

1 **Neurodynamic explanation of inter-individual and inter-trial variability in** 2 **cross-modal perception**

3 G. Vinodh Kumar^{*}, Shrey Dutta^{*}, Siddharth Talwar, Dipanjan Roy[§], Arpan Banerjee[§]

4 *Cognitive Brain Dynamics Lab, National Brain Research Centre,*

5 *NH-8, Manesar, Gurgaon-122051, Haryana, India*

6 ^{*} Authors have equal contribution

7 [§] Joint corresponding authors: dipanjan.nbrc@gov.in (Dr Dipanjan Roy); arpan@nbrc.ac.in

8 (Dr Arpan Banerjee)

9

10 **Abstract**

11 A widely used experimental design in multisensory integration is the McGurk paradigm that
12 entail illusory (cross-modal) perception of speech sounds when presented with incongruent
13 audio-visual (AV) stimuli. However, the distribution of responses across trials and
14 individuals is heterogeneous and not necessarily everyone in a given group of individuals
15 perceives the effect. Nonetheless, existing studies in the field primarily focus on addressing
16 the correlation between subjective behavior and cortical activations to reveal the neuronal
17 mechanisms underlying the perception of McGurk effect, typically in the “frequent
18 perceivers”. Additionally, a solely neuroimaging approach does not provide mechanistic
19 explanation for the observed inter-trial or inter-individual heterogeneity. In the current study
20 we employ high density electroencephalogram (EEG) recordings in a group of 25 human
21 subjects that allow us to distinguish “frequent perceivers” from “rare perceivers” using
22 behavioral responses as well as from the perspective of large-scale brain functional
23 connectivity (FC). Using global coherence as a measure of large-scale FC, we find that alpha
24 band coherence, a distinctive feature in frequent perceivers is absent in the rare perceivers.
25 Secondly, a decrease in alpha band coherence and increase in gamma band coherence occur
26 during illusory perception trials in both frequent and rare perceivers. Source analysis
27 followed up with source time series reconstructions reveals a large scale network of brain
28 areas involving frontal, temporal and parietal areas that are involved in network level
29 processing of cross-modal perception. Finally, we demonstrate that how a biophysically
30 realistic computational model representing the interaction among key neuronal systems
31 (visual, auditory and multisensory cortical regions) can explain the empirical observations.
32 Each system involves a group of excitatory and inhibitory Hindmarsh Rose neurons that are
33 coupled amongst each other. Large-scale FC between areas is conceptualized using coupling

34 functions and the identity of a specific system, e.g., visual/ auditory/ multisensory is chosen
35 using empirical estimates of the time-scale of information processing in these systems. The
36 model predicts that the disappearance of alpha band coherence observed in rare perceivers
37 stems from a negligible direct A-V (audio-visual) coupling however, an increase in indirect
38 interaction via multisensory node leads to enhanced gamma band and reduced alpha band
39 coherences observed during illusory perception. Overall, we establish the mechanistic basis
40 of large-scale FC patterns underlying cross-modal perception.

42 **Introduction**

43 Speech perception during face-to-face conversation inextricably involves multisensory
44 integration of auditory and visual cues. This is nicely demonstrated in laboratory settings by
45 the McGurk effect (McGurk & Macdonald, 1976), in which the video stimulus of a human
46 speaker with the sound of /ba/ superimposed on the lip movements /ga/ is perceived by the
47 listener as a completely different syllable /da/ (illusory/ cross-modal percept). Subsequently
48 several studies have identified the psychophysical parameters that play a dominant role in
49 eliciting cross-modal effects (Munhall et. al., 1996; van Wassenhove et. al., 2007, Thakur et
50 al 2016) and their underlying neural mechanisms (Jones & Callan, 2003; Kaiser, 2004; van
51 Wassenhove et. al., 2005; Saint-Amour et. al., 2007; Beauchamp, 2010; Keil et. al., 2012;
52 Kumar et al 2017). Nonetheless, the distribution of responses to McGurk stimulus is
53 heterogeneous and some individuals rarely perceive the illusion (Nath & Beauchamp, 2012a).
54 While the neural correlates underlying illusory/ cross-modal perception has been extensively
55 studied in a group of McGurk perceivers, the neurophysiology subserving the perceptual
56 heterogeneity as well as the brain network mechanisms across individuals remains unclear.

57

58 Recent evidences show that subject-wise variability in the illusory perception is contingent on
59 the McGurk stimulus and the response choice employed in the experimental paradigm
60 (Mallick et. al., 2015). Concurrently, neuroimaging evidences attribute the heterogeneity
61 across individuals to the extent of activation at the superior temporal sulcus (STS)
62 (Beauchamp, 2010; Nath & Beauchamp, 2012b). Neurophysiological studies highlight the
63 pre-stimulus activity in STS and its functional connectedness to front-parietal regions as a
64 neuromarker of illusory perception within a group of individuals (Keil et al., 2012). More
65 recent studies have indicated that beyond a specific region of interest, a large-scale network
66 of oscillatory brain networks are involved in effectuating cross-modal perception(Kumar et
67 al., 2016). A key question emerges how robust is this network across a group of individuals
68 and whether the organization of these networks contingent on the stimulus configurations or
69 the perceptual outcome, specifically in the case of McGurk incongruent stimulus. Secondly,
70 what are the neural mechanisms that give rise to the network level correlates? While the first
71 question needs to be answered empirically using a detailed neurophysiological study of
72 underlying brain networks, the more broader question of systems-level understanding or
73 functional brain network organization require a neurobiologically inspired computational
74 model. The existing models of multisensory integration are either motivated from the context

75 of response choices and probabilistic distribution of stimulus cues in the environment
76 (Körding et al., 2007) or explanation of behavior from neurally inspired models (Thakur et
77 al., 2016; Cuppini et al., 2017). Typically these models attempt to explain the firing rate
78 dynamics of single neurons or the local population using a combination of synaptic and
79 stimuli inspired parameters. Thus, the explanation of neurophysiological findings observed at
80 the macroscopic scale of EEG and MEG remains elusive because of the dearth of a network
81 model that captures the large-scale network dynamics.

82

83 In the current study we use the psychophysical variable of audio-visual (AV) lag that can
84 modulate the degree of illusory perceptual experience in a group of individuals. We estimate
85 the large-scale network underlying illusory perceptual experience in a group of individuals
86 who frequently perceive McGurk illusion as well as investigate the functional network
87 reorganization in individuals who rarely perceive the McGurk illusion. We find two distinct
88 large-scale mechanisms operation during the multisensory information processing: 1)
89 increase in gamma band global coherence and decrease in alpha band global coherence
90 during illusory perception trials in both frequent and rare perceivers and 2) absence of peak in
91 alpha band coherence across both illusory and unimodal perception trials in rare perceivers.
92 Both these mechanisms were validated at the sensor level data and from source connectivity
93 analysis using the LCMV beamformer (Van Veen et al., 1997). Subsequently, we designed a
94 neural mass model that captures the global coherence dynamics observed in the EEG data.
95 Previous studies have argued this kind of modeling is ideally suited to explain the emergence
96 of spontaneous rhythmic patterns in EEG (Becker et al., 2015). Here we illustrate that a large-
97 scale model of multisensory interactions involving distinct local neuronal populations e.g.,
98 unisensory areas (Heschl's gyrus/ STG and higher visual areas) and multisensory
99 convergence zones (STS) can generate the synchronization patterns in sensor and source
100 dynamics. Each local population consists of excitatory and inhibitory neural populations that
101 are interconnected using biophysically observed parameters and each neuron within the
102 population are capable of generating periodic spiking and bursting dynamics. Finally, we
103 could illustrate how direct auditory-visual coupling whose presence was reported in
104 neuroanatomical studies (Falchier et al., 2002; Rockland & Ojima, 2003; Wallace et al.,
105 2004) and indirect interactions between audio-visual areas via multisensory convergence sites
106 (Bizley & King, 2012) can bring forth distinct network mechanisms to facilitate perceptual
107 experience.

108 Results

109 *Inter-subject variability: Frequent and rare perceivers of illusory McGurk perception*

110 We employed the incongruent McGurk stimulus, visual /ka/ paired with auditory /pa/ to
111 induce the illusory response /ta/. Overall, we used four kinds of AV stimuli: three McGurk
112 incongruent pair with AV lags -450 ms (audio leads the articulation), 0 ms (synchronous),
113 +450 ms (articulation leads the audio) and one congruent AV stimulus (visual /ta/ with
114 auditory /ta/). Following a forced choice paradigm, the participants reported if they heard /ta/,
115 /pa/ or something else (others). Concurrently, the participants' eye gaze behavior was
116 recorded by an infra-red based eye tracking device. We characterized a participant as a
117 'frequent perceiver' (N=15) if they responded with 60% of /ta/ response to the McGurk
118 incongruent stimulus at any lag, -450, 0 or +450 ms, failing which the participants were
119 categorized as a 'rare perceiver' (N=10). **Figure 1B, C** illustrates the distribution of
120 perceptual categorization responses in frequent and rare perceivers to the McGurk
121 incongruent stimuli. At all AV lags, 80% of the rare perceivers reported /ta/ in <45% trials
122 (see **Figure 1 - figure supplement 1**). We ran a repeated-measures two-way ANOVA on the
123 percentage responses with AV lags and perceptual categories (/ta/ and /pa/) as the variables
124 within each group of participants and use $p < 0.05$ to evaluate statistical significance. For
125 frequent perceivers, we observed that AV lags had no influence on the percentage responses,
126 $F(2, 89) = 0.84, p = 0.44$. However, we observed a significant variation of percentage
127 responses between the two perceptual categories, $F(1, 89) = 19.90, p < 0.0001$. Also, the
128 interaction between perceptual categories and AV lags was significant, $F(2, 89) = 29.83, p <$
129 0.0001 . For rare perceivers, no influence of AV lags was observed, $F(2, 59) = 0.27, p = 0.76$.
130 However, variation of percentage responses between the two perceptual categories was
131 significant, $F(1, 59) = 64.47, p < 0.0001$. Also, no significant interaction was observed
132 between the perceptual categories and AV lags, $F(2, 59) = 0.47, p = 0.66$. We also performed
133 paired Student's t-test on the percentage of responses (/ta/ and /pa/) at each AV lag for
134 frequent and rare perceivers and use statistical threshold of $p = 0.05$ to evaluate significance.
135 In frequent perceivers, we find significantly higher percentage of /ta/ responses at 0 ms (t
136 (14) = 7.81, $p < 0.0001$) and +450 ms AV lag (t (14) = 2.12, $p = 0.04$). No significant
137 difference was observed at -450 ms (t (14) = 1.97, $p = 0.06$) AV lag. However, in rare
138 perceivers we observed a significantly higher percentage of /pa/ responses at -450 ms (t (9) =
139 -3.62, $p = 0.002$), 0 ms (t (9) = -4.93, $p < 0.0001$) and +450 ms (t (9) = -5.61, $p < 0.0001$) AV

140 lag. Unpaired Student's t test employed to compare the percentage of /ta/ responses during
141 the congruent /ta/ stimulus showed no significant difference between the groups ($t(23) =$
142 $2.02, p = 0.05$) **Figure 1 - figure supplement 2**.

143 Gaze fixations on the head and mouth areas of the speaker in the AV stimuli were converted
144 into percentage measures for each subject on a trial-by trial basis and sorted based on the
145 stimulus type and perceptual categories. The bar graphs in **Figure 1 - figure supplement 3**
146 illustrates the mean and the standard error of the percentage of gaze fixations on the mouth of
147 the articulator during /ta/ and /pa/ perception averaged across the participants. We performed
148 a repeated-measures two-way ANOVA on the percentage responses with AV lags and
149 perceptual categories (/ta/ and /pa/) as the variables in frequent and rare perceivers. In
150 frequent perceivers **Figure 1 - figure supplement 3A**, we observed that there was no
151 influence of AV lags, $F(2, 89) = 0.36, p = 0.70$ and perceptual categories, $F(2, 89) = 3.88, p$
152 $= 0.05$ on the percentage of gaze fixations on the mouth. Furthermore, the interaction effect
153 between them was also insignificant, $F(2, 89) = 0.07, p = 0.93$. Similarly, in rare perceivers
154 **Figure 1 - figure supplement 3B**, AV lags, $F(2, 59) = 2.54, p = 0.09$ and perceptual
155 categories, $F(2, 59) = 0, p = 0.97$ had no effect on the percentage of gaze fixations at the
156 mouth. Also, no evidence of an interaction effect between them was observed, $F(2, 59) =$
157 $0.2, p = 0.82$. We further performed unpaired Student's t-test to compare the percentage of
158 gaze fixations on mouth between frequent and rare perceivers i.e frequent /ta/ vs Rare /ta/ and
159 frequent /pa/ vs Rare /pa/. We observed that frequent perceivers elicited significantly higher
160 percentage of fixations at mouth during /ta/ perception at -450 ms ($t(23) = 3.42, p = 0.002$),
161 0 ms ($t(23) = 3.88, p = 0.0007$) and +450 ms ($t(22) = 2.79, p = 0.01$) AV lag. Similarly,
162 during /pa/ perception frequent perceivers elicited higher percentage of fixations on mouth at
163 -450 ms ($t(23) = 4.56, p < 0.001$), 0 ms ($t(23) = 2.95, p = 0.0071$) and +450 ms ($t(23) =$
164 $2.45, p = 0.02$) AV lag.

165 ***Large-scale functional connectivity dynamics***

166 To investigate the underlying differences in dynamic functional connectivity (FC) between
167 the perceptual categories we computed the global coherogram during /ta/ and /pa/ perception.
168 Global coherogram defined from the normalized vector sum of all pairwise coherences
169 amongst EEG sensors captures the evolution of global coherence in time and frequency
170 domain simultaneously (Lachaux et al., 1999). Mathematically, global coherence is the ratio
171 of the largest eigenvalue of the cross-spectral matrix to the sum of its eigenvalues (Mitra &

172 Bokil, 2008). Subsequently, we compared the global coherogram of /ta/ and /pa/ at AV lags: -
173 450ms (**Figure 2A, C**), 0ms (**Figure 2E, G**) and +450 (**Figure 2I, K**) using cluster based
174 permutation tests. The onset of first stimulus was considered the point of reference for time-
175 locking (zero). Positive clusters highlighted in black dashed rectangles and negative clusters
176 in red dashed boxes signify time-frequency islands of increased and decreased synchrony
177 respectively in the large-scale functional network. We also compared the presence of band-
178 specific peaks/enhancement in global coherence in the frequent (**Figure 2B, F, J**) and rare
179 perceivers (**Figure 2D, H, L**) during /ta/ and /pa/ perception using Silverman's bootstrapping
180 test for examining multimodality.

181

182 For frequent perceivers, during -450 ms AV lag (**Figure 2A**), we observed a negative cluster
183 in the theta ($z_{0.05} = -5.31$) in the temporal range of 650-800 ms and a positive cluster in the
184 beta band ($z_{0.95} = -3.84$) between 500 ms to 700ms. For videos at 0 ms AV lag (**Figure 2E**),
185 we observed a negative cluster in the theta ($z_{0.05} = -6.05$) and alpha band ($z_{0.05} = -5.81$) in the
186 temporal range of ~0-450 ms, and a positive cluster ($z_{0.95} = -4.32$) in gamma band between
187 800 and 900 ms. During +450 ms AV lag (**Figure 2I**), we observed three positive clusters,
188 (1) in the theta band ($z_{0.95} = -5.52$) in the 100-400 ms time window, (2) in the beta band (
189 $z_{0.95} = -4.40$) between 200 to 500 ms, and (3) in the gamma range ($z_{0.95} = -4.34$) in the
190 temporal window of 50 ms and 250 ms. Also, a negative cluster in the alpha band (
191 $z_{0.05} = -5.85$) was observed between 700 to 900 ms time window.

192

193 For rare perceivers, at -450 ms AV lag (**Figure 2C**), we observed two prominent negative
194 clusters ($z_{0.05} = -6.72$) spanning gamma band in the temporal range of 0-400 ms and ~500-
195 900 ms. For videos with 0 ms AV lag (**Figure 2G**), we observed two positive clusters in the
196 beta band ($z_{0.05} = -5.69$) and ($z_{0.05} = -5.68$) between ~0-150ms and ~300-500 ms
197 respectively. At +450 ms AV lag (**Figure 2K**), we observed two positive clusters (
198 $z_{0.95} = -6.16$) and ($z_{0.95} = -6.04$) in the theta band in the temporal window of ~0-300 ms and
199 ~400-700 ms respectively.

200

201 Cluster based permutation tests, performed to test the differences in global coherogram
202 between /ta/ and /pa/ elucidated the neural signatures in large-scale FC corresponding to
203 inter-trial variability observed within frequent and rare perceivers. Consequently, to address

204 if inter-individual heterogeneity stems from the differences in the inherent processing of
205 multisensory stimuli in the two groups of perceivers, we evaluated if any frequency specific
206 enhancement of global coherence occurs during /ta/ and /pa/ perception. Consequently, we
207 computed the global coherence during /ta/ and /pa/ perception for frequent and rare
208 perceivers that would provide a holistic picture of the underlying large-scale FC. In frequent
209 perceivers we observed qualitatively that the global coherence followed a bimodal
210 distribution during /ta/ and /pa/ perceptions across all AV lags (**Figure 2B, F, J**). The modes
211 were primarily centered around alpha (8-13 Hz) and gamma (30-40 Hz) bands, signifying
212 enhanced coherence. Silvermann's bootstrapping test employed to examine the statistical
213 significance of those peaks revealed significant bimodal peaks ($p < 0.05$) during /ta/ and /pa/
214 perception at -450 ms, 0 ms AV lags. However, there were no significant bimodal peaks
215 during /ta/ ($p = 0.07$) and /pa/ perception ($p = 0.13$) at +450 ms AV lag (**Figure 2J**).

216 In rare perceivers, Silvermann's bootstrapping test revealed significant bimodal peaks only
217 during /ta/ perception at 0 ms AV lag ($p < 0.05$) (**Figure 2H**). There were no significant
218 bimodal peaks during /pa/ perception at -450 ms ($p = 0.35$) (**Figure 2D**), 0 ms ($p = 0.30$)
219 (**Figure 2H**) and +450 AV lag ($p = 0.15$) (**Figure 2L**). Similarly, no significant bimodal
220 distribution were observed during /ta/ perception at -450 ms ($p = 0.21$) (**Figure 2D**) and +450
221 ms ($p = 0.20$) AV lags (**Figure 2L**). Importantly, the bimodal peaks during /ta/ perception at
222 0 ms AV lag were clustered around delta (1-4 Hz), theta (4-8 Hz) and gamma (30-40 Hz).
223 Notably, a desynchronization in the alpha band was observed in rare perceivers (**Figure 2D,**
224 **H, L**) across all AV lags and perceptual categories.

225
226 To further understand if these frequency specific coherence differences contingent on the
227 stimulus configurations, we computed the global coherogram and time averaged global
228 coherence during congruent /ta/ in frequent and rare perceivers and compared them using
229 cluster based permutation tests (**Figure 3**). Global coherogram differences in congruent /ta/
230 between frequent and rare perceivers computed employing cluster based permutation tests
231 revealed three negative clusters, (1) in the beta band between ~50-400 ms ($z_{0.05} = -5.87$)
232 temporal window, (2) in the beta band between the time window ~700-900 ms ($z_{0.05} = -5.79$)
233 and (3) in the gamma band from ~150-900 ms ($z_{0.05} = -6.54$). A positive cluster in the delta
234 and theta band was also observed between ~700-900 ms ($z_{0.95} = -6.24$) time window (**Figure**
235 **3C**). Conspicuously, the global coherence distribution for the congruent /ta/ in frequent and
236 rare perceivers followed a similar pattern (**Figure 3D**).

237 *Source analysis reveals cortical areas participating in functional connectivity dynamics*

238 To validate the role of the identified sources in the overall functional connectivity pattern
239 observed in the sensor EEG, we initially identified the cortical generators of the EEG time
240 series by employing linear constrained minimum variance (LCMV) beamformer algorithm
241 (Van Veen et al., 1997). Subsequently, we projected the epoched time series into the source
242 time space by multiplying them with the concordant spatial filter (constructed by LCMV
243 beamformer, for more info. see methods) of the source locations that showed statistical
244 significance in the ratio of source power between /ta/ and /pa/ trials. Finally, we computed
245 the global coherogram for the perceptual categories and compared them using cluster based
246 permutation tests. Elicitation of a similar trend in the global coherogram differences
247 essentially confirms the involvement of the identified sources in the large-scale FC
248 underlying McGurk perception. The sources eliciting statistical significance in the ratio of
249 source power between /ta/ and /pa/ are illustrated in **Figure 4A** and the source locations are
250 listed in **Table 1**. The source locations were consistent across all AV lags and between
251 frequent and rare perceivers. Cluster based permutation tests employed to compare the global
252 coherogram (/ta/ - /pa/) computed from the source time series revealed in frequent perceivers
253 at -450 ms AV lag (**Figure 4B**) one positive cluster in the alpha band ($z_{0.95} = -5.18$) in the
254 temporal range of ~200-700 ms,. During 0 ms AV lag (**Figure 4D**), three negative clusters,
255 two clusters in the alpha band in the temporal window of ~0–100ms ($z_{0.05} = -6.03$), ~600-
256 900ms ($z_{0.05} = -6.05$) and one ($z_{0.05} = -6.93$) in the low gamma band between ~150-350 ms.
257 At +450 ms AV lag (**Figure 4F**), one prominent positive cluster in the high beta and gamma
258 band ($z_{0.95} = -6.65$) spanning the entire stimulus duration, and two negative clusters
259 spanning the theta and alpha band in the time window of ~0-400 ms ($z_{0.05} = -5.57$) and
260 between ~500-650 ms ($z_{0.05} = -5.62$) was observed.

261 For rare perceivers, during -450 ms AV lag (**Figure 4C**), three positive clusters, (1) in the
262 theta and alpha band ($z_{0.95} = -4.56$) from the onset to ~500ms, (2) in the beta band between
263 ~300- 700 ms and (3) a prominent positive cluster ($z_{0.95} = -4.56$) in the gamma band
264 spanning the entire stimulus duration was observed. At 0 ms AV lag (**Figure 4E**), a negative
265 cluster ($z_{0.05} = -5.39$) in the alpha band between from stimulus onset to ~600 ms and a
266 negative cluster ($z_{0.95} = -4.57$) in the beta band in the temporal range of ~600-900 ms was
267 observed. During +450 ms AV lag (**Figure 4G**), three positive clusters, (1) in the theta band

268 and alpha band ($z_{0.95} = -4.60$) between ~0-220 ms, (2) in the beta band ($z_{0.95} = -4.16$)
 269 between ~600-900 ms, and (3) in the gamma band ($z_{0.95} = -4.38$) spanning the entire
 270 stimulus duration was observed.

271 **Table 1:** The table lists the cortical loci that elicited power higher than the set threshold (>
 272 99.5 percentile) in the source analysis

	Left hemisphere	Right hemisphere
Frontal lobe	Inferior frontal gyrus Middle frontal gyrus Superior frontal gyrus Cingulate gyrus	Inferior frontal gyrus Middle frontal gyrus Superior frontal gyrus Cingulate gyrus
Temporal lobe	Fusiform gyrus Middle temporal gyrus Superior temporal gyrus	Fusiform gyrus Middle temporal gyrus Superior temporal gyrus
Parietal Lobe	Precuneus	

273

274 *Network model comprising of 3 neural masses with fast, intermediate and slow time-*
 275 *constants generates alpha and gamma coherence*

276

277 We incorporated a neural mass model approach (Becker et al., 2015; Aerts et al., 2018) to
 278 investigate the alpha and gamma coherence dynamics associated with inter-individual and
 279 inter-trial variability respectively. Since EEG data does not necessarily reflect the local
 280 synaptic activity, neural mass model which operates to phenomenologically explain
 281 mesoscopic and macroscopic features in EEG/ MEG data offers an attractive tool to
 282 understand the underlying neural mechanisms (Lopes da Silva et al., 1974; Jansen & Rit,
 283 1995; David & Friston, 2003). A neural mass is essentially an abstraction of summed
 284 synapto-dendritic activity of several thousand neurons in an area which can be in a
 285 cooperative dynamical state such as synchronous firing that gives rise to low-frequency
 286 oscillations. Such shared dynamical states allow us to reduce the population dynamics in
 287 terms of coupled ordinary differential equations where explicit spatial effects can be ignored
 288 (Stefanescu & Jirsa, 2008). Armed with the knowledge of cortical sources underlying cross-
 289 modal perception (**Table 1**) we consider broadly a network of three neural masses as the
 290 underlying neuro-cognitive network comprising of auditory, visual and cross-modal masses
 291 (nodes). Each node can be further expanded as a population of excitatory and inhibitory
 292 Hindmarsh-Rose (HR) neurons (Hindmarsh & Rose, 1984) representing auditory, visual and

293 multisensory areas. The key parameters that govern the time scale of the oscillatory dynamics
294 come from physiologically motivated parameter values for each neural area. For instance, the
295 auditory node is assumed to be the most sensitive to ambient temporal fluctuations hence
296 operating with a fast time-scale, visual node the slowest in terms of sensitivity and somewhat
297 intermediate time-scale for multisensory node (see materials and methods for details, **Figure**
298 **5**). The existence of two time-scales facilitates the co-existence of synchronous states in alpha
299 and gamma oscillations when slow (visual) node is source of excitatory influence (EI) and
300 fast (auditory) node is sink of EI and when coherence was computed across all nodes. These
301 co-existent states emerge via two possible routes, 1) when visual node (V) interacts with the
302 auditory node (A) through direct coupling (W_{AV}) and 2) when indirect coupling (W_{AM} & W_{VM})
303 between A-V nodes via the multisensory node (M) range from 0.35 to 0.7 (**Figure 6**
304 **Supplement 1A**). We assume coupling strength less than 0.35 to be weak coupling (WC),
305 coupling strength between 0.35 and 0.7 to be moderate coupling (MC) and coupling greater
306 than 0.7 to be strong coupling (SC). We also observe high coherence around alpha band and
307 gamma band in SC range however, a distinct peak around alpha band is not clearly observed.
308 Any other model configuration is not able to create the co-existence of alpha and gamma
309 band coherence in MC range (**Figure 6 Supplement 2**). Further, when the fast-slow
310 interaction takes place via direct coupling alone (W_{AV} ranges from 0 to 1, W_{AM} & $W_{VM} = 0$) we
311 observe the existence of only alpha band coherence but not the gamma band coherence
312 (**Figure 6 Supplement 1B**). Here, the absence of gamma band coherence implies a
313 diminished indirect coupling of A-V nodes via multisensory node (W_{AM} & W_{VM}). Moreover,
314 we observe only gamma band coherence in MC range (**Figure 6 Supplement 1C**) when we
315 restricted the fast-slow (A-V) interactions via multisensory node alone (indirect A-V
316 coupling W_{AM} & W_{VM} range from 0 to 1, $W_{AV} = 0$). This observation clearly links alpha
317 coherence to direct A-V coupling whereas gamma coherence to indirect A-V coupling (A-M-
318 V) of neural masses.

319

320 *Direct Audio-Visual interaction underpins Inter-Individual Variability*

321 Our empirical results suggest that negligible alpha coherence is a hallmark of rare perceivers.
322 Since, direct A-V interaction generates a peak around alpha coherence (**Figure 6**
323 **Supplement 1 A & B**), we hypothesize that lesser amount of direct interaction or even
324 absence of it is associated with de-synchronization of alpha band coherence. To test this

325 hypothesis, we start with a balanced network coupling state, $W_{AV} = W_{AM} = W_{VM} = 0.35$, where
326 alpha and gamma band coherences co-exist (**Figure 6 Supplement 1A**) and study the change
327 in the coherence peaks as direct A-V coupling (W_{AV}) decreases. In **Figure 6A**, we observe a
328 suppression of alpha coherence peak as A-V coupling decreases; however gamma coherence
329 peak remains more or less intact. Further, when A-V coupling becomes negligible (
330 $W_{AV} < 0.05$) we observe disappearance of alpha coherence peak. This suggests that alpha de-
331 synchronization can stem from low direct A-V coupling in rare perceivers.

332

333 *Audio-Visual interaction via Multisensory node underpins Inter-Trial Variability*

334 Broadly speaking, enhanced gamma coherence and decreased alpha coherence is observed
335 unequivocally in frequent perceivers and rare perceivers when illusory and non-illusory trial
336 comparisons were extracted to study the inter-trial variability. Even though rare perceivers
337 exhibited overall lower alpha coherence, the differential decrease in alpha band coherence
338 was clearly observed at sensor and source level (**Figure 2 & 4**). As shown earlier, decrease in
339 direct A-V coupling causes a decrease in alpha band coherence (**Figure 6A**) in rare
340 perceivers and hence decrease in direct A-V coupling cannot be associated with illusory
341 perception. However, gamma band coherence peaks emerge as a coexistent state once
342 indirect A-V interactions via multisensory node are incorporated in the model (**Figure 6**
343 **Supplement 1 A & C**) allowing us to propose a dominant role of interactions between
344 multisensory and unisensory areas modulating cross-modal perception. To test this
345 hypothesis for frequent perceivers we start with a balanced network configuration that
346 generates co-existing alpha band and gamma band coherence ($W_{AV} = W_{AM} = W_{VM} = 0.35$,
347 **Figure 6 Supplement 1A**) and for rare perceivers we choose a network configuration that
348 generates peak only around gamma band ($W_{AV} = 0.05; W_{AM} = W_{VM} = 0.35$, **Figure 6A &**
349 **Figure 6 Supplement 1C**). Then, we track the change in gamma coherence as indirect A-V
350 interaction via multisensory node (W_{AM} & W_{VM}) increases simultaneously in MC range (0.35
351 to 0.7). As hypothesized, we observe an increase in gamma coherence in network
352 configurations for both frequent and rare perceivers. Interestingly, increasing indirect A-V
353 interactions not only increases gamma band coherence but also display a decrease around
354 alpha band coherence in network configurations of frequent as well as rare perceivers even
355 though rare perceivers exhibit overall weaker alpha band coherence (**Figure 6 B & C**). Thus,
356 our model implicates an increase in indirect A-V interaction via multisensory node leading to

357 an increase in gamma band coherence as well as a decrease in alpha band coherence and thus
358 facilitating illusory perception.

359

360 Discussion

361 A vast body of work has used the McGurk paradigm to study cross-modal perception and the
362 numbers are only increasing (Alsius et al., 2018). An ongoing challenge still remaining to the
363 community is accurate identification and characterization of possible neural mechanisms that
364 govern the behavioral variability. For example, why do some people perceive it so strongly,
365 whereas others do not? An approach taken by brain stimulation studies had earlier addressed
366 the issue of inter-individual variability, and identified the candidate brain areas that are
367 probably responsible (Beauchamp, 2010). A more emerging understanding suggest the
368 existence of networks of brain regions facilitating perceptual processing (Bressler & Menon,
369 2010), nonetheless the neurophysiological correlates of inter-individual variability are yet to
370 be understood. In this perspective, a recent review suggests neuronal oscillations as a key
371 substrate of neuronal information processing that needs to be fully explored to answer the
372 individual's perceptual experience (Keil & Senkowski, 2018). It is well known that robust
373 oscillations observed from macroscopic recordings such as EEG/ MEG are an outcome of
374 network interactions among local subpopulations of excitatory and inhibitory neurons
375 (Wilson & Cowan, 1972; Deco et al., 2010; Becker et al., 2015). Empirically such
376 interactions result in global coherence dynamics observed by earlier studies such as Kumar et
377 al (Kumar et al., 2017). In the current study we demonstrate how distinct coherence patterns
378 further become the hallmark of category specific perceptual experience such as the presence
379 of alpha band coherence became a group-labeling attribute for perceptual categorization.
380 Furthermore we find that across trials, the pattern of coherence dynamics determine the trial-
381 specific perceptual outcome. Finally, using computational models of interactive large-scale
382 brain networks, we capture the neural mechanisms through which coherence dynamics evolve
383 in the brain. Put together, we present an attractive mechanistic proposal that underlie the
384 observed inter-individual and inter-trial variability in multisensory speech perception.

385

386 The key empirical observations in our study are: (1) Rare perceivers exhibit a diminished
387 alpha band global coherence, indicating desynchronization of large-scale neural assemblies in
388 the alpha band (2) Both rare and frequent perceivers' cross-modal perception (such as /ta/)
389 involves an enhanced gamma band coherence and decrease in alpha band coherence
390 compared to unimodal perception (such as /pa/). The large-scale neuro-dynamic model of
391 cross-modal perception suggests de-synchrony in the alpha band, characteristic of rare
392 perceivers, is due to extremely weak direct A-V coupling ($W_{AV} < 0.05$). Furthermore, an

393 increase in indirect interaction between auditory and visual systems via multisensory node
394 (increase in W_{AM} & W_{VM}) facilitates high level of synchronization in gamma band and a
395 desynchronization at alpha band. We further elaborate on our empirical and modeling results
396 in the following subsections.

397

398 *Heterogeneous nature of illusory perception*

399 Trial-by-trial variation of perceptual experience within an individual has been previously
400 reported by several studies (Beauchamp, 2010; Keil et al., 2012; Roa Romero et. al., 2015,
401 Kumar et. al. 2016). Behavioral results (**Figure 1B, C**) also indicate that the entire population
402 of volunteers can be distinctly classified in two categorical groups: frequent perceivers and
403 rare perceivers. Similar inter-individual variability were observed and quantified by previous
404 studies (Nath & Beauchamp, 2012; Proverbio et al., 2016). We also presented the McGurk
405 incongruent video (*/pa/-/ka/*) with varying temporal asynchrony, AV lags of ± 450 ms.
406 Perceptual experience of frequent perceivers was modulated as a function of lags, however,
407 no such effect was observed in rare perceivers. The decrease in McGurk perception for
408 ± 450 ms AV lags is consistent with the existing studies (Munhall et al., 1996; van
409 Wassenhove et al., 2007). Also for ± 450 ms AV lagged videos, higher degree of illusory
410 perception was observed in frequent perceivers compared to rare perceivers. Furthermore,
411 irrespective of the perception (*/ta/* or */pa/*) the gaze fixations on the mouth of the articulator
412 were also significantly lower in rare than frequent perceivers. The distinctness in the behavior
413 of rare perceivers pinpoints a difference in the processing of multisensory speech. Therefore,
414 we expected to identify neurophysiological correlates that can characterize a rare perceiver
415 from the frequent perceivers as well as the cross-modal perceptual experience from the
416 unimodal perception that varies trial-by-trial within an individual. Ideally, a single measure
417 that captures these different kinds of heterogeneity, inter-individual and inter-trial can set the
418 ideal platform for discussing about network mechanisms.

419

420 *Neuromarkers of inter-individual and inter-trial variability*

421 Large-scale systems of distributed and interconnected neuronal populations organized to
422 perform specific cognitive tasks are referred to as neurocognitive networks (NCNs) (Bressler
423 & Menon, 2010). Multisensory speech perception that requires the integration of information
424 among spatially distinct sensory systems, components of which are often distributed over the
425 whole brain becomes an ideal candidate to explore from the perspective of NCNs. In

426 physiological signals NCNs can be studied by quantifying the extent of coordination among
427 neuronal assemblies over the whole brain (Bressler, 1995; Bressler & Kelso 2001). The most
428 significant achievement of our study was to capture the network correlates of inter-individual
429 and inter-trial variability with the same measure of global coherence at both sensor level and
430 source level EEG analysis. Our results show that frequent perceivers exhibit enhanced global
431 coherence in the alpha band than rare perceivers. Notably, the enhancement was consistent
432 across all AV lags in frequent perceivers. Previous evidences accentuate the modulations in
433 alpha band coherence to central executive processes (Klimesch, 1999; Sauseng et. al., 2005)
434 that are postulated to be involved in allocating working memory storage to phonological loop
435 that maintains verbal information, and the visuo-spatial sketchpad that maintains transient
436 visuo-spatial information (Baddeley, 1992). Therefore, we posit that the enhanced global
437 coherence in alpha band as a marker that characterizes the presence of specific NCN level
438 processing in frequent perceivers which is absent in rare perceivers.

439

440 Recent study by Fernández and colleagues demonstrates an increase in the power of theta
441 oscillations in response to an incongruent McGurk stimulus accentuating its role in the
442 prediction of the conflict (Fernández et al., 2018). Noticeably, we observed an enhanced
443 global coherence in the theta band in frequent and rare perceivers irrespective of the
444 perceptual experience which indicates even if theta band communication is present in both
445 group of perceivers, it is a not necessarily a marker of inter-individual differences or trial
446 specific perception. In general it is quite possible that different neuro-cognitive processes can
447 be operating simultaneously involving communication at various frequencies via coherence
448 (Senkowski et al., 2008). Hence, it is important to identify which of these are meaningful to
449 the ongoing task and the subtle differences that vary with the context in which the task
450 evolves. In an earlier study Kumar et al. (Kumar et al., 2016) have showed that global
451 coherogram captures the difference in processing of crossmodal (illusory /*ta*/) and unimodal
452 (non-illusory /*pa*/) perception in frequent perceivers from a subset of data that we present in
453 this manuscript. While the detailed pattern of coherogram differences between /*ta*/ and /*pa*/
454 trials in perceivers and rare perceivers are slightly different, there was an enormous similarity
455 in trend of coherence differences in distinct spectro-temporal locations that was conspicuous.
456 For example, both frequent and rare perceivers have enhanced gamma band coherence and
457 diminished alpha band coherence in /*ta*/ trials compared to /*pa*/ trials for temporally
458 synchronous AV stimuli. For asynchronous trials, broadband coherence enhancement in both
459 frequent and rare perceivers was observed. Based on these observations we argue that global

460 coherogram differences (*/ta/-/pa/*) present itself as a signature of the inter-trial perceptual
461 variability. Furthermore, frequency specific signature in the global coherence consistent
462 across the perceptual categories enhanced alpha and gamma band coherence in frequent
463 perceivers and desynchronization in alpha band coherence accompanied with enhanced
464 gamma band coherence pinpoints alpha band coherence as signature of inter-individual
465 variability. These observations further highlight a mechanistic difference in the processing of
466 cross-modal stimuli between frequent and rare perceivers. Nonetheless, such differences are
467 contingent on the stimulus as there was no difference in the global coherence pattern between
468 frequent and rare perceivers during congruent */ta/*. In retrospect of the global coherence
469 patterns during McGurk stimuli, an obvious question is, do cross-frequency couplings among
470 theta, alpha, beta and gamma band exist in a context specific way? Questions of such nature
471 become a prime candidate to answer for future studies. A detailed account of cross-frequency
472 coupling via coherence is currently out of scope of the present study.

473

474 *Characterization of NCN at source space*

475 Pairwise coherence is affected by volume conduction to a considerable degree, specifically
476 for local functional connectivity (Winter et al., 2007). The global coherence results are
477 affected to a lesser degree by volume conduction, simply because the functional connections
478 that can spuriously affect a distinct pattern of coherence are unlikely to survive the
479 normalized vector summation procedure that is undertaken. Nonetheless, we need to validate
480 if at least qualitatively the source and sensor level analysis are consistent. Subsequently, the
481 global coherogram computed from reconstructed sources, first estimated through LCMV
482 analysis were explored. The locations that showed statistical significance in the ratio of
483 source power between */ta/* and */pa/* trials were used for reconstruction of sources. Frequent
484 and rare perceivers showed a considerable overlap in brain areas involving right STS,
485 fusiform gyrus, left inferior frontal gyrus and bilateral superior frontal gyrus. When
486 coherogram was computed at the source level and the difference of global coherence between
487 */ta/* and */pa/* are plotted, we could identify a high degree of similarity with the sensor space
488 results (**Figure 4B-G**). Even though the exact spectro-temporal boundaries were slightly
489 different, the overall pattern of results of enhanced gamma coherence and decreased alpha
490 coherence at zero AV lag, and broadband coherence for ± 450 ms AV lag was observed.
491 Crucially, the major overlap of cortical sources across frequent and rare perceivers pinpoints
492 the significance of understanding the communication within network of cortical regions over
493 emphasizing role of isolated cortical loci in cognition

494 ***Mechanistic understanding of NCN dynamics using biologically realistic computational***
495 ***model***

496 Alpha and gamma band coherences are observed in processing of multisensory stimulus (
497 Hummel & Gerloff, 2005; Kanayama et al., 2007; Doesburg et al. 2008; Kayser et al., 2008;
498 Maier et al., 2008; Kayser & Logothetis, 2009; Kumar et al., 2016; also present results,
499 **Figure 2**). Interestingly, high gamma coherence is seen when the nature of multisensory
500 stimulus is complex (asynchronous, incongruent) (Doesburg et al., 2008; Kumar et al., 2016
501 and present results, **Figure 2**) which in some instances lead to illusory perception (Kanayama
502 et al., 2007; Kumar et al., 2016; and present results, **Figure 1**). Gamma coherence is also
503 observed in the communication involving higher order multisensory areas (Maier et al., 2008;
504 and present results, **Figure 4**). Our computational model explains that alpha band coherence
505 emerges when visual system has a direct influence on auditory node, while gamma coherence
506 was observed only with indirect A-V interactions via multisensory node (**Figure 6**
507 **Supplement 1**). From a theoretical perspective this is possible because the time scale of
508 processing is most disparate for the auditory and visual system, with auditory the fastest and
509 visual the slowest. Without the presence of an intermediate time-scale, one “mode of
510 communication” (alpha coherence) is sustained by the neural mass model within biologically
511 relevant parameter regimes. Once there is another neural mass of intermediate time-scale
512 participating in processing of information, the higher dimensionality of the resultant
513 dynamical system allows creation of another mode of communication. Hence, our model
514 suggests that gamma coherence could emerge due to the communication between primary
515 auditory and visual areas but routed indirectly via higher order areas such as pSTS or inferior
516 parietal or frontal areas. Our suggestion is in line with earlier observations of visual stimuli
517 modulating auditory perception either directly resulting in alpha coherence (Kayser et al.,
518 2008) or indirectly via higher order regions (STS) resulting primarily in gamma coherence
519 (Maier et al., 2008; Kayser & Logothetis, 2009).

520

521 Behavioral responses from rare perceivers indicate limited influence of visual stimulus in
522 shaping up the perceptual response since their response is akin to unisensory auditory
523 response. The neuromarker of inter-individual variability, alpha coherence was drastically
524 diminished (desynchronization) when A-V coupling was extremely weak ($W_{AV} < 0.05$)
525 (**Figure 6A**). This indicates that overall interaction between visual and auditory node (direct
526 and indirect via pSTS for example) is comparatively lesser in rare perceivers with respect to

527 frequent perceivers and thus, results more in unisensory perception. Subsequently, we can
528 also infer that direct A-V coupling is crucial for "frequently" perceiving the illusion of
529 McGurk stimulus as in the case of frequent perceivers. On the other hand, differences in
530 illusory perception and unisensory perception in both kinds of perceivers emerge from
531 indirect A-V coupling via multisensory node (**Figure 6 B & C**). As discussed before, high
532 gamma coherence is associated with multisensory processing involving interaction with
533 higher order multisensory areas (Maier et al., 2008). Supporting this observation, we show A-
534 V communication via multisensory node is crucial to generate gamma coherence during
535 illusory perception in frequent and rare perceivers.

536

537 Alpha and/or gamma coherences have been observed in other Audio-Visual perception
538 studies involving A-V speech phrases (Doesburg et al., 2008), natural A-V scenes (Kayser et
539 al., 2008) and also artificially generated A-V looming signals (Maier et al., 2008). Increase in
540 gamma coherence and reduction in alpha and beta coherences were observed during the
541 perception of incongruent (lagged) A-V speech phrases (Doesburg et al., 2008). Increase in
542 the interaction between fast and slow nodes via intermediate node increases the gamma
543 coherence and decreases coherences in alpha and beta band (**Figure 6B**). Therefore, a similar
544 mechanism that explains the observations of McGurk illusory perception is also applicable
545 for explaining observations during perception of incongruent (lagged) A-V speech phrases.
546 Increase in A-V interactions via multisensory node also explains the enhanced gamma
547 coherence between auditory cortex and Superior Temporal Sulcus during congruent A-V
548 looming signals in rhesus monkeys (Maier et al., 2008). Similarly, strong A-V interactions
549 that distinguish the two kinds of perceiver groups (**Figure 6A**) also explain the increase in
550 alpha phase consistency observed during natural A-V scenes in rhesus monkeys (Kayser et
551 al., 2008). A different configuration of the model, where fast (auditory) node is source of EI
552 and the slow (visual) node is sink of EI, generates peaks in beta band coherence (**Figure 6**
553 **Supplement 2B**) whereas the default configuration generates peaks in beta band as well as
554 alpha band coherences (**Figure 6 Supplement 1B**). Therefore, this difference in
555 configurations distinguishes visual perception of words (increase in beta band coherence and
556 decrease in alpha band coherence) from auditory perception of words (increase in alpha and
557 beta band coherence) suggesting that auditory (fast) node is the sink of EI during auditory
558 perception and visual (slow) node is the sink of EI during visual perception (von Stein et al.,
559 1999). Stretching to studies other than audio-visual perception, direct interactions between
560 fast and slow nodes also explain the observed high alpha coherence during good performance

561 while matching tactile Braille stimulus with its visual counterpart (Hummel & Gerloff, 2005)
562 and the fast-slow indirect interactions via intermediate time-scale node explains the high
563 gamma band coherence during rubber-hand illusion when visuo-tactile stimuli were
564 congruent (Kanayama et al., 2007).

565

566 We have speculated the specific interactions of neural masses with different time-constants
567 that generate band specific coherences and that are responsible for their enhancement and
568 diminution. Multi-parametric and unbounded nature of the parameter space results in myriads
569 of dynamics including chaos which is non-biological (Stefanescu & Jirsa, 2008). Therefore,
570 such models should not be used to directly fit the data by estimating model parameters that
571 minimize the error using optimization techniques. However, our model will be useful as a
572 phenomenological or minimalistic model in providing mechanistic insights into many findings
573 (Fries, 2015; Engel et al., 2012) including pathological conditions (Başar & Güntekin, 2008)
574 where relative changes in band specific coherences are observed.

575

576 **Materials and Methods**

577 *Participants*

578 Twenty nine normal healthy volunteers (16 males and 13 females, in the range of 21-29 years
579 of age; mean age 25, SD = 3) participated in the study. All participants gave written informed
580 consent in a format approved by the Institutional Human Ethics Committee of the National
581 Brain Research Centre, Gurgaon which is in agreement with the Declaration of Helsinki.
582 None of the participants had a history of neurological or audiological problems and were
583 compensated for their time devoted to the experiment. All had normal or corrected-to-normal
584 vision and were right-handed (tested using Edinburgh handedness inventory). The data from
585 four volunteers were not included in the study because the channel impedance values in EEG
586 exceeded 10 k Ω .

587 *Stimuli and trials*

588 The experiment composed of 360 trials in which videos of a native Hindi speaking male
589 articulating the syllables /ka/ and /ta/ (Fig. 1A) were presented. One-fourth (90 trials) of the
590 trials consisted of congruent videos (visual /ta/ auditory /ta/). The remaining three-fourths of
591 the trials comprised incongruent videos (visual /ka/ auditory /pa/) presented with AV lags: -
592 450 ms (audio leads the articulation), 0 ms (synchronous) and +450 ms (articulation leads the
593 audio), each encompassing one-fourth of the overall trials. The auditory object in the
594 incongruent trials was extracted from a video of the speaker articulating /pa/ using the
595 software Audacity (www.audacityteam.org). Subsequently, the extracted auditory /pa/ was
596 superimposed onto the muted video of the speaker articulating the syllable /ka/ using the
597 software Videopad Editor (www.nchsoftware.com). The composite multisensory stimuli were
598 rendered into an 800 x 600 pixels movie with a digitization rate of 29.97 frames per second.
599 Stereo soundtracks were digitized at 48 kHz with 32 bit resolution. Presentation software
600 (Neurobehavioral System Inc.) was used to present the stimulus videos using a 17" LED
601 monitor. Sound was delivered using sound tubes at an overall intensity of ~60 dB.

602 *Experimental design*

603 The experiment was divided into three blocks. Each block consisted of 120 trials comprising
604 all the four kinds of videos (30 trials of each). Inter-stimulus intervals were pseudo-randomly
605 varied between 1200 ms and 2800 ms to minimize expectancy effects. Using a forced choice

606 task, the participants had to indicate their choice by pressing a specified key on the keyboard
607 whether they heard /ta/, /pa/ or something else (others) while watching the videos.

608 *Eye Tracking*

609 Gaze fixations of participants on the computer screen were recorded by EyeTribe eye
610 tracking device (<https://theeyetribe.com/>). The gaze data were analyzed using customized
611 MATLAB codes. The image frame of the speaker video was divided into 2 parts, the head,
612 and the mouth. The gaze fixations at these locations over the duration of stimulus
613 presentation were converted into percentage measures for further statistical analysis.

614 *EEG recording*

615 Continuous EEG scans were acquired using a Neuroscan system (Synamps2, Compumedics,
616 Inc.) with 64 Ag/AgCl scalp electrodes sintered on an elastic cap in a 10-20 montage.
617 Recordings were made against the centre (near Cz) reference electrode on the Neuroscan cap
618 and digitized at a sampling rate of 1000 Hz. Channel impedances were monitored to be at
619 values $< 5\text{k}\Omega$. Four volunteers showing higher impedances ($\sim 10\text{k}\Omega$) were discarded from
620 further analysis.

621 *EEG Data processing*

622 In the preprocessing step, the acquired EEG data was filtered using a band pass of 0.2-45 Hz.
623 Subsequently, epochs of 900ms post the onset of first sensory object (auditory vocalization or
624 articulatory lip movement) was extracted. Epochs extracted from congruent and incongruent
625 videos were further sorted based on the perceptual experience: /ta/, /pa/ and 'others'. The
626 sorted epochs were then baseline corrected by removing the temporal mean of the EEG signal
627 on an epoch-by-epoch basis. Finally, in order to remove the response contamination from
628 ocular and muscle-related artifacts, epochs with maximum signal amplitude above $50\ \mu\text{V}$ or a
629 minimum below $-50\ \mu\text{V}$ were removed from all electrodes.

630 *Network analysis and global coherogram*

631 To investigate frequency specific FC that subserves cross-modal perception and characterizes
632 a frequent from a rare perceiver, we computed the global coherogram. Global coherogram
633 captures the global coherence dynamics and quantifies the strength of neural co-activation
634 across the whole brain at specific frequencies over time. In order to compute global

635 coherogram from the preprocessed time series sorted based on the perceptual categories, we
636 employed the Chronux (Mitra & Bokil, 2008) function cohgramc.m to obtain trial-wise time
637 frequency cross-spectral matrix for all the sensor combinations. The output variable 'S12' of
638 the function cohgramc.m yields the time frequency cross-spectrum density at a frequency f
639 between sensor pair i and j employing the formula:

$$640 \quad C_{ij}(f) = X_i(f)X_j(f)^* \quad (1)$$

641 where, $C_{ij}(f)$ represents the cross spectrum, $X_i(f)$ represent the tapered Fourier transform
642 of the time series from the sensor i and $X_j(f)^*$ represent the complex conjugate of the
643 tapered time series from the sensor j at frequency f . In our analysis, a 62 x 62 matrix of cross
644 spectra that represents all pairwise sensor combinations was computed. The time bandwidth
645 product and the number of tapers were set at 3 and 5, respectively, and a moving window of
646 0.4 s with a step size of 0.05s were employed in the computation. Thereafter, we computed
647 the global coherence at each time and frequency bin by computing the ratio of the largest
648 eigenvalue of the cross-spectral matrix to the sum of the eigenvalues on a trial-by-trial basis
649 employing the following equation:

$$650 \quad C_{Global}(f, t) = \frac{S_1^Y(f, t)}{\sum_{i=1}^n S_i^Y(f, t)} \quad (2)$$

651 where $C_{Global}(f, t)$ represent the global coherence at frequency f in the time window t ,
652 $S_1^Y(f, t)$ represent the largest eigenvalue and the denominator $\sum_{i=1}^n S_i^Y(f, t)$ represents the sum
653 of eigenvalues of the cross-spectral matrix at every time bin. Subsequently, the time-
654 frequency global coherogram computed for /ta/ and /pa/ responses were compared non-
655 parametrically using cluster based permutation tests for frequent and rare perceivers
656 explicitly (Maris et. al., 2007; Kumar et. al., 2016).

657 We computed the global coherence collapsed across the entire epoch to identify if there are
658 certain frequencies around which the network is most robust underlying cross-modal (illusory
659 /ta/) and unimodal (/pa/) perception in frequent and rare perceivers. Furthermore, to
660 investigate whether the organization of these networks dependent on the stimulus
661 configurations or the perceptual outcome, we also computed the global coherence during
662 congruent /ta/ perception in frequent and rare perceivers. We employed the Chronux function
663 CrossSpecMatc.m for computing the global coherence. The output variable 'Ctot' of the

664 function yields the global coherence value at frequency f by initially computing the cross-
665 spectrum for all sensor combinations following the Equation 1. Subsequently, global
666 coherence at every frequency bin is obtained by computing the ratio of the largest eigenvalue
667 of the cross-spectral matrix to the sum of the eigenvalues on a trial-by-trial basis employing
668 Equation 2. The time bandwidth product and the number of tapers were set at 3 and 5,
669 respectively, and a fixed window size of 0.9 s was employed in the computation. Finally, we
670 employed Silvermann's bootstrapping test for detecting the presence of a bimodal distribution
671 (Silverman, 1981). We performed Silvermann's bootstrapping bimodality test on the time
672 averaged global coherence separately on the perceptual categories across all AV lags in
673 frequent and rare perceivers.

674 The aforementioned analysis was further performed to compute the global coherogram and
675 coherence during the perception of congruent /*ta*/ in frequent and rare perceivers.
676 Subsequently, the global coherogram was compared employing cluster based permutation
677 tests.

678 ***Source Reconstruction and functional connectivity***

679 To investigate if the global coherogram patterns observed at the sensor level affected by
680 volume conduction, we constructed source time-series and computed the global coherogram
681 differences between /*ta*/ and /*pa*/ at all AV lags in frequent and rare perceivers. We
682 employed a linearly constrained minimum variance (LCMV) beamformer algorithm (Van
683 Veen et al., 1997) to identify the cortical generators of the time-series during /*ta*/ and /*pa*/
684 perception in frequent and rare perceivers. The entire epoch of 0.9s was employed in the
685 source analysis. Prior to source reconstruction, we constructed our personalized average
686 template from the individual MRIs of the subjects using the function
687 'antsMultivariateTemplateConstruction' developed by Advanced Normalization Tools
688 (ANTs) (<http://stnava.github.io/ANTs/>)(Avants et al., 2011). The pipeline initially involves
689 rigidly registering the participants T1 images to a MNI template while maintaining the
690 volume and size of the original structural images. The rigidly registered images are then
691 averaged to generate a temporary template. This template is then used as the first registration
692 target onto which each participants T1 image is non-linearly registered, transformed and
693 averaged. Iteratively, the T1 images are non-linearly registered to the new average,
694 transformed and re-averaged generating a relatively a more precise average for the next
695 iteration.

696 For source reconstruction we employed Fieldtrip toolbox. Firstly, we used
697 `ft_prepare_leadfield.m` and employed the Boundary Element Method (BEM) to generate the
698 leadfield matrix from the template we constructed. The leadfield matrix corresponds to the
699 tissue and geometrical properties of the brain represented as discrete grids or voxels.
700 Subsequently, we employed `ft_timelockanalysis.m` to evaluate the covariance matrix of the
701 epochs sorted based on perceptual categories in frequent and rare perceivers as the LCMV
702 adaptive spatial filters are constrained by the covariance and leadfield matrices. These spatial
703 filters regulate the amplitude of brain electrical activity passing from a specific location while
704 attenuating activity originating at other locations. The distribution of the output amplitude of
705 the spatial filters provides the metric for source localization. However, in order to compare
706 the source power during */ta/* and */pa/* perception, we computed an inverse ‘common spatial
707 filter’ employing `ft_sourceanalysis.m` from the dataset obtained by appending the datasets of
708 */ta/* and */pa/* post time-lock analysis. Eventually, based on the pre-computed common spatial
709 filter we evaluated the sources separately for */ta/* and */pa/* employing `ft_sourceanalysis.m`.
710 The difference in the source power between */ta/* and */pa/* were consequently compared by
711 taking the ratio of the source power of */ta/* and */pa/*. Finally, the grids eliciting power above
712 the 99.5th percentile were identified as sources and were interpolated onto the constructed
713 template for illustrative purposes.

714 For reconstructing time series from the thresholded sources, we projected single trial epoched
715 time series from sensors onto the source space by multiplying them to the concordant spatial
716 filters of the thresholded sources. There were overall 52 grids of the spatial filter
717 corresponding to the sources represented in **Figure 4A** onto which the sensor level data was
718 projected to obtain the source time series. Furthermore, each spatial filter is represented by
719 three components representing the unity moment in the x , y and z direction of the dipole at the
720 respective grid location. We estimated the global coherogram differences between the */ta/*
721 and */pa/* perception in frequent and rare perceivers from the source time series from the
722 component that best matched the sensor level global coherogram results.

723 *Large scale dynamical model of three neural masses*

724 Our objective was to construct a large-scale dynamical model which is biologically realistic
725 to explain the generative mechanisms underlying observed coherence spectra and frequency
726 specific functional connectivity during illusory and non-illusory perception in rare and
727 frequent perceivers based on empirical data. Our proposed model is a network of three neural

728 masses, each comprising of excitatory and inhibitory neurons representing auditory, visual
729 and higher order multisensory cortical regions (**Figure 5**). We follow a previously established
730 practice and convention in computational modelling by treating each cortical region as an
731 individual node as suggested by Stefanescu and Jirsa (Stefanescu & Jirsa, 2008).

732

733 Broadly we incorporate the following biophysically realistic factors in our model
734 construction.

735 1. The time-scale of processing of the visual system can be considered slowly varying in
736 comparison to auditory system (Williams et al., 2004; Rosen & Howell, 2011).
737 Multisensory system can be placed in between the auditory and visual systems in terms of
738 the processing time-scale.

739 2. Two of the ways visual inputs are directed to the auditory cortex are: 1) visual cortex
740 could directly influence the auditory cortex in a feedforward manner due to direct
741 projections (Falchier, et al., 2002; Rockland & Ojima, 2003; Wallace et al., 2004) and 2)
742 feedback from the higher multisensory association areas (Bizley & King, 2012). Hence,
743 in our proposed model visual node influences the auditory node in both manners: directly
744 and indirectly via multisensory node.

745 3. As post-synaptic potentials of pyramidal cells, which are excitatory, are shaped by their
746 connections with other excitatory cells and inhibitory cells (Kirschstein & Köhling,
747 2009). We use a population of excitatory and inhibitory neurons in each node where the
748 number of excitatory neurons are considerably higher (Olbrich & Braak, 1985). Thus, 150
749 excitatory neurons and 50 inhibitory neurons are selected to have a 3:1 ratio between
750 them, an approach previously followed by Stefanescu and Jirsa (Stefanescu & Jirsa,
751 2008). Inhibitory neurons in one neural area do not directly influence inhibitory neurons
752 within the same area since such connections are sparse in nature (Wilson & Cowan, 1972;
753 Stefanescu & Jirsa, 2008).

754

755 Incorporating these factors we define a dynamic mean field model that comprises of three
756 equations for an excitatory Hindmarsh Rose (HR) neuron (number of excitatory neurons are
757 150 within an area, $N_E = 150$) and three equations for an inhibitory HR neuron (number of
758 inhibitory neurons are 50 within an area, $N_I = 50$) (**Figure 5**). The three variables account
759 for the membrane dynamics and two kinds of gating currents, one fast and one slow

760 respectively. Thus, the entire network can be represented as a network of coupled non-linear
 761 differential equations comprising of

Excitatory Subpopulation

$$\begin{aligned} \tau_L \frac{dx_{n_E}^L}{dt} &= y_{n_E}^L - ax_{n_E}^{L^3} + bx_{n_E}^{L^2} - z_{n_E}^L + K_{EE}(E(x_{n_E}^L) - x_{n_E}^L) - K_{IE}(E(x_{n_I}^L) - x_{n_E}^L) + I_{n_E}^L + \sum_{M=1}^3 W_{ML} E(x_{n_E}^M) + \varepsilon \\ \tau_L \frac{dy_{n_E}^L}{dt} &= c - dx_{n_E}^{L^2} - y_{n_E}^L \\ \tau_L \frac{dz_{n_E}^L}{dt} &= r(s(x_{n_E}^L - x_0) - z_{n_E}^L); n_E = 1, \dots, N_E; L = 1(A), 2(M) \& 3(V) \end{aligned}$$

Inhibitory Subpopulation

(3)

$$\begin{aligned} \tau_L \frac{dx_{n_I}^L}{dt} &= y_{n_I}^L - ax_{n_I}^{L^3} + bx_{n_I}^{L^2} - z_{n_I}^L + K_{EI}(E(x_{n_E}^L) - x_{n_I}^L) + I_{n_I}^L \\ \tau_L \frac{dy_{n_I}^L}{dt} &= c - dx_{n_I}^{L^2} - y_{n_I}^L \\ \tau_L \frac{dz_{n_I}^L}{dt} &= r(s(x_{n_I}^L - x_0) - z_{n_I}^L); n_I = 1, \dots, N_I; L = 1(A), 2(M) \& 3(V) \end{aligned}$$

762

763 Where L : A, V and AV for auditory, visual and audio-visual areas that are driven by a
 764 common noise distribution (ε). In our model auditory node has the fastest time-constant (
 765 $\tau_A \sim 0.05ms$), visual node has the slowest time-constant ($\tau_V \sim 2.5ms$) and time-constant of
 766 multisensory node is chosen to be in between the two ($\tau_M \sim 1ms$) as it integrates information
 767 from both the modalities. The mean activity of excitatory neurons in a node (
 768 $E(x_{n_E}) = \frac{1}{N_E} \sum_{n_E=1}^{N_E} x_{n_E}$ influences neuronal activities of other nodes that is governed by coupling

$$E(x_{n_E}) = \frac{1}{N_E} \sum_{n_E=1}^{N_E} x_{n_E}$$

769 parameters: W_{AV} (auditory-visual coupling), W_{AM} (auditory-multisensory coupling) and W_{VM}
 770 (visual-multisensory coupling). Positive value of coupling parameters reflects excitatory
 771 influence and negative value reflects inhibitory influence. Inhibitory influences are chosen to
 772 maintain a balance with excitation. For example, visual node's excitatory influence of $+W_{AV}$
 773 on auditory node is balanced with inhibitory influence of the same strength ($-W_{AV}$) from the
 774 auditory node.

775

776 In this configuration, visual node is referred as source node as it is the source of excitatory
 777 influence whereas auditory node is referred to as sink node as all excitatory influences are
 778 directed towards auditory node and multisensory node behaves as both source and sink..

779 We place each individual neuron in a dynamical regime where both spiking and bursting
780 behavior is possible depending on the external input current (I) that enters the neuron when
781 other parameters are held constant at the following values:
782 $a = 1; b = 3; c = 1; d = 5; s = 4; r = 0.006; x_0 = -1.6$; (Stefanescu & Jirsa, 2008).

783

784 The coupling between the neurons within a node is linear and its strength is governed by the
785 following parameters: K_{EE} for excitatory-excitatory coupling, K_{EI} for excitatory-inhibitory
786 coupling and K_{IE} for inhibitory-excitatory coupling. As excitatory and inhibitory synapses
787 are not independent processes, their relation is captured by the ratio $n = \frac{K_{IE}}{K_{EE}}$. As alpha (8-12

788 Hz) and delta (1-4 Hz) rhythms are observed during resting state (Gold et al., 2006), the
789 inhibition to excitation ratio ($n = 3.39$) is chosen when the average activity of nodes in a
790 disconnected network has higher power at alpha and delta frequencies in the absence of
791 stimulus ($\mu(I_{A,V,M}) = 0.1$; baseline) (**Figure 5 Supplement 1**). The external currents to both
792 the excitatory and inhibitory subpopulations are drawn from a Gaussian distribution where μ
793 and σ are the mean and standard-deviation. As the input stimulus relays to auditory, visual
794 and multisensory regions via thalamus, we interpret lateral geniculate nucleus (LGN) and
795 medial geniculate nucleus (MGN) to be the source of external current (I_A , I_V and I_M) pulse
796 of 450 ms in the nodes when the model was simulated for 1 sec. In rhesus monkey, the
797 projections of MGN to pSTS were found to be sparse (Yeterian & Pandya, 1989). Therefore,
798 we choose lower mean value of external current to multisensory node ($\mu(I_M) = 0.85$) in
799 comparison to visual node ($\mu(I_V) = 2.8$) and auditory node ($\mu(I_A) = 2.8$) while keeping the
800 standard deviation of the external current at 0.4 for all nodes.

801 **References**

- 802 Aerts, H., Schirner, M., Jeurissen, B., Roost, D. Van, Achten, R., Ritter, P., & Marinazzo, D.
803 (2018). Modeling brain dynamics in brain tumor patients using the Virtual Brain.
804 <http://doi.org/10.1101/265637>
- 805 Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty Years After Hearing Lips and Seeing
806 Voices: the McGurk Effect Revisited. *Multisensory Research*, *31*(1–2), 111–144.
807 <http://doi.org/10.1163/22134808-00002565>
- 808 Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A
809 reproducible evaluation of ANTs similarity metric performance in brain image
810 registration. *NeuroImage*, *54*(3), 2033–2044.
811 <http://doi.org/10.1016/j.neuroimage.2010.09.025>
- 812 Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.
813 <http://doi.org/10.1126/science.1736359>
- 814 Başar, E., & Güntekin, B. (2008). A review of brain oscillations in cognitive disorders and
815 the role of neurotransmitters. *Brain Research*, *1235*, 172–193.
816 <http://doi.org/10.1016/j.brainres.2008.06.103>
- 817 Beauchamp, M. S. (2010). fMRI-guided TMS reveals that the STS is a Cortical Locus of the
818 McGurk Effect. *Journal of Neuroscience*, *30*(7), 2414–2417.
819 <http://doi.org/10.1523/JNEUROSCI.4865-09.2010.fMRI-guided>
- 820 Becker, R., Knock, S., Ritter, P., & Jirsa, V. (2015). Relating Alpha Power and Phase to
821 Population Firing and Hemodynamic Activity Using a Thalamo-cortical Neural Mass
822 Model. *PLoS Computational Biology*, *11*(9), e1004352.
823 <http://doi.org/10.1371/journal.pcbi.1004352>
- 824 Bizley, J. K., & King, A. J. (2012a). *What Can Multisensory Processing Tell Us about the*
825 *Functional Organization of Auditory Cortex? The Neural Bases of Multisensory*
826 *Processes*. CRC Press/Taylor & Francis. Retrieved from
827 <http://www.ncbi.nlm.nih.gov/pubmed/22593889>
- 828 Bizley, J. K., & King, A. J. (2012b). *What Can Multisensory Processing Tell Us about the*
829 *Functional Organization of Auditory Cortex? The Neural Bases of Multisensory*
830 *Processes*. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22593889>
- 831 Bressler, S. L. (1995). Large-scale cortical networks and cognition. *Brain Research Reviews*,
832 *20*(3), 288–304. [http://doi.org/10.1016/0165-0173\(94\)00016-I](http://doi.org/10.1016/0165-0173(94)00016-I)
- 833 Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging
834 methods and principles. *Trends in Cognitive Sciences*, *14*(6), 277–290.
835 <http://doi.org/10.1016/j.tics.2010.04.004>
- 836 Cuppini, C., Shams, L., Magosso, E., & Ursino, M. (2017). A biologically inspired
837 neurocomputational model for audiovisual integration and causal inference. *European*
838 *Journal of Neuroscience*, *46*(9), 2481–2498. <http://doi.org/10.1111/ejn.13725>
- 839 David, O., & Friston, K. J. (2003). A neural mass model for MEG/EEG: coupling and
840 neuronal dynamics. *NeuroImage*, *20*(3), 1743–55. Retrieved from
841 <http://www.ncbi.nlm.nih.gov/pubmed/14642484>
- 842 Deco, G., Rolls, E. T., & Romo, R. (2010). Synaptic dynamics and decision making.
843 *Proceedings of the National Academy of Sciences of the United States of America*,
844 *107*(16), 7545–9. <http://doi.org/10.1073/pnas.1002333107>
- 845 Doesburg, S. M., Emberson, L. L., Rahi, A., Cameron, D., & Ward, L. M. (2008).
846 Asynchrony from synchrony: long-range gamma-band neural synchrony accompanies
847 perception of audiovisual speech asynchrony. *Experimental Brain Research*, *185*(1), 11–
848 20. <http://doi.org/10.1007/s00221-007-1127-5>
- 849 Engel, A. K., Senkowski, D., & Schneider, T. R. (2012). *Multisensory Integration through*
850 *Neural Coherence. The Neural Bases of Multisensory Processes*. CRC Press/Taylor &

- 851 Francis. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22593880>
- 852 Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of
853 multimodal integration in primate striate cortex. *The Journal of Neuroscience: The*
854 *Official Journal of the Society for Neuroscience*, 22(13), 5749–59.
855 <http://doi.org/20026562>
- 856 Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1),
857 220–235. <http://doi.org/10.1016/j.neuron.2015.09.034>
- 858 Gold, C., Henze, D. A., Koch, C., & Buzsáki, G. (2006). On the Origin of the Extracellular
859 Action Potential Waveform: A Modeling Study. *Journal of Neurophysiology*, 95(5),
860 3113–3128. <http://doi.org/10.1152/jn.00979.2005>
- 861 Hindmarsh, J. L., & Rose, R. M. (1984). A model of neuronal bursting using three coupled
862 first order differential equations. *Proceedings of the Royal Society of London. Series B,*
863 *Biological Sciences*, 221(1222), 87–102. Retrieved from
864 <http://www.ncbi.nlm.nih.gov/pubmed/6144106>
- 865 Hummel, F., & Gerloff, C. (2005). Larger Interregional Synchrony is Associated with Greater
866 Behavioral Success in a Complex Sensory Integration Task in Humans. *Cerebral*
867 *Cortex*, 15(5), 670–678. <http://doi.org/10.1093/cercor/bhh170>
- 868 Jansen, B. H., & Rit, V. G. (1995). Electroencephalogram and visual evoked potential
869 generation in a mathematical model of coupled cortical columns. *Biological*
870 *Cybernetics*, 73(4), 357–66. Retrieved from
871 <http://www.ncbi.nlm.nih.gov/pubmed/7578475>
- 872 Jones, J. a, & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An
873 fMRI study of the McGurk effect. *Neuroreport*, 14(8), 1129–1133.
874 <http://doi.org/10.1097/01.wnr.0000074343.81633.2a>
- 875 Kaiser, J. (2004). Hearing Lips: Gamma-band Activity During Audiovisual Speech
876 Perception. *Cerebral Cortex*, 15(5), 646–653. <http://doi.org/10.1093/cercor/bhh166>
- 877 Kanayama, N., Sato, A., & Ohira, H. (2007). Crossmodal effect with rubber hand illusion and
878 gamma-band activity. *Psychophysiology*, 44(3), 392–402. <http://doi.org/10.1111/j.1469-8986.2007.00511.x>
- 880 Kayser, C., & Logothetis, N. K. (2009). Directed interactions between auditory and superior
881 temporal cortices and their role in sensory integration. *Frontiers in Integrative*
882 *Neuroscience*, 3, 7. <http://doi.org/10.3389/neuro.07.007.2009>
- 883 Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual Modulation of Neurons in
884 Auditory Cortex. *Cerebral Cortex*, 18(7), 1560–1574.
885 <http://doi.org/10.1093/cercor/bhm187>
- 886 Keil, J., Muller, N., Ihssen, N., & Weisz, N. (2012). On the Variability of the McGurk Effect:
887 Audiovisual Integration Depends on Prestimulus Brain States. *Cerebral Cortex*, 22(1),
888 221–231. <http://doi.org/10.1093/cercor/bhr125>
- 889 Keil, J., & Senkowski, D. (2018). Neural Oscillations Orchestrate Multisensory Processing.
890 *The Neuroscientist*, 107385841875535. <http://doi.org/10.1177/1073858418755352>
- 891 Kirschstein, T., & Köhling, R. (2009). What is the Source of the EEG? *Clinical EEG and*
892 *Neuroscience*, 40(3), 146–149. <http://doi.org/10.1177/155005940904000305>
- 893 Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory
894 performance: a review and analysis. *Brain Research Reviews*, 29, 169–195.
- 895 Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007).
896 Causal Inference in Multisensory Perception. *PLoS ONE*, 2(9), e943.
897 <http://doi.org/10.1371/journal.pone.0000943>
- 898 Kumar, G. V., Halder, T., Jaiswal, A. K., Mukherjee, A., Roy, D., & Banerjee, A. (2016).
899 Large Scale Functional Brain Networks Underlying Temporal Integration of Audio-
900 Visual Speech Perception: An EEG Study. *Frontiers in Psychology*, 7, 1558.

- 901 <http://doi.org/10.3389/fpsyg.2016.01558>
- 902 Kumar, G. V., Kumar, N., Roy, D., & Banerjee, A. (2018). Segregation and Integration of
903 Cortical Information Processing Underlying Cross-Modal Perception. *Multisensory*
904 *Research*, 31(5), 481–500. <http://doi.org/10.1163/22134808-00002574>
- 905 Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase
906 synchrony in brain signals. *Human Brain Mapping*, 8(4), 194–208.
907 [http://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:4<194::AID-HBM4>3.0.CO;2-C](http://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C)
- 908 Lopes da Silva, F. H., Hoeks, A., Smits, H., & Zetterberg, L. H. (1974). Model of brain
909 rhythmic activity. The alpha-rhythm of the thalamus. *Kybernetik*, 15(1), 27–37.
910 Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4853232>
- 911 Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of Bimodal
912 Looming Signals through Neuronal Coherence in the Temporal Lobe. *Current Biology*,
913 18(13), 963–968. <http://doi.org/10.1016/j.cub.2008.05.043>
- 914 Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the
915 McGurk effect: contributions of participants, stimuli, time, and response type.
916 *Psychonomic Bulletin & Review*, 22(5), 1299–307. <http://doi.org/10.3758/s13423-015-0817-4>
- 917
- 918 Maris, E., Schoffelen, J.-M., & Fries, P. (2007). Nonparametric statistical testing of
919 coherence differences. *Journal of Neuroscience Methods*, 163(1), 161–75.
920 <http://doi.org/10.1016/j.jneumeth.2007.02.011>
- 921 McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 691–811.
922 <http://doi.org/10.1038/264746a0>
- 923 Mitra, P., & Bokil, H. (2008). *Observed brain dynamics*. Oxford Univ Press, New York.
- 924 Morís Fernández, L., Torralba, M., & Soto-Faraco, S. (2018). Theta oscillations reflect
925 conflict processing in the perception of the McGurk illusion. *European Journal of*
926 *Neuroscience*, 1–12. <http://doi.org/10.1111/ejn.13804>
- 927 Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the
928 McGurk effect. *Perception & Psychophysics*, 58(3), 351–362.
929 <http://doi.org/10.3758/BF03206811>
- 930 Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the
931 McGurk effect, a multisensory speech illusion. *NeuroImage*, 59(1), 781–787.
932 <http://doi.org/10.1016/j.neuroimage.2011.07.024>
- 933 Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the
934 McGurk effect, a multisensory speech illusion. *NeuroImage*, 59(1), 781–787.
935 <http://doi.org/10.1016/j.neuroimage.2011.07.024>
- 936 Nath, A. R., & Beauchamp, M. S. (2012). A Neural Basis for Interindividual Differences in
937 the McGurk Effect, a Multisensory Speech Illusion, 59(1), 781–787.
938 <http://doi.org/10.1016/j.neuroimage.2011.07.024.A>
- 939 Olbrich, H.-G., & Braak, H. (1985). Ratio of pyramidal cells versus non-pyramidal cells in
940 sector CA1 of the human Ammon's horn. *Anatomy and Embryology*, 173(1), 105–110.
941 <http://doi.org/10.1007/BF00707308>
- 942 Proverbio, A. M., Massetti, G., Rizzi, E., & Zani, A. (2016). Skilled musicians are not subject
943 to the McGurk effect. *Scientific Reports*, 6, 30423. <http://doi.org/10.1038/srep30423>
- 944 Roa Romero, Y., Senkowski, D., & Keil, J. (2015). Early and Late Beta Band Power reflects
945 Audiovisual Perception in the McGurk Illusion. *Journal of Neurophysiology*,
946 jn.00783.2014. <http://doi.org/10.1152/jn.00783.2014>
- 947 Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in
948 macaque monkey. *International Journal of Psychophysiology: Official Journal of the*
949 *International Organization of Psychophysiology*, 50(1–2), 19–26. Retrieved from
950 <http://www.ncbi.nlm.nih.gov/pubmed/14511833>

- 951 Rosen, S., & Howell, P. (2011). *Signals and systems for speech and hearing / S. Rosen and P.*
952 *Howell - Details - Trove*. Retrieved from <https://trove.nla.gov.au/work/6339833>
- 953 Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices:
954 High-density electrical mapping and source-analysis of the multisensory mismatch
955 negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*(3), 587–97.
956 <http://doi.org/10.1016/j.neuropsychologia.2006.03.036>
- 957 Sauseng, P., Klimesch, W., Schabus, M., & Doppelmayr, M. (2005). Fronto-parietal EEG
958 coherence in theta and upper alpha reflect central executive functions of working
959 memory. *International Journal of Psychophysiology*, *57*(2), 97–103.
960 <http://doi.org/10.1016/j.ijpsycho.2005.03.018>
- 961 Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding
962 through neural coherence: implications for multisensory processing. *Trends in*
963 *Neurosciences*, *31*(8), 401–9. <http://doi.org/10.1016/j.tins.2008.05.002>
- 964 Silverman, B. W. (1981). Using Kernel Density Estimates to Investigate Multimodality.
965 *Journal of Royal Statistical Society*, *43*, 97–99. Retrieved from
966 <https://www.stat.washington.edu/wxs/Stat593-s03/Literature/silverman-81a.pdf>
- 967 Stefanescu, R. A., & Jirsa, V. K. (2008a). A Low Dimensional Description of Globally
968 Coupled Heterogeneous Neural Networks of Excitatory and Inhibitory Neurons. *PLoS*
969 *Computational Biology*, *4*(11), e1000219. <http://doi.org/10.1371/journal.pcbi.1000219>
- 970 Stefanescu, R. A., & Jirsa, V. K. (2008b). A low dimensional description of globally coupled
971 heterogeneous neural networks of excitatory and inhibitory neurons. *PLoS*
972 *Computational Biology*, *4*(11), e1000219. <http://doi.org/10.1371/journal.pcbi.1000219>
- 973 Thakur, B., Mukherjee, A., Sen, A., & Banerjee, A. (2016). A dynamical framework to relate
974 perceptual variability with multisensory information processing. *Scientific Reports*, *6*(1),
975 31280. <http://doi.org/10.1038/srep31280>
- 976 Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of
977 brain electrical activity via linearly constrained minimum variance spatial filtering.
978 *IEEE Transactions on Biomedical Engineering*, *44*(9), 867–880.
979 <http://doi.org/10.1109/10.623056>
- 980 van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural
981 processing of auditory speech. *Proceedings of the National Academy of Sciences of the*
982 *United States of America*, *102*(4), 1181–6. <http://doi.org/10.1073/pnas.0408949102>
- 983 van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in
984 auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607.
985 <http://doi.org/10.1016/j.neuropsychologia.2006.01.001>
- 986 von Stein, A., Rappelsberger, P., Sarnthein, J., & Petsche, H. (1999). Synchronization
987 between temporal and parietal cortex during multimodal object processing in man.
988 *Cerebral Cortex (New York, N.Y. : 1991)*, *9*(2), 137–50.
- 989 Wallace, M. T., Ramachandran, R., & Stein, B. E. (2004). A revised view of sensory cortical
990 parcellation. *Proceedings of the National Academy of Sciences of the United States of*
991 *America*, *101*(7), 2167–72. <http://doi.org/10.1073/pnas.0305697101>
- 992 Williams, P. E., Mechler, F., Gordon, J., Shapley, R., & Hawken, M. J. (2004). Entrainment
993 to Video Displays in Primary Visual Cortex of Macaque and Humans. *Journal of*
994 *Neuroscience*, *24*(38), 8278–8288. <http://doi.org/10.1523/JNEUROSCI.2716-04.2004>
- 995 Wilson, H. R., & Cowan, J. D. (1972). Excitatory and Inhibitory Interactions in Localized
996 Populations of Model Neurons. *Biophysical Journal*, *12*(1), 1–24.
997 [http://doi.org/10.1016/S0006-3495\(72\)86068-5](http://doi.org/10.1016/S0006-3495(72)86068-5)
- 998 Winter, W. R., Nunez, P. L., Ding, J., & Srinivasan, R. (2007). Comparison of the effect of
999 volume conduction on EEG coherence with the effect of field spread on MEG
1000 coherence. *Statistics in Medicine*, *26*(21), 3946–3957. <http://doi.org/10.1002/sim.2978>

1001 Yeterian, E. H., & Pandya, D. N. (1989). Thalamic connections of the cortex of the superior
1002 temporal sulcus in the rhesus monkey. *The Journal of Comparative Neurology*, 282(1),
1003 80–97. <http://doi.org/10.1002/cne.902820107>

1004

1005 **Acknowledgements**

1006 The study was supported by NBRC Core funds and by grants Ramalingaswami fellowship,
1007 (BT/RLF/Re-entry/31/2011) and Innovative Young Bio-technologist Award (IYBA),
1008 (BT/07/IYBA/2013) from the Department of Biotechnology (DBT), Ministry of Science and
1009 Technology, Government of India to AB. AB also acknowledges the support of Centre of
1010 Excellence in Epilepsy and MEG (BT/01/COE/09/08/2011) from DBT. DR was supported by
1011 the Ramalingaswami fellowship (BT/RLF/Re-entry/07/2014) from DBT.

Figures

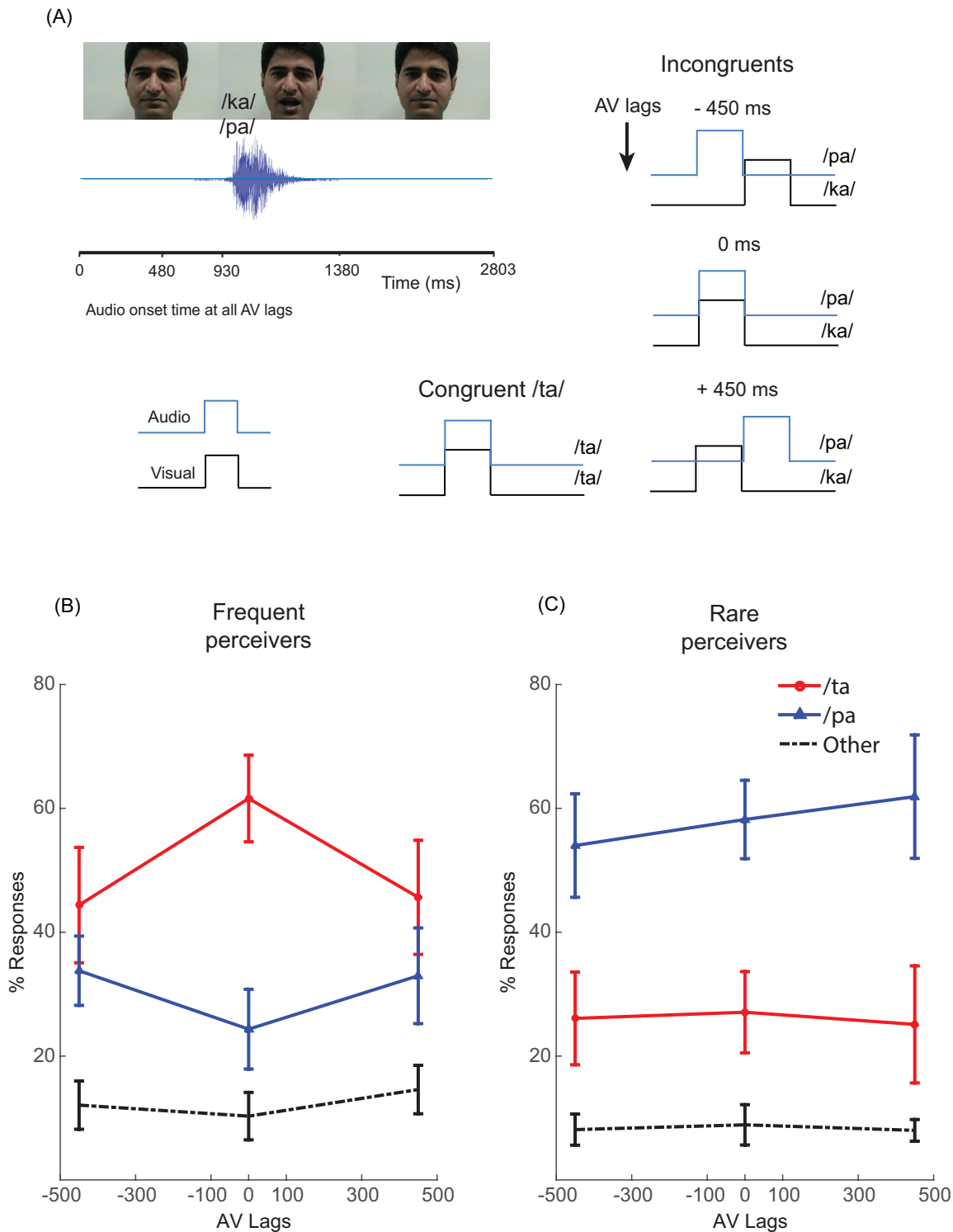


Figure 1: Experimental setup and behavior: (A) Video frames from the stimulus showing neutral face at the stimulus onset and the facial gesture during articulation (B) The McGurk stimuli: Audio /pa/ superimposed onto the lip movement /ka/ presented with AV lags -450 ms, 0 ms and +450 ms and the congruent stimulus: Audio /ta/ superimposed onto the lip movement /ta/. The location of the onset of the audio is place with respect to the articulator's initiation of lip movement. (C) Group percentage distribution of the perceptual responses (/ta/, /pa/ and others) in frequent and rare perceivers.

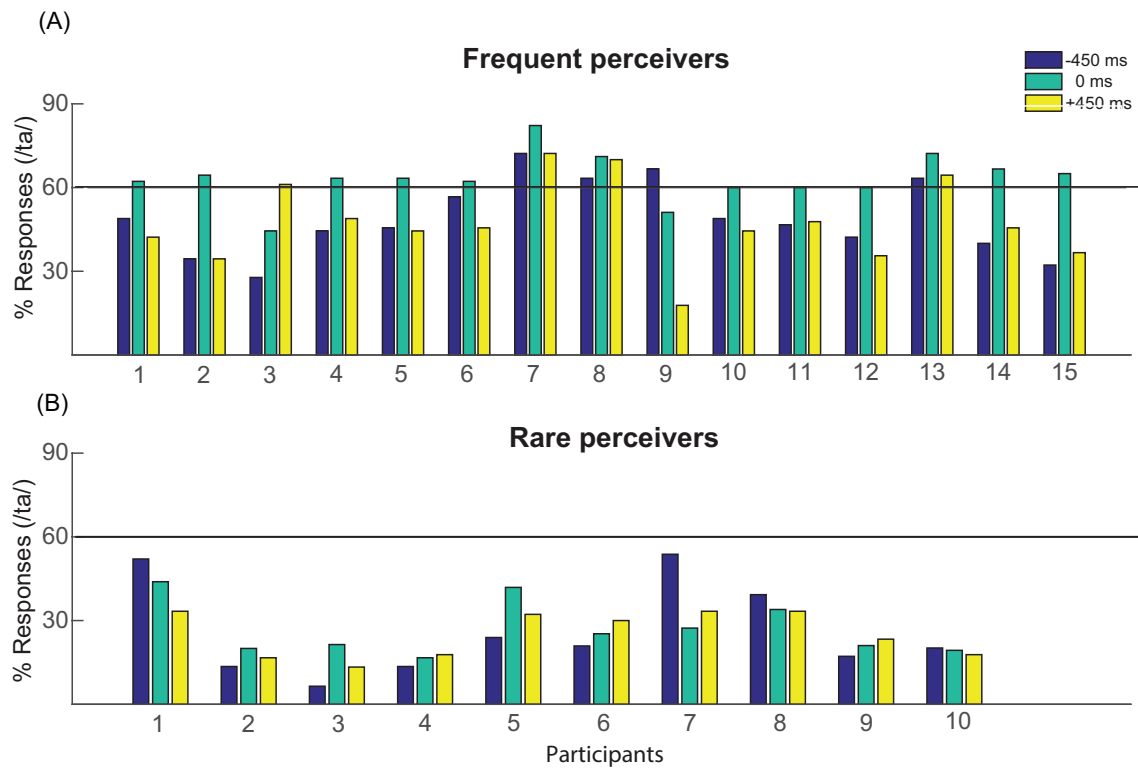


Figure 1 Supplement 1: Individualist participant behavior: Percentage of /ta/ responses during the 0 ms and 450 ms AV stimuli in (A) Frequent perceivers (B) Rare perceivers.

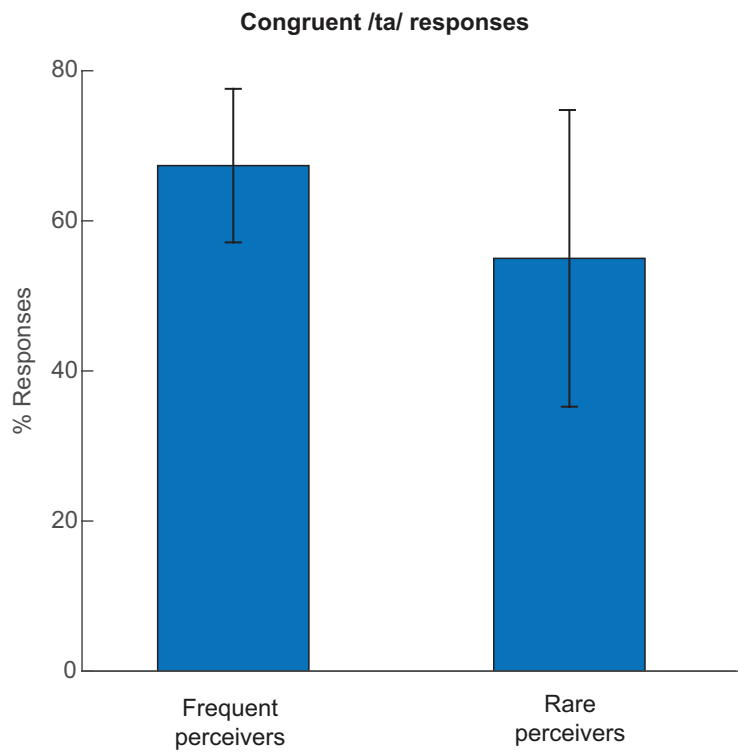


Figure 1 Supplement 2: Hit rate during congruent /ta/ stimulus: The percentage of /ts/ responses trail-by-trial across the participants.

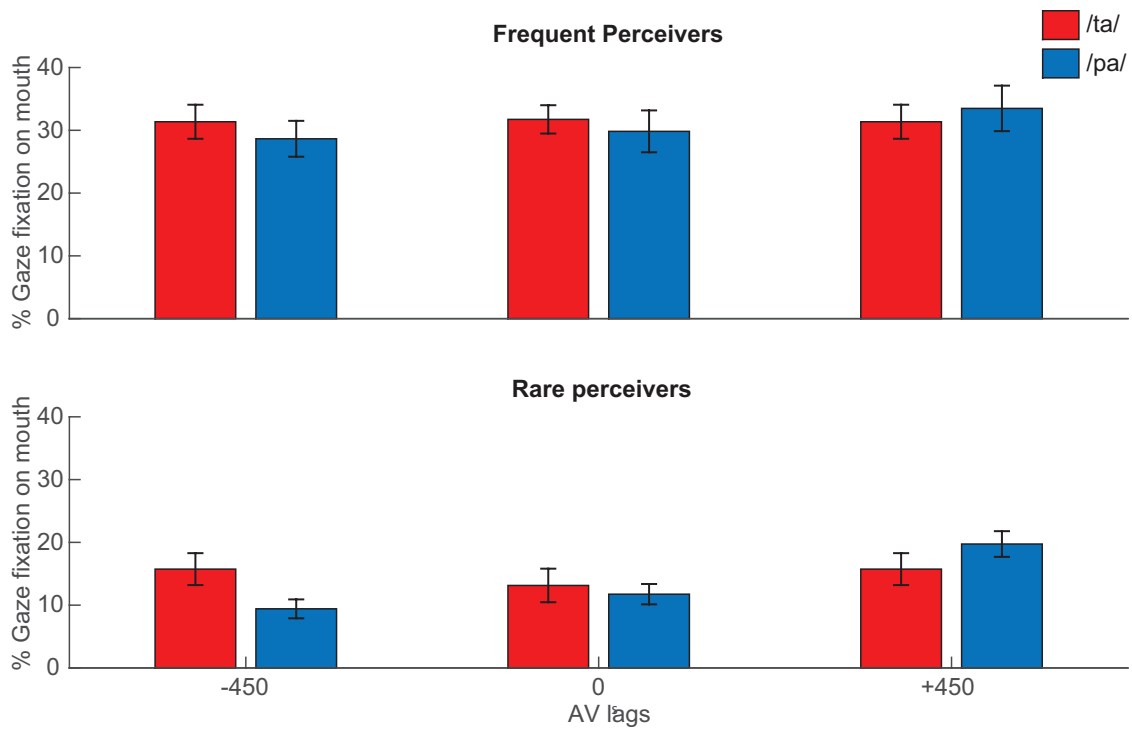


Figure 1 Supplement 3: Gaze behavior: Percentage of gaze fixations on the mouth of the articulator in the AV stimuli averaged trial-by-trial across the participants (A) Frequent perceivers (B) Rare perceivers.

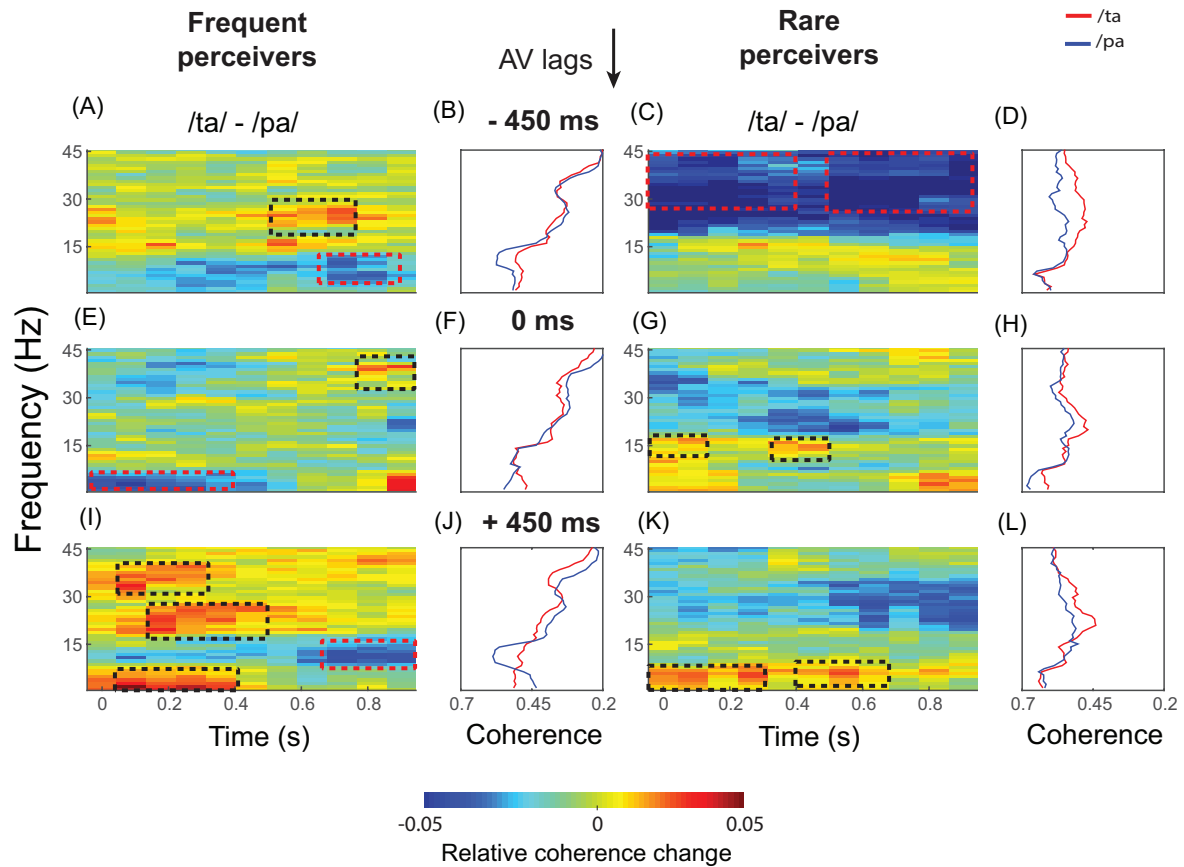


Figure 2: Large scale functional connectivity dynamics observed in sensor time series: Global coherence differences between the perceptual categories (/ta/ and /pa/) and time averaged global coherence respectively during /ta/ and /pa/ perception in frequent and rare perceivers at -450 ms (A,B,C,D), 0 ms (E,F,G,H) and +450 ms (I,J,K,L) AV lag.

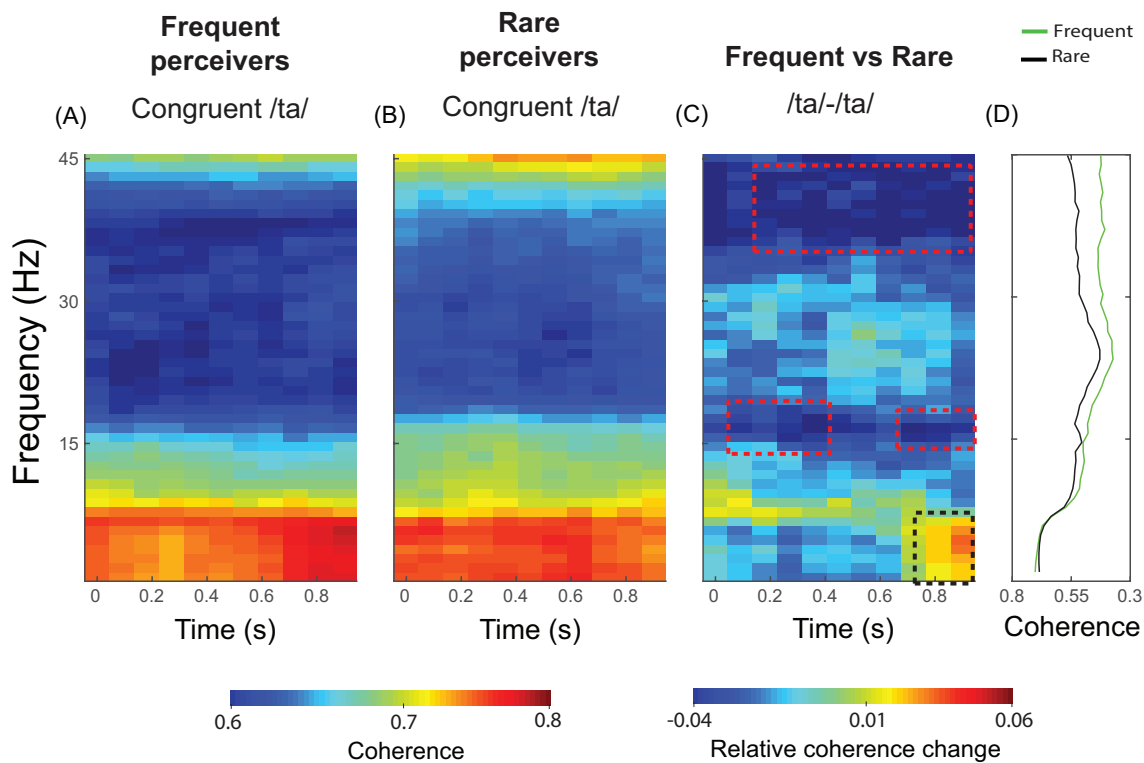


Figure 3: Large-scale functional connectivity dynamics during congruent /ta/: (A) Global coherogram during congruent /ta/ perception in (A) Frequent perceivers (B) Rare perceivers (B) Global coherogram difference between frequent and rare perceivers (D) Time averaged global coherence during /ta/ and /pa/ perception in frequent and rare perceivers.

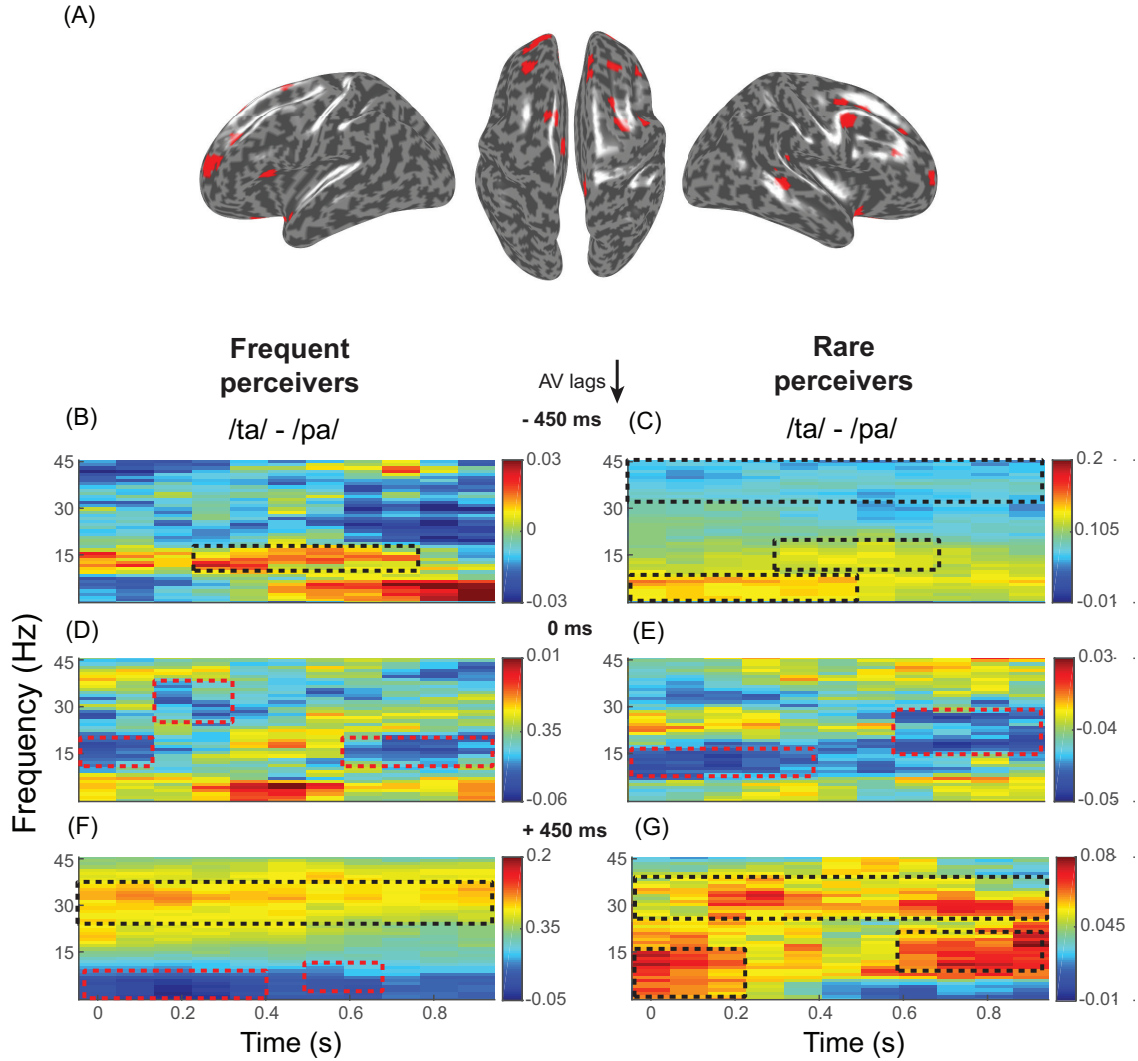


Figure 4: Source reconstruction: (A) Sources identified using the LCMV beamformer algorithm from the sensor time series. The source power of the ratio between /ta/ and /pa/ eliciting power than the set threshold (>99.5 percentile) are highlighted. Global coherogram differences between the perceptual categories (/ta/ and /pa/) computed from the source-time series in frequent and rare perceivers during - 450 ms (B,C), 0 ms (D,E) and +450 ms (F,G) AV lag.

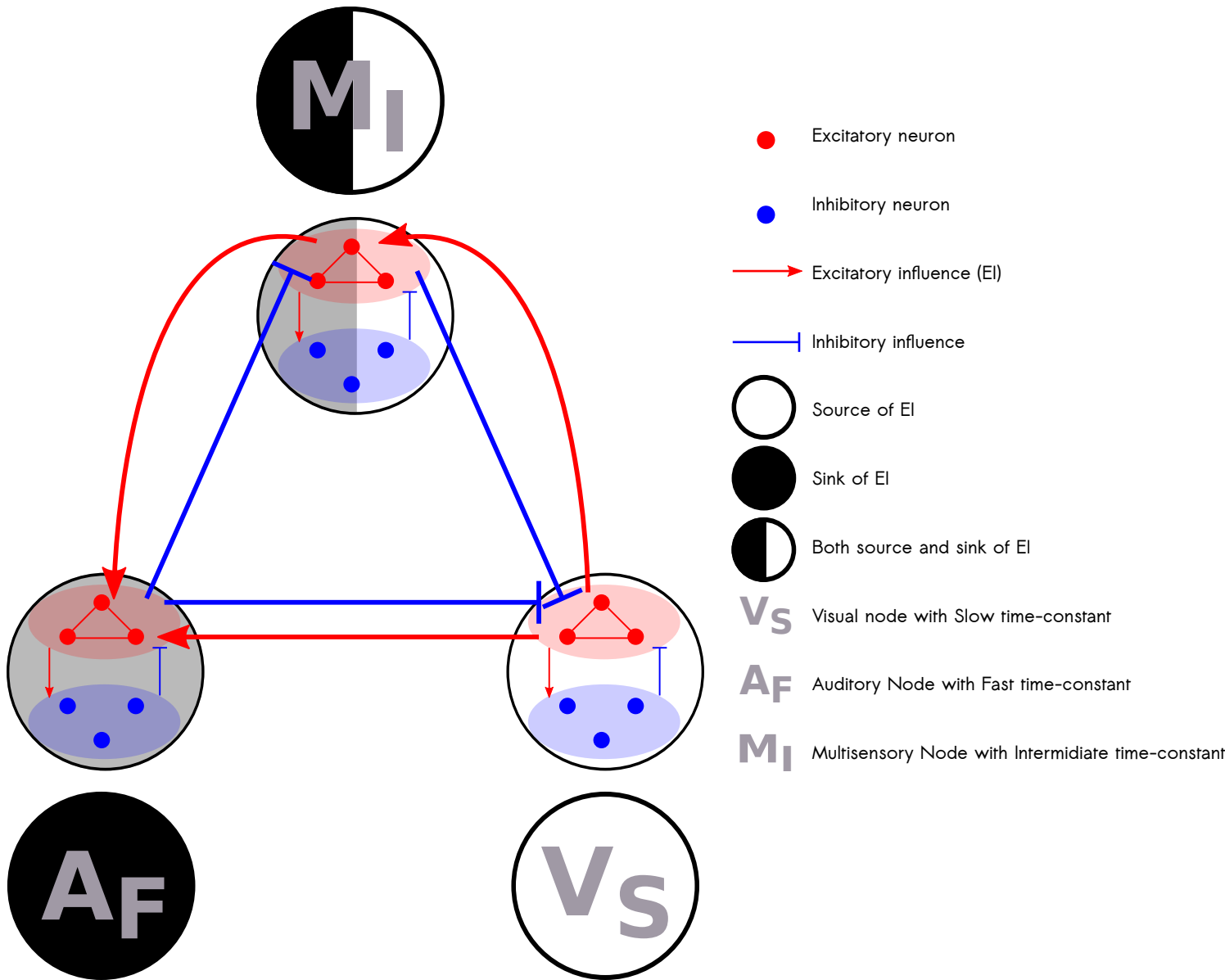


Figure 5: Large scale dynamical model consisting of a network three neural masses with different time-constants: The model comprises three nodes representing auditory (fast time-constant), visual (slow time-constant) and higher order multisensory regions (intermediate time-constant). Each node consists of network of 100 Hindmarsh-Rose excitatory and 50 inhibitory neurons. Each neuron can exhibit isolated spiking, periodic spiking and bursting behavior. Excitatory influences between the nodes are balanced by their inhibitory counterpart. The source and sink represent the flow of excitatory influence.

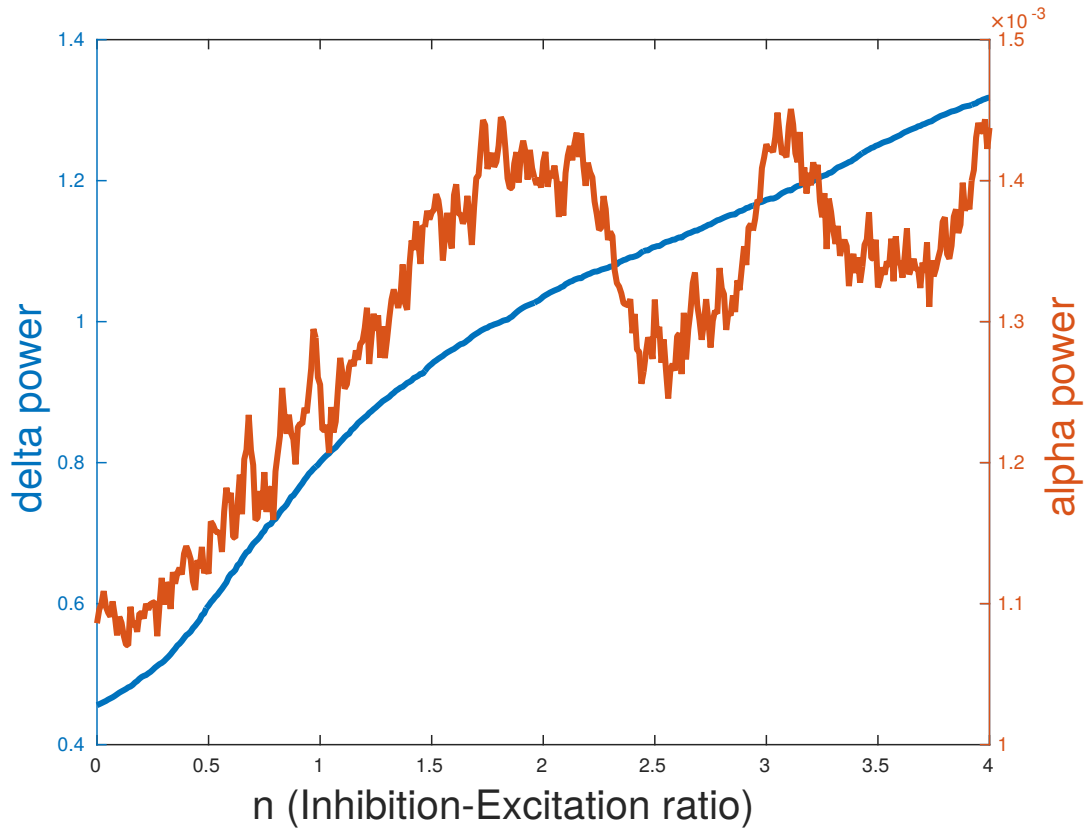


Figure 5 Supplement 1: Selection of Inhibition-Excitation ratio (n): Delta and alpha power when the nodes are disconnected and driven by baseline current ($I=0.1$). Selection of n was made in order to have comparatively higher delta and alpha power.

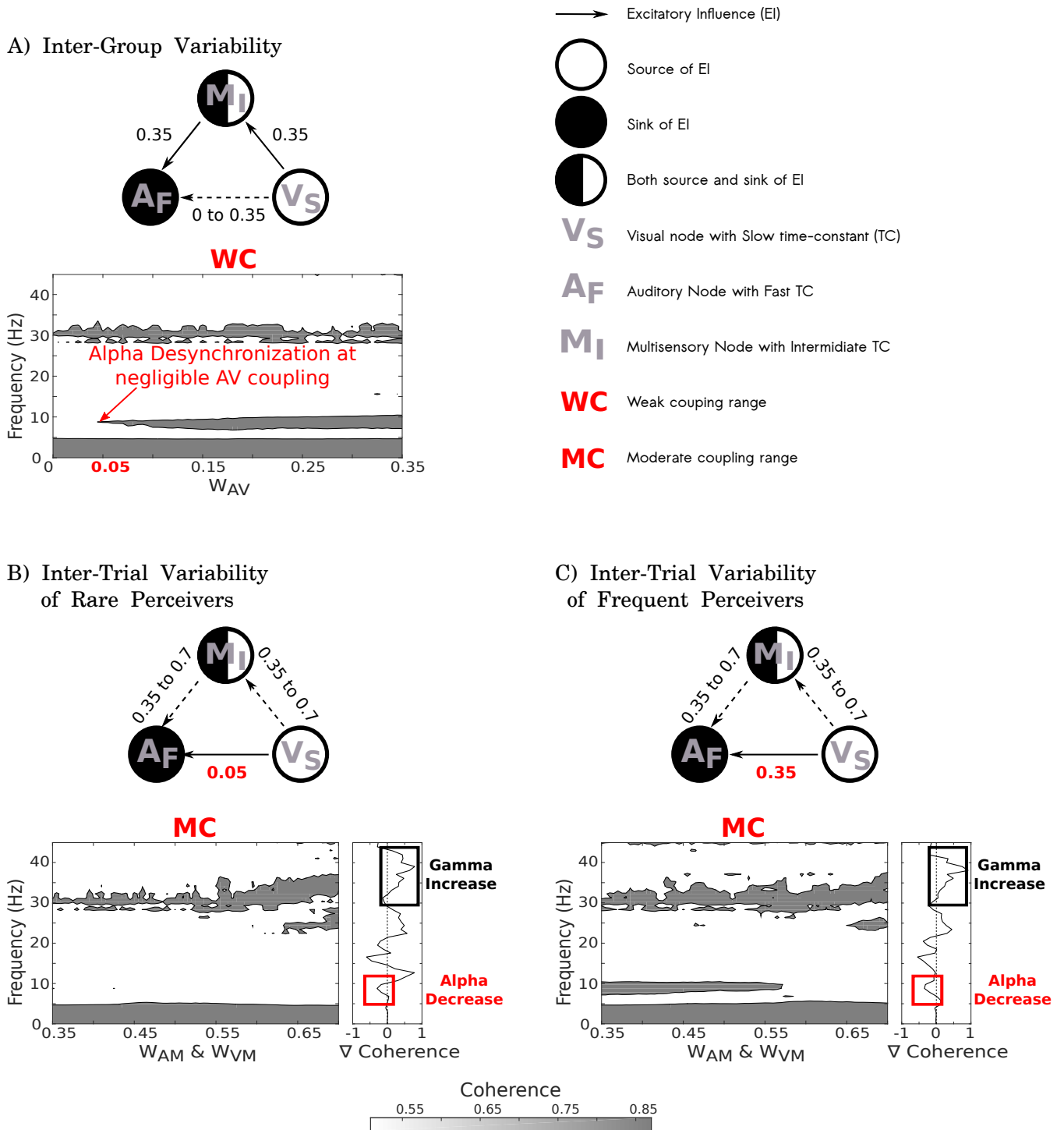


Figure 6: Mechanistic understanding of Inter-individual and inter-trial variability: A) Alpha de-synchronization characteristic of rare perceivers resulted due to negligible A-V coupling. B) & C) Enhanced gamma coherence and reduced alpha coherence observed in illusory perception is due to increase in indirect coupling involving multisensory node irrespective of the influence of direct A-V coupling.

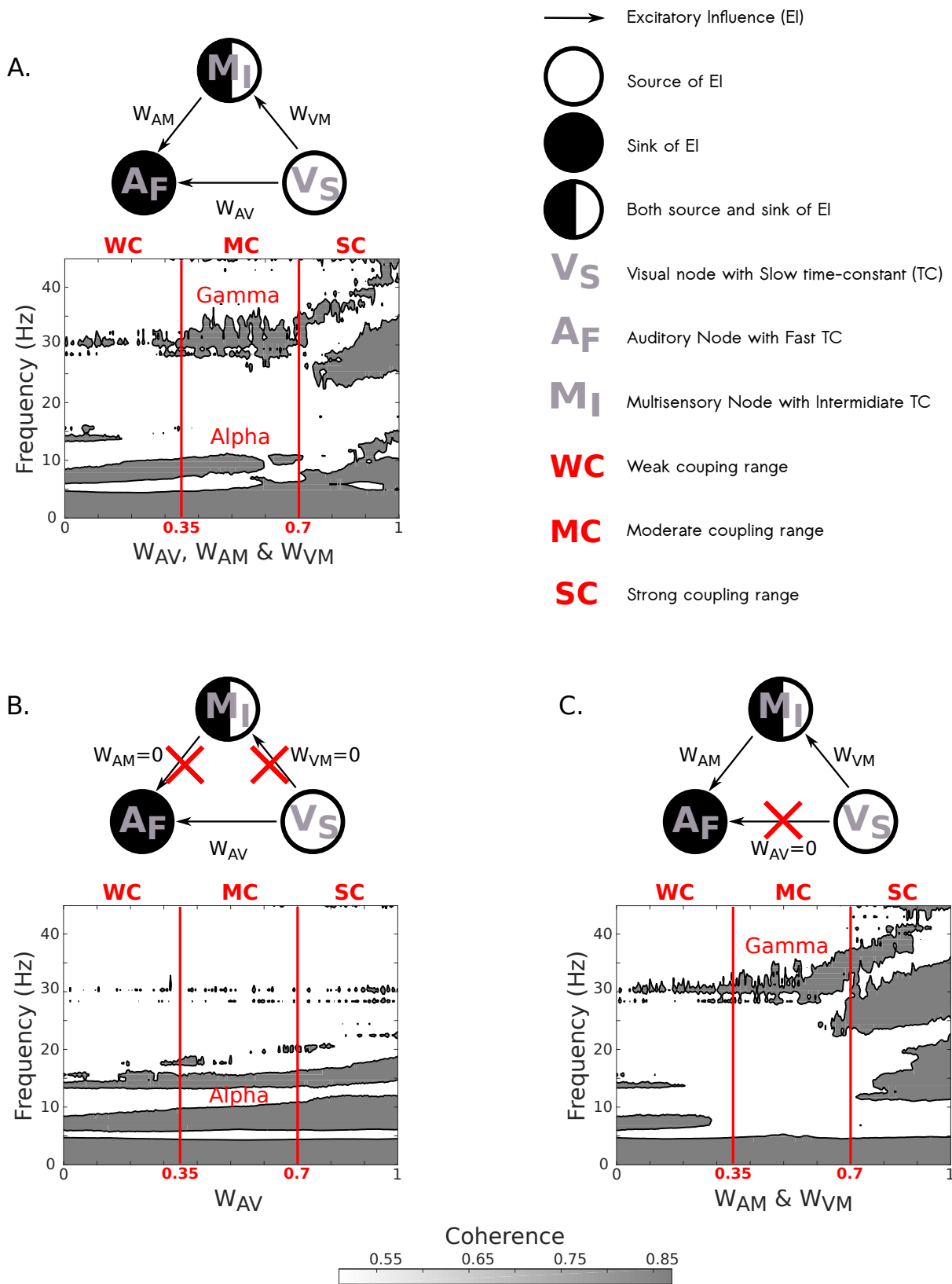


Figure 6 Supplement 1: Prediction of alpha and gamma coherences from neural mass model: A) Alpha and gamma band coherence co-exist in moderate coupling range. B) Only direct A-V coupling generates alpha coherence independently. C) Indirect A-V coupling via multisensory node generates gamma coherence at the limit case scenario of weak direct coupling.

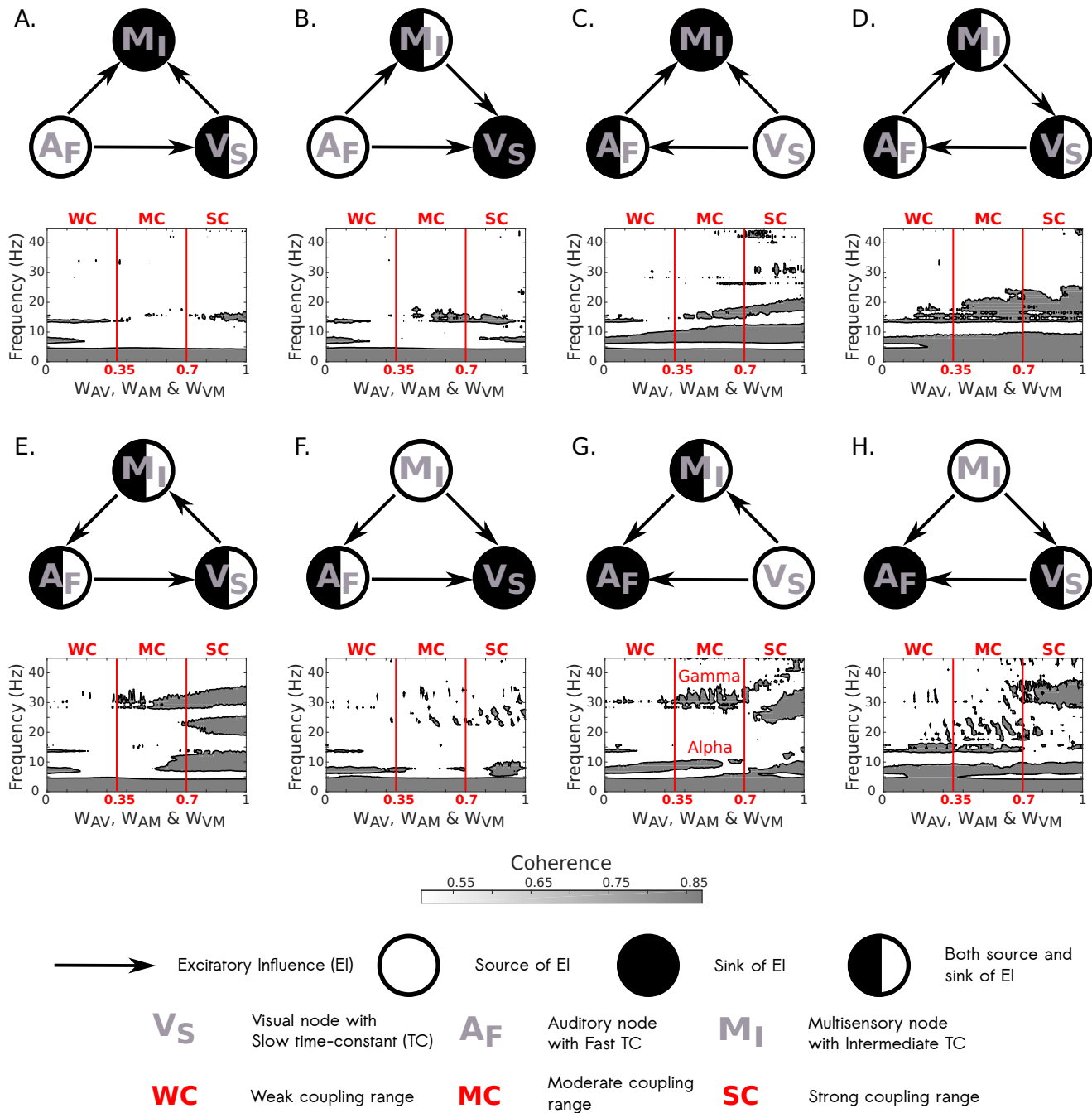


Figure 6 Supplement 2: Global coherence for different source-sink combinations: G) Only when visual node is source and auditory node is the sink (as in our model), we observe co-existence of alpha and gamma band coherence in moderate coupling range. A)-F) and H) Exploration of various coupling scenarios to identify if it is possible to generate alpha and gamma coherence in moderate coupling range.