# Characterizing Variability of Audiovisual Speech Perception Based on Prestimulus Oscillatory Features of Electrophysiological Brain Signals

**Vinsea A V Singh**

*A Dissertation Submitted to the*
*National Brain Research Centre*
*for the award of*
*Masters in Neuroscience*

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

राष्ट्रीय मस्तिष्क अनुसन्धान केंद्र
(समविश्वविद्यालय)
(जैव एवं प्रोद्योगिक विभाग का स्वायत संस्था,
विज्ञानं एवं तकनीकी मंत्रालय, भारत सरकार)
एन० एच०-8, नैनवाल मोड़, मानेसर- 122051, (हरियाणा)

National Brain Research Centre
(Deemed University)
(An Autonomous Institute of Deptt. of Biotechnology,
Ministry of Science and Technology, Govt. of India)
NH-8, Nainwal Mode, Manesar-122051(Haryana)

# CERTIFICATE

This is to certify that the dissertation entitled **"Characterizing variability of audiovisual speech perception based on prestimulus oscillatory features of electrophysiological brain signals"** is the result of work carried out by **Vinsea A V Singh** in National Brain Research Centre, Manesar, Haryana, India.

The work presented herein is original and has not been submitted previously for the award of any degree or diploma to **National Brain Research Centre (Deemed to be University)** or to any other University. This work is completely based on the guidelines given by **National Brain Research Centre** and is a record of the candidate's own efforts.

Dr. Dipanjan Roy
Associate Professor/ Scientist IV
NBRC, Manesar
(Supervisor)

Prof. Neeraj Jain 28/5/20
NBRC, Manesar
(Director, NBRC)

प्रो० नीरज जैन / Prof. Neeraj Jain
निदेशक / Director
राष्ट्रीय मस्तिष्क अनुसंधान केन्द्र
National Brain Research Centre
मानेसर-122 051 / Manesar-122 051
हरियाणा / Haryana
फोन नं: 0124-2338929 / Tel. No. 0124-2338929

Place: NBRC, Manesar

Date: 28/05/2020

# DECLARATION

I, **Vinsea A V Singh**, hereby declare that the work presented in the dissertation entitled **"Characterizing variability of audiovisual speech perception based on prestimulus oscillatory features of electrophysiological brain signals"** is carried out by me under the guidance of Dr. Dipanjan Roy, Associate Professor/ Scientist IV, **National Brain Research Centre (Deemed to be University),** Manesar, Haryana, India.

I also declare that no part of this dissertation has been previously submitted for the award of any degree or diploma to **National Brain Research Centre** or any other University.

Place : NBRC, Manesar

Date : 31/05/2020

(Vinsea A V Singh)
M.Sc. Student, NBRC

*Dedicated to my family: Mummy, Papa, and Brother*

# Acknowledgements

At the outset, I express my gratitude to my supervisor, Dr. Dipanjan Roy, for his constant support and guidance. The work contained herein would not have been possible without his persistent help. The discussions that we had related to this project or otherwise has not only augmented my knowledge in neuroscience but has also made me a better researcher. I am also highly thankful to Dr. Arpan Banerjee for his valuable inputs and feedbacks.

I am really indebted to my classmates, especially Rishika, Sharmistha, Mantosh, and Darshit, for they have been more like a family here in NBRC. A special thanks to my roommate Keerthana, who was with me during my ups and downs.

I gratefully acknowledge the feedback and support received from fellow lab mates at the Cognitive Brain Dynamics Lab, especially from Vinodh. His great advice for my study proved significant towards the success of this study.

Lastly, I would like to thank my family for constantly supporting and believing in me, regardless of what I do in my life.

Vinsea A V Singh

# Contents

# List of Figures

# List of Tables

# Abstract

Perception arises through interactions among specific neural ensembles. In this work, we have analyzed the pre-stimulus brain signals and identified specific oscillatory features to explain the inter-individual and inter-trial variability in perception among the participants. The experimental data was collected while the subjects were engaged in a perceptual task involving a sufficiently well-known experimental paradigm called McGurk illusion (incongruent visual and auditory syllables perceived as a completely different syllable). First, using behavioral perceptual responses we have characterized participants as the 'rare' and 'frequent' group of perceivers. Subsequently, we have studied the neural differences in the pre-stimulus electrophysiological (EEG) brain signals between the 'rare' and 'frequent' group of perceivers during the McGurk illusory perception. The oscillatory features we have characterized consist of the power spectrum, global coherence, and FOOOF (Fitting Oscillations and One Over Frequency) to study the cortical dynamics in the pre-stimulus EEG signals of the participants, that were subjected to both the incongruent (McGurk- audio- */pa/*, video-*/ka/*) and congruent (syllables */pa/, /ta/,* and */ka/*) AV-stimulus. Our results show that there is indeed a difference in the cortical dynamics of the two groups, suggesting that multisensory processing relies on different brain regions and that these regions have distinct processing time scales and developmental trajectory leading to a difference in the variability of perception at the individual level. Furthermore, to a large extent this perceptual variability could be predicted based on the pre-stimulus brain oscillatory features.

# Chapter 1

# Introduction

Language is the most remarkable system of communication that humans possess. It is expressed in written, spoken, and gestured form. It helps us to convey our thoughts and feelings, our desires, and our motives with one another. With language we can easily convey information about the past, present, and possible future. No other species (at least on our planet) has such an elaborate and complex means of communication. Even the closest of our primate relatives have a communication system far less inferior to that of humans. Due to this, it is tough to understand the evolution of human language. Since the human language arises from the ability of the brain, so it is called a natural language. Late British neurologist Oliver Sacks in his book Seeing Voices, states the importance of human language as-

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

The emergence of language is strongly linked to the evolution of human speech. Spoken language (or speech) is an audible form of communication built on the sounds that humans naturally produce. The study of language and sound is, hence, of great importance in various scientific fields including cognitive psychology, linguistics, evolutionary biology, computer sciences, and neuroscience to name a few. However, the development of such complex connections between the brain and the vocal apparatus, to understand and convey the linguistic needs, is impossible without a perceptual specialization machinery. Hence, it is important to understand the perceptual machinery which in the neuroscience field is called as the "neurobiology of speech perception", which will be discussed in the later sections of this chapter.

## 1.1 Speech Perception: A Multimodal System

Speech perception is a multisensory phenomenon that can be best understood in the case of face-to-face communication. While communicating face-wise, speech is perceived primarily by visual and auditory sensory input and a combination of the two. Decades worth of research have conclusively revealed that our perception and understanding of speech are influenced by both the speaker's face (a visual cue) with accompanying gestures, as well as the phoneme[1] of the speech. However, there are scenarios where a visual cue is absent[2] in which case there is interference with the way speech is perceived by the engaged listeners. Nonetheless, visual cue is important and accessibility of this input certainly enhances and facilitates the perception of speech [1].

Visual cues in which the lip movement is visible improve the intelligibility of speech processing and perception, especially when the auditory signal is hindered by the presence of noise or being masked intentionally by the experimenter as part of the experimental design [2][3]. Furthermore, it has been observed by works from Patricia Kuhl's group that connections between the auditory and visual processing of speech articulation are actually present very

[1]The fundamental unit of sound that distinguishes one word from the other in a particular language.
[2]For example, talking over a telephone

early in life. Infants who were barely 18 weeks old could successfully relate the auditory consequences of speech to the visible lip movements [4]. These observations clearly suggest that the ability to integrate the visual input to that of the auditory vocalization is part of the normal maturation and that speech perception is a multimodal system. With this idea in place, scientists now turned their attention to investigate specifically the question of how do the visual sensory input gets integrated with the auditory sensory input during speech perception and processing, in the brain.

## 1.2 The Audio-Visual (AV) Integration of Speech: A Literature Review

### 1.2.1 AV integration in infants

The audio-visual speech perception is a multi-sensory phenomenon that emerges early in infants experiencing language development. As the infants mature, they learn to identify, differentiate, and integrate perception that is arising from various senses. Numerous research work in this area further suggests that these various sub-components, commonly referred to as "multisensory processing", may rely on at least partially different brain regions and have distinct developmental trajectories [5][6]. However, the literature realm is divided into two halves on understanding multimodal development and processing in infants.

One group of researches suggests that some ability to associate and combine audio-visual sensory modalities appears to be present in the first year of life. For example, 10 to 18-week old infants can detect temporal synchrony between lip movements and speech sounds [7][4]. Also, infants that are 3 months old, can learn to randomly associate faces with voices [8][9], and by 4–7 months of age, they are able to correctly identify faces with matching voices based on age[10]. In addition to this, studies have also reported the presence of some degree of audio-visual integration in 4.5-5 months old infants, when they were exposed to McGurk illusion[3] [11]. Bristow and colleagues reported that 10 to 12-week old infants have a cross-modal representation of phonemes and that they are able to integrate auditory and visual information during early stages of perception, which had previously been reported for adults. They observed this by recording event-related potentials (ERPs) to the auditory pronunciation of a vowel that either matched or mismatched the earlier visual articulation [12]. Some research groups even claim that certain visual speech skills of young infants may be better than those of adults. For example, Weikum and colleagues have successfully demonstrated that 4- and 6-months-old (but not 8-months-old) infants are able to differentiate between two languages based on visual speech cues alone [13].

The other group suggests a prolonged developmental course of certain aspects of audiovisual processing. For example, Lewkowicz and colleagues reported that, unlike adults, infants require significantly longer temporal separation between the onsets of auditory and visual stimuli, both in speech and non-speech contexts, in order for them to detect temporal asynchrony [14]. Furthermore, even though there exists some ability to perceive McGurk illusion in early stages of life, multiple studies have claimed that there is a reduced susceptibility to the McGurk illusion in children than in adults, suggesting that the ability to fully integrate auditory and visual speech cues does not mature until late childhood and depends, partially, on children's experience with audio-visual speech [15][16].

Although there are two different schools of thought for AV-integration in infants, both theory (and evidences) suggest that there exists AV-integration prior to familiarization with specific nuances and experience in a specific context and settings with AV speech. Hence, although expository based on the current evidence it seems there must be something deeper and intrinsic about AV-integration and speech processing. AV integration certainly cannot be explained entirely based on the familiarity of specific type of speech sounds. It is also

[3]See section 1.3

clear, that the neural systems or anatomy supporting AV-integration emerge very early in life and could critically constrain perceptual variability (although not investigated in this thesis). However, experience much like other sensory sharpening also plays a fundamental role in enhancing and tuning these capabilities.

### 1.2.2 AV integration in adults

Over the past few decades, scientists have focused on the characterization of AV speech perception mechanisms. They have focused on the mapping between acoustic features and components of language like phonemes. However, this mapping has turned out to be convoluted, and a search for the explanation about the underlying processes (representation of the speech inputs and the outputs of such integrative process) has led to the rise of three main theoretical models on speech perception that frame much of the empirical work. These three models are discussed in turn.

***Motor Theory of Speech Perception:***

The motor theory of speech perception states that people perceive spoken words by identifying specific speech gestures, with which the words are pronounced, rather than by identifying the sound patterns of the incoming speech. Liberman and his colleagues at Haskins Laboratories, Yale University were one of the early pioneers to develop the motor theory (MT) of speech perception. The motor theory has undergone significant modifications since its initial formulation. In the revised theory, Liberman and Mattingly (1985) [17] proposed three points: (i) Speech processing is unique and special; (ii) Speech perception is directly linked to vocal gestures of the speaker; and (iii) The motor system is necessary for perceiving speech. Based on these three points, they claimed that both speech perception and production are part of one motor module that shares neural processes. However, the theory faces certain criticisms as there are studies on patients with damaged motor brain regions that have shown intact speech perception abilities.

***Direct-Perception Theory of Speech Perception:***

The theory of direct-perception of speech states that speech perception is strongly linked to speech producing mechanisms. Based on the Gibsonian theory of perception, the direct-perception theory holds that perceptual difference between speech and non-speech sounds is actually a consequence of meaningfulness. This was supported by the finding that multi-sensory perception occurs for meaningful non-speech sounds such as a slamming door [18]. Since the direct-perception theory does not claim that speech is special, evidence for human-like perception of speech that originates from non-human species/objects is not problematic for the theory. In addition to this, the theory also argues that speech-perception phenomena cannot be accounted for either by auditory or general learning explanation. This hypothesis is proved by understanding the multi-modality of speech perception. The McGurk effect shows audio-visual speech integration, but similar integration also arises when perceivers hear and feel - with their fingers on the lips of the speaker. Just as with audio-visual speech, the tactile speech also codetermines the speech perception. The lack of opportunity to learn the association between touch and hand gestures- we rarely touch other's lips during conversations- goes against a general learning and thus, in favor of the direct-perception theory.

***Acoustic Invariance Theory:***

The acoustic invariance theory assumes that for each distinct phoneme there is a complementary set of acoustic features. This means that a part of acoustic property is always present whenever a speech sound is produced, regardless of coarticulation and other contextual effects. This part can be referred to as a template with which the listener compares the

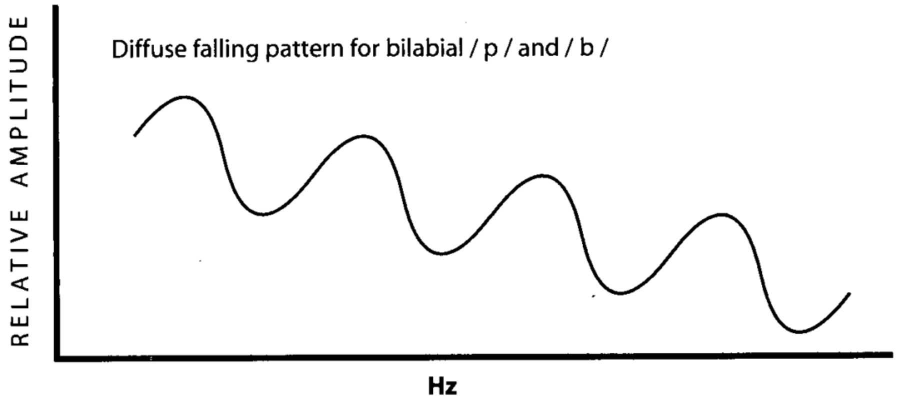


Figure 1.1: *Spectral patterns for bilabial stops* [20]

incoming sound. This model was initially proposed by Stevens and Blumstein [19]. In this model, the incoming acoustic signal is first processed to determine the special bursts in the signal. For example, the burst for bilabial[4] stops are diffuse and falling. This means that the acoustic energy in the spectrum is mainly concentrated in certain frequency locations across the frequency range, with the amplitude of consecutive peaks decreasing towards the higher frequencies. See Figure 1.1. Stevens states that coarticulation causes controlled, systematic, and predictable variation in the signal which the listener is able to deal with. Several of these acoustic features unambiguously identify phonetic segments (like phonemes, syllables, and words). These segments are part of the lexicon that is stored in the listener's memory. Thus, listeners reconstruct the articulatory events which were necessary to produce the perceived speech signal. This can be therefore described as an analysis-by-synthesis theory of speech perception.

[4](of a speech sound) Formed by the closure of the lips, e.g., p, b, m, w.

## 1.3 Methods And Techniques To Understand AV Speech Perception

A substantial amount of studies on speech perception have employed different imaging techniques[5] such as the functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG), Positron Emission Tomography (PET), Transcranial Magnetic Stimulation (TMS), transcranial Direct Current Stimulation (tDCS), and Electroencephalography (EEG), in order to explore the neural workings during AV- integration of speech. Even though each of these techniques has its own uniqueness, the nature of the information that each tool provides is worth noting. fMRI and PET studies offer information about the hemodynamic response state of the brain regions involved while performing a certain task, thereby providing the spatial information. Whereas, EEG and MEG provide electrophysiological data, with better temporal resolution. Finally, neuromodulation techniques like the TMS and tDCS allow us to understand the causal links between brain activity and corresponding behavioral responses.

## 1.4 The McGurk Effect: Paradigm to Understand AV-Integration

About forty years ago, Harry McGurk and John MacDonald published their paper titled *Hearing Lips and Seeing Voices*, Nature (1976) [16], a manuscript where they first described the remarkable audiovisual speech illusion known as the McGurk illusion or the McGurk effect. It is a phenomenon that demonstrates a fundamental interaction between hearing and vision in speech perception. Since its discovery, the McGurk effect has been predominantly employed as the quintessential paradigm to understand AV- integration. The illusion occurs

[5] Or, non-invasive techniques

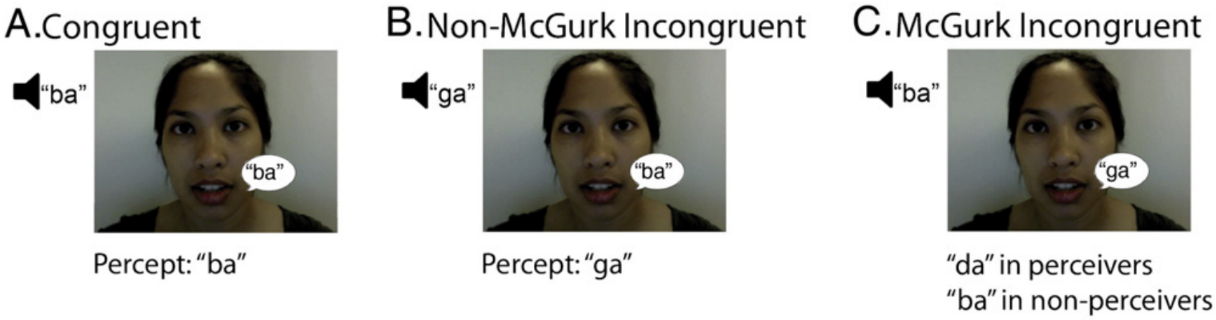


Figure 1.2: ***The McGurk stimuli:*** *A. Congruent audiovisual syllable, consisting of matching auditory "ba" (depicted by speaker icon) and visual "ba" (single frame of video shown). Percept (shown below picture) is "ba." B. Non-McGurk incongruent syllable, consisting of auditory "ga" and visual "ba." This stimulus does not result in an illusory percept; the resulting percept is most often "ba." C. McGurk incongruent syllable, consisting of auditory "ba" and visual "ga." For McGurk perceivers, this results in the percept of an illusory "da." For non-perceivers, the percept is "ba."* [21]

when an acoustic signal of one phoneme is dubbed onto a specific visual signal of a different phoneme. The observers of such incongruent audio-visual set (For example., audio /ba/ and visual /ga/) often fail to recognize the inter-modal differences and perceive (or hear) a completely different phoneme (in this case /da/) which is different from the audio-visual input, see Figure 1.2. Many researchers have called the McGurk effect solely as the fusion effect because in this illusion integration results in the perception of a third consonant, which is merging information from the audition and vision syllables [23]. The McGurk effect is often described as robust because it occurs even when the perceiver is aware of the manipulation [30], and also if the audio and visual inputs are not temporally aligned [31].

The McGurk illusion efficaciously demonstrates that speech perception is not only an auditory process but also can involve the processing of acoustic elements across modalities even when the auditory information is intact [22]. For these reasons, the McGurk effect has been extensively used as a paradigm in order to understand multisensory integration across decades (6982 Google Scholar citations of the original study, from 1976 to May 2020). It has also been used as a standard measure for audio-visual speech perception studies over other types of measures[6] because it allows researchers to present short and simple linguistic stimuli over a large number of trials, in open[7] and closed[8] set response tasks. In short, the McGurk illusion is a very useful research tool to understand speech perception as the strength of its effect can be directly linked to the strength of the AV- integration. Also, electrophysiological evidences, addressing the temporal dynamics, have explored the signatures in the event-related potentials (ERPs) and brain oscillations in different frequency ranges. Consequently, these findings have enabled scientists to conceptualize and design effective diagnostic markers for speech-related disorders.

### 1.4.1 McGurk effect in clinical populations

Impaired speech perception has been reported in people suffering from Schizophrenia, ASD (Autism Spectrum Disorder), and other neurodevelopmental and neuropsychiatric disorders. Under these clinical settings, the McGurk illusion paradigm offers scientists and clinicians to better understand the departure of neuronal dynamics of multisensory integration in these patients and alteration compared with normal and typical developing individuals.

For example, people with Autism Spectrum Disorder (ASD) are one of the most widely studied clinical cohorts in the field of sensory integration deprivation. It is a pervasive neurodevelopmental disorder wherein people with ASD have deficits in language development, social communication, and show motor and speech production impairments with neurological causes. It has been observed that individuals with ASD perceive the McGurk effect less frequently than typically developing controls [24]. The lesser incidence of McGurk perception has been attributed to the atypical gaze pattern for face stimuli and also prosody of language which is a known characteristic of a person with ASD. However, contrasting evidences show no significant difference in the eye movements in ASD patients and the controls

[6]For example., speech in noise tasks (or SPIN)
[7]Subjects respond with syllables that they perceive.
[8]Subjects are asked to choose from specific response options.

indicating a lack of integration. As people in the autism spectrum, children who have been identified with language learning disorders like dyslexia have also shown weaker McGurk effect than their typically developing peers because they are less accurate at speech reading and have a problem identifying speech sounds which may have implications on speech perception [25].

In the case of healthy individuals, the AV- integration function increases from childhood to adolescence and decreases from adolescence to early adulthood. However, in the case of people suffering from schizophrenia, the experience is poor in adolescence and adulthood. Therefore, it seems that schizophrenia is associated with early and persistent impairment in the development of the audiovisual integration ability [26].

## 1.5 Pre-stimulus: Gateway to Predicting Perception

The McGurk illusion is the most predominantly used audio-visual integration illusion paradigm. However, there are a number of clear limits on the use of the McGurk effect as a method. Numerous studies state that there is a significant individual variation even in healthy elderly in response to the McGurk stimuli, with some individuals perceiving it more frequently than the others [21]. Moreover, an individual may perceive the same illusion on some occasions (experimental trials) but not on others exhibiting inter-trial differences. This inter-individual and trial-wise variability in perception performance is likely to stem from the subjects' attentional capabilities that mirrors the ongoing oscillatory brain activity, that is already present prior to stimulus presentation (or the pre-stimulus duration). Earlier seminal studies carried out by Keil and colleagues have shed some insight into how pre-stimulus oscillatory brain activity may sculpt perception and reflect preparatory stages for the incoming stimuli. In their study, they have reported that activity primarily in the beta band (14-30 Hz) neural oscillations can predict the perception of the McGurk effect of the

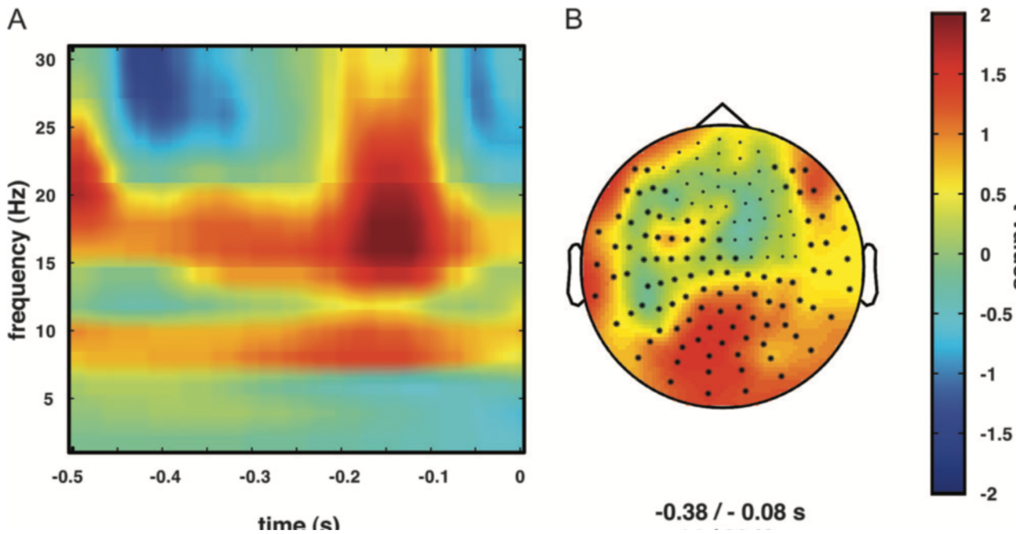


Figure 1.3: ***Prestimulus activity:*** *(A) Time-frequency representation of the prestimulus interval at sensor level for the comparison between 'fusion' and 'unimodel' trials. Time 0 ms indicates the onset of mouth movement and audio stream. (B) Topography (14-30 Hz, -380 to -80 ms) of the positive beta band cluster found in the prestimulus interval at sensor level for the comparison between fusion and unimodel trials* [23].

observers [23]. By doing an MEG recording of subjects that were exposed to the McGurk trials (fusion) and unimodal trails, they found a high beta band activity in parietal, frontal and temporal areas of the brain at a time duration of 300 ms (-380 to -80 ms) before the onset of the McGurk stimulus, see Figure 1.3. The broadband beta power activity was particularly prominent in the left superior temporal gyrus (lSTG), which is one of the key sites involved in speech and language processing and also involved in multisensory integration. The illusory perception was accompanied by a decrease in post-stimulus theta band activity over the cuneus, precuneus, and left superior frontal gyrus (lSFG). Thus, they claimed that the McGurk effect depends on the unsteady brain states suggesting that functional connectedness of the left STS (Superior Temporal Sulcus) at a pre-stimulus point is crucial for an audiovisual percept.

## 1.6 Scope of the Dissertation

Multisensory integration is a quintessential process in speech perception during a face-to-face conversation as it involves both the auditory and visual inputs. This phenomenon can efficaciously be studied during the McGurk effect, in which the auditory signal of one phoneme is integrated with the visual signal of a second phoneme which causes an individual to perceive a different third phoneme [16]. However, it has been clearly stated by several studies that there is a variation in the number of times the McGurk illusion is experienced by an individual, some perceiving it more frequently than the others [21]. Also, an individual may perceive the same illusion on some occasions (experimental trials) but not on others exhibiting inter-trial differences. This inter-individual and inter-trial variability might stem from an individual's ongoing oscillatory brain activity, that is already present in the absence of any stimulus also called as the pre-stimulus[9] activity. Moreover, these oscillatory brain activities are an outcome of network interactions among local subpopulations of excitatory and inhibitory neurons so macroscopic-recordings such as EEG can help us in capturing and analyzing these brain signals [35]. Therefore, we focused on understanding the brain oscillations in the pre-stimulus durations of two different epoch lengths: (a) 500 ms (closer to the preceding stimulus) and (b) 800 ms (closer to the succeeding stimulus), between the two groups of perceivers: rare and frequent (characterized by behavioral perceptual responses [27]), using different methods and proposed the following hypothesis for all the proposed quantitative measures:

1. First, we computed the power of the signal at distinct frequency ranges (theta: 4-7 Hz; alpha: 8-12 Hz; beta: 15-30 Hz, and; gamma: 31-45 Hz) and determined inter-trial and inter-individual variability between the two groups of perceivers. A recent study by Keil and colleagues on pre-stimulus oscillations (of duration -380 to -80 ms

[9] Pre-stimulus data carry spontaneous meaningful brain rhythm signatures

before the sound onset) have reported an increase in the beta-band activity which can predict the perception of the McGurk illusion in the observers [23]. For our current study, we hypothesize that we would also observe a higher beta-band activity in the participants' pre-stimulus durations (for both 500 ms and 800 ms duration). Furthermore, since the inter-trial variability correlates with the behavioral response, we hypothesize that trial-wise variability in power may indicate the robustness of the oscillations as a reliable neuronal marker of AV integration.

2. Second, we computed the time-averaged global coherence to measure the large-scale functional connectivity dynamics in the pre-stimulus duration during AV speech perception [40]. Previous studies by Kumar *et al.*, have reported that the global coherence patterns can actually be a functional connectivity marker for perceptual experience during the McGurk illusion at the inter-individual and inter-trial level of variability [27][34][40]. Therefore, we computed for the global coherence in both rare and frequent group of perceivers. In this study, we hypothesize that we would observe a higher beta-band activity in the participants' pre-stimulus coherence values (for both 500 ms and 800 ms duration), similar to the results of the power spectrum analysis.

3. Finally, we extracted the periodic and aperiodic components of the spectra for both the rare and frequent group using FOOOF (Fitting Oscillations and One Over $f$) algorithm. Some studies suggest that a signal typically has both the periodic and aperiodic oscillations. The periodic oscillations are linked to numerous cognitive, perceptual, and behavioral states, whereas the aperiodic component is said to be linked to the background neural noise. Hence, we performed the FOOOF analysis, developed by Voytex and group [33] to isolate the periodic components of the pre- and post-stimulus 800 ms epochs. For this study, we hypothesize that we would see a change in the significant oscillatory power between the two groups of perceivers.

# Chapter 2

# Materials and Methods

The electrophysiological (EEG) data of the participants, subjected to the following experiment, were both designed and recorded previously by Dr. G Vinodh Kumar from the Cognitive Brain Dynamics Lab (CBDL), NBRC.

## 2.1 Participants

Eighteen healthy right-handed participants (10 males, and 8 females) with a mean age of $24.9 \pm 2.8$ years had given a written informed consent under the experimental protocol approved by the Institutional Human Ethics Committee of the National Brain Research Centre (NBRC), Gurgaon which is in agreement with the Declaration of Helsinki.

## 2.2 Experiment Design and Trials

A digital video of a native Hindi speaking male articulating the syllables */pa/*, */ta/*, and */ka/* 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 du-

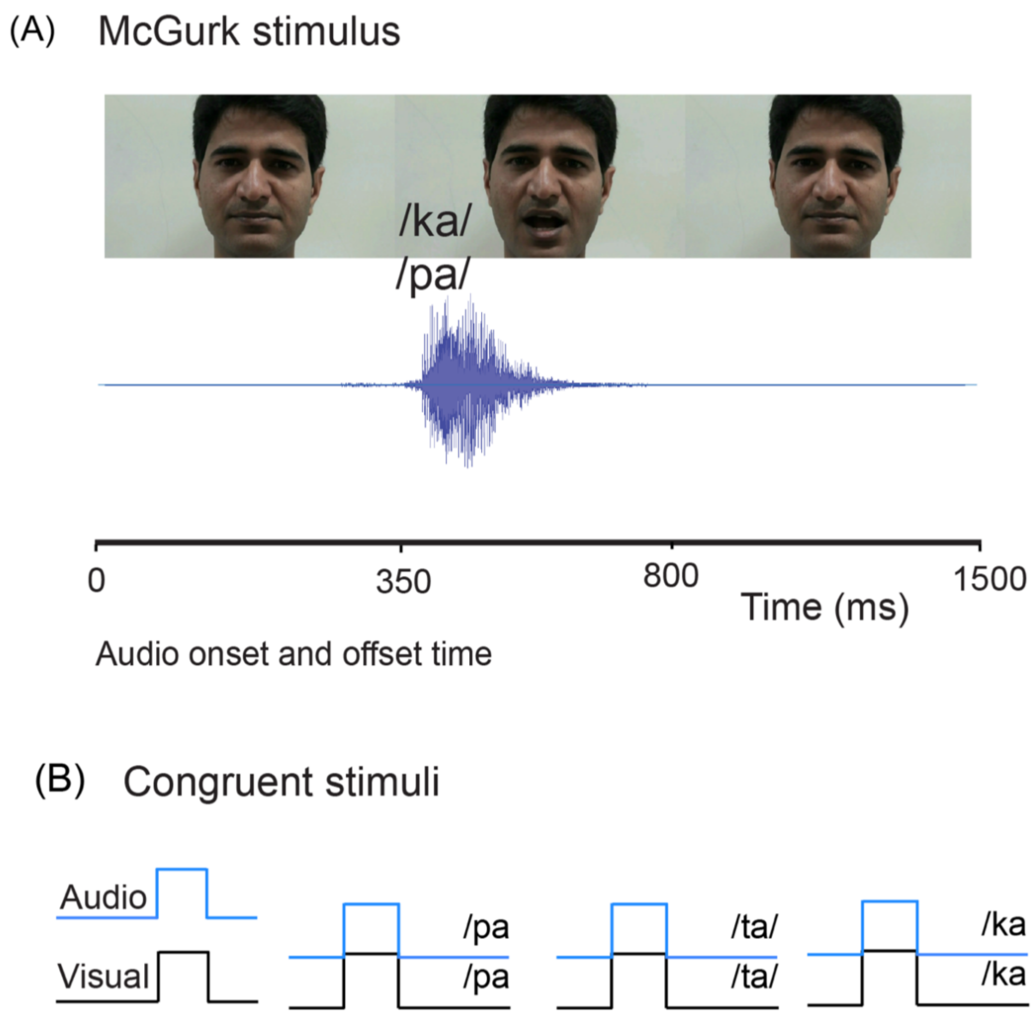


Figure 2.1: ***Stimuli:*** *(A) Sample trial with three video frames from McGurk stimulus (audio- /pa/ + video- /ka/) used in this experiment (top row), the audio trace of the syllable /pa/ presented simultaneously to the video (middle row) and the onset and offset time of the audio. (B) The congruent AV stimuli: each block represents a video with the audio /pa/, /ta/, and /ka/ dubbed onto a video of a person articulating /pa/, /ta/, and /ka/ respectively* [27].

ration of the auditory syllables in the videos ranged from 0.4 to 0.5 seconds. Also, the duration of each video clip ranged from 1.5 to 1.7 seconds to include the neutral, mouth closed position, mouth movement of articulation, and mouth closing. The stimuli consisted of four kinds of videos: three congruent[1] syllables */pa/*, */ta/*, and */ka/*; and one McGurk[2] syllable (auditory */pa/* with visual */ka/*) producing the illusion of syllable */ta/* or */tha/*. See Figure 2.1.

The experiment contained five blocks of experiments- each block was made up of 120 trials (30 trials of each video was presented at random). Inter-stimulus intervals were pseudo-randomly varied between 1200 ms (milliseconds) to 2800 ms to minimize prediction error. Using a forced-choice task on every trial, participants reported their subjective perception by pressing a specified key on the keyboard corresponding to */pa/*, */ta/*, */ka/,* or *something else* while watching the audio-visual videos.

## 2.3 Data Acquisition and Preprocessing

### 2.3.1 EEG

Continuous EEG scans were acquired using a Neuroscan system (Synamps2, Compumedics, Inc.) with 64 Ag/ AgCl scalp electrodes molded 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 center (near Cz) reference electrode on the Neuroscan cap and computed at a sampling rate of 1000 Hz. Channel impedances were maintained at values $< 10\text{k}\Omega$.

[1] Audio syllables matching with the video articulation

[2] Incongruent- audio and visual mismatch

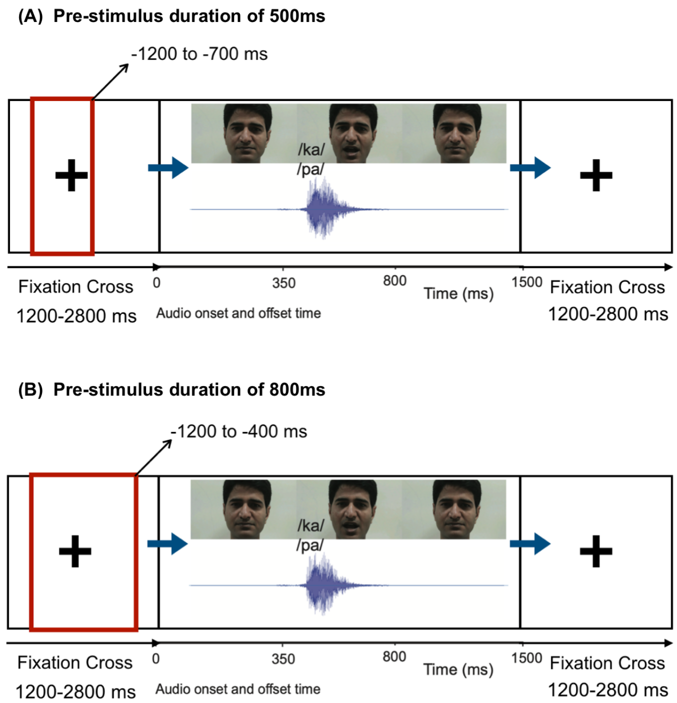


Figure 2.2: ***Pre-stimulus time-window:*** *Example trial with a fixation cross pseudorandomly varied between 1200 to 2800ms, before the onset of the audio-visual stimuli. (A) The first group of pre-stimulus epoch of 500ms duration was extracted between 1200ms to 700ms of the inter-stimulus time interval (ISI). (B) And, the second group of pre-stimulus epoch of 800ms duration was extracted between 1200ms to 400ms of the inter-stimulus time interval (ISI).*

### 2.3.2 Preprocessing

Preprocessing and off-line data analysis were performed using EEGLAB - an open-source MATLAB toolbox for EEG data analysis [28], Chronux - a MATLAB based platform for analyzing neural signals [29], and custom-made MATLAB scripts (The MathWorks, Inc., version R2019b). Continuous data were high-passed at 0.1 Hz with finite impulse response (FIR) filter, low-passed at 80 Hz FIR filter, and Notch-filtered between 46 to 54 Hz to remove the line noise ($9^{th}$ order 2-pass Butterworth filter). The noisy channels were removed and interpolated, following which the data was average re-referenced. For further data analysis, three different kinds of epochs[3] were extracted. The first two were the epochs of 0.5 (-1200

[3] Specific-time windows that are extracted from the continuous EEG signal. The epochs are usually time-locked with respect to an event.

ms to -700 ms) seconds and 0.8 (-1200 ms to -400 ms) seconds duration before the onset of the sound stimulus, and the last kind of epoch was of 0.8 seconds post the onset of the stimulus. See Figure 2.2. The extracted epoch data were sorted based on the AV stimuli - Congruent AV stimuli: */pa/*, */ta/*, and */ka/* and incongruent McGurk stimulus. The sorted pre-stimulus epochs were then baseline corrected by removing the temporal mean of the EEG signal on an epoch-by-epoch basis. However, for the post-stimulus 0.8s epoch, baseline correction was done by removing the temporal mean of the EEG signal 0.2s before the onset of the stimulus. Furthermore, in order to remove the response contamination from the ocular and muscle-related artifacts, epochs with amplitudes above and/or below $\pm 75\ \mu V$ were removed from all the electrodes. And, finally, the remaining artifacts were removed by self-inspection.

## 2.4 Data Analysis

### 2.4.1 Behavior

For the incongruent McGurk stimulus, visual */ka/* was paired with auditory */pa/* to induce the illusory percept */ta/*. Also, three congruent AV stimulus (*/pa/*, */ta/*, and */ka/)* with no lag between the audio and visual input was presented to the participants. As the participants observed the four stimuli presented to them at random, they reported if they heard either */pa/*, */ta/*, */ka/, or "something else,"* being unaware of the McGurk illusion. Kumar V. *et al.*, in their previous work classified the participants into two groups: (1) *rare perceivers* - $<$50% */ta/* percept, and (2) *frequent perceivers* - $>$50% */ta/* percept, based on their McGurk susceptibility[27]. For our study, we followed the same behavioral classification. This classification was for subsequent analysis associated with inter-individual variability between these two specific groups. Furthermore, with these between-group classification (rare and frequent groups), inter-trial variability was also analyzed. The inter-trial variability (or response tendency) was computed as the relative proportion of illusory (*/ta/*) responses in all the McGurk trials across all participants belonging to the respective perceiver group.

### 2.4.2 Spectral Analysis

#### Introduction

Power Spectral analysis is one of the most widely accepted method for the analysis of EEG signals. It is based on the Fourier theorem, which states that the waveform can be decomposed into the sum of sine waves at different frequencies with different amplitudes and phase relationships, and the summation of these waves reconstitutes the original waveform. The analysis of the time-frequency EEG signal begins with a Fourier transform. The Fourier transformation is a mathematical operation which provides the frequency, amplitude, and phase parameters of every sine wave components. Fourier coefficients represent the amplitude and phase relationship at each of the component sine wave frequencies. Squaring and summing these Fourier coefficients at every frequency point gives the power at that frequency. Finally, a plot of power at each of the component frequencies is called the power spectrum. It is this power spectrum that allows the determination of relative amounts of given frequencies in the waveform over the time segment analyzed. The fast Fourier transform method allows for real-time spectral analysis.

#### Method

Power spectra were computed to investigate inter-trial and inter-individual differences for all the three different epoch conditions (pre-stimulus 0.5s and 0.8s; post-stimulus 0.8s). The power spectra of the preprocessed EEG signals at each electrode were computed on a trial-by-trial basis. The spectral power was computed at different frequencies using customized MATLAB scripts and the Chronux toolbox function mtspectrumc.m - function for a multi-taper spectrum, continuous process [29]. Time bandwidth product and the number of tapers used were set to 3 and 5 respectively for all the three different epoch data. Subsequently, in the case of inter-individual variability, the difference in the power between the two groups at different frequency bands (theta, alpha, beta, and gamma) for all the stimulus conditions (incongruent McGurk, and congruent) were statistically compared by means of the cluster-based permutation test. However, in the case of inter-trial variability only the differences in the power during /ta/ and /pa/ responses of the two groups were compared for the same statistical test. See section 2.4.5 for further details.

### 2.4.3 Network analysis: Global Coherence

Global coherence was computed to investigate frequency-specific functional connectivity (FC) that subserves the cross-modal speech perception and also to characterize inter-trial and inter-individual variability differences between the rare and frequent perceivers [27][34][40]. The global coherence is a mathematical technique that is used to capture and quantify the strength of the covariation of neural oscillations at the global (or whole-brain) scale [27]. We employed the Chronux [29] function CrossSpecMatc.m to obtain trial-wise global coherence of the three kinds of epoch data (pre-stimulus 0.5s and 0.8s; post-stimulus 0.8s) that were already sorted based on stimuli and responses. Three orthogonal discrete prolate spheroidal sequences (or dpss) were also considered. The dpss is also known as Slepian tapers, which is used to reduce leakage of the spectral estimates into the nearby frequency bins. The time-bandwidth chosen was 5 that resulted in a frequency bandwidth of 0.25Hz. The output variable Ctot of the function yields the total global coherence value at frequency $f$. Initially, the function multiplies the given kind of epoch data with the set specified number of Slepian tapers and then performs the fast Fourier transform (FFT).

The resulting FFT values were then averaged and the cross-spectrum for all the sensor combinations at frequency $f$ was computed. The cross-spectral density between any two sensors was computed from the tapered Fourier transforms using the following equation

$$C_{ij} = conj\left(X_i\left(f\right).X_j\left(f\right)\right) \tag{2.1}$$

where $C_{ij}$ represents the cross-spectrum, $X_i$ and $X_j$ 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 the specified frequency $f$ which yields the following

$$D\left(f\right) = VCV^T \tag{2.2}$$

The diagonal matrix $D$ comprises of the values that are corresponding to the variance explained by the orthogonal set of eigenvectors $\left(V, V^T\right)$. Finally, the global coherence $C_{Global}\left(f\right)$ at frequency bin $f$ was computed by normalizing largest eigenvalue (or the first value of the $D\left(f\right)$ at frequency $f$) on a trial-by-trial basis for each participant in the two groups, employing the following equation

$$C_{Global}\left(f\right) = \frac{D_1\left(f\right)}{\sum_{i=1}^{n} D_i\left(f\right)} \tag{2.3}$$

Finally, the global coherence computed on a trial-by-trial basis was sorted based on the perceptual category *(/ta/* and */pa/)* and averaged across all the participants belonging to either one of the groups (rare or frequent). Furthermore, in the case of inter-individual variability, the coherences (or Ctot values) of the two groups for all the stimulus conditions (incongruent and congruent) were compared for the significant difference at different frequency bands (theta, alpha, beta, and gamma) explicitly by means of cluster-based permutation test [32]. However, in the case of inter-trial variability, only the difference in the Ctot values during */ta/* and */pa/* responses was compared for the same statistical test. See section 2.4.5 for further details.

In addition to this, we further analyzed if changes in global coherence values at specific frequency bands (theta: 4-7 Hz; alpha: 8-12 Hz; beta: 15-30 Hz; gamma: 31-45 Hz) of the two pre-stimulus epoch condition (0.5s and 0.8s) and the post-stimulus epoch condition of 0.8s, correlated with the participants' susceptibility to the McGurk perception. Participant-wise mean of the global coherence (Ctot values) in specific frequency bands, for both incongruent and congruent stimulus conditions, were computed and statistically analyzed using Spearman rank correlation and t-tests.

### 2.4.4 FOOOF analysis

#### Introduction

Electrophysiological signals exhibit both periodic and aperiodic features. Several studies on the periodic oscillations have linked it to numerous cognitive, perceptual, and behavioral states. Whereas for the aperiodic "background" 1/f component of the neural power, it is said to be dynamic and is linked to the relative excitation/inhibition of the underlying neuronal population (the background neural noise) [33]. This suggests that the spectral power one measures to determine relative amounts of given frequencies in the waveform does not necessarily imply oscillatory power. In fact, the changes in power ratios between bands may coincide with the aperiodic slope differences rather than a change in true oscillatory power. Therefore, the four features (frequency, power, aperiodic broadband offset, and aperiodic slope) must be carefully parameterized to avoid coalescing them with one another. Voytex and group have introduced an efficient algorithm to automatically parameterize neural power spectral densities (PSDs) into periodic and aperiodic components.

FOOOF (or Fitting Oscillations and One Over $f$) algorithm takes the original PSDs data and extracts the aperiodic signal and superimposes them on periodic oscillatory components, referred to as "peaks". These peaks are considered to be oscillations and are modeled individually as Gaussian functions. Each of these Gaussian has three parameters that are used to define an oscillation. The formula for fitting the power spectrum is as follows

$$P = L + \sum_{n=0}^{N} G_n \tag{2.4}$$

where $P$ is power. It is a linear combination of the aperiodic signal $L$ with the $N$ total number of Gaussians $G$. Each $G_n$ is a Gaussian fit to the peak for $N$ total number of peaks extracted from the power spectrum, modeled as

$$G_n = a * \exp\left(\frac{-(F-c)^2}{2w^2}\right) \tag{2.5}$$

where $a$ is the amplitude, $c$ is the center frequency, $w$ is the bandwidth of the Gaussian, and $F$ is the vector of input frequencies. Furthermore, the aperiodic signal $L$ is modeled using an exponential function in *semilog-power* space (linear frequencies and logged power values) as the following

$$L = b - \log\left(k + F^{\chi}\right) \tag{2.6}$$

where $b$ is the aperiodic broadband offset, $\chi$ is the aperiodic slope, and $k$ is the "knee" parameter, controlling for the bend in the aperiodic signal, with $F$ as the vector of input frequencies. The above equation is equivalent to fitting a line in *log-log* space, when $k = 0$, referred to as the "fixed" model.

The final output of the algorithm is the parameters defining the best model fit for the aperiodic signal and the $N$ Gaussians. Along with the Gaussian parameters, transformed parameters are also returned. These parameters involve (1) center frequency which is the mean of the Gaussian; (2) amplitude of the peak which is the distance between the peak of the Gaussian and the aperiodic fit, and; (3) bandwidth as two standard deviations.

**Method**

The periodic and the aperiodic components of the spectral data was extracted for only two epoch conditions (pre-stimulus 0.8s, and post-stimulus 0.8s) using FOOOF (or Fitting Oscillations and One Over $f$) algorithm. The algorithm is an open-source Python Package with package dependencies limited to *NumPy* and *scipy* ($\geq$ version 0.19). The FOOOF algorithm (version 1.0.0) was used to parameterize neural power spectra. Power spectra were parameterized across the frequency range 0.1 to 45 Hz using customized Python scripts on PyCharm 2019.3.4 Python environment. FOOOF was only performed to the inter-individual spectral data of the two groups (rare and frequent) for all the stimulus conditions (incongruent and congruent). The settings for the algorithm were set as: (1) *peak_width_limits = [0.5, 7]*; (2) *min_peak_height = 0.075*; (3) *max_n_peaks = 2*; (4) *peak_threshold = 1*, and; (5) *aperiodic_mode = "fixed"*. See Figure 2.3 for the details on FOOOF output. Subsequently, the difference in the periodic components (or *fooofed_spectrum*) and the aperiodic components (or *aperiodic_values*) between the two groups for all the different frequency bands were statistically compared using two-sample $t$-tests.

### 2.4.5 Statistical analysis

The statistical test used for the above two analyses: (1) Power spectral analysis (section 2.4.2) and (2) Global Coherence (section 2.4.3) was the cluster-based permutation testing [32]. For every frequency bin at each time-point, the corresponding differences (of power and coherence values) between the two groups the rare and the frequent perceivers were

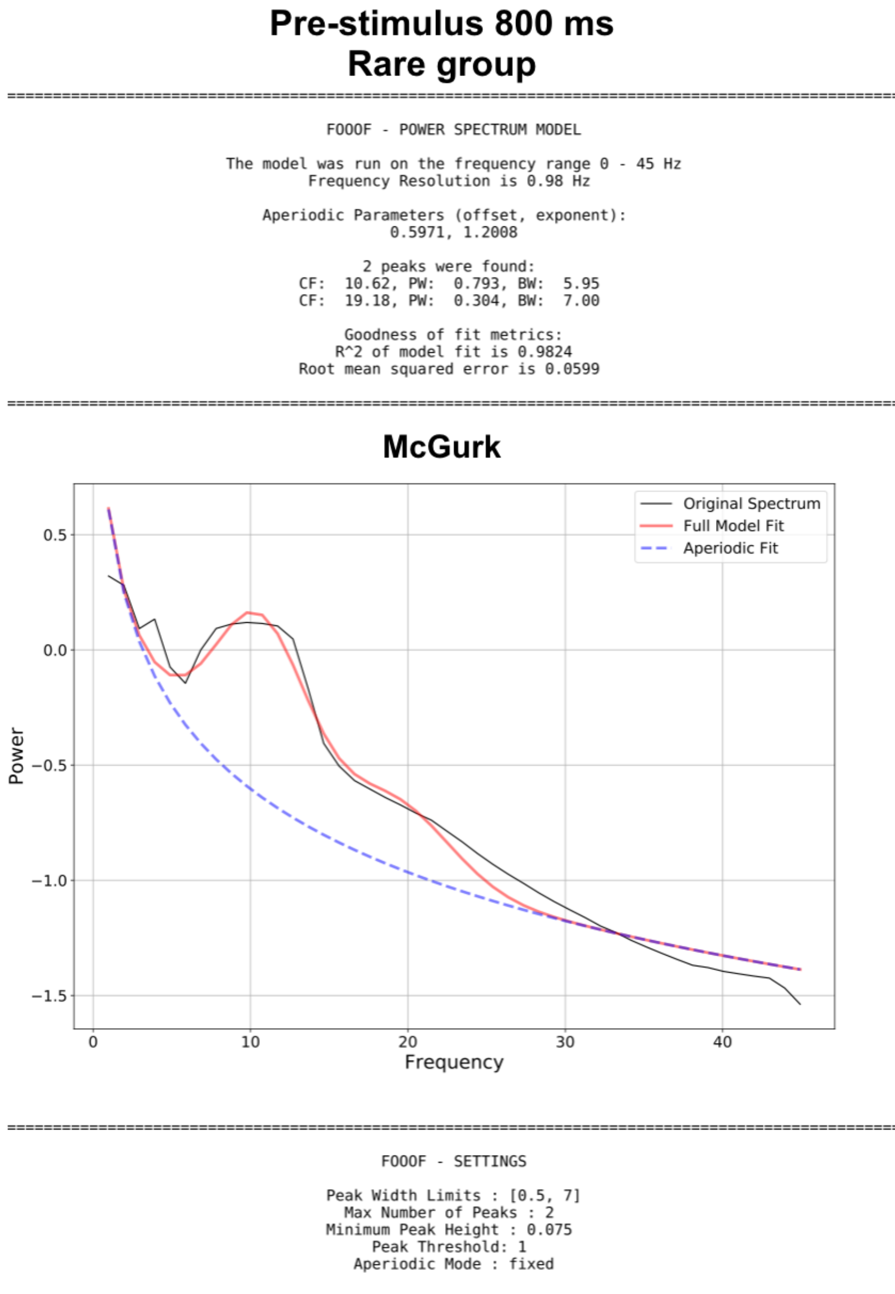


Figure 2.3: ***FOOOF report:*** *(of pre-stimulus 800 ms rare group's McGurk stimulus) This report consists of the original power spectrum, with the periodic (full model fit) and aperiodic fit. The parameters used for the analysis are mentioned below the graph. The output parameters of the model include the aperiodic components (offset and exponents), goodness of fit ($R^2$ and root mean square error), and the peak parameters (center frequency (CF), peak width (PW), and bandwidth (BW)).*

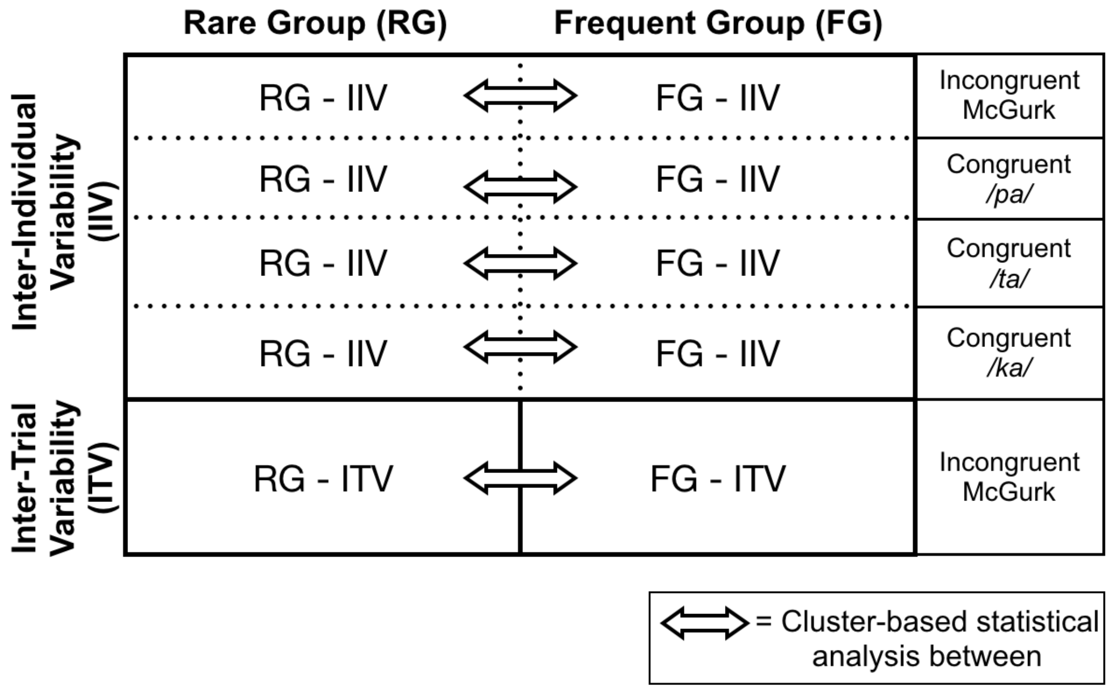


Figure 2.4: *Schematic representation of the aspects (IIV and ITV), groups (RG and FG), and stimuli (McGurk, congruent /pa/, /ta/, and /ka/) that were statistically analyzed in the study.*

evaluated using the Fisher's Z transformation:

$$Z(f) = \frac{\tanh^{-1}(C_1(f)) - \tanh^{-1}(C_2(f)) - \left(\frac{1}{2m_1-2} - \frac{1}{2m_2-2}\right)}{\sqrt{\frac{1}{2m_1-2} + \frac{1}{2m_2-2}}} \tag{2.7}$$

where $2m_1$, $2m_2$ = degrees of freedom in the first and second condition; $Z(f) \approx N(0,1)$ a unit normal distribution; and $C_1$ and $C_2$ are the corresponding differences (of either power or coherence) at frequency $f$. The *Z-statistic* matrix obtained from the above computation formed the observed *Z-statistics.* Subsequently, clusters from observed *Z-statistics* matrix were selected based on oscillatory frequencies (4-7 Hz, theta; 8-12 Hz, alpha; 15-30 Hz, beta; 31-45 Hz, gamma). Following the computation of the cluster-level statistics of the observed *Z-statistics,* 1000 iterations of trial randomization were carried out in the case of power spectral and global coherence data. For every iteration, a cluster-level statistic was computed on the totally randomized trials to generate the permutation distribution.

Subsequently, the values of the observed cluster-level statistics were compared with $5^{th}$ and $95^{th}$ quantile values of the respective permutation distribution. The observed cluster-level statistics value that were above the $95^{th}$ and below the $5^{th}$ quantile subsequently for two time-points formed the positive and negative clusters respectively.

# Chapter 3

# Results

## 3.1 Behavior

The participants were subjected to four kinds of AV stimuli presented to them at random: (1) Incongruent McGurk pair (visual */ka/* paired with auditory */pa/* to induce the illusory percept */ta/)*; (2) Congruent (video and audio synched) */pa/*; (3) Congruent */ta/*, and; (4) Congruent */ka/*. Participants observing the four stimuli reported if they heard either */pa/*, */ta/*, */ka/, or "something else,"* being unaware of the McGurk illusion. Kumar V. *et al.*, observed a high degree of inter-individual variability in McGurk susceptibility[27]. See Figure 3.1 for further details. Based on their McGurk susceptibility, the participants were classified into two groups: (1) *rare perceivers* - $<50\%$ */ta/* percept, and (2) *frequent perceivers* - $>50\%$ */ta/* percept. This classification was for subsequent analysis associated with inter-individual variability between these two specific groups. Eight of the eighteen participants showed $<50\%$ propensity towards the McGurk effect. The rest of the participants showed $>50\%$ propensity towards the McGurk effect.

Furthermore, with these two group classification (rare and frequent groups) inter-trial variability was also analyzed. The inter-trial variability (or response tendency) was computed

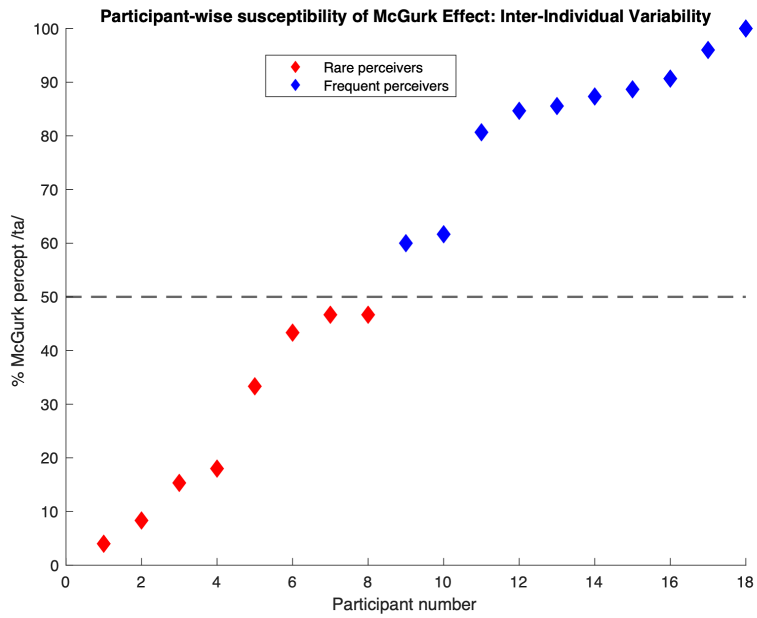


Figure 3.1: ***Behavior:*** *Inter-individual variability - Propensity of McGurk effect for all the 18 participants expressed as the percentage of* /ta/ *percept during the presentation of the McGurk illusion. The participants were categorized into two groups: (red diamonds, <50%) rare perceivers and (blue diamonds, >50%) frequent perceivers [27].*

as the relative proportion of illusory (*/ta/*) responses in all the McGurk trials across all participants belonging to the respective perceiver group (rare and frequent). It was found that rare group participants on an average reported an illusory */ta/* percept 26.95% (SD = 17.62) of trials, whereas a unisensory */pa/* percept was reported in 69.35% (SD = 15.64) of trials, for the incongruent McGurk stimulus condition. Contrastingly, the frequent group participants on an average reported an illusory */ta/* percept 83.52% (SD = 13.18) of trials, and a unisensory */pa/* percept was reported in 11.82% (SD = 10.14) of trials for the incongruent McGurk stimulus. Congruent AV stimuli (*/pa/*, */ta/*, and */ka/*), in the case of rare group, were correctly identified in 95.72% (SD = 3.07) of trials. Whereas for the frequent group, the congruent AV stimuli were 97.19% (SD = 2.09) reported of the trials. The difference between the percentage of */ta/* and */pa/* percept was significant in both rare ($t\,(14) = -5.08, p < 0.001$) and frequent ($t\,(18) = 13.63, p < 0.001$) group of perceivers.

See Figure 3.2 for more details.

## 3.2 Power of Oscillatory Activity

### Inter-Individual Variability

We were interested to differentiate between the distinct perceptual states in both the pre-stimulus (0.5s and 0.8s) and the post-stimulus (0.8s) epoch data, in terms of brain oscillations. Therefore, spectral power at different frequency bands (theta: 4-7 Hz; alpha: 8-12 Hz; beta: 15-30 Hz, and; gamma: 31-45 Hz) during different perceptual conditions were compared. For this purpose, we categorized our participants into two groups based on their susceptibility to the McGurk effect: (1) *rare perceivers* - <50% */ta/* percept, and (2) *frequent perceivers* - >50% */ta/* percept. Firstly, we computed the power spectrum of the pre-stimulus 0.5s (or 500 ms) epoch data for all the four stimulus conditions (incongruent McGurk, and Congruent */pa/*, */ta/*, and */ka/*), and then the two groups were statistically compared by means of cluster-based permutation test (Section 2.4.5). We observed that for all the stimulus conditions, the rare group of perceivers elicited an enhanced power spectrum than the frequent perceivers in the theta and alpha bands. However, the frequent perceivers were characterized by an enhanced power spectrum in the beta and gamma bands (Figure 3.3A and supplementary table 5.1A).

Secondly, we computed the power spectrum of the pre-stimulus 0.8s (or 800 ms) epoch data for all the four stimulus conditions, followed by a statistical comparison between the two groups. We observed that for all the stimulus conditions the rare group of perceivers elicited an enhanced power spectrum in the beta band. Moreover, the frequent perceivers were characterized by an enhanced power spectrum in the alpha and gamma bands (Figure 3.3B and supplementary table 5.1B).

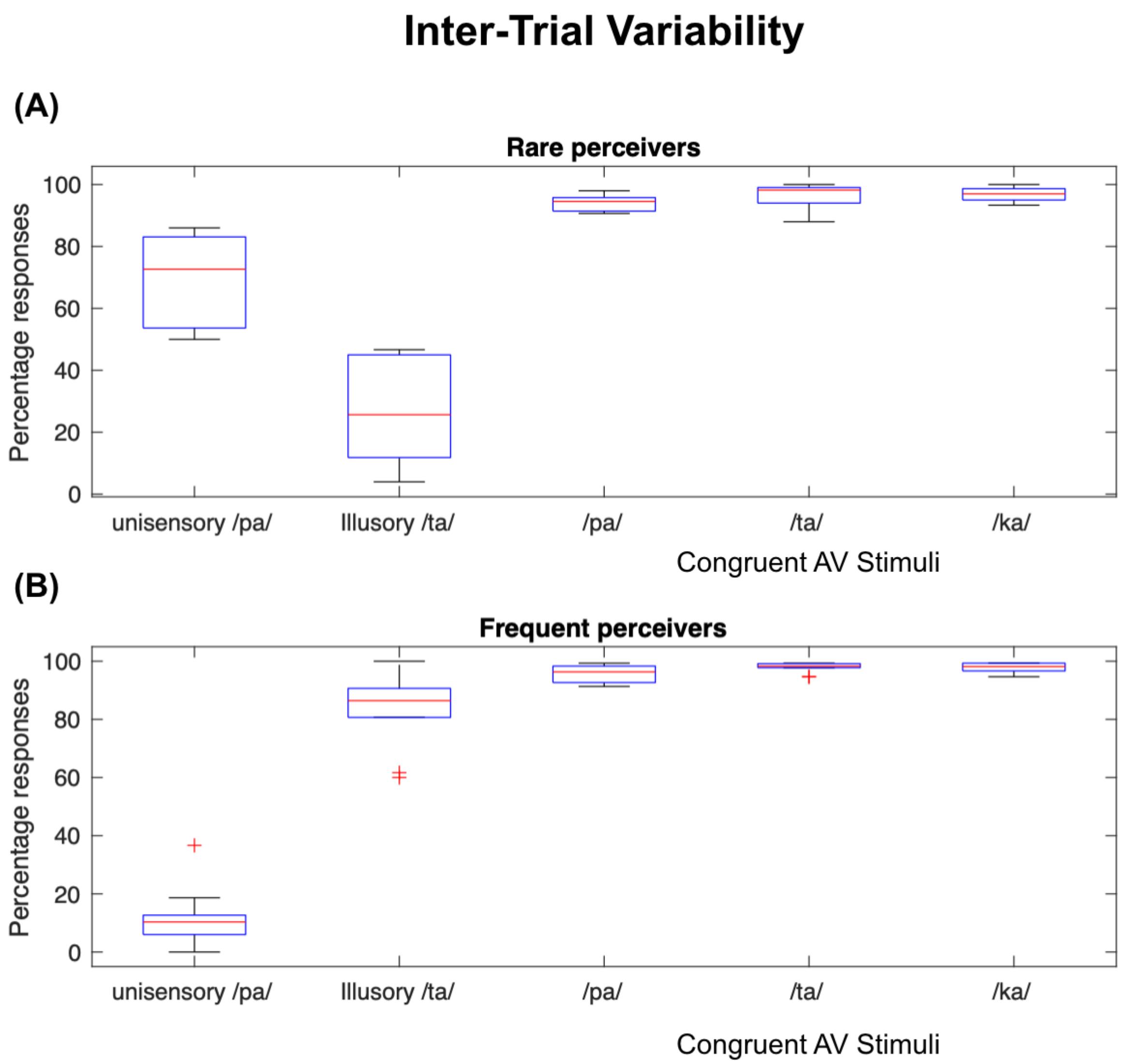


Figure 3.2: ***Inter-Trial Variability:*** *Percentage of* /ta/ *(illusory) and* /pa/ *(unisensory) percept during the presentation of McGurk stimulus and the congruent AV stimuli (* /pa/, /ta/, *and* /ka/ *) averaged over participants in (A) Rare group of perceivers, and; (B) Frequent group of perceivers.*

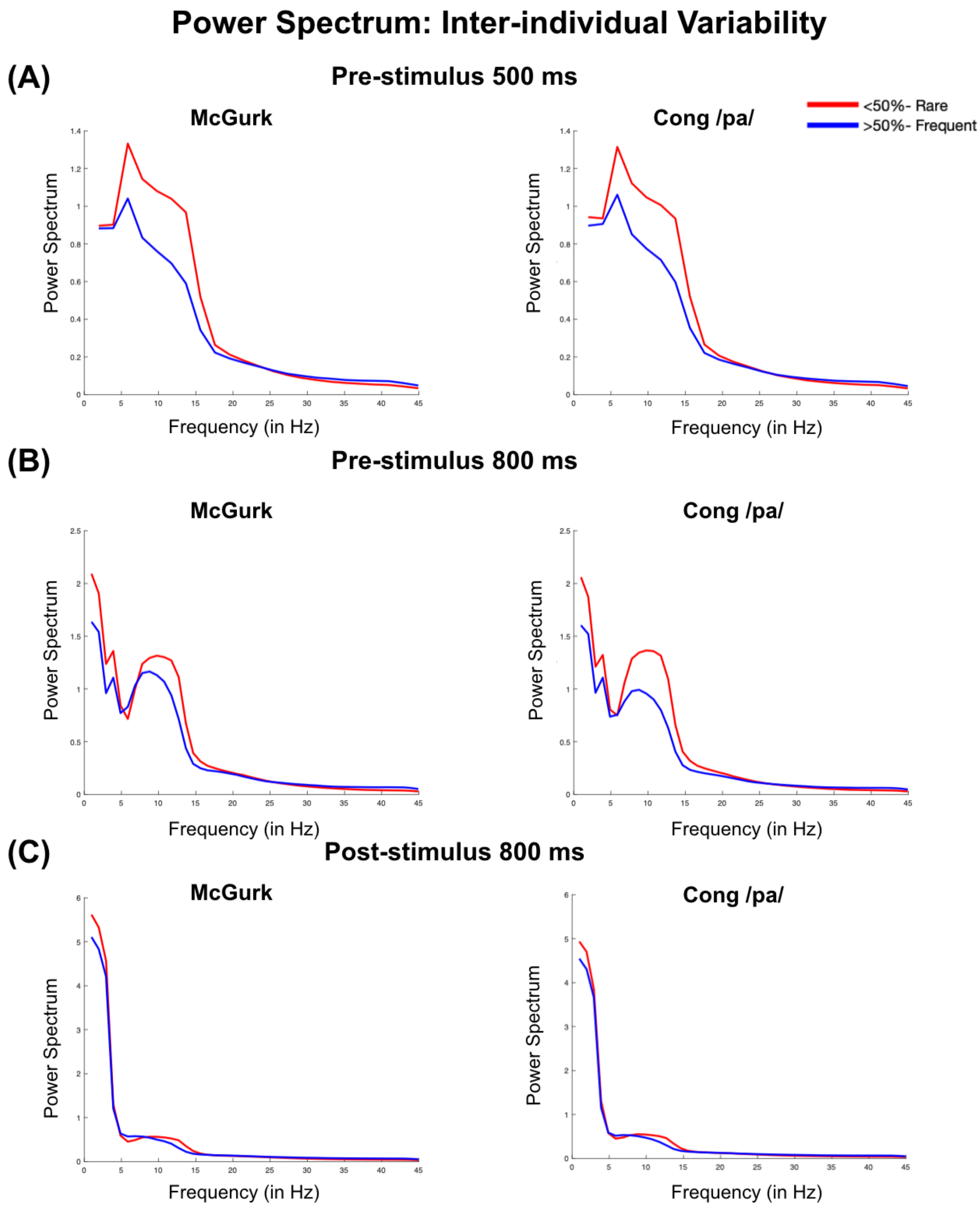


Figure 3.3: ***Power Spectrum:*** *Inter-individual variability - Power Spectrum (with X-axis as Frequency (in Hz) and Y-axis as Power Spectrum) for incongruent McGurk and Congruent /pa/ stimulus conditions of (A) Pre-stimulus 500 ms, (B) Pre-stimulus 800 ms, and (C) Post-stimulus 800 ms epoch data.*

Lastly, we computed the power spectrum of the post-stimulus 0.8s (or 800 ms) epoch data for all the stimulus conditions, followed by a statistical comparison between the two groups. We observed that the rare group of perceivers elicited an enhanced power spectrum in the alpha band for all the stimulus conditions, and in the beta band for only Congruent */ta/* and */ka/* conditions. However, the frequent perceivers were characterized by an enhanced power spectrum in the theta and gamma bands (Figure 3.3C and supplementary table 5.1C).

## Inter-Trial Variability

The power spectrum was computed on trials sorted based on the perceptual categories (*/ta/* and */pa/*) over the participants categorized into the two groups (rare and frequent) and were statistically compared by means of the cluster-based permutation test. For the pre-stimulus 0.5s (500 ms) epoch data, we observed that the */ta/* perception was characterized by a significant increase in the theta band for both the rare and the frequent group. The */ta/* perception was also significant in the alpha band for the frequent perceivers (Figure 3.4A and supplementary table 5.2A). For the pre-stimulus 0.8s (800 ms) epoch data, we observed that the */ta/* perception was significantly higher in the theta, alpha, and beta bands for the rare group of perceivers. Whereas, for the frequent group, the */ta/* perception was significantly higher in only the theta band (Figure 3.4B and supplementary table 5.2B). Finally, in the case of post-stimulus 0.8s (800 ms) epoch data, we observed that the */ta/* perception was significantly higher in the alpha band for the rare group. For the frequent group, no significant difference was observed in the */ta/* perception (Figure 3.4C and supplementary table 5.2C).

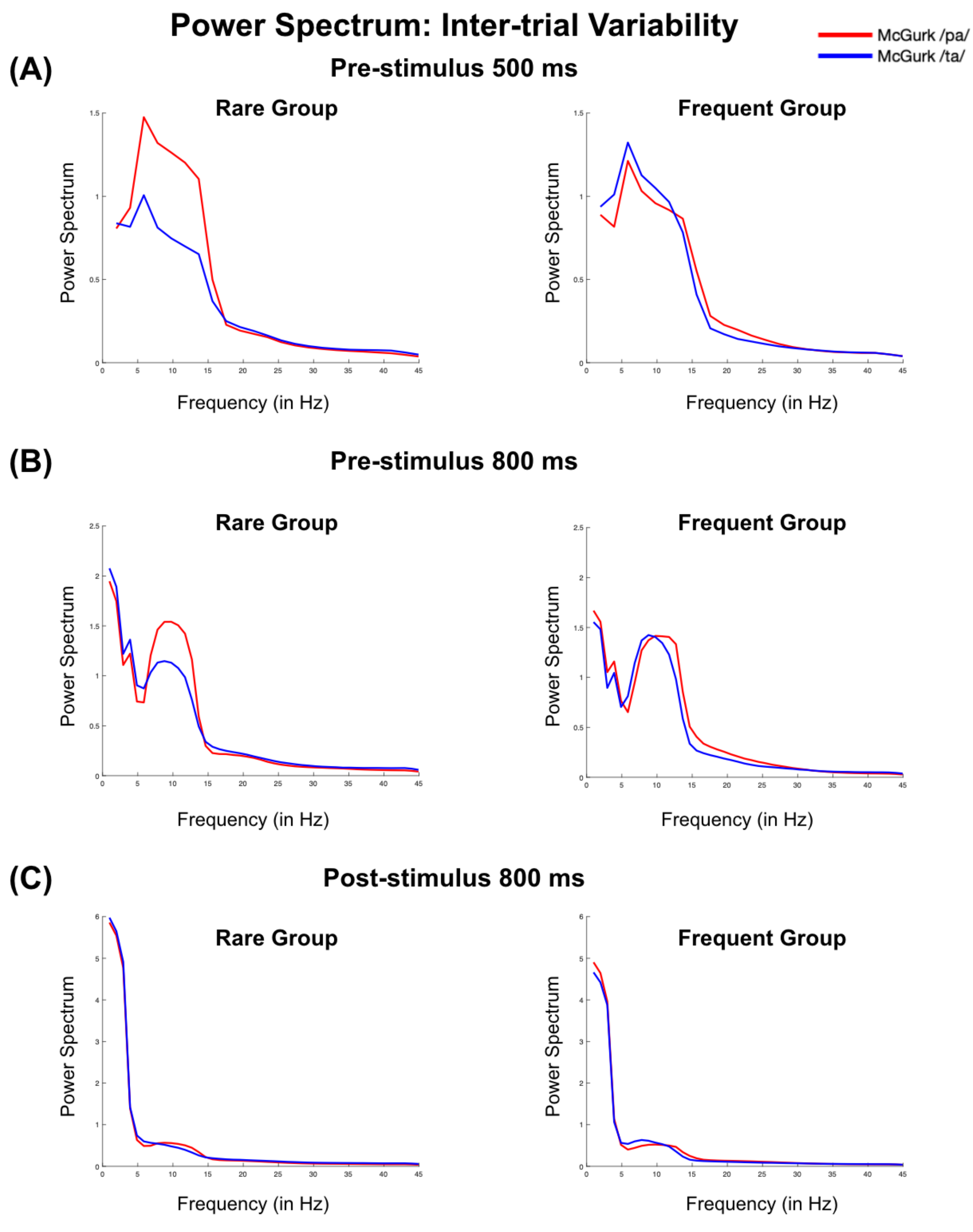


Figure 3.4: ***Power Spectrum***: *Inter-trial variability - Power Spectrum of trials during /ta/ (illusory) and /pa/ (unisensory) percept averaged over participants of the respective groups (rare and frequent) of (A) Pre-stimulus 500 ms, (B) Pre-stimulus 800 ms, and (C) Post-stimulus 800 ms epoch data.*

## 3.3 Functional Connectivity

### Inter-Individual Variability

We were also interested to assess the influence of large-scale functional connectivity on inter-individual differences in the perception of McGurk effect. Therefore, global coherence (or Ctot values) at different frequency bands (theta: 4-7 Hz; alpha: 8-12 Hz; beta: 15-30 Hz, and; gamma: 31-45 Hz) during different perceptual conditions were compared. Participants were categorized into two groups: rare and frequent perceivers, based on their susceptibility to the illusory */ta/* percept. It allowed us to interrogate if inter-individual variability stems from the differences in the inherent processing of multisensory stimuli between the two groups of perceivers. First, we computed the time-averaged global coherence on the pre-stimulus 0.5s (500 ms) epochs for both the incongruent and congruent stimulus conditions, and then the two groups were statistically compared. We observed that for all the stimulus conditions the rare perceivers elicited an enhanced global coherence than the frequent group in the theta and alpha bands. Frequent perceivers were characterized by enhanced global coherence in the beta and gamma bands (see Figure 3.5A and supplementary table 5.3A). Secondly, we computed the time-average global coherence on the pre-stimulus 0.8s (800 ms) epochs for both the incongruent and congruent stimulus conditions. We observed that for all the stimulus conditions, the rare perceivers elicited an enhanced global coherence than the frequent group in the theta, alpha, and beta bands, except for McGurk stimulus in the alpha-band and for congruent */ka/* stimulus in the beta-band which was "insignificant". Frequent perceivers were characterized by enhanced global coherence in only the gamma band (see Figure 3.5B and supplementary table 5.3B). Lastly, post-stimulus 0.8s (800 ms) epoch data for all the stimulus conditions were computed and analyzed. We observed that the rare perceivers showcased an enhanced global coherence in the theta, alpha, and beta bands except for congruent */pa/* stimulus in the theta-band which was "insignificant." The frequent perceivers elicited a higher gamma-band activity for all the stimulus conditions (see Figure 3.5C and supplementary table 5.3C).

Furthermore, we correlated the global coherence values at specific frequency bands (theta: 4-7 Hz; alpha: 8-12 Hz; beta: 15-30 Hz, and; gamma: 31-45 Hz) with the participants' susceptibility to the illusory */ta/* percept. The correlation was carried out using Spearman rank correlation and t-tests. As previously reported by Kumar *et al.,* [27], in the case of post-stimulus 0.8s (800 ms), across all the participants, a significant negative correlation was observed only between the participants' alpha band global coherence and their McGurk susceptibility during incongruent McGurk stimulus condition ($r_s\,(18) = -0.52, p = 0.027$). Interestingly, however, no such significant difference was observed in the case of pre-stimulus 0.5s (500 ms) and 0.8s (800 ms) epoch data condition indicating that the pre-stimulus global coherence value is not correlated to the behavior. Moreover, to make sure that the correlation is a result of cross-modal aspects and not because of the stimulus-specific sensory processing global coherence dynamics for congruent AV stimuli (*/pa/, /ta/,* and */ka/*) were investigated. For post-stimulus 0.8s (800 ms) epochs we observed a significant negative correlation only between participants' alpha band global coherence and McGurk susceptibility in only Congruent */ka/* condition ($r_s\,(18) = -0.47, p = 0.048$). No such significant correlation was observed during congruent */pa/* ($r_s\,(18) = -0.22, p = 0.36$) and */ta/* ($r_s\,(18) = -0.44, p = 0.06$) conditions. Also, there was no significant correlation observed in the case of pre-stimulus (0.5s and 0.8s) epoch data also indicating that the pre-stimulus global coherence value is not associated with the behavior.

## Inter-Trial Variability

The time-averaged global coherence was computed on trials based on the perceptual categories (*/ta/* and */pa/*) over the participants categorized into the two groups (rare and frequent) and were statistically compared by means of the cluster-based permutation test.

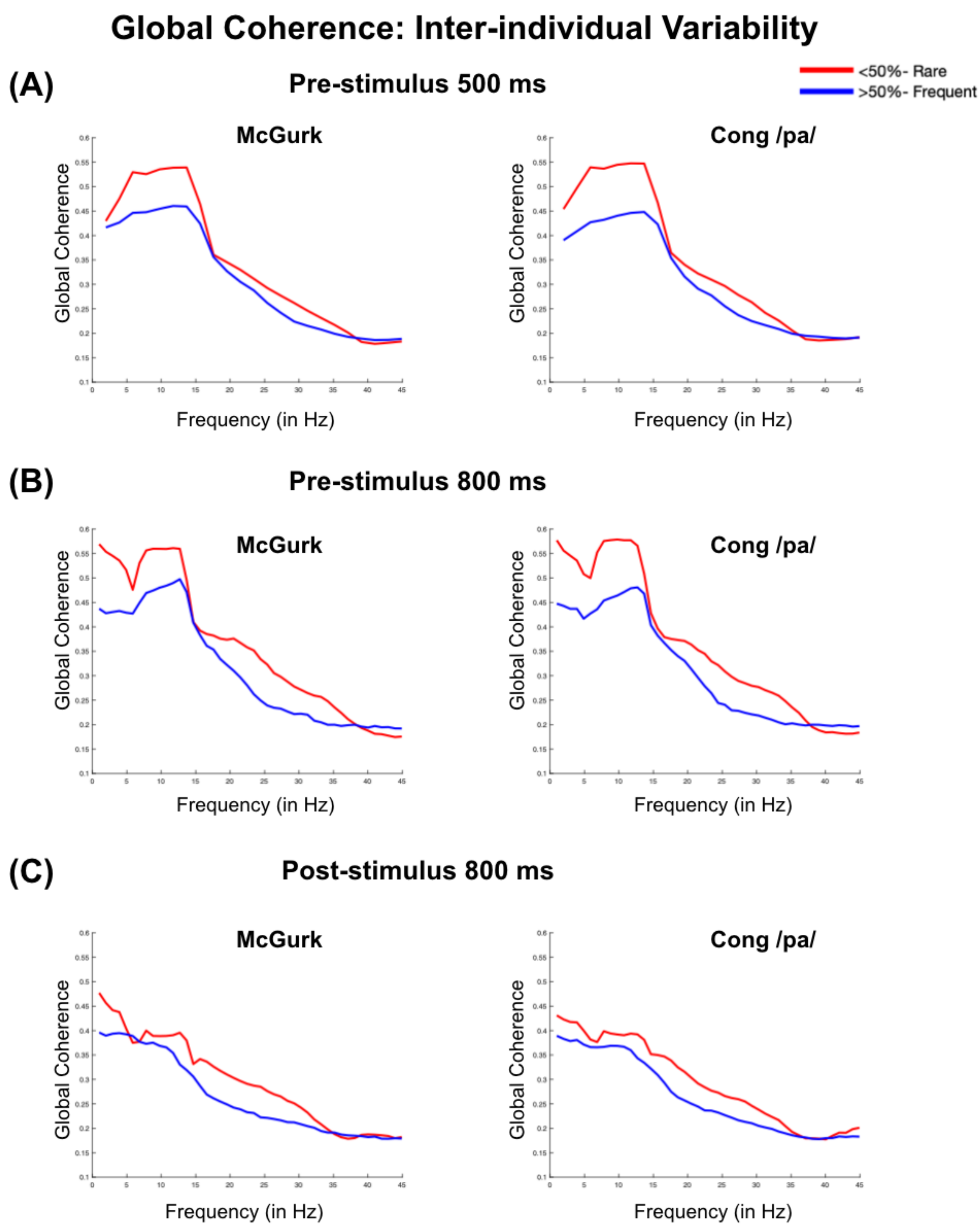


Figure 3.5: ***Global Coherence:*** *Inter-individual variability - Time averaged global coherence during McGurk and Congruent /pa/ stimulus condition of (A)Pre-stimulus 500 ms, (B) Pre-stimulus 800 ms, and (C) Post-stimulus 800 ms epoch data.*

For the pre-stimulus 0.5s (500 ms) epoch data, we observed that */ta/* perception was characterized by a significant increase in the theta and alpha bands for only the rare group of perceivers (Figure 3.6A and supplementary table 5.4A). For the pre-stimulus 0.8s (800 ms) epoch data, we observed that the */ta/* perception was significantly higher in the theta, alpha bands for the rare group of perceivers. However, in the case of the frequent perceivers, the beta and gamma bands elicited the */ta/* perception significance (Figure 3.6B and supplementary table 5.4B). Finally, in the case of post-stimulus 0.8s (800 ms) epoch data, we observed that the */ta/* perception was significantly higher in all the frequency ranges (theta, alpha, beta, and gamma) for the frequent group of perceivers. No such significant difference was observed in the rare group (Figure 3.6C and supplementary table 5.4C).

## 3.4 Periodic and Aperiodic Components of Oscillatory Activity

In order to differentiate between periodic and aperiodic components in both the pre-stimulus 0.8s and post-stimulus 0.8s epoch data of the rare and frequent group of perceivers, FOOOF algorithm was performed. Power spectra were parameterized across the frequency range 0.1 to 45 Hz. Subsequently, the difference in the periodic components (or *fooofed_ spectrum*) and the aperiodic components (or *aperiodic_ values*) between the two groups for all the different frequency bands were statistically compared using two-sample $t$-tests. First, we parameterized the pre-stimulus 0.8s (800 ms) epochs to extract periodic component. We observed that for the McGurk stimulus, the rare perceivers were characterized by a significant decrease in periodic component in gamma band ($t\,(28) = -7.58, p < 0.001$). For Congruent */pa/*, */ta/*, and */ka/* stimuli, a significant increase in alpha band (Cong */pa/:* $t\,(8) = 5.12, p =< 0.001$; Cong */ta/:* $t\,(8) = 4.63, p = 0.0017$; Cong */ka/:* $t\,(8) = 4.58, p = 0.0018$) followed by a significant decrease in the gamma band (Cong */pa/:* $t\,(28) = -7.12, p =< 0.001$; Cong

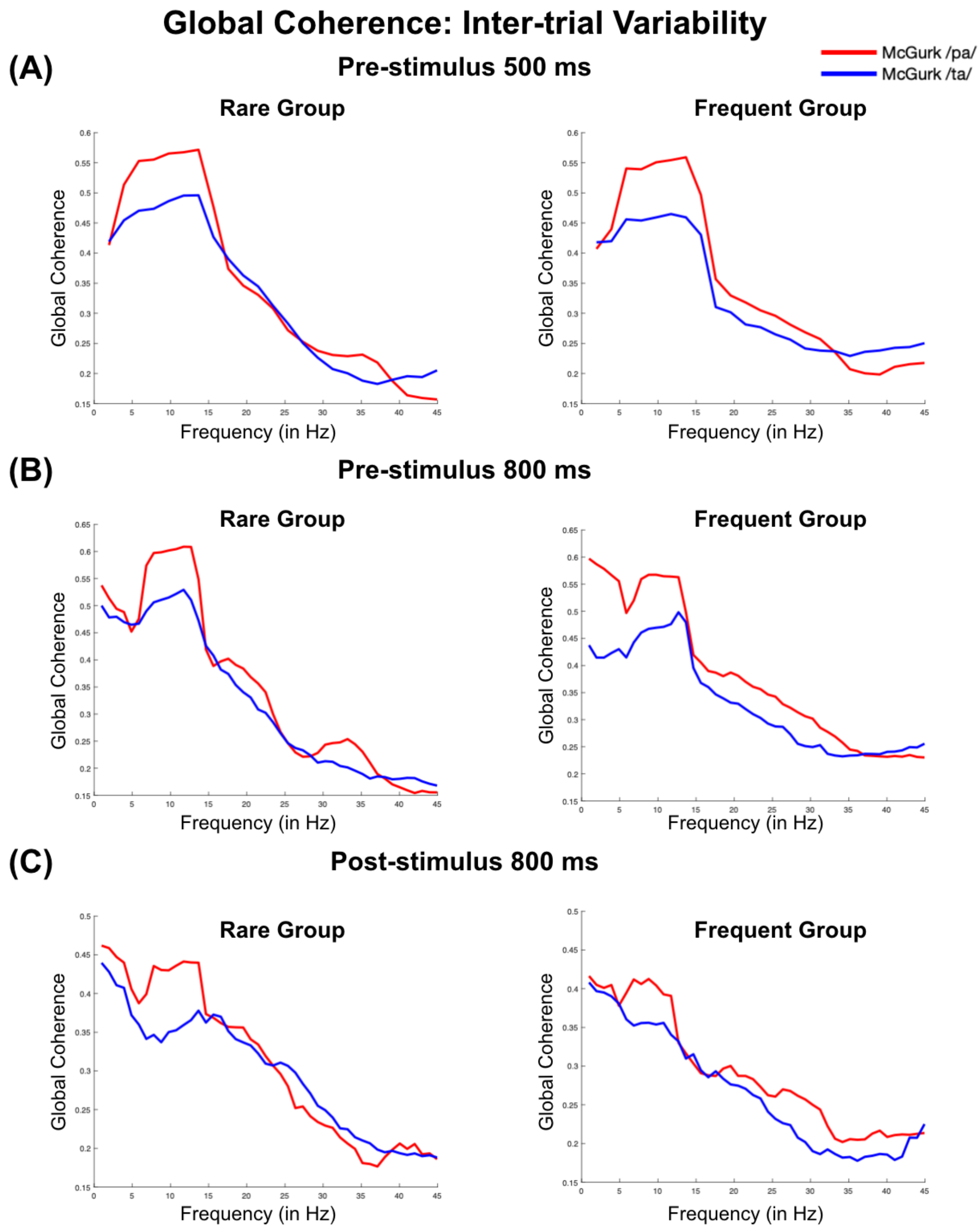


Figure 3.6: ***Global Coherence:*** *Inter-trial variability - Time averaged global coherence of trials during /ta/ (illusory) and /pa/ (unisensory) percept averaged across all the participants of the respective groups (rare and frequent) of (A) Pre-stimulus 500 ms, (B) Pre-stimulus 800 ms, and (C) Post-stimulus 800 ms epoch data.*

*/ta/:* $t\,(28) = -6.44, p < 0.001$; Cong */ka/:* $t\,(28) = -6.84, p < 0.001$) was observed (Figure 3.7A). In the case of the post-stimulus 0.8s (800 ms) epoch data's periodic component, we observed that for the all the stimulus conditions, the rare perceivers were characterized by a significant decrease in the gamma band (McGurk: $t\,(28) = -5.89, p < 0.001$; Cong */pa/:* $t\,(28) = -5.32, p < 0.001$; Cong */ta/:* $t\,(28) = -4.56, p < 0.001$; Cong */ka/:* $t\,(28) = -4.23, p < 0.001$) (Figure 3.7B).

Furthermore, the aperiodic component of the pre-stimulus 0.8s (800 ms) epochs was also statistically analyzed. We observed that for both the incongruent and congruent stimuli, the rare perceivers were characterized by a significant decrease in the beta (McGurk: $t\,(30) = -2.85, p = 0.0078$; Cong */pa/:* $t\,(30) = -2.69, p = 0.011$; Cong */ta/:* $t\,(30) = -2.27, p = 0.03$; Cong */ka/:* $t\,(30) = -2.45, p = 0.02$) and gamma (McGurk: $t\,(28) = -7.61, p < 0.001$; Cong */pa/:* $t\,(28) = -7.27, p < 0.001$; Cong */ta/:* $t\,(28) = -6.47, p < 0.001$; Cong */ka/:* $t\,(28) = -6.87, p < 0.001$) bands (Figure 3.8A). Finally, in the case of post-stimulus 0.8s (800 ms) epochs, we observed that for McGurk and Congruent */pa/* stimulus conditions, the rare perceivers elicited a significant decrease in the beta (McGurk: $t\,(30) = -2.43, p = 0.021$; Cong */pa/:* $t\,(30) = -2.11, p = 0.043$) and gamma (McGurk: $t\,(28) = -5.89, p < 0.001$; Cong */pa/:* $t\,(28) = -5.32, p < 0.001$) bands, whereas for Congruent */ta/* and */ka/* stimuli, a significant decrease in the gamma (Cong */ta/:* $t\,(28) = -4.56, p < 0.001$; Cong */ka/:* $t\,(28) = -4.29, p < 0.001$) band was observed (Figure 3.8B).

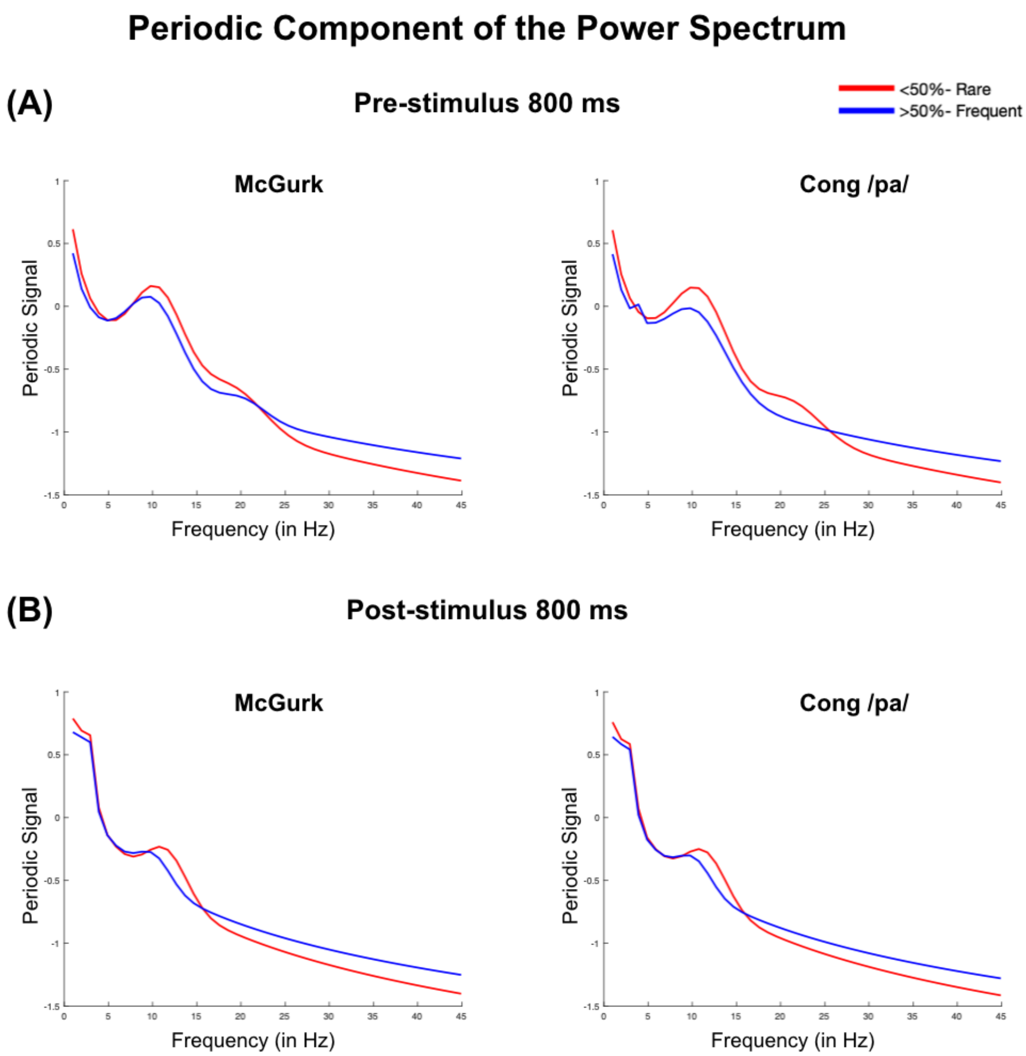


Figure 3.7: ***Periodic Oscillations:*** *Inter-individual variability - FOOOF analysis of the power spectrum data to extract the periodic components during McGurk and Congruent /pa/ stimulus conditions of (A) Pre-stimulus 0.8s (800 ms) and (B) Post-stimulus 0.8s (800 ms).*

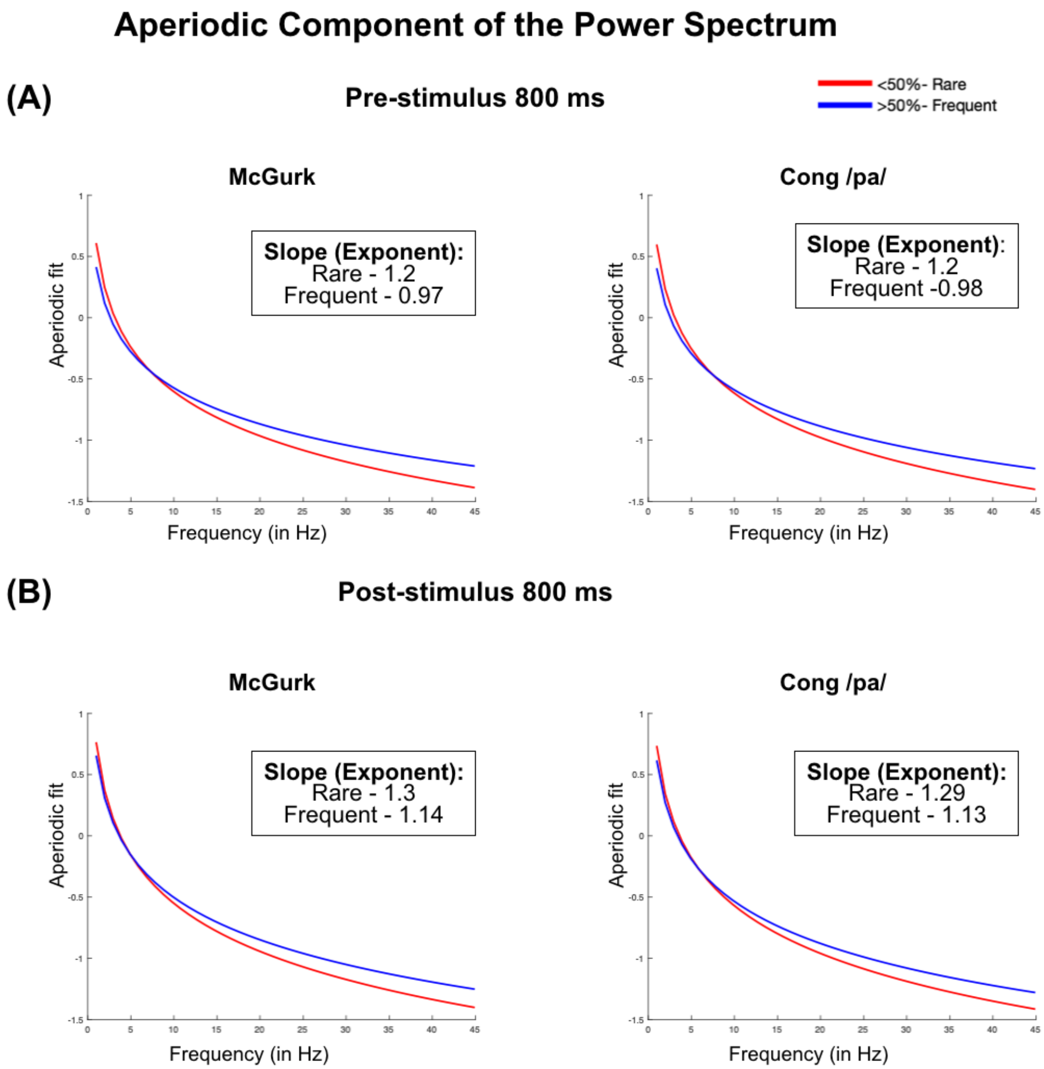


Figure 3.8: ***Aperiodic fit:*** *Inter-individual variability - FOOOF analysis of the power spectrum data to extract the aperiodic components during McGurk and Congruent /pa/ stimulus conditions of (A) Pre-stimulus 0.8s (800 ms) and (B) Post-stimulus 0.8s (800 ms).*

# Chapter 4

# Discussions

In the present study, we analyzed the pre-stimulus EEG data of the participants that were subjected to the incongruent McGurk and the congruent (*/pa/*, */ta/*, and */ka/*) stimuli. Based on the percentage of illusory */ta/* percept, the participants were categorized into two groups: rare (<50% percept) and frequent (>50% percept) [27]. We employed three distinct computational techniques to extract and quantify the differences in the neural oscillations between the two groups at the pre-stimulus and post-stimulus region. Along with this, the behavioral data was also analyzed. The main findings of this investigation are as follows:

1. The participants in the rare group subjected to the incongruent McGurk stimulus (audio - */pa/* and video - */ka/*) responded to the illusory */ta/* percept far less than the frequent group as previously reported by Kumar *et al.,* [27], indicating on the fact that frequent perceivers of the McGurk effect favor the visual input far more than the auditory input. This reasoning is corroborated by a recent study by Beauchamp and colleagues where they have reported that there is a significant correlation between McGurk frequency and mouth looking time [36].

2. In both the 0.5s and 0.8s pre-stimulus duration, we not only see an enhanced beta-

band activity as mentioned by Keil and colleagues [23] but rather a whole spectrum of activity in theta, alpha, and gamma bands for both the power and coherence values of the rare and frequent group of perceivers. Our study suggests that rather than looking into a tiny portion within the pre-stimulus duration to predict the perception of the McGurk illusion in an individual, it is quite important to look at the entire pre-stimulus range. The significant frequency differences in coherence values also suggest that there is, in fact, a myriad of regions along with the left Superior Temporal Gyrus (lSTG)[1] [23] responsible for the AV- integration. In short, this study questions the robustness of the prestimulus beta-band activity reported by Julian Keil and colleagues.

The pre-stimulus duration closer to the preceding stimulus (of 0.5s duration: -1.2s to -0.7s) showed an enhanced power in the theta and alpha bands in rare group of perceivers for both the incongruent and congruent stimuli. Keil and others have found reduced frontal theta band power following the McGurk illusion, indicating that an increase in theta-band power may correlate with the reduced number of McGurk perception [23]. Moreover, during multisensory integration, changes in the local alpha-band power may reflect shifting in spatial attention to either a sensory modality or towards a specific stimulus feature. Keller and others have recently reported that alpha-band power is crucial for selective attention in task-specific information, whereas theta-band power modulations are likely a marker of multisensory divided attention, independent of task difficulty [37]. Our study on the pre-stimulus 0.5s epochs indicates that perhaps the increase in theta-band for the rare group might be linked to the reduced perception of the McGurk effect. And, the alpha is related to the attention domain suggesting that maybe the participants of the rare group had their attention channeled towards the auditory information. The frequency trend was similar for the functional connectivity (global coherence) of the rare group indicating that

[1]Plays an important role in the multi-modal integration.

the pre-stimulus of 0.5s duration's coherence values can be used as a neuronal marker of inter-individual differences or trial-specific perception.

The pre-stimulus duration closer to the succeeding stimulus (of 0.8s duration: -1.2s to -0.4s) showed an enhanced power in the beta band in rare group of perceivers for both incongruent and congruent stimuli. Our results partially coincide with the work of Julian Keil and the group on the pre-stimulus activity. They found increased beta-band power in the lSTG, precuneus, and right frontal cortex prior to an AV integration [23]. Our findings suggest that for both the rare and frequent group of perceivers, the lSTG plays an important role in the multi-modal integration, however, the integration might be happening much more strongly in the rare group as compared to the frequent group. Moreover, we observed an enhanced global coherence in the theta, alpha, and beta bands for the rare group of perceivers which indicates that plausibly the theta and alpha band communications are arising from regions other than lSTG, precuneus, right frontal cortex, and in the case of the rare group, these regions are strongly communicating with the lSTG. However, the source localization of these brain regions is currently out of the scope of the present study.

The post-stimulus epochs (0.8s duration) showed an enhanced power in the alpha band in the rare group of perceivers for both incongruent and congruent stimuli. This can be correlated to the shift in the spatial attention domain stated by Keller and the group [37]. However, for the frequent group, we observed an enhanced power in the theta and gamma bands for all the stimulus conditions. The increase in the theta band after stimulus presentation for the frequent group in our findings may indicate on a recent study by Fernández and colleagues stating that an increase in the power of central theta-band oscillations in response to a McGurk stimulus, may be linked to a general-purpose conflict detection mechanism [38]. On the other hand, the role of increased gamma-band for the frequent group may be explained by the work of Rhone and colleagues. They investigated the processing stages of AV integration and found that gamma-band power in the STG (a site known for multimodal integration) was enhanced following meaningful AV speech stimuli [39]. Our study on the post-stimulus epochs indicates that the gamma-band oscillations might be arising from regions responsible for differentiating speech with non-speech sounds in general, and also helps in integrating the speech sound with the visual input. A higher gamma-band power of frequent group might suggest that this differentiation and integration of AV speech input happens by engaging more neuronal regions for the frequent group as compared to the rare perceivers. However, the source analysis for these neuronal oscillations is currently beyond the scope of the study. Furthermore, for the post-stimulus epochs, we observed an enhanced global coherence in the theta, alpha, and beta bands of the rare group of perceivers which indicates that the inter-individual variability between the two groups of perceivers still persists after the stimulus is presented.

Furthermore, inter-trial variability based on the perceptual categories (illusory response */ta/* and unisensory response */pa/* to the McGurk stimulus) at spectral and network-level was explored in the present study. We wanted to understand the change in the neuronal oscillations on a trial-by-trial basis in both the rare and frequent group of McGurk perceivers. Overall the results of pre-stimulus and post-stimulus epochs highlight the robustness of these oscillations as a reliable marker of AV integration as the trial-wise variability is related to the behavior itself.

Finally, FOOOF analysis was performed on the spectral data to extract periodic and aperiodic components of the electrophysiological signal. Several studies on the periodic oscillations have linked it to numerous cognitive, perceptual, and behavioral states. Whereas for the aperiodic "background" 1/f component of the neural power, it is said to be dynamic and is linked to the relative excitation/inhibition of the underlying neuronal population (the background neural noise) [33]. This suggests that the spectral power one measures to determine relative amounts of given frequencies in the waveform does not necessarily imply oscillatory power. In fact, the changes in power ratios between bands may coincide with the aperiodic slope differences rather than a change in true oscillatory power. Therefore, it is necessary that the four features (frequency, power, aperiodic broadband offset, and aperiodic slope) are carefully parameterized to avoid coalescing them with one another. The FOOOF analysis on our spectral data revealed a difference in the enhanced oscillatory frequencies as compared to the original power spectrum data between the rare and the frequent group implying the importance of separating the aperiodic 1/f component before analyzing the spectral value for the cognitive deduction.

# Chapter 5

# Conclusions

In the present study, we have demonstrated that the pre-stimulus neuronal activity might carry spontaneous meaningful brain rhythm signatures. These neuronal signatures might help us in differentiating the rare group of McGurk perceivers from that of the frequent group even in the absence of any McGurk stimulus. We employed both the power spectrum and global coherence methods to determine inter-individual variability and inter-trial variability between the two groups. At the spectral level, we show that there is indeed a significant difference in the oscillatory power between the two groups suggesting that neuronal oscillations in different frequency bands might reflect distinct mechanisms in multisensory processing in the rare and frequent perceivers. We also performed global coherence to understand the difference in the large-scale functional connectivity that facilitates AV speech perception between the two groups. We observed that global coherence acts as a robust functional connectivity marker and thus can be used to differentiate perceivers based on their pre-stimulus coherence values. Overall, this study suggests that perception, in general, is a subjective experience and whether an individual will perceive a stimulus or not greatly depends on the underlying strength of the neuronal connections in the brain regions responsible for the perception.

## Future Directions

Our study indicates that different brain regions are involved simultaneously during speech perception. Therefore, an immediate extension of this study can be to employ source-analysis to identify the possible cortical generators of the power spectrum and the global coherence dynamics that characterize inter-individual variability in the perception of McGurk and congruent stimuli. Also, an enhancement to the observed global coherence in different frequency bands can include exploring phase-amplitude and phase-frequency coupling that may underlie the cognitive processes. Finally, designing a modular paradigm that can help us in understanding the workings of the brain and cognition.

# Bibliography


[1] Macdonald J, McGurk H (1978) Visual influences on speech perception processes. Perception & Psychophysics 24:253–257.

[2] W. H. Sumby and I. Pollack, "Visual contribution to speech intelligibility in noise," Journal of Acoustical Society of America, vol. 26, no. 2, pp. 212–215, 1954.

[3] A. Q. Summerfield, "Use of visual information for phonetic perception," Phonetica, vol. 36, no. 4-5, pp. 314–331, 1979.

[4] Kuhl PK, Meltzoff AN (1996) Infant vocalizations in response to speech: Vocal imitation and developmental change. The Journal of the Acoustical Society of America 100:2425–2438.

[5] Burr, D., & Gori, M. (2012). Multisensory integration develops late in humans. In The neural bases of multisensory processes. CRC Press/Taylor & Francis.

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

[7] Dodd, B. (1979). Lip reading in infants: Attention to speech presented in-and out-of-synchrony. Cognitive psychology, 11(4), 478-484.

[8] Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face-voice relations. Developmental psychology, 41(3), 541.

[9] Brookes, H., Slater, A., Quinn, P. C., Lewkowicz, D. J., Hayes, R., & Brown, E. (2001). Three-month-old infants learn arbitrary auditory–visual pairings between voices and faces. Infant and Child Development: An International Journal of Research and Practice, 10(1-2), 75-82.

[10] Bahrick, L. E., Netto, D., & Hernandez-Keif, M. (1998). Intermodal perception of adult and child faces and voices by infants. Child development, 69(5), 1263-1275.

[11] Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. Perception & Psychophysics, 59(3), 347-357.

[12] Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (2008). Hearing faces: how the infant brain matches the face it sees with the speech it hears. Journal of cognitive neuroscience, 21(5), 905-921.

[13] Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. Science, 316(5828), 1159-1159.

[14] Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. Developmental psychology, 46(1), 66.

[15] Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. Journal of experimental child psychology, 41(1), 93-113.

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

[17] Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21(1), 1-36.

[18] Fowler, C. A. (1996). Listeners do hear sounds, not tongues. The Journal of the Acoustical Society of America, 99(3), 1730-1741.

[19] Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. The Journal of the Acoustical Society of America, 64(5), 1358-1368.

[20] Ferrand CT (2013) Speech Science: An Integrated Approach to Theory and Clinical Practice. Pearson Education.

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

[22] Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: the McGurk effect revisited. Multisensory research, 31(1-2), 111-144.

[23] Keil, J., Müller, N., Ihssen, N., & Weisz, N. (2012). On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. Cerebral Cortex, 22(1), 221-231.

[24] Zhang, J., Meng, Y., He, J., Xiang, Y., Wu, C., Wang, S., & Yuan, Z. (2019). McGurk effect by individuals with autism spectrum disorder and typically developing controls: a systematic review and meta-analysis. Journal of autism and developmental disorders, 49(1), 34-43.

[25] Knowland, V. C., Evans, S., Snell, C., & Rosen, S. (2016). Visual speech perception in children with language learning impairments. Journal of Speech, Language, and Hearing Research, 59(1), 1-14.

[26] Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., ... & Kikinzon, L. (2009). Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. Comprehensive psychiatry, 50(2), 186-192.

[27] Kumar, G. V., Dutta, S., Talwar, S., Roy, D., & Banerjee, A. (2020). Biophysical mechanisms governing large-scale brain network dynamics underlying individual-specific variability of perception. European Journal of Neuroscience.

[28] Brunner, C., Delorme, A., & Makeig, S. (2013). Eeglab–an open source matlab toolbox for electrophysiological research. Biomedical Engineering/Biomedizinische Technik, 1(aop).

[29] Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S., & Mitra, P. P. (2010). Chronux: a platform for analyzing neural signals. Journal of neuroscience methods, 192(1), 146-151.

[30] Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. Journal of Experimental Psychology: Human Perception and Performance, 22(2), 318.

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

[32] Maris, E., Schoffelen, J. M., & Fries, P. (2007). Nonparametric statistical testing of coherence differences. Journal of neuroscience methods, 163(1), 161-175.

[33] Haller, M., Donoghue, T., Peterson, E., Varma, P., Sebastian, P., Gao, R., & Voytek, B. (2018). Parameterizing neural power spectra. BioRxiv, 299859.

[34] Kumar, G. V., Kumar, N., Roy, D., & Banerjee, A. (2018). Segregation and integration of cortical information processing underlying cross-modal perception. Multisensory research, 31(5), 481-500.

[35] Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. Biophysical journal, 12(1), 1-24.

[36] Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. Attention, Perception, & Psychophysics, 77(4), 1333-1341.

[37] Keller, A. S., Payne, L., & Sekuler, R. (2017). Characterizing the roles of alpha and theta oscillations in multisensory attention. Neuropsychologia, 99, 48-63.

[38] Morís Fernández, L., Torralba, M., & Soto-Faraco, S. (2018). Theta oscillations reflect conflict processing in the perception of the McGurk illusion. European Journal of Neuroscience, 48(7), 2630-2641.

[39] Rhone, A. E., Nourski, K. V., Oya, H., Kawasaki, H., Howard III, M. A., & McMurray, B. (2016). Can you hear me yet? An intracranial investigation of speech and non-speech audiovisual interactions in human cortex. Language, cognition and neuroscience, 31(2), 284-302.

[40] Kumar, G. V., Halder, T., Jaiswal, A. K., Mukherjee, A., Roy, D., & Banerjee, A. (2016). Large scale functional brain networks underlying temporal integration of audio-visual speech perception: An EEG study. Frontiers in psychology, 7, 1558.

# Supplementary Tables

## Power Spectrum: Inter-Individual Variability

### (A) Pre-stimulus 500 ms

| Stimuli (500 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *Incongruent McGurk* | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' | 'Significantly lower' |
| *Congruent /pa/* | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' | 'Significantly lower' |
| *Congruent /ta/* | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' | 'Significantly lower' |
| *Congruent /ka/* | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' | 'Significantly lower' |

### (B) Pre-stimulus 800 ms

| Stimuli (800 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *Incongruent McGurk* | 'Significantly lower' | 'Significantly lower' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /pa/* | 'Significantly higher' | 'Significantly lower' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /ta/* | 'Insignificant' | 'Significantly lower' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /ka/* | 'Significantly lower' | 'Significantly lower' | 'Significantly higher' | 'Significantly lower' |

### (C) Post-stimulus 800 ms

| Stimuli (800 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *Incongruent McGurk* | 'Significantly lower' | 'Significantly higher' | 'Insignificant' | 'Significantly lower' |
| *Congruent /pa/* | 'Significantly lower' | 'Significantly higher' | 'Insignificant' | 'Significantly lower' |
| *Congruent /ta/* | 'Significantly lower' | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /ka/* | 'Significantly lower' | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' |

Table 5.1: *Table summarizing the significant inter-individual power differences between the two groups of perceivers: rare and frequent of (A) Pre-stimulus 500 ms, (B) Pre-stimulus 800 ms, and (C) Post-stimulus 800 ms epoch data.*

## Power Spectrum: Inter-Trial Variability

### (A) Pre-stimulus 500 ms

| Stimuli (500 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *McGurk-Rare Group* | 'Significantly higher' | 'Insignificant' | 'Insignificant' | 'Insignificant' |
| *McGurk-Frequent Group* | 'Significantly higher' | 'Significantly higher' | 'Insignificant' | 'Insignificant' |

### (B) Pre-stimulus 800 ms

| Stimuli (800 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *McGurk-Rare Group* | 'Significantly higher' | 'Significantly higher' | 'Significantly higher' | 'Insignificant' |
| *McGurk-Frequent Group* | 'Significantly higher' | 'Insignificant' | 'Insignificant' | 'Insignificant' |

### (C) Post-stimulus 800 ms

| Stimuli (800 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *McGurk-Rare Group* | 'Insignificant' | 'Significantly higher' | 'Insignificant' | 'Insignificant' |
| *McGurk-Frequent Group* | 'Insignificant' | 'Insignificant' | 'Insignificant' | 'Insignificant' |

Table 5.2: *Table summarizing the power of trials during /ta/ (illusory) and /pa/ (unisensory) percept averaged across all the participants of the respective groups (rare and frequent) of (A) Pre-stimulus 500 ms, (B) Pre-stimulus 800 ms, and (C) Post-stimulus 800 ms epoch data.*

## Global Coherence: Inter-Individual Variability

### (A) Pre-stimulus 500 ms

| Stimuli (500 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *Incongruent McGurk* | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' | 'Significantly lower' |
| *Congruent /pa/* | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' | 'Significantly lower' |
| *Congruent /ta/* | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' | 'Significantly lower' |
| *Congruent /ka/* | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' | 'Significantly lower' |

### (B) Pre-stimulus 800 ms

| Stimuli (800 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *Incongruent McGurk* | 'Significantly higher' | 'Insignificant' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /pa/* | 'Significantly higher' | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /ta/* | 'Significantly higher' | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /ka/* | 'Significantly higher' | 'Significantly higher' | 'Insignificant' | 'Significantly lower' |

### (C) Post-stimulus 800 ms

| Stimuli (800 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *Incongruent McGurk* | 'Significantly higher' | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /pa/* | 'Insignificant' | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /ta/* | 'Significantly higher' | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' |
| *Congruent /ka/* | 'Significantly higher' | 'Significantly higher' | 'Significantly higher' | 'Significantly lower' |

Table 5.3: *Table summarizing the significant inter-individual time-averaged global coherence differences between the two groups of perceivers: rare and frequent of (A) Pre-stimulus 500 ms, (B) Pre-stimulus 800 ms, and (C) Post-stimulus 800 ms epoch data.*

## Global Coherence: Inter-Trial Variability

### (A) Pre-stimulus 500 ms

| Stimuli (500 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *McGurk-Rare Group* | 'Significantly higher' | 'Significantly higher' | 'Insignificant' | 'Insignificant' |
| *McGurk-Frequent Group* | 'Insignificant' | 'Insignificant' | 'Insignificant' | 'Insignificant' |

### (B) Pre-stimulus 800 ms

| Stimuli (800 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *McGurk-Rare Group* | 'Significantly higher' | 'Significantly higher' | 'Insignificant' | 'Insignificant' |
| *McGurk-Frequent Group* | 'Insignificant' | 'Insignificant' | 'Significantly higher' | 'Significantly higher' |

### (C) Post-stimulus 800 ms

| Stimuli (800 ms) | Theta (4-7 Hz) | Alpha (8-12 Hz) | Beta (15-30 Hz) | Gamma (31-45 Hz) |
|---|---|---|---|---|
| *McGurk-Rare Group* | 'Insignificant' | 'Insignificant' | 'Insignificant' | 'Insignificant' |
| *McGurk-Frequent Group* | 'Significantly higher' | 'Significantly higher' | 'Significantly higher' | 'Significantly higher' |

Table 5.4: *Table summarizing the time-averaged global coherence of trials during /ta/ (illusory) and /pa/ (unisensory) percept averaged across all the participants of the respective groups (rare and frequent) of (A) Pre-stimulus 500 ms, (B) Pre-stimulus 800 ms, and (C) Post-stimulus 800 ms epoch data.*