aforementioned studies result from the interactions among subpopulations of local excitatory and inhibitory neurons. Although, our investigations did offer a global picture of the dynamics of cortical network, it does not provide insights on the neuronal mechanisms in terms of specific interactions between cortical areas that facilitate cross-modal speech perception. Moreover, the neural basis of inter-individual and inter-trial heterogeneity from network perspective remains unclear. Therefore, we performed a psychophysical experiment involving incongruent McGurk stimulus and congruent AV stimuli while recording EEG from participants. We show that subjective differences observed in the susceptibility of McGurk effect was negatively correlated with their alpha band global coherence. Notably, we observed this effect to be consistent even during congruent stimuli indicating subjective differences in the inherent processing of AV stimuli. Also, we demonstrate that inter-trial variability is characterized by decreased alpha and increased gamma band global coherence. These findings in addition to validating the employment of McGurk effect as a proxy for AV speech perception, also strongly indicate large-scale functional connectivity as the most pertinent representational space of information processing.

Finally, to gain insights on the mechanism underlying cross-modal perception, we employed a biophysically realistic neural mass model. Since EEG data does not necessarily reflect the local synaptic activity, neural mass model which is essentially an abstraction of summed synapto-dendritic activity of several thousand neuron in an area, would phenomenologically explain the macroscopic features of EEG data. Our parsimonious model of three nodes (auditory, visual and multisensory) each comprising a population of excitatory and inhibitory Hindmarsh-Rose neurons revealed that a stronger coupling of the auditory and the visual node to the multisensory node than the direct coupling between auditory and visual node facilitate cross-modal perception effectuating inter-individual heterogeneity.

Overall, we present comprehensive analyses and insights on cross-modal speech perception from the perspective of large-scale functional connectivity. In addition to that, our findings primarily indicate towards a shift from the modular paradigm (looking at brain as isolated component processes)to understanding brain as a functional network dynamics of which are amenable to various cognitive processes.

Future Directions

As in any other examples of biological information processing, understanding structure-function relationships is crucial to develop mechanistic insights. Hence, an immediate extension of this work in the future can be to explore the effect of underlying anatomical network to the large-scale functional connectivity patterns observed.

Elicitation of enhancement in global coherence in multiple frequency bands as observed in our results raises an obvious question - do cross-frequency couplings between theta, alpha, beta and gamma bands exist in a context specific way ? Future analyses can explore phase-amplitude, phase-frequency coupling that may underlie these processes.

Finally, a concurrent fMRI-EEG employing similar psychophysical experiments can delimit the spatio-temporal boundaries over which the cross-modal speech perception related dynamics unfold .

Bibliography

[1] Karen E Adolph and Kari S. Kretch. Learning to Perceive or Perceiving to Learn? *International Encyclopedia of the Social and Behavioral Sciences*, pages 127–134, 2015.

[2] Guzmán Alba, Ernesto Pereda, Soledad Mañas, Leopoldo D. Méndez, Ma Rosario Duque, Almudena González, and Julián J. González. The variability of EEG functional connectivity of young ADHD subjects in different resting states. *Clinical Neurophysiology*, 127(2):1321–1330, 2016.

[3] Thomas D Albright. On the perception of probable things: neural substrates of associative memory, imagery, and perception. *Neuron*, 74(2):227–245, apr 2012.

[4] Michelle A. Aldridge, Erika S. Braga, Gail E. Walton, and T. G. R. Bower. The intermodal representation of speech in newborns. *Developmental Science*, 2(1):42–46, mar 1999.

[5] Brian L Allman, Leslie P Keniston, and M Alex Meredith. Adult deafness induces somatosensory conversion of ferret auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14):5925–30, apr 2009.

[6] Agnès Alsius, Martin Paré, and Kevin G. Munhall. Forty Years after Hearing Lips and Seeing Voices: The McGurk Effect Revisited. *Multisensory Research*, 31(1-2):111–144, 2018.

[7] Edward T Auer and Jr. Investigating speechreading and deafness. *Journal of the American Academy of Audiology*, 21(3):163–8, mar 2010.

[8] A Baddeley. Working memory. *Science*, 255(5044):556–559, jan 1992.

[9] Susanne Balazs, Kazem Kermanshahi, Heinrich Binder, Frank Rattay, and Ivan Bodis-Wollner. Gamma-Band Modulation and Coherence in the EEG by Involuntary Eye Movements in Patients in Unresponsive Wakefulness Syndrome. *Clinical EEG and neuroscience*, sep 2015.

[10] Marcel Bastiaansen and Peter Hagoort. Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, 159:179–196, 2006.

[11] Mireille Bastien-Toniazzo, Aurélie Stroumza, and Christian Cavé. Audiovisual perception and integration in developmental dyslexia: An exploratory study using the McGurk effect., 2010.

[12] Michael S Beauchamp. fMRI-guided TMS reveals that the STS is a Cortical Locus of the McGurk Effect. *Journal of Neuroscience*, 30(7):2414–2417, 2010.

[13] Michael S. Beauchamp. The social mysteries of the superior temporal sulcus. *Trends in Cognitive Sciences*, 19(9):489–490, sep 2015.

[14] Robert Becker, Stuart Knock, Petra Ritter, and Viktor Jirsa. Relating Alpha Power and Phase to Population Firing and Hemodynamic Activity Using a Thalamo-cortical Neural Mass Model. *PLoS computational biology*, 11(9):e1004352, sep 2015.

[15] Jennifer K. Bizley and Andrew J. King. *What Can Multisensory Processing Tell Us about the Functional Organization of Auditory Cortex?* CRC Press/Taylor & Francis, 2012.

[16] S L Bressler, R Coppola, and R Nakamura. Episodic multiregional cortical coherence at multiple frequencies during visual task performance. *Nature*, 366(6451):153–156, 1993.

[17] Steven L. Bressler. Large-scale cortical networks and cognition. *Brain Research Reviews*, 20(3):288–304, 1995.

[18] Steven L. Bressler and J.A.Scott Kelso. Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, 5(1):26–36, 2001.

[19] Steven L. Bressler and Vinod Menon. Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, 14(6):277–290, 2010.

[20] Steven L Bressler and Craig G Richter. Interareal oscillatory synchronization in top-down neocortical processing. *Current opinion in neurobiology*, 31C:62–66, sep 2014.

[21] David a. Bulkin and Jennifer M. Groh. Seeing sounds: visual and auditory interactions in the brain. *Current Opinion in Neurobiology*, 16:415–419, 2006.

[22] György Buzsáki and Erik W Schomburg. What does gamma coherence tell us about interregional neural communication? *Nature Publishing Group*, 18, 2015.

[23] Gemma A. Calvert and Thomas Thesen. Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, 98(1):191–205, 2004.

[24] Aylin Cimenser, Patrick L. Purdon, Eric T. Pierce, John L. Walsh, Andres F. Salazar-Gomez, Priscilla G. Harrell, Casie Tavares-Stoeckel, Kathleen Habeeb, and Emery N. Brown. Tracking brain states under general

anesthesia by using global coherence analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 108(21):8832–8837, 2011.

[25] Adam R Clarke, Robert J Barry, Amrit Indraratna, Franca E Dupuy, Rory McCarthy, and Mark Selikowitz. EEG activity in children with Asperger's Syndrome. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 127(1):442–51, jan 2016.

[26] Patrick S. Cooper, Aaron S.W. Wong, W.Ross Fulham, Renate Thienel, Elise Mansfield, Patricia T. Michie, and Frini Karayanidis. Theta frontoparietal connectivity associated with proactive and reactive cognitive control processes. *NeuroImage*, 108:354–363, 2015.

[27] Maurizio Corbetta and Gordon L. Shulman. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3):215–229, mar 2002.

[28] Cristiano Cuppini, Ladan Shams, Elisa Magosso, and Mauro Ursino. A biologically inspired neurocomputational model for audiovisual integration and causal inference. *European Journal of Neuroscience*, 46(9):2481–2498, nov 2017.

[29] Olivier David and Karl J Friston. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage*, 20(3):1743–55, nov 2003.

[30] Gustavo Deco, Edmund T Rolls, and Ranulfo Romo. Synaptic dynamics and decision making. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16):7545–9, apr 2010.

[31] Arnaud Delorme and Scott Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, mar 2004.

[32] Ophelia Deroy, Yi-chuan Chen, and Charles Spence. Multisensory constraints on awareness. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1641):20130207, 2014.

[33] Barbara Dodd. Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, 11(4):478–484, oct 1979.

[34] Sam M. Doesburg, Lauren L. Emberson, Alan Rahi, David Cameron, and Lawrence M. Ward. Asynchrony from synchrony: long-range gamma-band neural synchrony accompanies perception of audiovisual speech asynchrony. *Experimental Brain Research*, 185(1):11–20, feb 2008.

[35] I. Dominic and W. Massaro. Speechreading: Illusion or window into pattern recognition. *Trends in Cognitive Sciences*, 3(8):310–317, 1999.

[36] P. E. Downing, Y Jiang, M Shuman, and N Kanwisher. A Cortical Area Selective for Visual Processing of the Human Body. *Science*, 293(5539):2470–2473, sep 2001.

[37] A. Engel. Role of the temporal domain for response selection and perceptual binding. *Cerebral Cortex*, 7(6):571–582, sep 1997.

[38] A K Engel, P Fries, and W Singer. Dynamic predictions: oscillations and synchrony in top-down processing. *Nature reviews. Neuroscience*, 2(10):704–16, oct 2001.

[39] Andreas K Engel and Pascal Fries. Beta-band oscillations—signalling the status quo? *Current Opinion in Neurobiology*, 20(2):156–165, 2010.

[40] Andreas K. Engel, Daniel Senkowski, and Till R. Schneider. *Multisensory Integration through Neural Coherence*. CRC Press/Taylor & Francis, 2012.

[41] Arnaud Falchier, Simon Clavagnier, Pascal Barone, and Henry Kennedy. Anatomical evidence of multimodal integration in primate striate cortex.

The Journal of neuroscience : the official journal of the Society for Neuroscience, 22(13):5749–59, jul 2002.

[42] Lineu Corrêa Fonseca, Gloria M A S Tedrus, Ana Laura R A Rezende, and Heitor F Giordano. Coherence of brain electrical activity: a quality of life indicator in Alzheimer's disease? *Arquivos de neuro-psiquiatria*, 73(5):396–401, may 2015.

[43] Pascal Fries. Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1):220–235, oct 2015.

[44] Joaquín M Fuster. The Module. *Neuron*, 26(1):51–53, apr 2000.

[45] A Ghazanfar and C Schroeder. Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6):278–285, jun 2006.

[46] Carl Gold, Darrell A. Henze, Christof Koch, and György Buzsáki. On the Origin of the Extracellular Action Potential Waveform: A Modeling Study. *Journal of Neurophysiology*, 95(5):3113–3128, may 2006.

[47] K P Green, P K Kuhl, A N Meltzoff, and E B Stevens. Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect. *Perception & psychophysics*, 50(6):524–36, dec 1991.

[48] J Gross, J Kujala, M Hä Mä Lä Inen, L Timmermann, A Schnitzler, and R Salmelin. Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):694–9, 2001.

[49] Adrian G. Guggisberg, Susanne M. Honma, Anne M. Findlay, Sarang S. Dalal, Heidi E. Kirsch, Mitchel S. Berger, and Srikantan S. Nagarajan. Mapping functional connectivity in patients with brain lesions. *Annals of Neurology*, 63(2):193–203, feb 2008.

[50] Demet Gurler, Nathan Doyle, Edgar Walker, John Magnotti, and Michael Beauchamp. A link between individual differences in multisensory speech perception and eye movements. *Attention, perception & psychophysics*, 77(4):1333–41, may 2015.

[51] Simon Hanslmayr, Joachim Gross, Wolfgang Klimesch, and Kimron L Shapiro. The role of alpha oscillations in temporal attention, 2011.

[52] Uri Hasson, Jeremy I Skipper, Howard C Nusbaum, and Steven L Small. Abstract coding of audiovisual speech: beyond sensory representation. *Neuron*, 56(6):1116–26, dec 2007.

[53] Karen S. Helfer. Auditory and Auditory-Visual Perception of Clear and Conversational Speech. *Journal of Speech Language and Hearing Research*, 40(2):432, apr 1997.

[54] Christoph S Herrmann and Robert T Knight. Mechanisms of human attention : event-related potentials and oscillations. 25, 2001.

[55] Joerg F. Hipp, Andreas K. Engel, and Markus Siegel. Oscillatory Synchronization in Large-Scale Cortical Networks Predicts Perception. *Neuron*, 69(2):387–396, 2011.

[56] Barry Horwitz. Integrating neuroscientific data across spatiotemporal scales. 2005.

[57] Friedhelm Hummel and Christian Gerloff. Larger Interregional Synchrony is Associated with Greater Behavioral Success in a Complex Sensory Integration Task in Humans. *Cerebral Cortex*, 15(5):670–678, may 2005.

[58] Aditya Jain, Ramta Bansal, Avnish Kumar, and K D Singh. A comparative study of visual and auditory reaction times on the basis of gender and

physical activity levels of medical first year students. *International journal of applied & basic medical research*, 5(2):124–7, 2015.

[59] B H Jansen and V G Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological cybernetics*, 73(4):357–66, sep 1995.

[60] Piaget Jean. The origins of intelligence in children. *Journal of Consulting Psychology*, 17(6):467–467, 1953.

[61] J a Jones and D E Callan. Brain activity during audiovisual speech perception: An fMR1 study of the McGurk effect. *Neuroreport*, 14(8):1129–1133, 2003.

[62] Jochen Kaiser ′, Werner Lutzenberger, and Jochen Kaiser. Cortical Oscillatory Activity and the Dynamics of Auditory Memory Processing. *Reviews in the Neurosciences*, 16(3):239–254, 2005.

[63] J. Kaiser. Hearing Lips: Gamma-band Activity During Audiovisual Speech Perception. *Cerebral Cortex*, 15(5):646–653, 2004.

[64] Jochen Kaiser, Ingo Hertrich, Hermann Ackermann, and Werner Lutzenberger. Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. *NeuroImage*, 30(4):1376–1382, 2006.

[65] Jochen Kaiser and Werner Lutzenberger. Human gamma-band activity: a window to cognitive processing. *Neuroreport*, 16(3):207–211, feb 2005.

[66] Noriaki Kanayama, Atsushi Sato, and Hideki Ohira. Crossmodal effect with rubber hand illusion and gamma-band activity. *Psychophysiology*, 44(3):392–402, may 2007.

[67] C. Kayser, C. I. Petkov, and N. K. Logothetis. Visual Modulation of Neurons in Auditory Cortex. *Cerebral Cortex*, 18(7):1560–1574, 2008.

[68] Christoph Kayser and Nikos K Logothetis. Directed interactions between auditory and superior temporal cortices and their role in sensory integration. *Frontiers in Integrative Neuroscience*, 3:7, 2009.

[69] Andreas Keil, Stefan Debener, Gabriele Gratton, Markus Junghöfer, Emily S Kappenman, Steven J Luck, Phan Luu, Gregory A Miller, and Cindy M Yee. Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, 51(1):1–21, jan 2014.

[70] J. Keil, N. Muller, N. Ihssen, and N. Weisz. On the Variability of the McGurk Effect: Audiovisual Integration Depends on Prestimulus Brain States. *Cerebral Cortex*, 22(1):221–231, 2012.

[71] Julian Keil and Daniel Senkowski. Neural Oscillations Orchestrate Multisensory Processing. *The Neuroscientist*, page 107385841875535, feb 2018.

[72] S.P. Kelly, P. Dockree, R.B. Reilly, and I.H. Robertson. EEG alpha power and coherence time courses in a sustained attention task. In *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings.*, pages 83–86. IEEE, 2003.

[73] Timo Kirschstein and Rüdiger Köhling. What is the Source of the EEG? *Clinical EEG and Neuroscience*, 40(3):146–149, jul 2009.

[74] Sayeed A. D. Kizuk and Kyle E. Mathewson. Power and Phase of Alpha Oscillations Reveal an Interaction between Spatial and Temporal Visual Attention. *Journal of Cognitive Neuroscience*, 29(3):480–494, mar 2017.

[75] Wolfgang Klimesch. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29:169–195, 1999.

[76] Wolfgang Klimesch. α-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12):606–617, 2012.

[77] Konrad P Kö Rding, Ulrik Beierholm, Wei Ji Ma, Steven Quartz, Joshua B Tenenbaum, and Ladan Shams. Causal Inference in Multisensory Perception. *PLoS ONE*, 2(9), 2007.

[78] G Vinodh Kumar, Tamesh Halder, Amit K Jaiswal, Abhishek Mukherjee, Dipanjan Roy, and Arpan Banerjee. Large Scale Functional Brain Networks Underlying Temporal Integration of Audio-Visual Speech Perception: An EEG Study. *Frontiers in psychology*, 7:1558, 2016.

[79] G Vinodh Kumar, Neeraj Kumar, Dipanjan Roy, and Arpan Banerjee. Segregation and Integration of Cortical Information Processing Underlying Cross-Modal Perception. (Special issue):1–20, 2017.

[80] J P Lachaux, E Rodriguez, J Martinerie, and F J Varela. Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208, jan 1999.

[81] Alvin M Liberman and Ignatius G. Mattingly. The motor theory of speech perception revised. *Cognition*, 21:1–36, 1985.

[82] F H Lopes da Silva, A Hoeks, H Smits, and L H Zetterberg. Model of brain rhythmic activity. The alpha-rhythm of the thalamus. *Kybernetik*, 15(1):27–37, may 1974.

[83] Aleksandr Romanovich Luria. *Higher Cortical Functions in Man*. Springer US, Boston, MA, 1995.

[84] J MacDonald and H McGurk. Visual influences on speech perception processes. *Perception & psychophysics*, 24(3):253–257, 1978.

[85] Joost X. Maier, Chandramouli Chandrasekaran, and Asif A. Ghazanfar. Integration of Bimodal Looming Signals through Neuronal Coherence in the Temporal Lobe. *Current Biology*, 18(13):963–968, jul 2008.

[86] Debshila Basu Mallick, John F Magnotti, and Michael S Beauchamp. Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. *Psychonomic bulletin & review*, 22(5):1299–307, oct 2015.

[87] Eric Maris, Jan-Mathijs Schoffelen, and Pascal Fries. Nonparametric statistical testing of coherence differences. *Journal of neuroscience methods*, 163(1):161–75, 2007.

[88] Lucas Murrins Marques, Olivia Morgan Lapenta, and Thiago Leiros Costa. Multisensory integration processes underlying speech perception as revealed by the McGurk illusion the McGurk illusion. 3798(June), 2016.

[89] Dominic W. Massaro. Categorical partition: A fuzzy-logical model of categorization behavior. 1989.

[90] Harry McGurk and John Macdonald. Hearing lips and seeing voices. *Nature*, 264:691–811, 1976.

[91] Anthony Randal McIntosh. Contexts and Catalysts: A Resolution of the Localization and Integration of Function in the Brain. *Neuroinformatics*, 2(2):175–182, 2004.

[92] M. Marsel Mesulam. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28(5):597–613, 1990.

[93] Wolfgang H. R. Miltner, Christoph Braun, Matthias Arnold A, Herbert Witte, and Edward Taub. Coherence of gamma-band EEG activity as a basis for associative learning. *Nature*, 397(6718):434–436, feb 1999.

[94] P Mitra and H Bokil. *Observed brain dynamics*. Oxford Univ Press, New York, 2008.

[95] Luis Morís Fernández, Mireia Torralba, and Salvador Soto-Faraco. Theta oscillations reflect conflict processing in the perception of the McGurk illusion. *European Journal of Neuroscience*, pages 1–12, jan 2018.

[96] K G Munhall, P Gribble, L Sacco, and M Ward. Temporal constraints on the McGurk effect. *Perception & psychophysics*, 58(3):351–362, 1996.

[97] R. F. Murray, P. J. Bennett, and A. B. Sekuler. Optimal methods for calculating classification images: Weighted sums. *Journal of Vision*, 2(1):6–6, feb 2002.

[98] Audrey R Nath and Michael S Beauchamp. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(5):1704–14, feb 2011.

[99] Audrey R. Nath and Michael S. Beauchamp. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage*, 59(1):781–787, jan 2012.

[100] Jordi Navarra, H Henny Yeung, Paris Descartes, and Janet F Werker. Multisensory interactions in speech perception. (January), 2012.

[101] Guido Nolte, Ou Bai, Lewis Wheaton, Zoltan Mari, Sherry Vorbach, and Mark Hallett. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10):2292–2307, oct 2004.

[102] Erika Nyhus and Tim Curran. Functional role of gamma and theta oscillations in episodic memory. *Neuroscience and biobehavioral reviews*, 34(7):1023–35, jun 2010.

[103] Hans-Georg Olbrich and Heiko Braak. Ratio of pyramidal cells versus non-pyramidal cells in sector CA1 of the human Ammon's horn. *Anatomy and Embryology*, 173(1):105–110, 1985.

[104] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011:156869, dec 2011.

[105] Lisa Payne, Sylvia Guillory, and Robert Sekuler. Attention-modulated alpha-band oscillations protect against intrusion of irrelevant information. *Journal of cognitive neuroscience*, 25(9):1463–76, sep 2013.

[106] Michael A. Pitts, Jennifer Padwal, Daniel Fennelly, Antígona Martínez, and Steven A. Hillyard. Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *NeuroImage*, 101:337–350, 2014.

[107] Ferran Pons, David J. Lewkowicz, Salvador Soto-Faraco, and Núria Sebastián-Gallés. Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences*, 106(26):10598–10602, jun 2009.

[108] F Pulvermüller, W Lutzenberger, H Preissl, and N Birbaumer. Spectral responses in the gamma-band: physiological signs of higher cognitive processes? *Neuroreport*, 6(15):2059–64, oct 1995.

[109] Henry Railo, Mika Koivisto, and Antti Revonsuo. Tracking the processes behind conscious perception: A review of event-related potential correlates of visual consciousness. *Consciousness and Cognition*, 20(3):972–983, 2011.

[110] Yadira Roa Romero, Daniel Senkowski, and Julian Keil. Early and Late Beta Band Power reflects Audiovisual Perception in the McGurk Illusion. *Journal of neurophysiology*, page jn.00783.2014, 2015.

[111] Kathleen S Rockland and Hisayuki Ojima. Multisensory convergence in calcarine visual areas in macaque monkey. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 50(1-2):19–26, oct 2003.

[112] Stuart Rosen and Peter Howell. *Signals and systems for speech and hearing / S. Rosen and P. Howell - Details - Trove.* 2011.

[113] R. Rutiku, M. Martin, T. Bachmann, and J. Aru. Does the P300 reflect conscious perception or its consequences? *Neuroscience*, 298:180–189, 2015.

[114] Dave Saint-Amour, Pierfilippo De Sanctis, Sophie Molholm, Walter Ritter, and John J. Foxe. Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, 45:587–597, 2007.

[115] Paul Sauseng, Markus Conci, Benedict Wild, and Thomas Geyer. Predictive coding in visual search as revealed by cross-frequency EEG phase synchronization. *Frontiers in Psychology*, 6:1655, oct 2015.

[116] Paul Sauseng, Wolfgang Klimesch, Manuel Schabus, and Michael Doppelmayr. Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International Journal of Psychophysiology*, 57(2):97–103, 2005.

[117] Kaoru Sekiyama, Iwao Kanno, Shuichi Miura, and Yoichi Sugita. Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47(3):277–287, 2003.

[118] Daniel Senkowski, Till R Schneider, John J Foxe, and Andreas K Engel. Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in neurosciences*, 31(8):401–9, aug 2008.

[119] Jose Shelton and Gideon Praveen Kumar. Comparison between Auditory and Visual Simple Reaction Times. *Neuroscience & Medicine*, 01(September):30–32, 2010.

[120] Rodrigo Sigala, Sebastian Haufe, Dipanjan Roy, Hubert R Dinse, and Petra Ritter. The role of alpha-rhythm states in perceptual learning: insights from experiments and computational models. *Frontiers in Computational Neuroscience*, 8(April):36, 2014.

[121] Richard Simson, Herbert G Vaughan, and Walter Ritter. The scalp topography of potentials in auditory and visual Go/NoGo tasks. *Electroencephalography and Clinical Neurophysiology*, 43(6):864–875, dec 1977.

[122] J. I. Skipper, V. van Wassenhove, H. C. Nusbaum, and S. L. Small. Hearing Lips and Seeing Voices: How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception. *Cerebral Cortex*, 17(10):2387–2399, 2007.

[123] Roxana A Stefanescu and Viktor K Jirsa. A low dimensional description of globally coupled heterogeneous neural networks of excitatory and inhibitory neurons. *PLoS computational biology*, 4(11):e1000219, nov 2008.

[124] B E Stein, M A Meredith, W S Huneycutt, and L McDade. Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. *Journal of cognitive neuroscience*, 1(1):12–24, jan 1989.

[125] Ryan A Stevenson, Nicholas A Altieri, Sunah Kim, David B Pisoni, and Thomas W James. Neural processing of asynchronous audiovisual speech perception. *NeuroImage*, 49(4):3308–18, feb 2010.

[126] W. H. Sumby. Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, 26(2):212, jun 1954.

[127] Durk Talsma. Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, 09:19, mar 2015.

[128] Natalie Taylor, Claire Isaac, and Elizabeth Milne. A Comparison of the Development of Audiovisual Integration in Children with Autism Spectrum Disorders and Typically Developing Children. *Journal of Autism and Developmental Disorders*, 40(11):1403–1411, nov 2010.

[129] Bhumika Thakur, Abhishek Mukherjee, Abhijit Sen, and Arpan Banerjee. A dynamical framework to relate perceptual variability with multisensory information processing. *Scientific Reports*, 6(1):31280, nov 2016.

[130] Nienke M van Atteveldt, Elia Formisano, Leo Blomert, and Rainer Goebel. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral cortex (New York, N.Y. : 1991)*, 17(4):962–74, apr 2007.

[131] Virginie van Wassenhove, Ken W Grant, and David Poeppel. Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4):1181–6, 2005.

[132] Virginie van Wassenhove, Ken W. Grant, and David Poeppel. Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3):598–607, 2007.

[133] Rufin VanRullen. Perceptual Cycles. *Trends in Cognitive Sciences*, 20(10):723–735, 2016.

[134] A. Vatakis, A. A. Ghazanfar, and C. Spence. Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of Vision*, 8(9):14–14, jul 2008.

[135] Argiro Vatakis and Charles Spence. Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & psychophysics*, 69(5):744–56, jul 2007.

[136] Ch. von der Malsburg and W. Schneider. A neural cocktail-party processor. *Biological Cybernetics*, 54(1):29–40, may 1986.

[137] M. T. Wallace, M. A. Meredith, and B. E. Stein. Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J Neurophysiol*, 69(6):1797–1809, jun 1993.

[138] Mark T Wallace, Ramnarayan Ramachandran, and Barry E Stein. A revised view of sensory cortical parcellation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7):2167–72, feb 2004.

[139] X J Wang. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological reviews*, pages 1195–1268, 2010.

[140] Whitney M Weikum, Athena Vouloumanos, Jordi Navarra, Salvador Soto-Faraco, Núria Sebastián-Gallés, and Janet F Werker. Visual language discrimination in infancy. *Science (New York, N.Y.)*, 316(5828):1159, may 2007.

[141] Sabine Weiss and Horst M Mueller. “ Too many betas do not spoil the broth ”: the role of beta brain oscillations in language processing. *Frontiers in Psychology*, 3(June):1–15, 2012.

[142] Justin H.G. Williams, Dominic W. Massaro, Natalie J. Peel, Alexis Bosseler, and Thomas Suddendorf. Visual–auditory integration during-

ing speech imitation in autism. *Research in Developmental Disabilities*, 25(6):559–575, nov 2004.

[143] Hugh R. Wilson and Jack D. Cowan. Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. *Biophysical Journal*, 12(1):1–24, jan 1972.

[144] Edward H. Yeterian and Deepak N. Pandya. Thalamic connections of the cortex of the superior temporal sulcus in the rhesus monkey. *The Journal of Comparative Neurology*, 282(1):80–97, apr 1989.

Publications

1. **Kumar, V.G.**, Kumar, N., Roy, D. & Banerjee, A. ‘Segregation and integration of cortical information processing underlying cross-modal perception." 2017. *Multisensory research*,
   doi: 10.1163/22134808-00002574.

2. **Kumar, V.G.**, Halder, T., Jaiswal, A. K., Mukherjee, A., Roy, D. & Banerjee, A. ‘Large scale functional brain networks underlying temporal integration of audio-visual speech perception: An EEG study." 2016. *Frontiers in Psychology.*,
   doi:7:1558.journal.frontiersin.org/article/10.3389..

3. Ghosh, S., **Kumar, V. G.** , Basu, A. & Banerjee, A ‘Graph theoretic network analysis reveals protein pathways underlying cell death following neurotropic viral infection." 2015. *Scientific Reports*,
   doi: 5: 14438 http://www.nature.com/articles/srep14438

Papers in Preparation

1. **Kumar, V.G.**, Dutta, S., Talwar, S., Roy, D., Banerjee, A. ‘Neurodynamic explanation of inter-individual and inter-trial variability in cross-modal perception."
   doi: https://doi.org/10.1101/286609