Synergistic balance of local segregation and global integration mechanisms in large-scale brain networks underlies perceptual experience

Abhishek Mukherjee[†], Soibam Shyamchand Singh[†], Dipanjan Ray,
Partha Raghunathan, and Arpan Banerjee[*]

National Brain Research Centre, NH8, Manesar, Gurgaon 122052, India.

[†] *These authors contributed equally to this work (co-first authors).*

[*] *To whom correspondence should be addressed. E-mail: arpan@nbrc.ac.in.*

Perceptual processes unfold along spatial and temporal hierarchies in brain dynamics. In daily lives, speech perception requires binding of spatiotemporally disjoint auditory and visual cues by distributed brain areas. The network-level information processing involving these brain areas can be characterized by two complementary mechanisms, functional segregation which refers to localization of processing within few areas, and integration which pertains to the cooperation among relevant functional modules. Here, we demonstrate using fMRI recordings that subjective perceptual experience of multisensory speech stimuli—real and illusory—are represented in differential states of segregation–integration balance. We were able to parametrically control the inter-subject variability of illusory/cross-modal perception by introducing temporal lags in the incongruent auditory–visual articulations of speech sounds using the McGurk paradigm. The functional segregation–integration balance is captured using two alternative computational approaches. First, the segregation of brain areas as modules responsible for illusory speech perception is captured using standardized parametric statistical approaches. The functional integration was quantified from exploring the temporal correlation of the relevant functional modules with the dynamics of all other brain areas. We observed that the regions with higher activation during illusory perception tended to have lower levels of cooperation with the rest of the brain indicating high levels of local segregation, but reduced global integration. Second, using graph theoretic measures of local clustering, modularity, and participation coefficient, which capture the contribution of segregation and integration among functional modules, we further validated the observation of enhanced local segregation and diminished global integration during the illusory/cross-modal perception. Further, the perceptual binding network that exhibited this observation was anti-correlated with other intrinsic functional brain networks, such as dorsal attention and default mode networks, and most importantly, inter-subject variability of perceptual categorization could be explained from the altered balance of local segregation–global integration.

Keywords

multisensory, whole-brain functional connectivity, pSTS, fMRI, segregation, integration, homeostasis, perception, predictive coding, Bayesian

Introduction

The illusion of a puppet vocalizing human speech that in reality comes from the sounds of a ventriloquist has intrigued mankind for ages (Alais and Burr, 2004). While spatial incongruency is a well-studied aspect of ventriloquism, perceptual binding of temporally dissonant stimuli is subtle (Bertelson and Aschersleben, 2003), such as when viewing a poorly dubbed video our perceptual system immediately perceives the asynchrony in sound and lip movements even when we do not know the language. Some previous studies have also reported that temporal dissonance within a certain window between multiple sensory components can be integrated to facilitate illusory perceptual experience (Bertelson and Aschersleben, 2003; Munhall et al., 1996; Shams et al., 2002; Stevenson et al., 2012), but do collective behavior of brain networks exhibit patterns through which integrative mechanisms of brain information processing underlying perceptual binding can be understood is an unanswered question.

In this article, we hypothesized that cooperation among relevant brain areas via functional integration mechanisms (Friston et al., 1993) and their individual role in local information processing need to be at a critically balanced state to facilitate cross-modal perceptual binding, and this relationship may be parametrically controlled by behavioral contexts. Previous studies have used illusory perceptual experience from the paradigms of sound-flash illusion (Watkins et al., 2007) and McGurk illusion (McGurk and Macdonald, 1976) to suggest posterior superior temporal sulcus (pSTS) as the key hub for cross-modal percept binding (Beauchamp et al., 2004; Nath and Beauchamp, 2011), whereas other studies have not found pSTS activations to be enhanced specific to temporal synchrony of auditory and visual modalities (Jones and Callan, 2003). From a large-scale network perspective, noisy stimulus in one modality has been shown to increase functional brain connectivity between pSTS and areas processing in the other modality; for example, a blurry video will increase strength of pSTS–auditory connectivity (Nath and Beauchamp, 2011). Such shifts in connectivity arising from stimulus associations can be linked to causal inference models of perception (Koerding et al., 2007; Magnotti et al., 2018) and lately biophysically realistic network mechanisms have been proposed (Cuppini et al., 2017; Kumar et al., 2020; Thakur et al., 2016). Audio-visual speech perception also involves active participation of the motor network (Dehaene-Lambertz et al., 2006; Liberman and Whalen, 2000; Morís Fernández et al., 2017; Scott et al., 2009), and there is increased interaction between motor representational area and inferior frontal gyrus during presentation of incongruent (McGurk) speech stimuli, as compared to congruent ones. This motor network was found to contribute toward resolving the audiovisual conflicts, thus avoiding the McGurk perception (Murakami et al., 2018). The manifestation of audio-visual illusory speech perception can thus be viewed as a critical balance between functional domains promoting the integration (such as the pSTS network) and domains with more abstract functions involved in conflict resolution (such as the motor network). Finally, the existence of an altered balance in network mechanisms can explain why perception is relatively stable for a wide range of behavioral contexts (Zopf Jr, 1963), e.g., perception of illusory sound in the McGurk paradigm for different vocalizers and post-learning (Mallick et al., 2015), perceivers across cross-cultural boundaries (Magnotti et al., 2015), even though inter-individual variability exists. An empirical demonstration on how subjective perceptual binding emerges from synergistic functional integration and segregation mechanisms in brain networks (Tomasi et al., 2014) is currently elusive.

To elicit instances of cross-modal perceptual experience, the McGurk incongruent video—auditory /pa/ superimposed on the lip movement of /ka/ from a human vocalizer (McGurk and Macdonald, 1976)—was shown to volunteers (Fig 1*A*). Subsequently, subjective perception was parametrically controlled by using different versions of this video created by introducing temporal lags between the

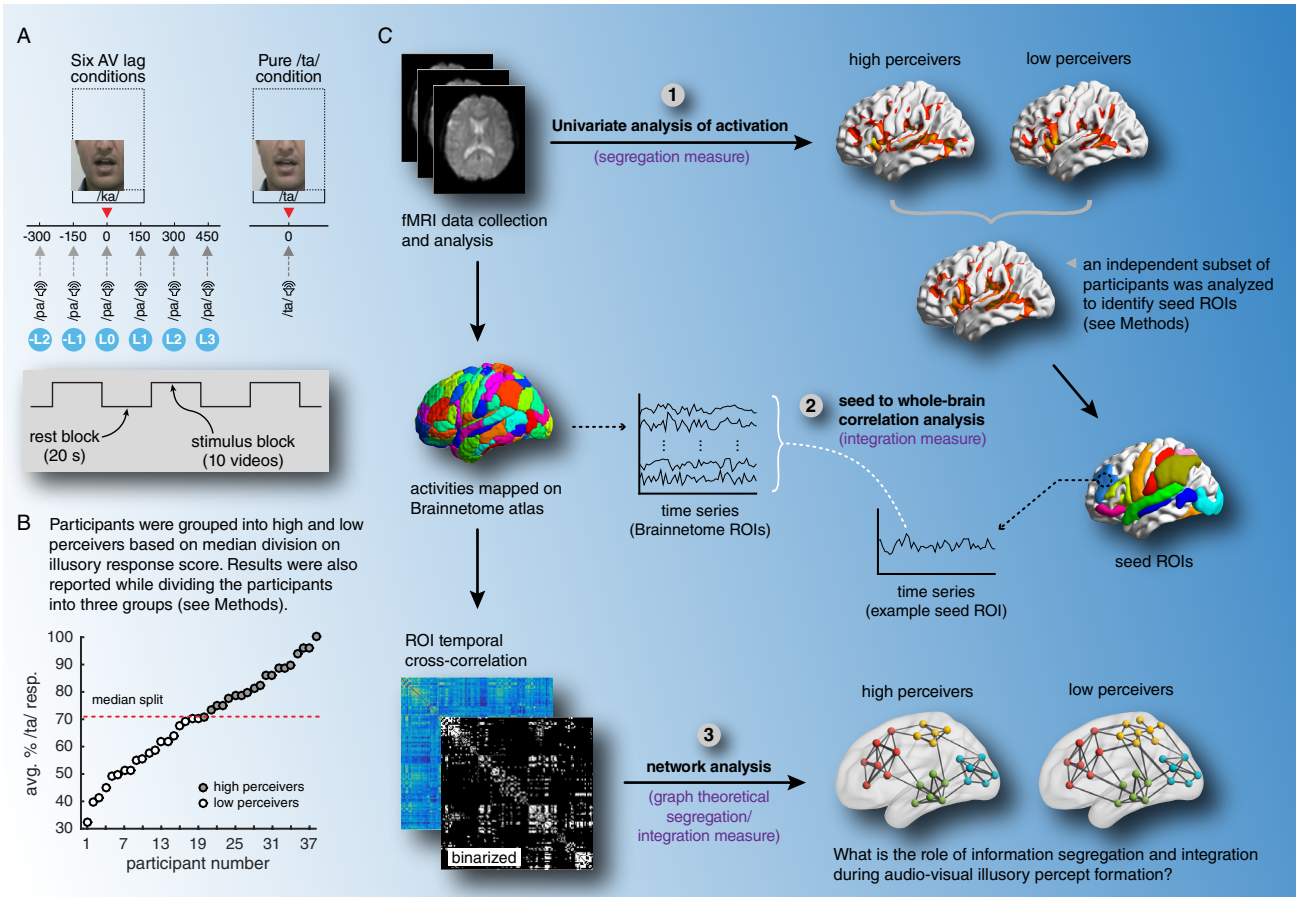

Figure 1. (*A*) Stimulus design using the McGurk paradigm (*B*) Intersubject variability in susceptibility for illusory perception for all 38 selected volunteers; the susceptibility was evaluated as the average of two maximal percentage /ta/ responses in any two AV lags (dotted line indicates the division between high and low perceivers). (*C*) Schematic showing various analysis steps, in particulars the dfifferent measures used to identify the integration/ segregation balance of cortical networks

auditory and visual sensory streams. The temporal lag ranges from −300 (A leading V) to 450 ms (V leading A) in intervals of 150 ms (Fig. 1*A*); also shown are matched number of congruent /ta/ (auditory /ta/–visual /ta/) presented at 0 ms lag. The choosing of an asymmetrical range of audio-visual (AV) lags (−L2, . . . , L0, . . . , L3) was guided by findings from a previous study (Munhall et al., 1996) reporting about a bias in perceptual experience toward visual-leading scenarios. Volunteers reported whether they heard /ta/, /pa/, or "others" in a forced-choice paradigm. Thus, a behavioral control of subjective illusory /ta/ perception, the hallmark of cross-modal perceptual binding, was achieved with AV lags as the control parameter (Fig. 2*A*; Supplementary Figs. 1 and 2). To understand the underlying mechanism, the fMRI data were analyzed through both stimulus-driven as well as perceptual-based approaches (see *Methods* for details).

Cognitive tasks generally involve distributed brain areas, and identifying integration and segregation characteristics (Tomasi et al., 2014; Tononi et al., 1994) of large-scale brain networks is increasingly becoming important to understand the information processing (Cohen and D'Esposito, 2016; Mohr et al., 2016). A more riveting hypothesis is that normative human cognition can be defined as a homeostatic balance between integrative and segregative mechanisms with imbalances

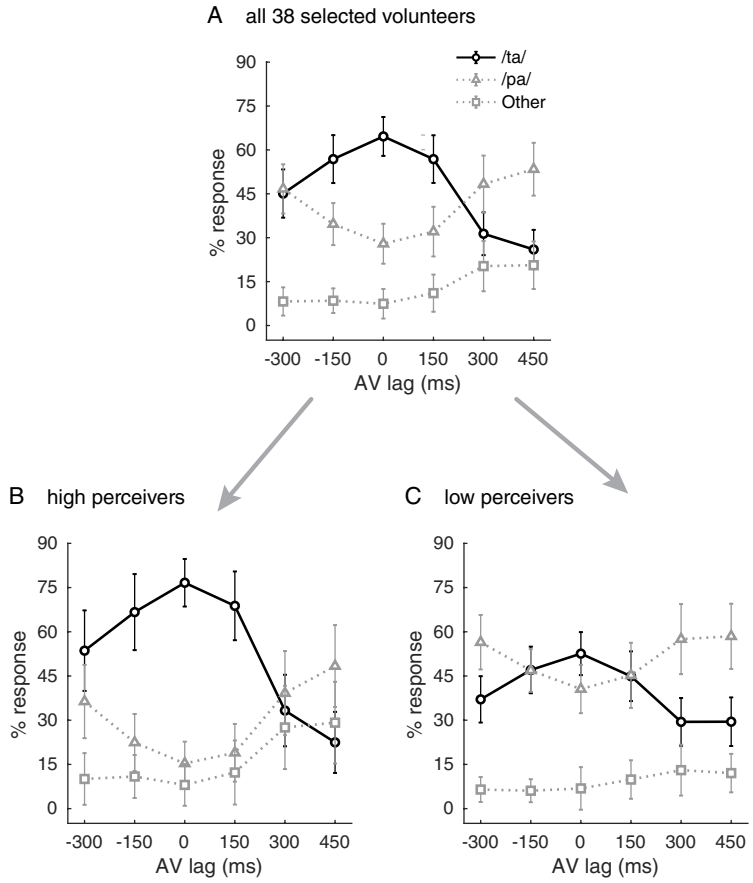

Figure 2. (*A*) Behavioral responses averaged across all 38 selected volunteers. Behavioral responses averaged across (*B*) high perceivers ($N = 19$) and (*C*) low perceivers ($N = 19$). Error bars represent the standard error of the mean (95% CI) across the respective population.

leading to neuropsychiatric disorders (Lord et al., 2017). The present study demonstrates how such balanced states of network segregation and integration facilitates perceptual binding of incongruent audio-visual/multisensory stimuli. This is a significant departure from the earlier understanding of multisensory integration from the perspective of inverse effectiveness which has been reinforced by several studies using different neuroimaging modalities (Meredith and Stein, 1983; Molholm et al., 2002; Stevenson and James, 2009).

Results

Parametric Control of Illusory Perception. More than chance level responses to illusory /ta/ (cross-modal perception) were observed in all participants at multiple AV lags for the McGurk incongruent stimulus, demonstrating that indeed the degree of AV temporal synchrony affects subjective perception (Fig. 2 and Supplementary Figs. 1 and 2). In order to study the neural basis of subjective variability, we categorize the participants into two equally sized groups ($N = 19$) while sorting them based on their *susceptibility* to audio-visual illusion. The susceptibility is measured as the average of two maximal percentage /ta/ responses in any two AV lags. The group reporting higher illusory perception is termed as *high perceivers* and the other group at the lower end as *low perceivers* (Fig. 1*B*).

A two-way analysis of variance (ANOVA) was further conducted to reveal the effect of AV lag and subject grouping factors on participants' perceptual responses. In the case of illusory /ta/ responses, there were significant main effects of AV lag ($F(5, 213) = 17.9202, p < 0.0001$), as well as of subject grouping ($F(1, 213) = 20.1903, p < 0.0001$) at 95% confidence levels (a confidence interval used in all statistical significance tests in the article). We also observed statistically significant interaction between the two factors ($F(5, 213) = 2.9257, p = 0.0141$); Tukey post-hoc comparison further showed that high perceivers were more succeptible to illusory /ta/ perception specifically at L0 ($p = 0.0459$) and L1 ($p = 0.0493$) conditions. For the unisensory /pa/ response, there was no significant interaction between the two factors ($p = 0.7087$). However, significant main effects of AV lags ($F(5, 213) = 6.6862, p < 0.0001$), as well as of subject grouping ($F(1, 213) = 40.7071, p < 0.0001$) were observed. Similarly, no significant interaction between the two factors was observed while reporting "other" responses also ($p = 0.3774$), but we again observed significant main effects of AV lags ($F(5, 213) = 3.5001, p = 0.0046$) and participant grouping ($F(1, 213) = 7.3264, p = 0.0073$).

Brain Mapping of Perceptual Binding. Analysis of fMRI images using statistical parametric mapping (SPM) software (Friston et al., 1995) was performed with the task (AV lag blocks) and baseline (inter-trial rest blocks) conditions (Fig. 1). Group-level inference from random effect analysis revealed that maximal brain areas were active during L0 and L1 conditions, where propensities to cross-modal illusory perception were also maximally reported (Fig. 3). The extent of task-activated brain areas diminished bilaterally with increasing AV separation (Fig. 3 and Supplementary Tables 1 and 2). Among those perceptual-sorted categories, Max /ta/ condition showed maximal brain activation. Furthermore, when compared with high perceivers, the low perceivers showed relatively weaker activation specially in primary motor, right frontal, anterior cingulate cortex, superior temporal region including pSTS, rostral inferior parietal, and occipital areas; on the other hand, orbitofrontal, insula, and thalamic regions showed higher activation (see Supplementary Table 2).

To study the differential functional roles of associated brain regions during audio-visual cross-modal perception, 32 seed regions of interest (ROIs) were defined using another independent subset of 29 participants, each of whom reported 60% or more maximal illusory /ta/ response in at least one AV lag condition (Fig. 1 *C*; Supplementary Tables 3 and 4; see also *Methods* for more details). We will henceforth term these 32 seed ROIs which were together active during the audio-visual illusory (cross-modal) percept formation as the perceptual binding network (PBN). The number of active voxels and percentage signal difference (PSD) were estimated in all ROIs of PBN for both high and low perceivers (Supplementary Figs. 3–6). Peak PSDs for high perceivers were observed at either L0 or L1 condition (except for bilateral MVOcC and thalamus), and at Max /ta/ condition (except for left MFG, bilateral ACC, MVOcC, and thalamus). Moreover, for low perceivers, relatively higher and rather invariant PSDs were observed toward video-leading stimulus conditions, and peak values were observed at either L1 or L2 (Supplementary Fig. 6). When compared with the high perceivers, low perceivers showed overall decrease of PSDs bilaterally in STG and LOcC, and in the left MVOcC; left hemispheric ROIs, such as MFG, IFG, FuG, IPL, and bilateral pSTS also displayed decreased PSD at L0, L1, and L2. Right PrG seed region on the other hand showed increased PSD in low perceivers. Interestingly, for the perceptually sorted categories minimal difference in seed region PSDs were observed between Max /ta/ and unisensory Pure /ta/ condition in the case of low perceivers (Supplementary Fig. 6).

Synergistic Network Organization. To investigate the functional integration mechanisms in the 32 seed regions of PBN, the seed-based whole-brain functional connectivity (SWFC) for each 32 seed was computed from parcellated time series (Fig. 1*C*; see also *Methods* for details). The aver-

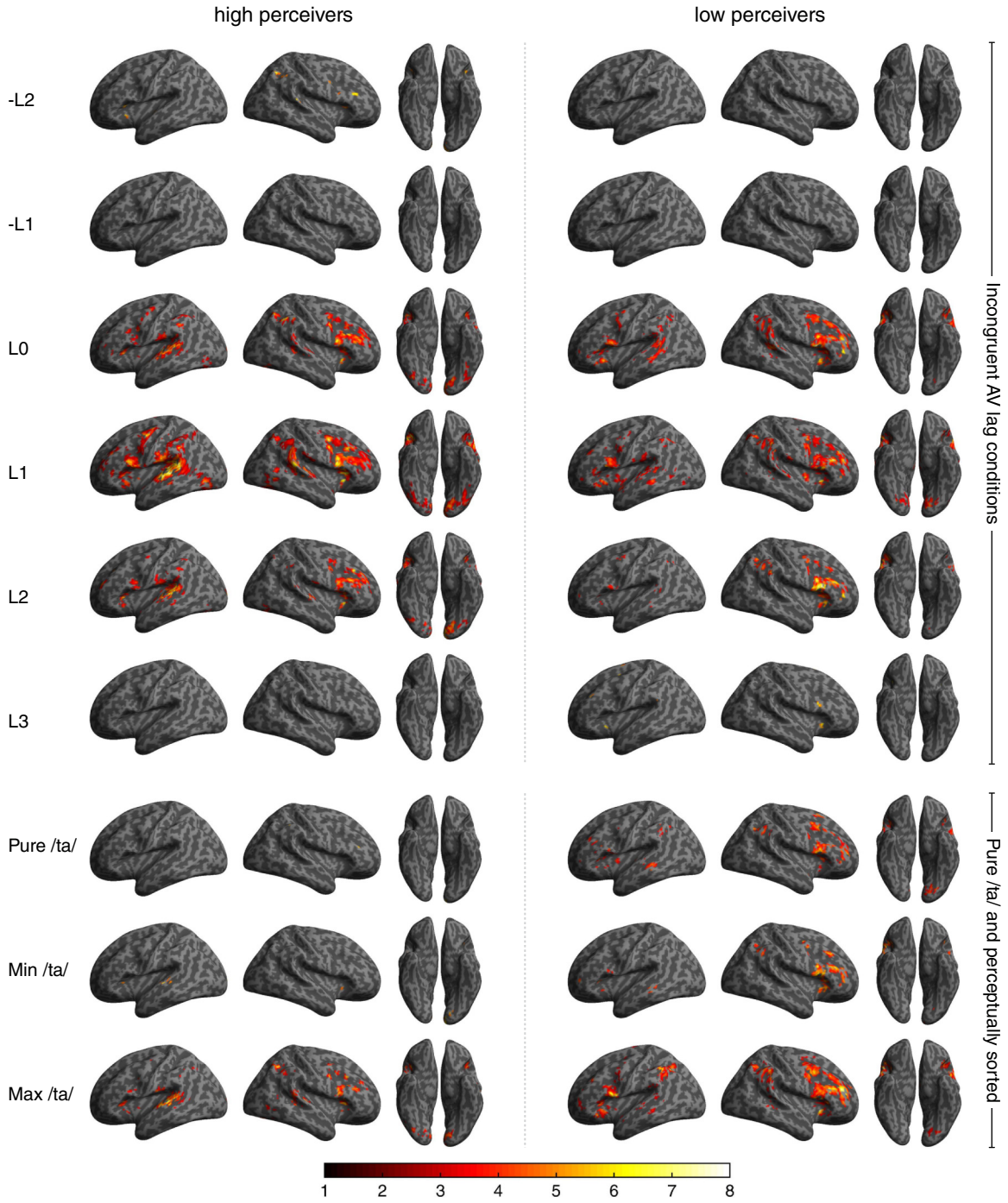

Figure 3. Group-level task-enhanced (task > rest) active brain areas for high and low perceivers (peak $p \leqslant 0.05$, FDR corrected). Color bar represents the T-score.

aged SWFC decreased maximally at L0 followed by L1 (Fig. 4 *A* and *C* shows example pSTS seeds). For high perceivers, among perception-sorted categories, Max /ta/ exhibited lowest SWFC followed by Min /ta/ and Pure /ta/ (Fig. 4*B*). However, for low perceivers, Max /ta/ and unisensory Pure /ta/ displayed comparable SWFC values (Fig. 4*D*). The spatial organization of SWFC with pSTS as example seed parcels is also shown in Supplementary Figs. 7 and 8. The question now is whether this decrease in SWFC can be meaningfully associated with the propensity for illusory perception. To meet this

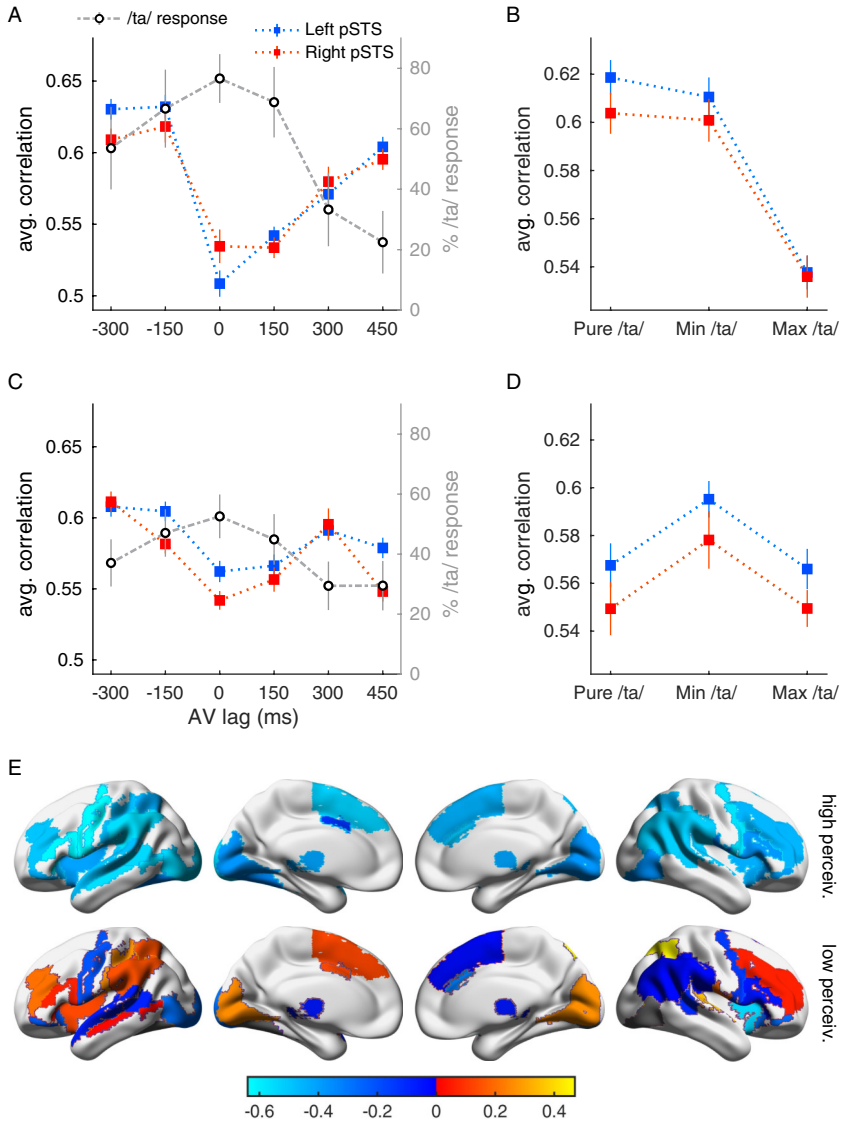

Figure 4. Average seed-based whole-brain functional connectivity (SWFC) of example pSTS seeds for high perceivers: across AV lag conditions (*A*), and Pure /ta/ and perceptually sorted categories (*B*). Panels (*C*) and (*D*) represent the corresponding SWFC for low perceivers. Error bars represent the standard error of the mean (95% CI). (*E*) Surface plot showing all 32 seed ROIs of the PBN, color mapped according to SWFC–behavior correlation (SBC) values for the high perceivers (upper) and low perceivers (lower) (Supplementary Table 5; see also *Methods* for details).

end, SWFC–behavior correlation (SBC; see *Methods*) was computed for each PBN seed regions. We observed that for high perceivers all the 32 parcels displayed negative SBC, which means the SWFC as a function of AV lags is anti-correlated with the propensity for illusory /ta/ perception. Seed ROIs in primary motor areas, left inferior frontal region, superior temporal region including pSTS, inferior parietal area, and lateral occipital area displayed the strongest anti-correlation (Supplementary Fig. 9 and Supplementary Table 5). In the case of low perceivers, several seed regions displayed positive SBC values or comparatively weaker negative SBC (Supplementary Fig. 9 and Supplementary Table 5). When the participants are regrouped into three smaller categories (high perceivers, $N = 13$; medium perceivers, $N = 12$; low perceivers, $N = 13$; see also Supplementary Fig. 2*B*), the SBC

values also mostly scaled with the propensity for illusory perception in each category (Supplementary Fig. 10). This also further emphasized that globally decreased information integration through the PBN seed regions was essential for the illusory audio-visual percept formation.

To further address the main hypothesis that perceptual binding is effectuated by an altered balance of network mechanisms, functional segregation and integration, we focus on the topological features of the whole-brain functional network using metrics from graph theory. For this analysis, 50 binarized whole-brain network subsamples were constructed for each stimulus condition by randomly permutating 70% of the participants in each perceiver's group—high and low. For estimating the network segregation, averaged local clustering coefficient and modularity measures were used; the integration measures include the characteristic path length and averaged node participation coefficient. Each network measure was further normalized with estimated measures from randomized networks (see *Methods* for more details). In the case of high perceivers, the normalized clustering coefficient (C_n) increased with decreasing AV lag and peaked at L0 where the reported average propensity to illusory perception is also the highest (Fig. 5*A*). This indicates that there is increased local cohesiveness in the network with increasing illusory perception. However, contrary to this effect is the increase in normalized characteristic path length (L_n), indicating a decrease in global cohesiveness. The manifestation of audio-visual illusion might thus require a critical balance between the segregrative (local cohesion) and integrative (global cohesion) information processing. Consequently, higher degree of illusory experience is associated with increasing dissociation among the functional modules, as reflected from the increased normalized modularity score (Q_n). The functional dissociation can also be interpreted from the decrease in average node participation coefficient (P) with increasing illusory perception (Fig. 5*A*). Thus, the balance between functional segregation and integration can be characterized from the node-level measure (e.g., local clustering coefficient, participation coefficient), and also from the whole-network-level measure (e.g., normalized modularity score). Low perceivers also displayed a similar trend, however the maximum values of clustering coefficient, characteristic path length, and modularity, and also the minimum value of participation coefficient were observed in L1 condition. The network differences between high and low perceivers is summarized in Fig. 5*B*, which displays only the left hemispheric brain networks at L0 condition: for example, the modules in the frontal and temporal brain regions show increased within-module connectivity both in number and strength (cf. blue arrow), and decrease in-between-module connectivity (cf. magenta and black arrows). However, except for the L0 condition, the increased illusory effect in high perceivers could not be interpreted from the difference in the estimated network measures. One possible explanation for this inconsistency is the difference in the activation patterns between the two groups. More specifically, reduced activation was observed in many sensory as well as associative cortical regions of low perceivers, and this difference in activation pattern might shift the segregation–integration balance accordingly. The graph theoretic measures for high perceivers overall delivered a similar take-home message, as observed for SWFC with behavior, that L0 condition triggers an increased local segregation realized via increase in clustering coefficient, characteristic path length, and modularity; and also the minimum participation coefficient. When sorted across perceptual categories of high perceivers, Max /ta/ was always significantly different from Pure /ta/ in all the estimated network measures: clustering coefficient ($p < 0.0001$), characteristic path length ($p < 0.0001$), modularity ($p < 0.0001$) and participation coefficient ($p < 0.0001$). In the case of low perceivers, the graph theoretic network metrics were similar between Max /ta/ and Pure /ta/, and no significance differences were obeserved especially for the clustering coefficient ($p = 0.2626$) and modularity ($p = 0.0747$).

Finally, functional interactions between the PBN and the resting-state networks (RSNs) were estimated by performing a between-network analysis. Pre-processed task-block BOLD time series for the PBN were projected onto their first principal component (Friston et al., 1993) and the same procedure

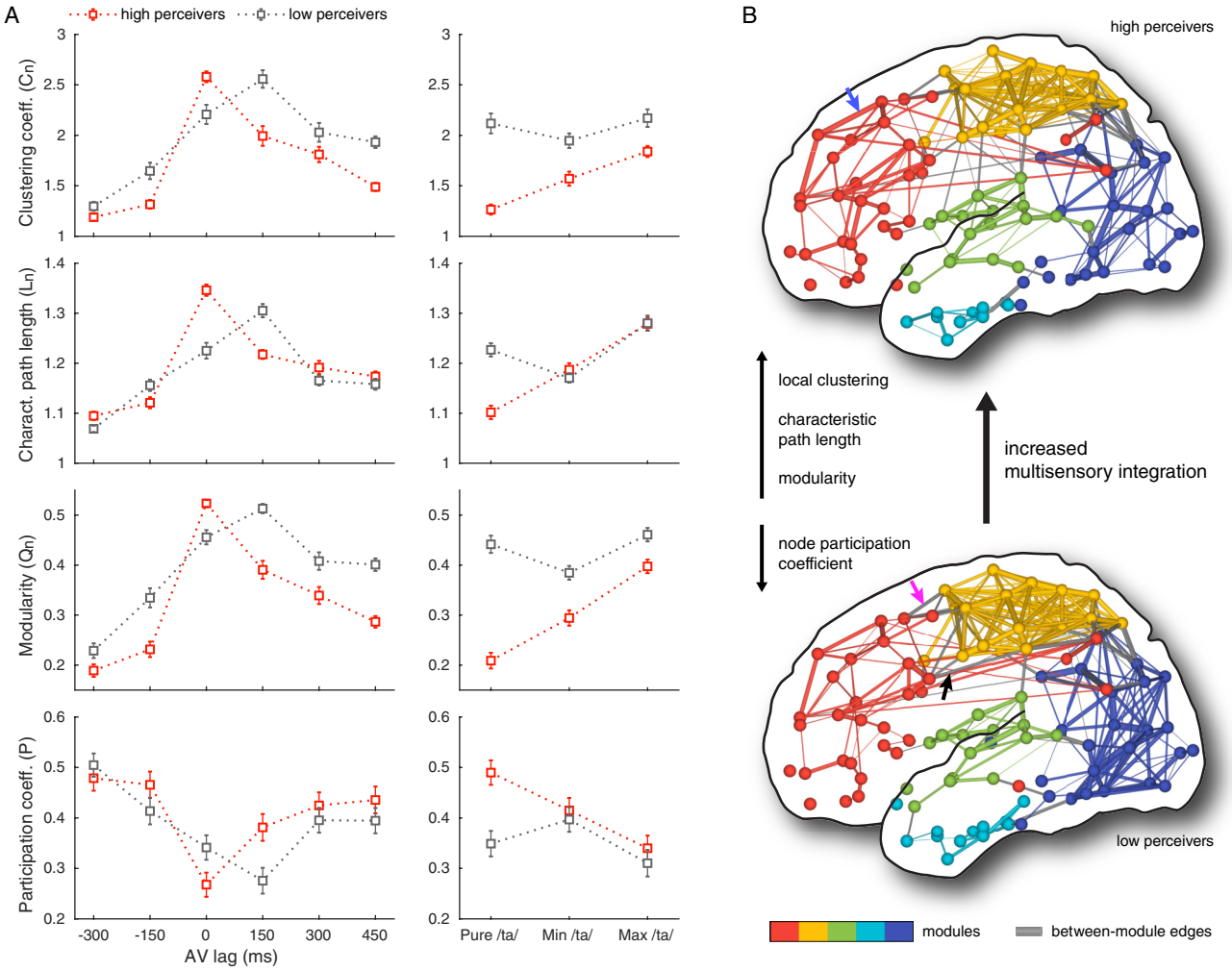

Figure 5. (*A*) Change in the network topological measures for high and low perceivers. (*B*) Left hemispheric brain network differences between high and low perceivers during L0 (synchronous and incongruent) condition.

was followed for each RSN: dorsal attention network (DAN), default mode network (DMN), saliency network (SN), fronto-parietal network (FPN), and ventral attention network (VAN), to obtain a lower dimensional representation of network dynamics. Correlation coefficients were computed between temporal projections of the PBN and each RSN for all AV lags and perception-sorted categories. Interestingly, all the RSNs showed high correlation with the PBN, but the correlations got weaker with increasing AV stimuli synchrony for DAN, DMN, and FPN (Fig. 6*A*). The observed correlation trend when factored with degree of audio-visual illusion closely resembled the SBC patterns in the case of high perceivers; however, a reversed trend was observed for FPN in the case of low perceivers (Fig. 6*B*). Thus, the strength of DAN–PBN and DMN–PBN functional interaction is anti-correlated with propensity to audio-visual illusion for both high and low perceivers. Furthermore, a trend of decreasing anti-correlation with increasing illusory perception was observed when the participants are more finely grouped into three categories (Supplementary Fig. 12).

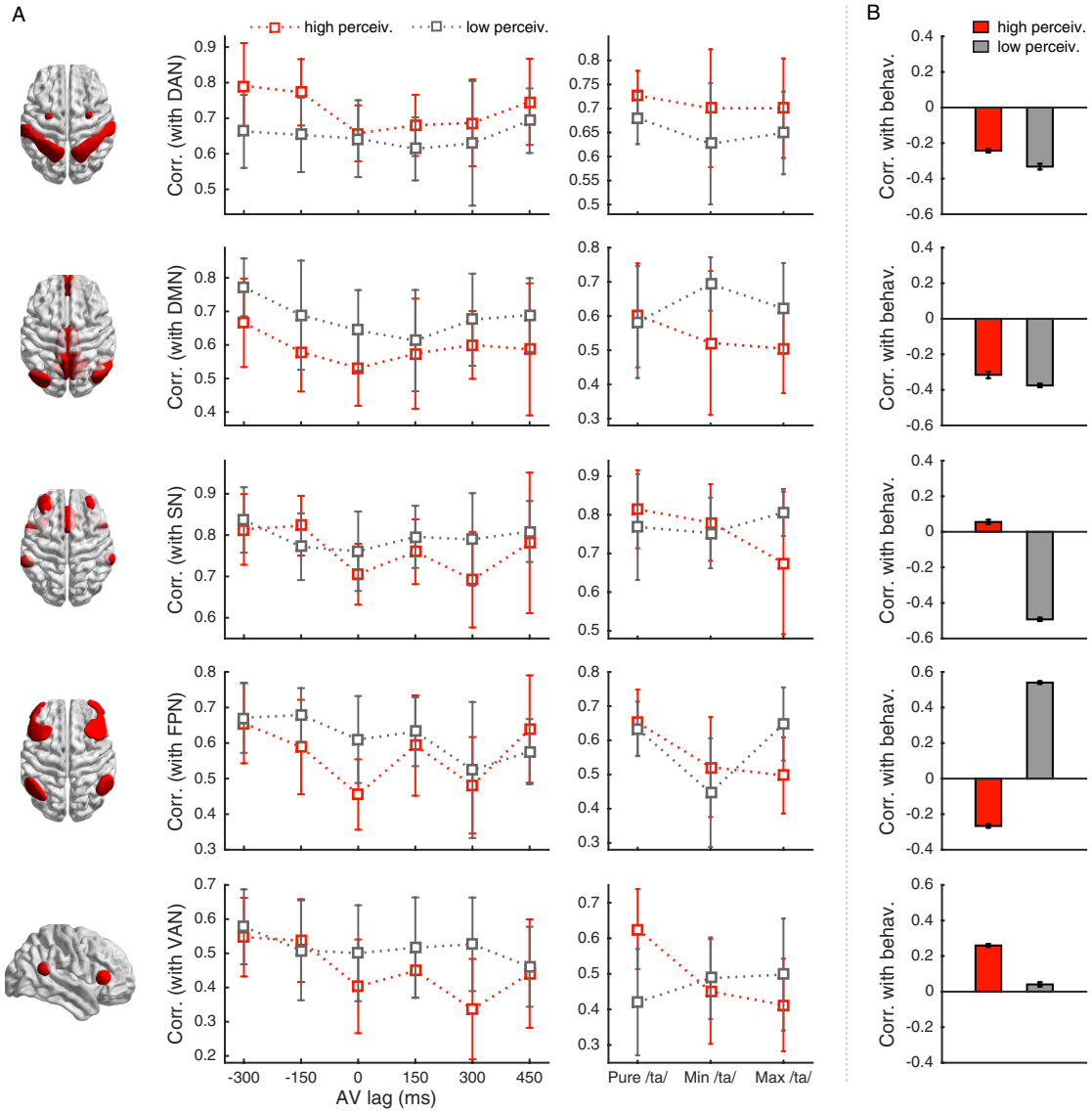

Figure 6. (*A*) Between-network connectivity computed from projections on first principal component for the perceptual binding network (PBN) and the resting-state brain networks—dorsal attention network (DAN), default mode network (DMN), saliency network (SN), fronto-parietal network (FPN), and ventral attention network (VAN). (*B*) The correlation between the connectivity trend described in panel (*A*) and the corresponding trend in propensity to audio-visual illusion across AV lags.

Discussion

We demonstrated that patterned relationships between segregative and integrative information processing mechanisms employed by distributed brain areas is critical for the binding of auditory (A) and visual (V) signals into a coherent perceptual experience. A distributed ensemble of brain regions defined as the perceptual binding network (PBN) exhibited heightened activity when A and V were either temporally congruent (L0) or nearly congruent (L1), as well as when trial blocks were sorted in terms of perceptual categories (Fig). The heightened activity in PBN nodes can be interpreted as local segregation whereas the decreased functional connectivity with rest of the brain at L0 and L1 (as well as Max /ta/) reflects diminished global integration. Hence, simultaneous occurrence of

both increased local segregation and diminished global integration results in higher propensity to perceive illusory-McGurk (cross-modal) sounds. PBN interaction with other resting-state networks, such as default mode (DMN) and dorsal attention networks (DAN), decreased during enhancement of cross-modal bindings. The pattern of decreasing coupling between nodes of PBN and the whole brain (SWFC) was concomitant with rise in illusory perception, i.e., SWFC was anti-correlated with behavior. Follow-up explorations of these patterns in smaller groups of high, medium, and low perceivers further emphasizes the theory of reduced integrative processing and increased segregation involving the nodes of PBN for higher cross-modal perception. Finally, the increased thalamic activity in low perceivers also supported an enhanced role of subcortical level processing in cross-modal perception (Gau et al., 2020; Noesselt et al., 2010; Tyll et al., 2011). In the following, we discuss these issues in more detail.

Previous studies focused on participation of individual regions such as pSTS for multisensory processing and integration (Nath and Beauchamp, 2011; Stevenson and James, 2009). However, several unanswered questions related to multisensory integration and more subtly cross-modal binding (how stimulation of one sensory stream affects others) requires a theoretical understanding of distributed network mechanisms (Johnston et al., 2022; Yau et al., 2015). Our paradigm allowed us to distinguish multisensory processing (involved in all AV lags, as well as during congruent AV stimuli) from the cross-modal perceptual binding (predominantly in L0 and L1), and the importance of the relationship of pSTS with other distributed brain regions which together was defined as PBN. Confirming previous studies (Beauchamp, 2005; Beauchamp et al., 2004), our data clearly establishes pSTS activation is enhanced during cross-modal illusory /ta/ perception. However, as was also observed in L0 stimulus condition where the degree of pSTS activation could not solely explain the perceptual difference between high and low perceivers (Supplementary Figs. 3 and 4, and Supplementary Tables 1 and 2), the interaction among distributed brain areas encompassing the primary sensory areas, motor representaional areas, associative areas such as IPL, V5/MT+, IFG, etc., is essential for understanding audio-visual illusory percept formation. Hence, departing from the earlier studies, our results underscore the need for building an understanding of cross-modal perception at the whole-brain network level. Interestingly, we found more thalamic activation in low perceivers which indicated a major role played by the cortico-thalamic circuitry in perceptual binding. Recent studies have demonstrated that decrease in levels of alpha coherence in electroencephalographic (EEG) activity is indicative of propensity to McGurk-illusion perception (Kumar et al., 2020), wherein the cortico-thalamic basis of alpha rhythms is well-known (da Silva et al., 1973) as well as cortico-thalamic influences on early sensory cortices (Noesselt et al., 2010). fMRI time-scale may not allow the temporal resolution to identify the dynamic mechanisms of the cortico-thalamic circuitry, but future electrophysiological studies can surely address this question.

The second important finding of this study was the decrease in global integration of information effectuated by PBN nodes with rest of the brain with increasing propensity to cross-modal perceptual experience. This was captured via the decrease in temporal correlation among several task-activated ROIs and rest of the brain areas as a function of increasing cross-modal perceptual experience (Fig. 4 and Supplementary Figs. 9 and 10). From the whole-brain network perspective also, increased cross-modal binding is associated with decreased global integration simultaneously balanced by increased local cohesiveness. Subsequently, PBN was also found to be correlated specially with DAN and DMN; but the correlation pattern varied across AV lags reaching a minimum at the peak of illusory perception (Fig. 6). Several researchers have already argued for the necessity to explore the coupling between task-positive and task-agnostic networks to understand behavior and inter-subject variability (Fox et al., 2007; Kelly et al., 2008), with some arguing that anti-correlations may reveal the strength of cognitive performance (Hampson et al., 2010). We argue that the competitive balance

of integrative and segregative processing delimits an individual's perceptual experience, and alteration of the patterns defined by the individual's functional connectome can explain perceptual variability (Finn et al., 2015). A preliminary attempt in this direction was to divide the participants into three smaller groups—high, medium, and low perceivers—that resulted in higher-to-lower anti-correlation between seed functional connectivity and behavior, respectively. High perceivers also displayed relatively higher local clustering and larger characteristic path length (Supplementary Fig. 11). To the best of our knowledge, this is the first demonstration of competitive segregation and integration mechanisms in multisensory integration tasks. Some integrative aspects of multisensory processing such as enhanced functional connectivity (FC) among the multisensory and unisensory regions, observed by earlier studies (Nath and Beauchamp, 2011) can be explained from the spatial topographies of functionally connected regions with pSTS (Supplementary Figs. 7 and 8) at L0 and L1. Thus, an argument that integrative mechanisms are predominantly compensatory to the task-driven haemodynamic oxygen metabolism can be put forward while contextualizing our results with previous work on multisensory integration. Temporal binding windows (Stevenson et al., 2012) can define the boundaries of the compensatory regimes where integration and segregation can be synergistically altered, giving rise to inter-trial variability.

The two major findings we presented are also consonant with the theoretical frameworks that are often used to understand multisensory integration, Bayesian causal inference (Koerding et al., 2007; Magnotti et al., 2018; Rohe et al., 2019) and predictive coding hypothesis (Friston and Kiebel, 2009; Mumford, 1991). First, enhanced segregation and decreased integration in distributed brain areas alludes to a mechanism of perceptual binding of disparate sensory objects by imposition of a common causal structure as suggested by earlier studies (Rohe et al., 2019). Second, in harmony with the predictive coding hypothesis, the decreased FC between the context-driven PBN and the context-free RSNs can be interpreted as signatures of feedback functional projections from hierarchically higher order "dynamic core" (Deco et al., 2017) that suppresses the SWFC within the temporal binding window. However, our study departs significantly from these earlier theoretical frameworks in providing a more neurobiologically grounded mechanism of multisensory integration based on the existence of balanced states of higher segregation and decreased global integration (Lord et al., 2017) employed by nodes of the PBN. In light of our current results and previous modeling studies (Abeysuriya et al., 2018; Naskar et al., 2021; Vattikonda et al., 2016), we propose that balanced states of functional integration and segregation may emerge from the underlying excitatory–inhibitory (E-I) balance in local neuronal populations of involved brain areas and may even be related to neurotransmitter homeostasis (Naskar et al., 2021; Shine et al., 2021). While spontaneous brain activity such as during resting state has been typically associated with the E-I balance (Abeysuriya et al., 2018; Deco et al., 2017; Naskar et al., 2021; Vattikonda et al., 2016), the potential of altered E-I balance—in guiding the formation of perceptual experience from interactions between inclement sensory signals with stored associative memories (Barron et al., 2017), as a catalyst for volitional control effectuating perceptual stability (Kondo et al., 2018), in guiding decision making strategies during multi-attribute choice (Pettine et al., 2021); all have been proposed recently. According to this line of reasoning, an altered balance between segregation and integration observed during the illusory perception may be the outcome of an energy consumption by the motor networks, reported in an earlier study (Murakami et al., 2018) and supported by underlying neuromodulatory systems which are otherwise adapted to lesser energy consuming congruent auditory–visual input. The existence of the stable illusory perception across groups of people from wide geographical–linguistic backgrounds (Magnotti et al., 2015), as well as its robust presence post-learning (Mallick et al., 2015) also hints toward a neuromodulatory origin. An immediate goal of subsequent studies should be to employ neurostimulation techniques such as TMS and tDCS (Yau et al., 2015) with a goal to perturb the balanced states of integration–segregation

mechanisms between PBN and context independent RSNs and identify their deeper causal architecture. Further extending the paradigms and analytic approaches that we have used here, studies on neuropsychiatric and affective disorders can explore whether an alteration of the balance of segregation and integration can be a viable framework to characterize patient-specific symptoms.

MATERIALS AND METHODS

Participants and Stimuli

Fifty-two healthy right-handed human volunteers (25 female, 27 male; mean age = 24.5 years, SD = 3.12) participated in the functional magnetic resonance imaging (fMRI) study. The study was approved by the Institutional Human Ethics Committee (IHEC), National Brain Research Centre, India, and all participants provided informed consents in a format approved by the IHEC.

Seven video stimuli each of 2 second duration were prepared. Six of them were incongruent audio-visual (AV) objects where audio recordings of a male speaker vocalizing /pa/ is dubbed on a video showing the lip movement while articulating the vocalization of /ka/ (/pa/–/ka/). The remaining one video was of a congruent AV object (/ta/–/ta/). In the six incongruent videos, the gap between the onset of auditory and visual streams were varied from −300 to 450 ms in steps of 150 ms (Fig. 1*A*). Negative sign indicates that the auditory stimulus onset precedes that of visual lip movement (A leading), and the positive sign on the other hand implies the auditory stimulus succeeds the lip articulation (V leading). An asymmetric range of AV lags were chosen, because Munhall and colleagues (Munhall et al., 1996) reported the dominance of illusory perception toward positive lags where the onset of lip articulation precedes auditory onset. The congruent /ta/–/ta/ stimulus had synchronous onset of AV stimuli. In the stimulus video, the articulation of /ka/ or /ta/ always started from a neutral lip-closed position. Videos were created using VideoPad video-editing software (NCH Software, CO, USA) at a frame rate of 25 frames/second and a resolution of 1280 × 720 pixels. The auditory /pa/ and /ta/ syllables were of 0.549 and 0.531 s duration and were edited in Audacity software (Sourceforge.net) to minimize the background noise. The audio sampling rate was 44 kHz and had a bit rate of 128 kbps. The resulted seven stimuli thus have five asynchronous and two synchronous stimuli—McGurk stimuli (/pa/–/ka/) with −300, −150, 0, 150, 300, 450 ms AV lags, and Pure /ta/ (/ta/–/ta/) stimulus with 0 ms AV lag. Henceforth, we shall also refer the AV lag conditions in shorter notation as −L2, −L1, L0, L1, L2, L3 for −300, −150, 0, 150, 300, 450 ms lag, respectively.

Task Design

The task design is illustrated in Fig. 1*A*. Inside the scanner, the stimuli were presented in a block design, and each activation block of length 20 s comprised 10 video stimuli of one kind. In total there were 28 activation blocks in the whole experiment: 4 activation blocks for each of the 7 stimulus categories. The 28 activation blocks were alternated with another 28 inter-trial resting blocks, each also of 20 s duration. We randomized the presentation of the 7 stimulus categories with each category comprising 4 alternated activation blocks of one kind. The stimuli were presented through an INVIVO MR-compatible CRT screen attached to the head-coil and an MRI-compatible headphones (Philips, The Netherlands). Presentation software (Neurobehavioral Systems, CA) was used to display the stimuli. During the presentation, participants were instructed to attend to the AV stimuli watching at the speaker all the times and report their perceptual experience of whether they heard "/pa/", "/ta/," or "any other" through a three-button fibre-optic response device by Curdes (Current Designs, PA, USA), which was used to record the responses of participants.

fMRI Data Acquisition

Structural T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) scan was performed in the beginning of each participant's scanning session using a 3 Tesla Philips Achieva MRI scanner. For each participants, seven functional T2* scans (TR = 2 s, TE = 35 ms, flip angle = 90°) corresponding to the seven different stimuli (−L2 to L3 and Pure /ta/) were acquired. An 8-channel head coil was used to acquire the T2*-weighted images. The voxel size of the acquired images was $3.59 \times 3.59 \times 5$ mm^3 and the voxel size of the reconstructed images was $1.80 \times 1.80 \times 5$ mm^3. The slice thickness was 5 mm and inter-slice distance was 1 mm.

Behavioral Data Analysis

The button press responses from the 52 participants were converted into normalized percentage measures of each perceptual response category—/pa/, /ta/, or "other". A maximum of 40 responses were expected from each task. Tasks with less than 80% responses were rejected, however all participants met this threshold. The volunteers were further screened based on their /ta/ response percentage during congruent Pure /ta/ (/ta/–/ta/) stimulus condition. The minimum threshold of /ta/ response was set to 40%, and seven participants were rejected from this criterion. Two more participants were rejected because one had reported >95% /ta/ response in five AV lag conditions and another >95% "other" response in two AV lag conditions. Thus, only 43 participants were selected based on their behavioral responses. Furthermore, five more participants were rejected because of excessive movement during MRI scanning. Finally, 38 participants were thus selected for further analyses. For each participant the *susceptibility* to audio-visual illusion was estimated by averaging maximal percentage /ta/ responses at any two AV lags. A median split to their susceptibility scores allowed the participants to be binned into two equal-sized groups (19–19): the group representing higher susceptibilities will be called *high perceivers* and the other at the lower end as *low perceivers* (Fig. 1*A*). Statistical analyses of each perceptual responses—/pa/, /ta/, and "other"—were performed using two-way analysis of variance (ANOVA), with AV lags and paticipant grouping as independent variable. A more finer grouping with 13–12–13 participants was also discussed in this paper. The procedure was similar with that of double split, however we selected the 13 highest susceptibility scores as high perceivers, the next 12 middle scores are selected as medium perceivers, and the lowest 13 scores are grouped as low perceivers.

fMRI Data Analysis

Pre-processing. For each participant, all the acquired functional brain volumes were realigned and co-registered with the structural scan using a six-parameter rigid-body transformation (three translations and three rotations). Those participants with maximum movement greater than 2 units (i.e., 2 mm in translational axes, or 2° in rotational axes) in any of the six degrees of freedom were rejected from further analysis. The structural image of each participant's brain was then segmented into grey and white matter maps, and using which the functional brain volumes were spatially normalized to Montréal Neurological Institute (MNI) standard brain. The normalized functional images were then spatially smoothed minimally with a 4 mm full-width at half-maximum (FWHM) isotropic Gaussian kernel. All the preprocessing was done using Statistical Parametric Mapping Software 8 (SPM8) in MATLAB (MathWorks, MA). The first brain volume of each participant was removed to allow the magnetization to reach to a steady state.

Stimulus-based activation. After pre-processing, we performed the first-level intra-fixed-effects analysis using t-test on the activation blocks versus the inter-trial rest blocks, with the experimental design (event onset and their time course) as essential regressors and the realignment parameters for head movements as nuisance regressors. To accommodate the inter-participant variability within each AV lag and perceptual condition, the contrast images (task block $>$ inter-trial rest block) from the first-level analysis were entered into a second-level random-effects group analysis, using one-sample t-tests. Subsequently, the fMRI data were investigated via two alternative routes. First, the group activation maps (Friston et al., 1995) were extracted for each AV lag, capturing stimulus-driven changes (stimulus-based categorization). For each participant we further selected the two AV lags that correspond to maximum and minimum percentage /ta/ responses (Max /ta/ and Min /ta/). Once compiled, the group activation maps in Max /ta/ and Min /ta/ were compared against Pure /ta/, thus capturing the brain activations sorted vis-à-vis perceptual response (perceptual categorization). Second, BOLD time series data were extracted for both stimulus-based and perception-based categories, to further estimate the seed-based correlations and other network characteristics (details are provided in the next section). When finally reporting the significantly activated clusters, the identification as well as labeling of the brain areas were done according to the parcels in Brainnetome atlas (Fan et al., 2016) (http://atlas.brainnetome.org/; Supplementary Table 3).

For defining seed regions of interest (ROIs), we use another subset of participants who reported perceiving at least 60% illusory /ta/ in any AV lag. Each seed parcel was located by mapping the task-active clusters during L0 and L1 conditions on brain volumes specified with Brainnetome atlas. We particularly chose L0 and L1 AV lag conditions for there were maximal significant activation of brain volumes as well as maximal average illusory /ta/ response during these lags. These active brain parcels were also majorly found to be active during L0, L1, and MaxTa conditions of high and low perceivers. The active parcels were then grouped into 32 ROIs based on major anatomical landmarks of Brainnetome atlas (Supplementary Table 4).

To report the number of within-ROI task-active voxels for different task conditions, we created 19 leave-one-out samples separately for each high and low perceiver group. The average within-ROI activation volume from these 19 subsamples is reported in Supplementary Figs. 3 and 4.

Percentage signal difference (PSD). For each event in the block design, the ROI percentage signal change was evaluated by estimating maximal height of the HRF fitted event time course (beta $\times$ event regressor), relative to the actual mean signal of the ROI. The signal was estimated for an event duration of 2 seconds. The percentage signal change was extracted using a MarsBaR batch script (http://marsbar.sourceforge.net). For each of the 32 ROIs, the PSD in each AV lag and perceptual condition was then evaluated by subtracting the percentage signal change of inter-trial rest blocks from that of the activation blocks. The changes in average PSD for high and low perceivers across conditions are reported in Supplementary Figs. 5 and 6, respectively.

Functional Connectivity Analysis

Seed-based functional connectivity. Each participant's grey matter volume was parcellated using the Brainnetome atlas (http://atlas.brainnetome.org/) for the seed-based connectivity analysis. Among the 210 cortical regions of the atlas, the two subregions of pSTS in each hemisphere were merged into a single ROI. In the subcortical region, only the whole thalamic nuclei in each hemisphere were considered. Taken together, the final set of 210 ROIs covered the entire cortical region and the thalamus (Supplementary Table 3). For each participant, the mean BOLD time series from each

210 ROI as well as from 32 seed ROIs (Supplementary Table 4) were extracted using REX toolbox (Whitfield-Gabrieli, 2009) (https://web.mit.edu/swg/software.htm). From each extracted time series, stimulus blocks corresponding to each AV lag and "Pure /ta/" condition were selected based on the experimental design, and then concatenated. The concatenated time series were further preprocessed through linear detrending (linear or linear breakpoint detrending). Functional correlation between each seed region and the remaining of 210 ROIs (i.e., exclusive of seed parcels) were then evaluated for each AV lag and perceptual condition using the Pearson correlation measure (Lee Rodgers and Nicewander, 1988; Pearson, 1895). Finally, the averaged correlation for each seed ROI was evaluated employing Fisher's z-transformed FDR-corrected ($p \leqslant 0.05$) correlation values. We will term this measure as seed-based whole-brain functional connectivity (SWFC).

The relationship between the trend that SWFC changes across AV lag conditions and the corresponding changes in the degree of illusory /ta/ perception was further evaluated by estimating the Pearson correlation coefficient between the two trends. Hereafter, we will refer it as SWFC–behavior correlation (SBC). Each reported seed SBC is the averaged significant correlation ($p \leqslant 0.05$, Bonferroni corrected) value from 50 repeats of 50 bootstrap samples created by random sampling 70% participants within each group (high and low perceivers).

Whole-brain graph theoretic network analysis. For each participant, the functional interaction across all the 210 brain parcels were estimated using pairwise Pearson correlation between corresponding task-block BOLD time series. At the group level, the functional connectomes were generated as follows. First, distinct subsamples were generated by random sampling 70% of the total participants in each group. For each subsample, the averaged cross-correlation matrix was then evaluated incorporating Fisher's z-transformation, and tested for statistical significance through one-sample t-test ($p \leqslant 0.05$, FDR corrected). We then applied a threshold of 0.48 to obtain the binarized network matrix; the threshold value of 0.48 was chosen to ensure a statistical power of ≈0.9. We also ensured that each selected network form a single connected component to allow comparisons across same network size. Out of the total 210 nodes (210 ROIs), one node in the anterior right parahippocampal area that was frequently disconnected irrespective of stimulus conditions was excluded from the network study. Repeating the same procedure, we finally generated 50 binarized connected brain networks of size 209×209 for each stimulus condition. Furthermore, to later estimate normalized network parameters, for each generated binarized brain network we also generated 50 randomized connected network samples with same degree sequence. We used the MATLAB script *randmio_und_connected.m* available with Brain Connectivity Toolbox (BCT) for the randomization, keeping edge iterations to 30 (Rubinov and Sporns, 2010).

To study the functional integration and segregation at different scales of brain network, we calculated average local clustering coefficient (C_{real}), characteristics path length (L_{real}), modularity (Q_{real}), and node participation coefficients (P). C_{real} and L_{real} estimate the network characteristics at the resolution of nodes (local scale), and Q_{real} and P estimates at the level of the entire network (global scale). In this study, we used Blondel et al. (Blondel et al., 2008) method (Louvain method; *community_louvain.m* available with BCT) for patitioning a network into nonoverlapping functional modules. We also applied normalized graph measures for comparing across the networks (Guimera et al., 2007; Van Wijk et al., 2010).

The local clustering coefficient of a node estimates the fraction of connections among the first neighbors of the node against all possible connection, and the normalized averaged clustering coefficient (C_n) is defined as

$$C_n = \frac{C_{\text{real}}}{C_{\text{rand}}} \tag{1}$$

where C_{real} and C_{rand} are the average local clustering coefficients (Rubinov and Sporns, 2010; Rubinov et al., 2009) of the empirical and randomized networks, respectively.

Similarly, for the characteristics path length that measures the average shortest path across all node pairs, the normalized characteristics path length is

$$L_n = \frac{L_{\text{real}}}{L_{\text{rand}}} \tag{2}$$

where L_{real} and L_{rand} represent the characteristics path length (Rubinov and Sporns, 2010) of the empirical and randomized networks, respectively.

Modularity on the other hand is a quality function that measures how well a network is divided into distinct functional units, and it can be conceptualized as another network segregation measure at the global scale (Newman and Girvan, 2004). The normalized modularity is further estimated as (Kashtan and Alon, 2005)

$$Q_n = \frac{Q_{\text{real}} - Q_{\text{rand}}}{Q_{\text{max}} - Q_{\text{rand}}} \tag{3}$$

Here Q_{max} is the maximal possible network modularity value with same degree sequence and can be approximated as $Q_{\text{max}} = 1 - 1/m$, where m represents the total number of modules in the real network (Parter et al., 2007). Q_{real} and Q_{rand} are modularity computed in the empirical and randomized network (Newman and Girvan, 2004) respectively.

Node participation coefficient (P) measures how equally the connections of a node are distributed among all available modules of a network (Guimera and Amaral, 2005). Thus, it estimates the capacity of a node in intermodular information integration.

$$P = 1 - \sum_{i=1}^{m} \left(\frac{k_i}{k} \right)^2 \tag{4}$$

where k is the degree of the node and k_i represents the number of connections the node have with the module i. Here, m is the number of available modules.

Interaction between the PBN and resting-state networks (RSNs). We also investigated the functional interaction between the PBN and human default mode network (DMN), saliency network (SN), fronto-parietal network (FPN), dorsal attention network (DAN), and ventral attention network (VAN). The ROI masses for DMN, SN, FPN, and DAN were taken from CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012) which elucidated these RSNs through ICA analyses of 497 participants from HCP dataset (www.nitrc.org/projects/conn). For the VAN, two spherical ROI masses each of 10 mm diameter were created using MarsBaR toolbox: one for the right temporo-parietal junction (Fox et al., 2006) (centered around MNI coordinates: 54, -50, 19) and another for the right ventrolateral prefrontal cortex (Sylvester et al., 2013) (centered around MNI coordinates: 51, 27, 8). Task-block BOLD time series for the respective ROIs were extracted using REX toolbox. Next, through principal component decomposition the projected time series from each network constituents to their first principal component was evaluated. Finally, network interactions were estimated by using Pearson correlation measure between corresponding first principal component time series. We reported the evolution of the network interaction across AV lag and perceptual conditions in Fig. 6.

References

R. G. Abeysuriya, J. Hadida, S. N. Sotiropoulos, S. Jbabdi, R. Becker, B. A. E. Hunt, M. J. Brookes, and M. W. Woolrich. A biophysical model of dynamic balancing of excitation and inhibition in fast oscillatory large-scale networks. *PLoS Comput. Biol.*, 14:e1006007, Feb. 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006007.

D. Alais and D. Burr. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.*, 14:257–262, 2004.

H. C. Barron, T. P. Vogels, T. E. Behrens, and M. Ramaswami. Inhibitory engrams in perception and memory. *Proc. Natl. Acad. Sci. U.S.A.*, 114:6666–6674, June 2017. ISSN 1091-6490. doi: 10.1073/pnas.1701812114.

M. S. Beauchamp. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr. Opin. Neurobiol.*, 15:145–153, 2005.

M. S. Beauchamp, B. D. Argall, J. Bodurka, J. H. Duyn, and A. Martin. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.*, 7:1190–1192, Nov. 2004. ISSN 1097-6256. doi: 10.1038/nn1333.

P. Bertelson and G. Aschersleben. Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *Int. J. Psychophysiol.*, 50: 147–155, Oct. 2003. ISSN 0167-8760. doi: 10.1016/s0167-8760(03)00130-2.

V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech.*, 2008(10):P10008, 2008.

J. R. Cohen and M. D'Esposito. The segregation and integration of distinct brain networks and their relationship to cognition. *J. Neurosci.*, 36:12083–12094, Nov. 2016. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.2965-15.2016.

C. Cuppini, L. Shams, E. Magosso, and M. Ursino. A biologically inspired neurocomputational model for audiovisual integration and causal inference. *Eur. J. Neurosci.*, 46:2481–2498, Nov. 2017. ISSN 1460-9568. doi: 10.1111/ejn.13725.

F. H. L. da Silva, T. H. M. T. van Lierop, C. F. Schrijer, and W. S. van Leeuwen. Organization of thalamic and cortical alpha rhythms: Spectra and coherences. *Elec. Clin. Neuro.*, 3(6):627 – 639, 1973.

G. Deco, M. L. Kringelbach, V. K. Jirsa, and P. Ritter. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. *Sci. Rep.*, 7:3095, June 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-03073-5.

G. Dehaene-Lambertz, L. Hertz-Pannier, J. Dubois, S. Mériaux, A. Roche, M. Sigman, and S. Dehaene. Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci. U.S.A.*, 103(38):14240–14245, 2006.

L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, P. T. Fox, S. B. Eickhoff, C. Yu, and T. Jiang. The human Brainnetome Atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex*, 26:3508–3526, Aug. 2016. ISSN 1460-2199. doi: 10.1093/cercor/bhw157.

E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.*, 18:1664–1671, Nov. 2015. ISSN 1546-1726. doi: 10.1038/nn.4135.

M. D. Fox, M. Corbetta, A. Z. Snyder, J. L. Vincent, and M. E. Raichle. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl. Acad. Sci. U.S.A.*, 103(26): 10046–10051, 2006.

M. D. Fox, A. Z. Snyder, J. L. Vincent, and M. E. Raichle. Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron*, 56:171–184, Oct. 2007. ISSN 0896-6273. doi: 10.1016/j.neuron.2007.08.023.

K. Friston and S. Kiebel. Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 364:1211–1221, May 2009. ISSN 1471-2970. doi: 10.1098/rstb.2008.0300.

K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. Frackowiak. Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.*, 13:5–14, Jan. 1993. ISSN 0271-678X. doi: 10.1038/jcbfm.1993.4.

K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.*, 2:189–210, 1995. ISSN 1065-9471. doi: 10.1002/hbm.460020402.

R. Gau, P.-L. Bazin, R. Trampel, R. Turner, and U. Noppeney. Resolving multisensory and attentional influences across cortical depth in sensory cortices. *Elife*, 9:e46856, 2020.

R. Guimera and L. A. N. Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005.

R. Guimera, M. Sales-Pardo, and L. A. N. Amaral. Classes of complex networks defined by role-to-role connectivity profiles. *Nat. Phys.*, 3(1):63–69, 2007.

M. Hampson, N. Driesen, J. K. Roth, J. C. Gore, and R. T. Constable. Functional connectivity between task-positive and task-negative brain areas and its relation to working memory performance. *Magn. Reson. Imaging*, 28:1051–1057, Oct. 2010. ISSN 1873-5894. doi: 10.1016/j.mri.2010.03.021.

P. R. Johnston, C. Alain, and A. R. McIntosh. Individual differences in multisensory processing are related to broad differences in the balance of local versus distributed information. *J. Cogn. Neurosci.*, 34(5):846–863, 2022.

J. A. Jones and D. E. Callan. Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, 14:1129–1133, June 2003. ISSN 0959-4965. doi: 10.1097/00001756-200306110-00006.

N. Kashtan and U. Alon. Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci. U.S.A.*, 102(39):13773–13778, 2005.

A. M. C. Kelly, L. Q. Uddin, B. B. Biswal, F. X. Castellanos, and M. P. Milham. Competition between functional brain networks mediates behavioral variability. *Neuroimage*, 39:527–537, Jan. 2008. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.08.008.

K. P. Koerding, U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum, and L. Shams. Causal inference in multisensory perception. *PLoS One*, 2:e943, Sept. 2007. ISSN 1932-6203. doi: 10.1371/journal.pone.0000943.

H. M. Kondo, D. Pressnitzer, Y. Shimada, T. Kochiyama, and M. Kashino. Inhibition-excitation balance in the parietal cortex modulates volitional control for auditory and visual multistability. *Sci. Rep.*, 8:14548, Sept. 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-32892-3.

G. V. Kumar, S. Dutta, S. Talwar, D. Roy, and A. Banerjee. Biophysical mechanisms governing large-scale brain network dynamics underlying individual-specific variability of perception. *Eur. J. Neurosci.*, Apr. 2020. ISSN 1460-9568. doi: 10.1111/ejn.14747.

J. Lee Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. *Am. Stat.*, 42(1):59–66, 1988.

A. M. Liberman and D. H. Whalen. On the relation of speech to language. *Trends Cogn. Sci.*, 4(5): 187–196, 2000.

L.-D. Lord, A. B. Stevner, G. Deco, and M. L. Kringelbach. Understanding principles of integration and segregation using whole-brain computational connectomics: implications for neuropsychiatric disorders. *Philos. Trans. A Math. Phys. Eng. Sci.*, 375:20160283, June 2017. ISSN 1364-503X. doi: 10.1098/rsta.2016.0283.

J. F. Magnotti, D. B. Mallick, G. Feng, B. Zhou, W. Zhou, and M. S. Beauchamp. Similar frequency of the mcgurk effect in large samples of native mandarin chinese and american english speakers. *Exp. Brain Res.*, 233:2581–2586, Sept. 2015. ISSN 1432-1106. doi: 10.1007/s00221-015-4324-7.

J. F. Magnotti, K. B. Smith, M. Salinas, J. Mays, L. L. Zhu, and M. S. Beauchamp. A causal inference explanation for enhancement of multisensory integration by co-articulation. *Sci. Rep.*, 8:18032, Dec. 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-36772-8.

D. B. Mallick, J. F. Magnotti, and M. S. Beauchamp. Variability and stability in the mcgurk effect: contributions of participants, stimuli, time, and response type. *Psychon. Bull. Rev.*, 22:1299–1307, Oct. 2015. ISSN 1531-5320. doi: 10.3758/s13423-015-0817-4.

H. McGurk and J. Macdonald. Hearing lips and seeing voices. *Nature*, 264:746–748, 1976.

M. A. Meredith and B. E. Stein. Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608):389–391, Jul 1983.

H. Mohr, U. Wolfensteller, R. F. Betzel, B. Misic, O. Sporns, J. Richiardi, and H. Ruge. Integration and segregation of large-scale brain networks during short-term task automatization. *Nat. Commun.*, 7:13217, Nov. 2016. ISSN 2041-1723. doi: 10.1038/ncomms13217.

S. Molholm, W. Ritter, M. M. Murray, D. C. Javitt, C. E. Schroeder, and J. J. Foxe. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res. Cogn. Brain Res.*, 14:115–128, June 2002. ISSN 0926-6410. doi: 10.1016/s0926-6410(02)00066-6.

L. Morís Fernández, E. Macaluso, and S. Soto-Faraco. Audiovisual integration as conflict resolution: the conflict of the mcgurk illusion. *Hum. Brain Mapp.*, 38(11):5691–5705, 2017.

D. Mumford. On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biol. Cybern.*, 65:135–145, 1991. ISSN 0340-1200. doi: 10.1007/BF00202389.

K. G. Munhall, P. Gribble, L. Sacco, and M. Ward. Temporal constraints on the McGurk effect. *Atten. Percept. Psychophys.*, 58:351–362, 1996.

T. Murakami, M. Abe, W. Wiratman, J. Fujiwara, M. Okamoto, T. Mizuochi-Endo, T. Iwabuchi, M. Makuuchi, A. Yamashita, A. Tiksnadi, et al. The motor network reduces multisensory illusory perception. *J. Neurosci.*, 38(45):9679–9688, 2018.

A. Naskar, A. Vattikonda, G. Deco, D. Roy, and A. Banerjee. Multi-scale dynamic mean field model (mdmf) relates resting-state brain dynamics with local cortical excitatory–inhibitory neurotransmitter homeostasis. *Netw. Neurosci.*, pages 1–55, 2021.

A. R. Nath and M. S. Beauchamp. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J. Neurosci.*, 31(5):1704–1714, 2011.

M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):026113, 2004.

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

M. Parter, N. Kashtan, and U. Alon. Environmental variability and modularity of bacterial metabolic networks. *BMC Evol. Biol.*, 7(1):1–8, 2007.

K. Pearson. Vii. note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.*, 58(347-352):240–242, 1895.

W. W. Pettine, K. Louie, J. D. Murray, and X.-J. Wang. Excitatory-inhibitory tone shapes decision strategies in a hierarchical neural network model of multi-attribute choice. *PLoS Comput. Biol.*, 17:e1008791, Mar. 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008791.

T. Rohe, A.-C. Ehlis, and U. Noppeney. The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nat. Commun.*, 10:1907, Apr. 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09664-2.

M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

M. Rubinov, S. A. Knock, C. J. Stam, S. Micheloyannis, A. W. F. Harris, L. M. Williams, and M. Breakspear. Small-world properties of nonlinear brain activity in schizophrenia. *Hum. Brain Mapp.*, 30(2):403–416, 2009.

S. K. Scott, C. McGettigan, and F. Eisner. A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.*, 10(4):295–302, 2009.

L. Shams, Y. Kamitani, and S. Shimojo. Visual illusion induced by sound. *Brain. Res. Cogn. Brain. Res.*, 14:147–152, 2002.

J. M. Shine, E. J. Müller, B. Munn, J. Cabral, R. J. Moran, and M. Breakspear. Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics. *Nat. Neurosci.*, pages 1–12, 2021.

R. A. Stevenson and T. W. James. Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*, 44:1210–1223, Feb. 2009. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2008.09.034.

R. A. Stevenson, R. K. Zemtsov, and M. T. Wallace. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *J. Exp. Psychol. Hum. Percept. Perform.*, 38:1517–1529, Dec. 2012. ISSN 1939-1277. doi: 10.1037/a0027339.

C. M. Sylvester, D. M. Barch, M. Corbetta, J. D. Power, B. L. Schlaggar, and J. L. Luby. Resting state functional connectivity of the ventral attention network in children with a history of depression or anxiety. *J. Am. Acad. Child Adolesc. Psychiatry*, 52(12):1326–1336, 2013.

B. Thakur, A. Mukherjee, A. Sen, and A. Banerjee. A dynamical framework to relate perceptual variability with multisensory information processing. *Sci. Rep.*, 6:31280, Aug. 2016. ISSN 2045-2322. doi: 10.1038/srep31280.

D. Tomasi, R. Wang, G.-J. Wang, and N. D. Volkow. Functional connectivity and brain activation: a synergistic approach. *Cereb. Cortex*, 24:2619–2629, Oct. 2014. ISSN 1460-2199. doi: 10.1093/cercor/bht119.

G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U.S.A.*, 91:5033–5037, May 1994. ISSN 0027-8424. doi: 10.1073/pnas.91.11.5033.

S. Tyll, E. Budinger, and T. Noesselt. Thalamic influences on multisensory integration. *Commun. Integr. Biol.*, 4(4):378–381, 2011.

B. C. Van Wijk, C. J. Stam, and A. Daffertshofer. Comparing brain networks of different size and connectivity density using graph theory. *PloS One*, 5(10):e13701, 2010.

A. Vattikonda, B. R. Surampudi, A. Banerjee, G. Deco, and D. Roy. Does the regulation of local excitation-inhibition balance aid in recovery of functional connectivity? a computational account. *Neuroimage*, 136:57–67, Aug. 2016. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2016.05.002.

S. Watkins, L. Shams, O. Josephs, and G. Rees. Activity in human V1 follows multisensory perception. *Neuroimage*, 37:572–578, Aug. 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.05.027.

S. Whitfield-Gabrieli. Region of interest extraction (rex) toolbox. *Boston, MA*, 497, 2009.

S. Whitfield-Gabrieli and A. Nieto-Castanon. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.*, 2(3):125–141, 2012.

J. M. Yau, G. C. DeAngelis, and D. E. Angelaki. Dissecting neural circuits for multisensory integration and crossmodal processing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 370:20140203, Sept. 2015. ISSN 1471-2970. doi: 10.1098/rstb.2014.0203.

G. W. Zopf Jr. Sensory homeostasis. In *Progress in Brain Research*, volume 2, pages 114–121. Elsevier, 1963.

Acknowledgments This study was supported by NBRC Core funds and Computing Facility. AB was supported by Ramalingaswami Fellowship (BT/RLF/Re-entry/31/2011) and Innovative Young Biotechnologist Award (IYBA, BT/07/IYBA/2013), both from Department of Biotechnology, Ministry of Science and Technology, Government of India. SSS was supported by a National Post-Doctoral Fellowship from Science and Engineering Research Board, Government of India (SERB-NPDF; No. PDF/2016/003188).

Competing interests The authors declare no competing interest.