High segregation and diminished global integration in large-scale brain functional networks enhances the perceptual binding of cross-modal stimuli

Soibam Shyamchand Singh[1,2,†], Abhishek Mukherjee[1,†], Partha Raghunathan[1,‡], Dipanjan Ray[2], and Arpan Banerjee[1,*]

[1] *National Brain Research Centre, NH8, Manesar, Gurgaon 122052, India.*

[2] *Department of Psychology, Ashoka University, Sonepat 131029, India.*

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

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

[‡] *Deceased*

Speech perception requires the binding of spatiotemporally disjoint auditory–visual cues. The corresponding brain network-level information processing can be characterized by two complementary mechanisms: functional segregation which refers to the localization of processing in either isolated or distributed modules across the brain, and integration which pertains to 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. We controlled the inter-subject variability of illusory/cross-modal perception parametrically, by introducing temporal lags in the incongruent auditory–visual articulations of speech sounds within the McGurk paradigm. The states of segregation–integration balance were captured using two alternative computational approaches. First, the module responsible for cross-modal binding of sensory signals defined as the perceptual binding network (PBN) was identified using standardized parametric statistical approaches and their temporal correlations with all other brain areas were computed. With increasing illusory perception, the majority of the nodes of PBN showed decreased cooperation with the rest of the brain, reflecting states of high segregation but reduced global integration. Second, using graph theoretic measures, the altered patterns of segregation–integration were cross-validated.

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, without even understanding the language. Previous studies have also reported that within a certain window, the temporal dissonance between multiple sensory components can still facilitate illusory perceptual experience (Bertelson and Aschersleben, 2003; Munhall et al., 1996; Shams et al., 2002; Stevenson et al., 2012); but whether large-scale brain networks exhibit specific patterns that can explain the degree of cross-modal integration during perceptual binding is an unanswered question.

Several studies have highlighted that studying the interplay of segregative and integrative mechanisms offer a comprehensive theoretical framework to interpret large-scale distributed brain activations that are often reported in complex cognitive tasks (Cohen and D'Esposito, 2016; McIntosh, 2004; Mohr et al., 2016; Tomasi et al., 2014; Tononi et al., 1994). In this article, we hypothesized that the cooperation among relevant brain areas via functional integration mechanisms (Friston et al., 1993) and their segregative local information processing work harmoniously to facilitate the cross-modal perceptual binding, and this relationship may be measured using network neuroscience metrics and be driven by behavioral contexts which we aim to control parametrically. Previous studies have used 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 enhanced pSTS activation toward temporal synchrony of auditory and visual modalities (Jones and Callan, 2003). From a large-scale network perspective, noisy stimulus in one modality also increases functional connectivity between pSTS and areas of other modalities; e.g., a blurry video will enhance 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 during the presentation of incongruent (McGurk) speech stimuli, there is increased interaction between motor representational area and inferior frontal gyrus. The motor network was however identified to be contributing towards resolving the audio-visual conflicts (Murakami et al., 2018). The manifestation of audio-visual illusory speech perception thus involves functional domains promoting the integration, such as the pSTS network, and domains with more abstract functions, such as the motor network. The present study empirically demonstrates how subjective perceptual binding emerges from synergistic functional integration and segregation mechanisms of brain (Tomasi et al., 2014). Moreover, 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 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.

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 during an fMRI experiment (Fig. 1A). Subsequently, subjective perception was parametrically controlled by using different versions of this video created by introducing

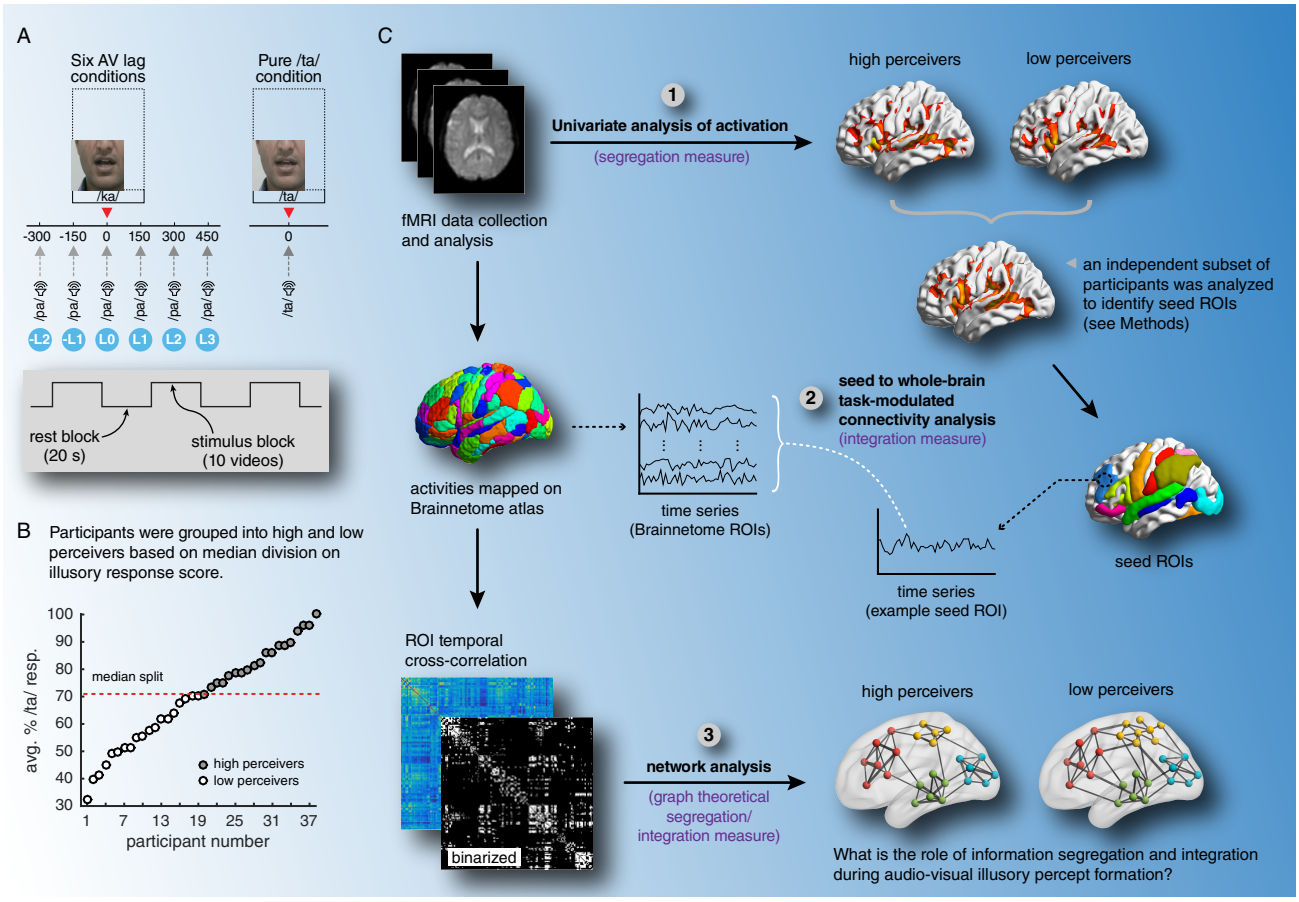

Figure 1. (*A*) Stimulus design using the McGurk paradigm. (*B*) Intersubject variability in susceptibility for illusory perception for all 38 selected volunteers; the susceptibility was evaluated as the average of two maximal percentage /ta/ responses in any two AV lags (dotted line indicates the division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

temporal lags between the 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. 1A); also shown are matched number of synchronous congruent /ta/ (auditory /ta/–visual /ta/). The choosing of an asymmetrical range of audio-visual lags (−L2, . . . , L0, . . . , L3) was guided by previously reported perceptual bias toward visual-leading scenarios (Munhall et al., 1996). Volunteers reported whether they heard /ta/, /pa/, or "others" in a forced-choice paradigm.

Cognitive tasks generally involve distributed brain areas, and a riveting hypothesis for normative human cognition is the maintenance of homeostatic balance between integrative and segregative mechanisms (Cohen and D'Esposito, 2016; McIntosh, 2004; Mohr et al., 2016; Tomasi et al., 2014; Tononi et al., 1994), with imbalances leading to neuropsychiatric disorders (Lord et al., 2017). The present study also demonstrates how alterations of balance affects the perceptual binding of incongruent 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).

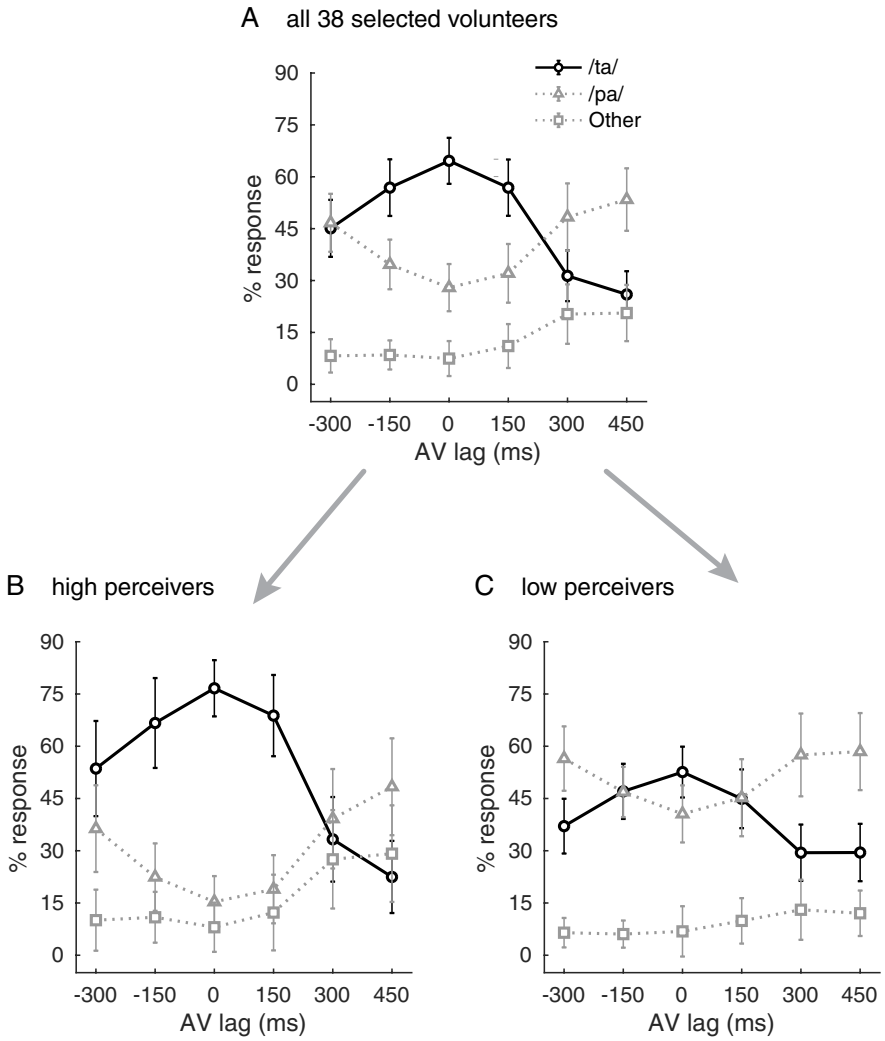

Figure 2. (*A*) Behavioral responses averaged across all 38 selected volunteers. Behavioral responses averaged across (*B*) high perceivers ($N = 19$) and (*C*) low perceivers ($N = 19$). Error bars represent the standard error of the mean (95% CI) across the respective population.

Results

Parametric Control of Illusory Perception. For the McGurk incongruent stimulus, more than a chance level of varied responses to illusory cross-modal /ta/ perception were observed in all participants at multiple AV lags, demonstrating that indeed the degree of AV temporal synchrony affects subjective perception (Fig. 2 and Supplementary Fig. 1). In order to study the neural basis of subjective variability, we categorized the participants into two equally sized groups ($N = 19$) after being sorted based on their *susceptibility* to audio-visual illusion. The susceptibility was 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. 1B).

A two-way 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 participants' 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 a statistically significant interaction between the two factors ($F(5, 213) = 2.9257, p = 0.0141$); Tukey's post-hoc comparison further showed that, between the same time lag condition, high perceivers were more susceptible to illusory /ta/ perception than the low perceivers 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. fMRI images were analyzed using SPM software (Friston et al., 1995) (see Fig. 1C for the schematic of the analysis pipeline). 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. 3A). The extent of task-activated brain areas diminished bilaterally with increasing AV separation (Fig. 3A). Furthermore, when compared with high perceivers, the low perceivers showed relatively weaker activation especially 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 in low perceivers.

We further performed a full factorial analysis of the activations with time lag and participants' grouping as independent factors. The analysis was done in SPM with the default settings. We observed significant main effects of participants' grouping ($F(1, 213) = 7.3491, p = 0.007$) as well as for AV lag ($F(5, 213) = 2.256, p = 0.0049$) (Fig. 3 B and C). The grouping factor had a significant effect on clusters located in all the brain parcels that were found active during L0 and L1 conditions of high and low perceivers. Significant main effect of AV lags was observed in clusters located dominantly in inferior frontal gyrus, superior temporal gyrus including pSTS, caudal cingulate cortex, inferior parietal lobule, insula, lateral occipital cortex, and thalamus. Interestingly, a significant interaction between the two factors was also observed ($F(1, 213) = 2.256, p < 0.0048$), and the active clusters contributing to this interaction were found majorly in superior and inferior frontal gyri, precentral gyrus, inferior parietal lobule (rostroventral area 40), caudal cingulate cortex, and thalamus (Fig. 3D). The clusters with significant effects are reported in Supplementary Table 1, with the corresponding brain subregions mapped according to Brainnetome parcellation (http://atlas.brainnetome.org/; Supplementary Table 2).

To study the differential functional roles of associated brain regions during audio-visual cross-modal perception, 29 seed regions were identified from the group-level analysis of all the 38 participants during synchronous congruent Pure /ta/ condition (Fig. 1A; Supplementary Fig. 2A; Supplementary Table 3). We will henceforth term these 29 seed ROIs which were also found active during L0 and L1 conditions as the perceptual binding network (PBN). The rationale behind using the congruent Pure /ta/ condition for the seed ROI selection are: (1) variable illusory perceptual response from the participants (although the stimulus is congruent) reflects the underlying cross-modal sensory processes. The active brain regions will thus reflect the areas responsible for both unisensory and multisensory processes. (2) The congruent condition also helped us to avoid circularity in the seed ROI selection.

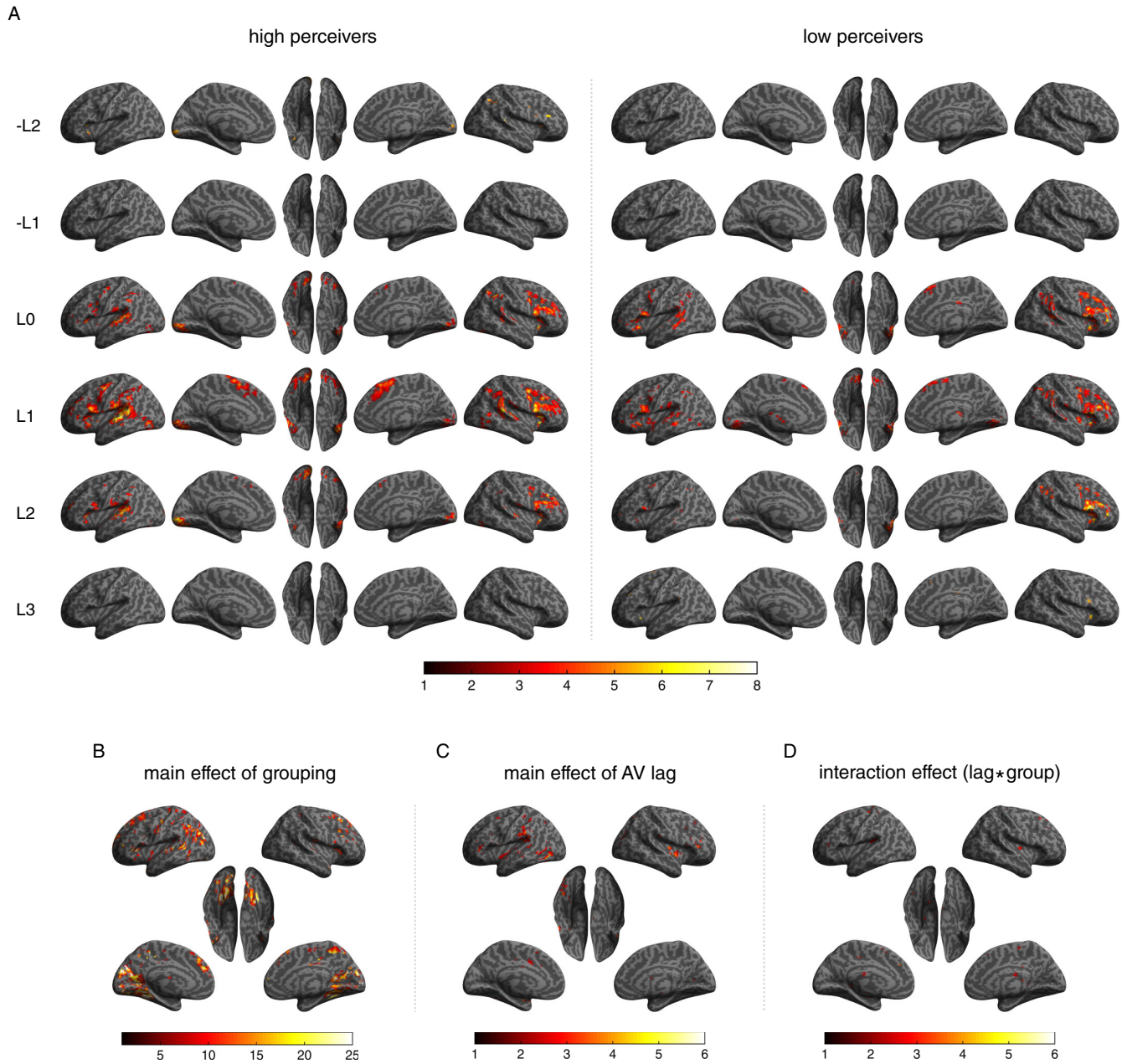

Figure 3. (*A*) Group-level task-enhanced (task > rest) active brain areas for high and low perceivers (peak $p \leqslant 0.05$, FDR corrected). Color bar represents the T-score. (*B*) Main effect of participants' grouping factor. (*C*) Main effect of AV lag factor. (*D*) The interaction between the groupings and AV lag. Color bars in panels (*B–D*) represent the F-score. The activations during congruent /ta/ stimuli are reported in Supplementary Figure 2.

Synergistic Network Organization. The process of cognition involves interactions among associated neural units (brain regions). To investigate the synergistic interactions involving the 29 seed regions of PBN, the seed-based whole-brain functional connectivity (SWFC) for each seed was computed through gPPI modeling (Fig. 1C; see also *Methods* for details) (Friston et al., 1997; McLaren

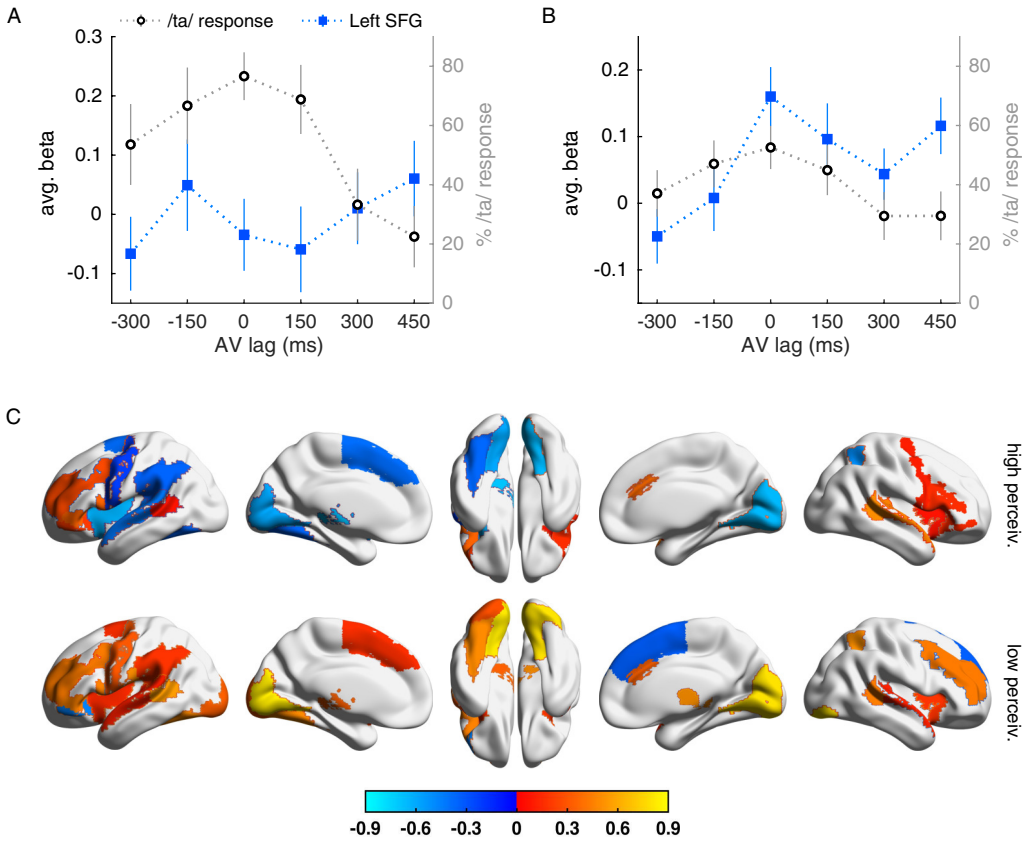

Figure 4. (*A*) Average seed-based whole-brain functional connectivity (SWFC) changes across AV lag conditions for an example left SFG seed in the case of high perceivers, compared against their degree of illusory /ta/ perception (black dotted line). (**B**) The corresponding plot for the low perceivers. Error bars represent the standard error of the mean (95% CI). (**C**) Surface plot showing the strength of SWFC–behavior correlation (SBC) for all 29 seed ROIs of PBN in high perceivers (upper) and low perceivers (lower) (cf. Supplementary Table 4; see also *Methods* for details).

et al., 2012). It was observed that the averaged SWFC mostly decreased at L0 or L1 for high perceivers (see the example left SFG seed regions, Fig. 4A), and the trend is either reversed or weakly followed in the case of low perceivers (Fig. 4B). The question now is whether this decrease in SWFC can be meaningfully associated with the propensity for illusory perception. To meet this end, SWFC–behavior correlation (SBC; see *Methods*) was computed for each PBN seed region. We observed that for high perceivers the majority of the seed ROIs displayed negative SBC, with the positively correlated seed regions mostly concentrated in frontal brain areas and in the right hemisphere (see Fig. 4C and Supplementary Table 4). On the other hand, the low perceivers displayed either a positive or a weakly negative SBC values. These findings emphasized that globally decreased information integration through the PBN seed regions was essential for the illusory audio-visual percept formation. Furthermore, to investigate if the results are independent of the choices of any particular brain parcellation scheme, the procedures were repeated using Schaefer atlas having 300 ROIs (https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal/Parcellations/MNI) (Lawrence et al., 2021; Schaefer et al., 2018). We observed a similar pattern of decreasing global information integration from the seed regions with increasing audio-visual illusory perception (Supplementary Fig. 3).

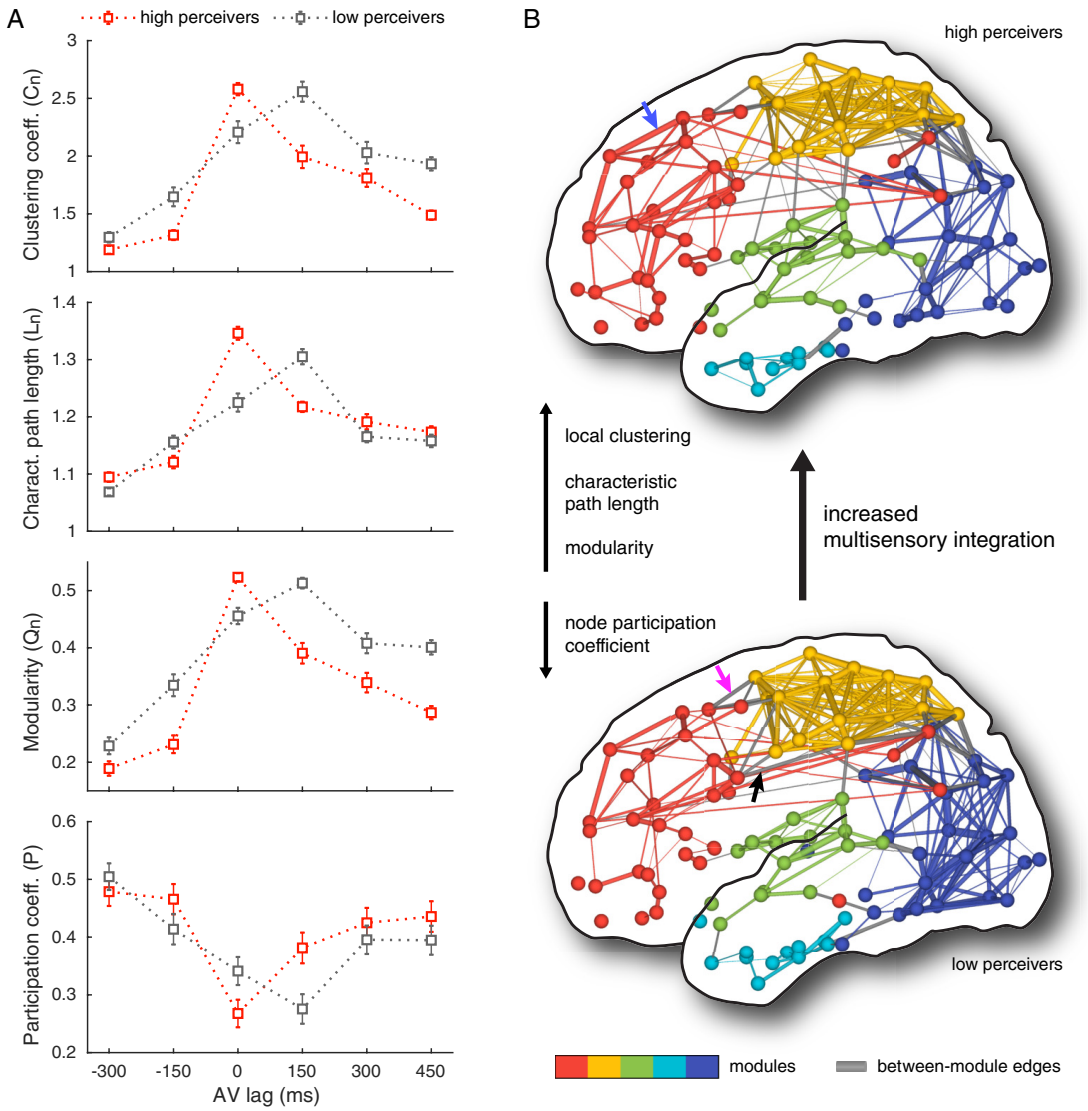

Figure 5. (*A*) Change in the network topological measures for high and low perceivers. (*B*) Left hemispheric brain network differences between high and low perceivers during L0 (synchronous and incongruent) condition.

To validate the above finding that hints toward an altered information segregation–integration balance during the cross-modal perceptual binding, we conducted an independent investigation on the topological features of whole-brain functional networks using metrics from graph theory. For this analysis, 50 binarized whole-brain network subsamples were constructed at the group level for each stimulus condition by randomly permutating 70% of the participants in each group—high and low perceivers. 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. 5A). This indicates 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 characteristic path length may be defined as the minimum steps each node pair will take to reach each other, averaged across all the available node pairs in the network. So, it does not however include the factor of actual structural connection lengths between brain regions. Hence, the manifestation of audio-visual illusion might thus require an altered balance between the segregative (local cohesion) and integrative (global cohesion) information processing. Increased illusory experience was also 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. 5A). Thus, the altered 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 above results were also stable against the choice of a different parcellation scheme (Supplementary Fig. 4).

Taking the left hemispheric brain networks at L0 condition as an example, the network differences between high and low perceivers are illustrated in Fig. 5B: e.g., 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) in the case of high perceivers. However, other AV lag conditions did not necessarily show similar trends. One possible explanation for this inconsistency is the difference in the activation patterns between the two groups, and also the decreasing effect of illusory perception with increasing AV lag. The reduced activation was specifically 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 (Fig. 3A). The graph theoretic measures thus delivered a similar take-home message, as observed for SWFC with behavior, that increased cross-modal illusory perception triggers increased local segregation realized via an increase in clustering coefficient, characteristic path length, and modularity; and also the minimum participation coefficient.

Discussion

We demonstrated that patterned relationships between segregative and integrative information processing mechanisms employed by distributed brain areas are 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) (cf. Fig. 3A). The heightened activity in PBN nodes can be interpreted as local segregation whereas the decreased functional connectivity with the rest of the brain at L0 and L1 reflects diminished global integration. Hence, the simultaneous occurrence of both increased local segregation and diminished global integration can be defined as an altered state of balance between integration and segregation mechanisms that results in higher propensity to perceive illusory-McGurk (cross-modal) sounds. When the level of binding traverses from low to high states, the pattern of balance between segregation and integration mechanisms undergoes an alteration. Supporting this emerging view, we also observed that Pure /ta/ activations in low perceivers (Supplementary Fig. 2B) were similar to the incongruent McGurk stimuli activations at L0/L1 whereas activation levels in high perceivers for Pure /ta/ were overall very low. The corre-

sponding states of altered balanced can also be characterized exclusively from graph theoretic network measures. Finally, the increased thalamic activity in low perceivers also supported an enhanced role of subcortical level processing during cross-modal perception (Gau et al., 2020; Noesselt et al., 2010; Tyll et al., 2011).

Previous studies focused on the 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) require 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 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 (Fig. 3A). However, as was also observed in L0 and L1 stimulus conditions where the degree of pSTS activation could not solely explain the perceptual difference between high and low perceivers, the interaction among distributed brain areas encompassing the primary sensory areas, motor representational 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 a decrease in levels of alpha coherence in electroencephalographic (EEG) activity is indicative of the 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 the rest of the brain with increasing propensity to cross-modal perceptual experience. This was captured via the decrease in task-induced function dependence among several task-activated ROIs and the rest of the brain areas as a function of increasing cross-modal perceptual experience (Fig. 4). From the whole-brain network perspective also, increased cross-modal binding is associated with decreased global integration simultaneously co-occurring with increased local cohesiveness. 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). To the best of our knowledge, this is the first demonstration of competitive segregation and integration mechanisms in multisensory integration tasks. We thus argue that integrative mechanisms are predominantly compensatory compared to the task-driven haemodynamic oxygen metabolism, while contextualizing our results with previous work on multisensory integration (Nath and Beauchamp, 2011). 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.

Our study departs significantly from the 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 energy consumption by the motor networks as well, 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 focusing on the PBN regions. An important limitation of the present study was that, with decreasing AV lag, although the graph theoretic measures such as characteristic path length that measures the functional segregation showed an increasing trend, simultaneously with a decreasing trend for participation coefficient that quantifies the functional integration, when these network measures are correlated with the degree of illusory perception (percentage /ta/ response) across all participants the trends were not significant even for the L0 condition. Hence, in future, a larger sample size will be required to evaluate the participant specific behavior. This brings us to another important limitation of the current study, that due to the block design which have inadequate repetitions of blocks, a subject-specific activation analysis and correlation with behavior couldn't be undertaken. However, in future studies, an event related design augmented by collection of fMRI data at fewer lags will give us the statistical power to evaluate the finer reorganization of functional connectivity differences. Extending this line of reasoning, studies on schizophrenia and other schizoaffective disorders can explore whether an alteration of the balance of segregation and integration can be a viable framework to characterize patient-specific symptoms (Silverstein and Lai, 2021; Waters et al., 2014).

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 consent 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 video stimuli, the gap between the onset of auditory and visual streams varied from −300 to 450 ms in steps of 150 ms (Fig. 1A). 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 was 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. In the stimulus video, the articulation of /ka/ or /ta/ always started from a neutral lip-closed position. The congruent /ta/–/ta/ stimulus had synchronous onset of AV stimuli. 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 to minimize the background noise in Audacity software (Sourceforge.net). The audio sampling rate was 44 kHz and had a bit rate of 128 kbps. The resulting 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 to 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. 1A. Inside the scanner, the stimuli were presented in a block design. Each activation block of length 20 s comprised 10 video stimuli of one catgory (L0/L1/..., etc.). In total there were 28 activation blocks in the whole experiment: 4 repeats for each of the 7 stimulus categories. The 28 activation blocks were alternated with another 28 inter-trial resting blocks or similar duration (i.e., 20 s) where participants attended to a fixation cross. We randomized the presentation of the 7 stimulus categories. The stimuli were presented through an INVIVO MR-compatible CRT screen attached to the head-coil and an MRI-compatible headphone (Philips, The Netherlands). Presentation software (Neurobehavioral Systems, CA) was used to display the stimuli. The participants were instructed to attend to the AV stimuli watching at the male speaker and report their perceptual experience of whether they heard "/pa/", "/ta/," or "any other" through a three-button fiber-optic response device by Curdes as soon as possible (Current Designs, PA, USA). They were also instructed that in case they missed a response, not to worry about the performance and rather concentrate on the current video being played. Most importantly, the volunteers were not given any cue on which AV lag condition will appear next. They were also not given any feedback on their response, since there is no "right" or "wrong" response in the incongruent conditions.

fMRI Data Acquisition

Structural T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) scan was performed at the beginning of each participant's scanning session using a 3T Philips Achieva MRI scanner. The T1-weighted rapid images were acquired with TR=8.4 ms, TE=3.715 ms, and flip angle=8°; the field of view was $250\times250\times170$ mm^3 with a resolution of $0.977\times0.977\times1$ mm^3. For each participant, seven functional T2* scans (TR = 2000 ms, 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 $3.59\times3.59\times5$ mm^3 and the voxel size of the reconstructed images was $1.80\times1.80\times5$ 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, on the basis of their behavioral responses we selected 43 participants. Five more participants had to be 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. 1A). Statistical analyses of the perceptual responses—/pa/, /ta/, and "other"—were performed using two-way analysis of variance (ANOVA) with AV lags and participant grouping as independent variables.

fMRI Data Analysis

Pre-processing. For each participant, 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 T1-weighted brain image was then segmented into grey and white matter maps, and the generated forward deformation field was used to spatially normalize the functional brain volumes to the Montréal Neurological Institute (MNI) standard brain employing the standard trilinear interpolation method. The images were resampled to $2\times2\times2$ mm^3 resolution. The normalized images were then spatially smoothed minimally with a 4 mm full-width at half-maximum (FWHM) isotropic Gaussian kernel. The preprocessing was done using Statistical Parametric Mapping Software (SPM) in MATLAB (MathWorks, MA). The first six brain volumes of each participant were removed to allow the magnetization to reach to a steady state.

Stimulus-based activation. After pre-processing, we performed the first-level fixed-effects analysis 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. Using the contrast images (task block > inter-trial rest block) from the first-level analysis, a second-level random-effects analysis was performed to generate group activation maps (Friston et al., 1995) for each AV lag condition, which captured the stimulus-driven changes. To study the main and possible interaction effects on the BOLD activations, a full factorial analysis was also performed with AV lags and participants' grouping as independent factors (Supplementary Table 1).

The present study also identifies a set of seed regions of interest (ROIs) which together could be responsible for the audio-visual illusory percept formation. To define this set, we performed group-level activation analysis of the congruent synchronous Pure /ta/ condition (cf. Fig. 1A), inclusive of all the 38 participants. We used the Pure /ta/ stimulus condition after considering the fact that participants reported variable illusory perceptual experiences during this stimulus condition, reflecting an underlying cross-modal sensory integration process. The involved brain areas will thus represent the necessary unisensory and multisensory processing components. Another advantage of using the congruent condition is to avoid the circularity in ROI selection. These active brain parcels were also majorly found to be active during L0 and L1 conditions of high and low perceivers. The active parcels were then grouped into 29 ROIs based on major anatomical landmarks of Brainnetome atlas (Fan et al., 2016) (http://atlas.brainnetome.org/; Supplementary Table 2 and Table 3).

Functional Connectivity Analysis

The 29 seed ROIs identified from SPM analysis were used to study functional integration mechanisms by computing the pairwise functional connectivities between each seed ROI and the rest of the brain (*seed-based functional connectivity*). While this approach identifies the global functional integration of a given seed, which when related to seed activation patterns identifies the segregation–integration balance; a more direct way to evaluate this key aspect can come from the *whole-brain graph theoretic network analysis*. Hence, we used both these approaches to study the balance of functional integration–segregation mechanisms.

Seed-based functional connectivity. For the seed-based connectivity analysis, each participant's grey matter volume was parcellated using the Brainnetome atlas. Among the 210 cortical regions of the atlas, the two subregions of pSTS in each hemisphere were merged into a single ROI. From the subcortical regions, 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 2). To measure the task-modulated seed-based functional connectivity changes across AV lag conditions, the functional interactions between each 29 seed region and the remaining 210 ROIs (i.e., exclusive of seed parcels) were evaluated using the generalized psychophysiological interaction (gPPI) modeling approach (Friston et al., 1997; McLaren et al., 2012). We used the CONN toolbox to implement the gPPI model (Whitfield-Gabrieli and Nieto-Castanon, 2012). For each participant, the averaged beta value representing each seed's functional connectivity during a particular AV lag condition was evaluated. 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 to it as SWFC–behavior correlation (SBC). Each seed ROI will thus get one SBC value. However, because of the very noisy characteristics of both fMRI and behavioral data, getting a significant SBC for every participant is problematic. Instead, we focused on the group-level SBC for which a subsample of 70% of the participants from each group were selected by simple random sampling without replacement. The group-averaged SWFC and illusory /ta/ perception trends were estimated for each subsample. The SBC for was then calculated as the Pearson correlation between these two averaged trends. Finally, the sampling was repeated 50 times, thus generating 50 SBC values for each seed ROI. The reported seed SBC is the averaged significant correlation value from the 50 subsamples (one-sample t-test; $p \leqslant 0.05$, FDR corrected). Taking 50 subsamples was necessary because of the experimental design where we had only 6 lag conditions (6 data points) on which the correlation was calculated, and the subsamples thus created the necessary distribution from which a population mean could be estimated.

Whole-brain graph theoretic network analysis. The required BOLD time series from 210 ROIs were extracted using REX toolbox (Whitfield-Gabrieli, 2009) (https://web.mit.edu/swg/software.htm). From each time series, task blocks were extracted based on the experimental design and then concatenated, and also further preprocessed through linear detrending (linear or linear breakpoint detrending). The functional interaction across all the 210 brain parcels were estimated using pairwise Pearson correlation between corresponding task-block BOLD time series (Lee Rodgers and Nicewander, 1988; Pearson, 1895). At the group level, the functional connectomes were generated as follows. First, distinct subsamples were generated by simple 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. From those generated distinct subsamples, we finally selected 50 binarized networks which form a single connected component, i.e., there are no disconnected nodes in these networks. This condition was necessary to allow comparisons across the same network size, as the binarized networks generated with a fixed thresholding sometimes produced isolated nodes (or groups of isolated nodes). Out of the total 210 nodes, one node in the anterior right parahippocampal area that was frequently disconnected irrespective of stimulus conditions was excluded from the network study. Hence, the network size of the finally generated 50 brain networks was 209×209. Furthermore, 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). The randomized networks were used to estimate the normalized network parameters, for a better comparison between networks of same sizes (especially when network size > 200 nodes) (Van Wijk et al., 2010).

To study the functional integration and segregation at different scales of brain network, we calculated the normalized average local clustering coefficient (C_{n}), characteristics path length (L_{n}), modularity (Q_{n}), and node participation coefficients (P). C_{n} and L_{n} estimate the network characteristics at the resolution of nodes (local scale), and Q_{n} and P estimates at the level of the entire network (global scale). The estimation of modularity and participation coefficients required partitioning each brain network into nonoverlapping functional modules. For this purpose we used Blondel et al. (Blondel et al., 2008) method (Louvain method; *community_louvain.m* code available with BCT). The graph measures were then normalized and compared across networks (Guimera et al., 2007; Van Wijk et al., 2010). See also Supplementary Methods.

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.

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.

K. J. Friston, C. Buechel, G. R. Fink, J. Morris, E. Rolls, and R. J. Dolan. Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3):218–229, 1997.

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, 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.

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.

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.

R. M. Lawrence, E. W. Bridgeford, P. E. Myers, G. C. Arvapalli, S. C. Ramachandran, D. A. Pisner, P. F. Frank, A. D. Lemmer, A. Nikolaidis, and J. T. Vogelstein. Standardizing human brain parcellations. *Sci. Data*, 8(1):78, 2021.

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.

A. R. McIntosh. Contexts and catalysts: a resolution of the localization and integration of function in the brain. *Neuroinformatics*, 2:175–181, 2004.

D. G. McLaren, M. L. Ries, G. Xu, and S. C. Johnson. A generalized form of context-dependent psychophysiological interactions (gppi): a comparison to standard approaches. *Neuroimage*, 61 (4):1277–1286, 2012.

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.

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.

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.

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.

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

A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cereb. Cortex*, 28(9):3095–3114, 2018.

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.

S. M. Silverstein and A. Lai. The phenomenology and neurobiology of visual distortions and hallucinations in schizophrenia: an update. *Front. Psychiatry*, page 904, 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.

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.

F. Waters, D. Collerton, D. H. Ffytche, R. Jardri, D. Pins, R. Dudley, J. D. Blom, U. P. Mosimann, F. Eperjesi, S. Ford, et al. Visual hallucinations in the psychosis spectrum and comparative information from neurodegenerative disorders and eye disease. *Schizophr. Bull.*, 40(Suppl_4):S233–S245, 2014.

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's and DR's research was made possible through the financial support provided by Ashoka university and also by Axis Bank. SSS was also 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.