

# SCIENTIFIC REPORTS



OPEN

## A dynamical framework to relate perceptual variability with multisensory information processing

Received: 08 February 2016

Accepted: 15 July 2016

Published: 09 August 2016

Bhumika Thakur<sup>1</sup>, Abhishek Mukherjee<sup>2</sup>, Abhijit Sen<sup>1</sup> & Arpan Banerjee<sup>2</sup>

Multisensory processing involves participation of individual sensory streams, e.g., vision, audition to facilitate perception of environmental stimuli. An experimental realization of the underlying complexity is captured by the “McGurk-effect”- incongruent auditory and visual vocalization stimuli eliciting perception of illusory speech sounds. Further studies have established that time-delay between onset of auditory and visual signals (AV lag) and perturbations in the unisensory streams are key variables that modulate perception. However, as of now only few quantitative theoretical frameworks have been proposed to understand the interplay among these psychophysical variables or the neural systems level interactions that govern perceptual variability. Here, we propose a dynamic systems model consisting of the basic ingredients of any multisensory processing, two unisensory and one multisensory sub-system (nodes) as reported by several researchers. The nodes are connected such that biophysically inspired coupling parameters and time delays become key parameters of this network. We observed that zero AV lag results in maximum synchronization of constituent nodes and the degree of synchronization decreases when we have non-zero lags. The attractor states of this network can thus be interpreted as the facilitator for stabilizing specific perceptual experience. Thereby, the dynamic model presents a quantitative framework for understanding multisensory information processing.

Multisensory signals seamlessly enrich our knowledge of the world<sup>1</sup>. For example, while driving a car or attending to a vocalizer in a noisy background we cannot rely on only one sensory modality be it visual or auditory, rather a harmonious interaction of visual and somatosensory or visual and auditory systems is required. In recent years, multisensory nature of perception has been the focus of much behavioral and neuroscientific research<sup>2</sup>. Merging information from different senses confers distinct behavioral advantages, for example, identification of audio-visual (AV) objects is more rapid<sup>3</sup> than with unimodal stimuli<sup>4,5</sup>, especially when the signals are ambiguous<sup>6,7</sup>. To realize these advantages, the brain continually coordinates sensory inputs across the audiovisual<sup>8,9</sup>, visual-tactile<sup>10,11</sup> and audio-somatic<sup>12</sup> domains and combines them into coherent perceptual objects. However, operational principles of how environmental and neural variables modulate multisensory processing underlying perception are poorly understood<sup>13</sup>. The goal of this paper is to propose a theoretical framework that explains the empirical observations from a paradigmatic framework widely acknowledged to be an entry point in studying multisensory processes.

McGurk and MacDonald<sup>14</sup> demonstrated that the sound of articulating /pa when superimposed on a video of lip movement during articulation of /ka, resulted in an illusory experience of /ta for the perceiver—a phenomenon now known as the McGurk-illusion/effect. Similar triads were also reported such as auditory /ba and visual /ga resulted in perception of /da. Several researchers have used this paradigm to study the key behavioral and neural variables that can modulate perception<sup>15–17</sup>. Ambiguity in one of the sensory streams arising from noisy stimuli affects the neuronal processing of multisensory stimuli<sup>15</sup>. For example, adding white noise to auditory stimuli resulted in an enhanced functional connectivity between the visual cortex and the posterior superior temporal sulcus (pSTS), a brain area populated with multisensory neurons<sup>15</sup>. On the other hand, an increase in ambiguity of visual stimuli (e.g., blurry video) resulted in an enhanced functional connectivity between auditory cortex and

<sup>1</sup>Institute for Plasma Research (IPR), Bhat, Gandhinagar 382428, Gujarat, India. <sup>2</sup>National Brain Research Centre (NBRC), NH 8, Manesar, Gurgaon 122051, Haryana, India. Correspondence and requests for materials should be addressed to B.T. (email: bhumika@ipr.res.in) or A.B. (email: arpan@nbr.ac.in)

pSTS. The relative timing of different sensory input signals is another important factor in multisensory information processing<sup>16</sup>. Temporal proximity is a critical determinant for cross-modal integration by multisensory neurons<sup>2,18</sup>. For instance, the audiovisual integration of speech breaks down if the asynchrony between the visual lip movements and the auditory speech sounds becomes too long<sup>16,19,20</sup>. However a large temporal window exists over which successful integration may occur<sup>16,21–23</sup>. The McGurk illusion, for example persists even when the visual information leads (by up to 240 ms), or lags (by up to 60 ms) the auditory input<sup>16</sup>. In fact, Munhall and colleagues reported the peak of illusory response happens when the visual stimuli leads the auditory stimuli by 180–200 s<sup>16</sup>.

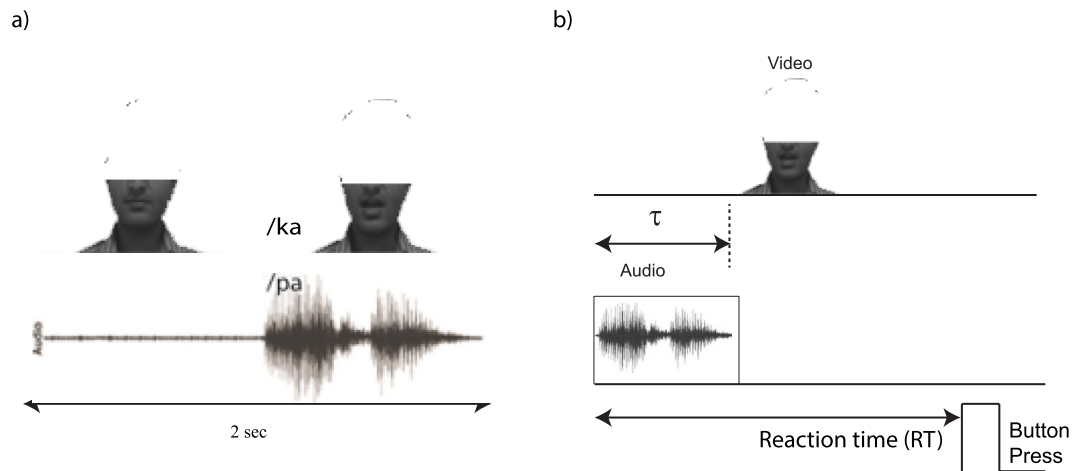
One important component not much explored in the literature is that of an integrative framework/ model that can elucidate the dynamic interactions among environmental and neural variables underlying multisensory processing of stimuli that shapes perception. The existing computational modeling literature address the spatial and temporal integration of incoming multisensory stimuli either at behavioral level or at the neural level<sup>24–30</sup>. Typically most models have 3 components, two unisensory and one multisensory module and attempt to address the integration mechanisms at the level of single units and populations at the neural level (inspired by the presence of multisensory neurons)<sup>31</sup> for constructing a model. Nonetheless, the modeling approaches can be broadly classified into primarily two classes. A Bayesian framework was used by few researchers to explain how unisensory streams of audition and vision can integrate to facilitate perception by solving the spatial localization problem<sup>24,25</sup>. The second approach uses dynamic models of underlying neural systems that can be further subdivided in two classes. The first is biologically inspired modelling of neural dynamics to understand the role of predictive coding<sup>26</sup> as well as spatial and temporal aspects of multisensory processing<sup>27,30</sup> (for a review on neuro-computational approaches to modelling multisensory integration in the brain, see Ursino *et al.*<sup>32</sup>). The second is that of using minimal models with least amount of parameters and variables to identify the key variables affecting dynamic changes in behavior, e.g., acoustic parameters such as gap duration facilitate phonemic categorization<sup>33</sup>. For non-speech sounds and multisensory processing, a dynamic model was introduced by Dhamala and colleagues<sup>23</sup> that explained the phenomenon of drift when slightly asynchronous audio-visual stimuli are presented. The current study incorporates the key environmental variables affecting the McGurk paradigm in a minimal model to understand their relationships with neurally relevant parameters such as the connectivity between the unisensory and multisensory systems and their potential role in oscillatory brain dynamics.

Electrophysiological signals can be conceptualized into patterned oscillations and decomposed into five frequency bands that are physiologically meaningful, delta (0.5–3.5 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (13–30 Hz) and gamma (> 30 Hz)<sup>34</sup>. Beta band power is enhanced in fronto-parietal areas in trials where subjects perceive McGurk illusion accompanied by a reduction in theta power<sup>17</sup>. In general, beta band synchronization has been associated with audio-visual stimulus perception<sup>35</sup>. Phase synchrony and phase modulation of oscillations across the different frequency bands have been suggested to play a key role in the organization of cortical networks engaged in complex cognitive functions such as speech processing<sup>36</sup> and constitute a critical component of auditory-articulatory alignment<sup>37</sup>. There is mounting evidence that coherence of oscillatory neural signals across cortical areas might be a crucial mechanism involved in multisensory processing<sup>34,38–40</sup>. Theta rhythms have been particularly associated with auditory perception of syllabic speech<sup>41</sup>. A recent article emphasizes the role of coherence between lip movement and brain oscillations at low frequency for intelligibility of speech stimuli<sup>42</sup>. Nonetheless, how perceptual categorization parametrically varies with the window of temporal integration and how oscillatory cortical activity observed by EEG/ MEG studies using McGurk-paradigm<sup>3,17,43</sup> relates to perception is unclear.

Developed in 1970's, the Kuramoto model<sup>44</sup> of coupled phase oscillators is commonly used for constructing theoretical models of neurobiological networks with oscillatory dynamics<sup>45,46</sup>. In this model, the phase of an oscillation exhibited by any node of a network becomes the key variable of interest, affected by free parameters such as coupling/ connectivity terms. Thus, each Kuramoto oscillator can depict the oscillatory state of a sub-network (e. g., sections of cortical columns), captured by their circular phase alone and the overall synchronization states can capture the collective dynamics of the network. A network of Kuramoto phase oscillators provides a dynamic framework to explain the functional connectivity changes in the brain electromagnetic data<sup>47</sup>. The original Kuramoto model and its extensions have been used to explore mechanisms underlying oscillations in the human cortex<sup>45</sup>. In this article we propose a dynamical model comprising of three coupled Kuramoto oscillators, coupled via electric coupling and time-delay. Electric coupling captures physiological constraints of the audio-visual system and time-delay captures the environmental factor of temporal asynchrony (see also)<sup>48</sup>. Empirically such networks can be imaged non-invasively from EEG/ MEG studies<sup>49</sup>. Computationally, presence of time-delays enriches the attractor space of a dynamical system<sup>50</sup>. Our proposed model predicts the behavior of McGurk perceivers reported under ambiguous stimuli scenarios as well as during variation of AV onset lags. The rest of the paper is organized as follows. In the Methods section, we present the behavioral experimental paradigm using McGurk like stimuli, where one can study the relationship between perceptual experience and psychophysical parameters such as audio-visual onset lags. Secondly, we propose a theoretical model of multisensory perception using symmetry arguments. In the Results section we perform statistical analysis on the behavior and report the model performance under various parameter set-ups. Finally, in the Discussion section, we discuss the theoretical results in context of the experimental paradigm and argue how this modelling framework captures the key features of complex multisensory integration processes and can potentially be helpful for explaining other experimental paradigms as well.

## Methods

**Experimental paradigm.** *Subjects and stimuli.* Fifty-two healthy right-handed adult participants (25 female, 27 male) of age range 20–35 years (mean age = 24.5 years, SD = 3.12) participated in a behavioral study of duration of about 45 minutes. The undertaken study design was approved by Institutional Human Ethics



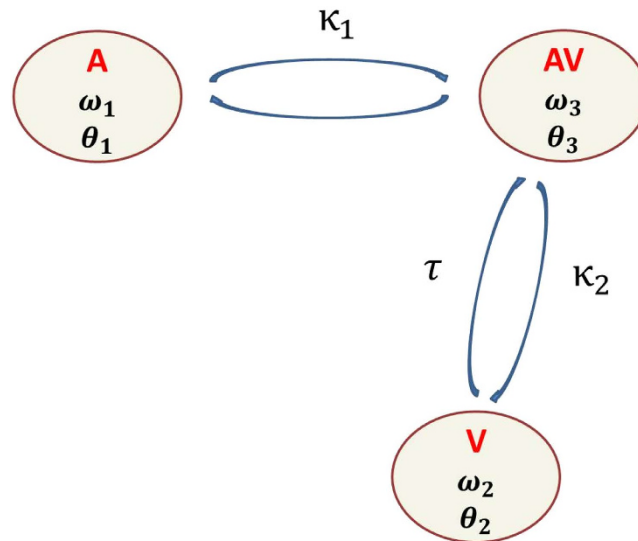
**Figure 1.** (a) Stimulus videos are created using visual lip movement of /ka superimposed on auditory /pa. Participants reported the auditory object they heard while watching the video using a button press response box. b) Stimuli videos were created at different audio-visual lags  $\tau$ , the timing difference between the onset of sound and lip movement, with values ranging from [−300, 450] ms.

Committee (IHEC), National Brain Research Centre (NBRC) and the study was carried out in accordance with the guidelines set by IHEC, NBRC and in strict adherence to the declaration of Helsinki. All participants provided written informed consent in a format approved by IHEC, NBRC and reported normal vision and hearing and no history of neurological disorders. 7 video stimuli, each of 2 seconds(s) duration were prepared, of which 6 were incongruent audio-visual objects where audio recordings of a human speaker vocalizing /pa is dubbed on the lip movement of vocalization /ka (/pa-/ka) and 1 was a congruent audio-visual object /ta (/ta-/ta). The gap in onset of auditory and visual streams was varied from −300 to 450 ms in steps of 150 ms in the 6 incongruent videos. Negative sign implies that auditory stimulus onset preceded that of lip movement onset and positive implies that lip movement starts before the sound. An asymmetric range of AV lags was chosen for incongruent trials, because a previous study by Munhall and colleagues<sup>16</sup> reported that the dominance of illusory perception was skewed towards positive lags where the start of lip movement precede sound onset. The congruent /ta-/ta video had synchronous onset of AV stimuli. The male speaker's lips were in a neutral closed position during when not engaged in utterance of the syllables /ka or /ta and the articulation always started from a neutral position. Videos were created/ edited using VideoPad video-editing software (NCH Software, CO) 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 s 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 study was done inside a 3T MRI scanner as part of a brain imaging investigation, the results of which will be presented elsewhere.

**Task.** The task design is illustrated in Fig. 1. Inside the MR scanner, the stimuli were presented in a block design with 20 s activation blocks consisting of 10 videos of one kind of AV lag. In total there were 28 activation blocks in the whole experiment inclusive of 4 activation blocks for each stimuli category (AV lag). There were alternating 28 resting blocks, each of 20 s duration. The order of presentation of activation blocks was randomized and the same kind of block never appeared consecutively. Within a block the trial videos comprised of one kind of lag, and this is a limitation of the fMRI block design. AV stimuli were presented through a INVIVO MR - compatible CRT screen attached to the head-coil and MRI-compatible headphones (Philips, The Netherlands). Presentation software (Neurobehavioral Systems, CA) was used to display the stimuli. Participants were presented with the stimuli (Fig. 1) and asked to indicate their response based on their perception via three buttons designated for /ta, /pa and for “any other” perceptual categories. They were instructed to attend to the audio-visual stimuli and watch the speaker at all times. A fibre-optic button-pad by Curdes (Current Designs, PA, USA) was used to record the responses of the participants.

**Theoretical framework.** Multisensory systems research has been built on analogies drawn from constituent unisensory processing modules. For example multisensory processing of audio-visual inputs was studied in comparison to standalone auditory and visual processing<sup>3,31,51</sup>. However, increasing evidences suggest that a network of unisensory and multisensory systems may be involved in multisensory task processing<sup>15,17,23</sup>. Thus, it is reasonable to assume that the most elemental model of multisensory perception will involve at least two unisensory systems and one multisensory system<sup>11,24,26,27</sup> even though a recent hypothesis suggests two unisensory streams can in principle give rise to multisensory effects<sup>13</sup>. Furthermore, changes in oscillatory brain rhythms such as pre-stimulus enhancement of beta power and post-stimulus depreciation of theta band power have been identified as the hallmarks of illusory perception in pre-stimuli or post stimuli regimes<sup>17,52</sup> respectively.

As discussed in the Introduction, Kuramoto oscillators<sup>44,53</sup> are increasingly becoming a handy tool to model oscillatory brain dynamics. The Kuramoto model was originally developed as a system of phase oscillators, each rotating at different intrinsic frequencies and coupled through the sine of their phase differences. Since then the



**Figure 2. The dynamical model for multisensory speech perception consisting of three Kuramoto oscillators for positive AV lags when visual leads auditory stimulus.** Oscillators A, V and AV represent the audio, visual and multisensory systems respectively. Their corresponding intrinsic frequencies are  $\omega_1$ ,  $\omega_2$  and  $\omega_3$ , and the respective phases are  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ . The coupling between A and AV is instantaneous and its strength is  $\kappa_1$  whereas the coupling between V and AV is time delayed by the parameter  $\tau$  and its strength is  $\kappa_2$ . There is no direct coupling between oscillators A and V.

model and its modifications have been used to study a number of biological and other phenomena such as choring crickets, flashing fireflies, pacemaker cells in heart. In the Kuramoto model, the phases of the individual oscillators evolving over time obey the following equations

$$\dot{\theta}_n = \omega_n + \sum_{m=1}^N \kappa_{nm} \sin(\theta_m - \theta_n), \quad (1)$$

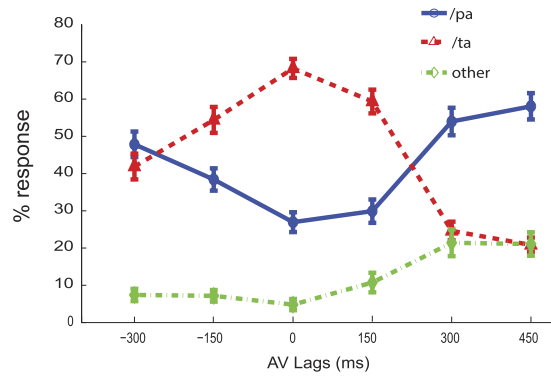
where  $\theta_n$  is the phase and  $\omega_n$  is the intrinsic frequency of the  $n^{\text{th}}$  oscillator and  $\kappa_{nm}$  is the strength of the coupling between the  $n^{\text{th}}$  and  $m^{\text{th}}$  oscillator. Here the over-dot over  $\theta_n$  represents the time derivative.

We consider a system of three coupled Kuramoto phase oscillators configured in the manner shown in Fig. 2. The phase of oscillator A representing the auditory system is  $\theta_1$ , the phase of the oscillator V representing the visual system is  $\theta_2$  and the phase of the oscillator AV representing the multisensory system is  $\theta_3$ . Each oscillator has a distinct natural frequency of oscillation, denoted by  $\omega_1$ ,  $\omega_2$  and  $\omega_3$  respectively. Thus, our model captures the functional connectivity between unisensory systems such as auditory, visual cortices and the multisensory system e.g., posterior superior sulcus (pSTS) that has been observed in functional imaging studies<sup>15</sup> via the coupling parameters  $\kappa_1$  and  $\kappa_2$ .

Furthermore, our experimental results as well as from Munhall and colleagues<sup>16</sup> indicate that a crucial parameter in creation of perceptual states is the AV lag which we capture through the time delay  $\tau$ . In normal hearing and visual circumstances one can expect the coupling between individual sensory systems and multisensory system to be balanced. However, perturbations to one of the sensory streams such as unreliable visual or auditory signals can lead to a situation of unbalanced coupling as shown by Nath and Beauchamp<sup>15</sup>. There is no direct coupling between A and V oscillators in our model though in the neural system there are evidences of connectivity between A1 and V1<sup>54</sup>. This is primarily because we are interested to understand the key symmetries of the most simple multisensory dynamical model catered towards understanding oscillatory states of brain. Nonetheless, A and V do interact functionally in our model, since they are both coupled to AV, the functional unit of multisensory system. To model the experimental situation of AV lags (positive or negative) between the audio and visual stimuli the coupling between oscillators 2 and 3 are time delayed by the parameter  $\tau$ . When the visual stimulus precedes the auditory stimulus (positive lag) the dynamics of phase oscillators is expressed as

$$\begin{aligned} \dot{\theta}_1(t) &= \omega_1 + \kappa_1 \sin(\theta_3(t) - \theta_1(t)), \\ \dot{\theta}_2(t) &= \omega_2 + \kappa_2 \sin(\theta_3(t - \tau) - \theta_2(t)), \\ \dot{\theta}_3(t) &= \omega_3 + \kappa_1 \sin(\theta_1(t) - \theta_3(t)) + \kappa_2 \sin(\theta_2(t - \tau) - \theta_3(t)), \end{aligned} \quad (2)$$

where  $\kappa_1$  is the strength of interaction between the auditory(A) and the AV oscillator and  $\kappa_2$  is the coupling strength between the visual (V) and the AV oscillator. When the auditory precedes the visual stimulus, one can consider the possibility of  $\tau$  being negative. However, physically it does not make sense to make the present dynamics dependent on the future. Alternatively, one can intuitively relate the situation of negative  $\tau$  to a situation where oscillator A has a time delayed coupling with AV while the visual oscillator V is instantaneously coupled. Hence, when the auditory stimulus precedes the visual stimulus, the phase oscillators can be represented as



**Figure 3. Normalized behavioral responses from 34 subjects.** Mean response for each perceptual category is presented as a function of AV lag  $\tau$ . The error bars reflect 95% significance thresholds.

$$\begin{aligned}
 \dot{\theta}_1(t) &= \omega_1 + \kappa_1 \sin(\theta_3(t - \tau) - \theta_1(t)), \\
 \dot{\theta}_2(t) &= \omega_2 + \kappa_2 \sin(\theta_3(t) - \theta_2(t)), \\
 \dot{\theta}_3(t) &= \omega_3 + \kappa_1 \sin(\theta_1(t - \tau) - \theta_3(t)) + \kappa_2 \sin(\theta_2(t) - \theta_3(t)).
 \end{aligned} \quad (3)$$

Equations 2 and 3 are similar in form but are not mirror images due to the difference in the individual frequencies of the oscillators that gives rise to asymmetry. When the oscillators are allowed to interact the frequencies of oscillations of all three units synchronize with various phase relationships (analytically derived in Results section for a simple case).

States of synchronized oscillations across a chain of phase oscillators can be mathematically identified by defining a complex order parameter,  $Z$ , which gives a quantitative measure of the synchronization among the oscillators and can be expressed as,

$$Z = Re^{i\Phi} = \frac{1}{3}[e^{i\theta_1} + e^{i\theta_2} + e^{i\theta_3}]. \quad (4)$$

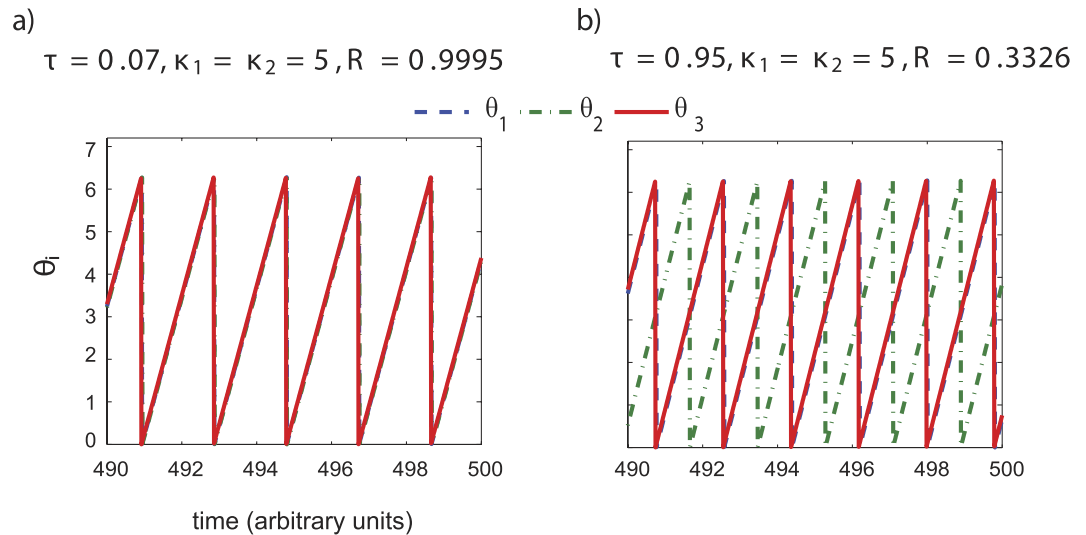
Here the amplitude  $R(t)$  is a measure of both the synchronization of the frequencies of the three oscillators as well as the phase coherence of the oscillators and  $\Phi(t)$  measures the average phase of the system. The system can get synchronized in frequency as the coupling strengths  $\kappa_1$  and  $\kappa_2$  are increased but the oscillators can still be separated in their mutual phases while oscillating at the same frequency. For phase synchronized state a value of  $R = 1$  represents maximum synchronization (0 phase difference) among the oscillators and  $R = 0$  means maximally separated phases among individual oscillators.

Earlier studies have proposed that perceptual experience can be qualitatively conceptualized as attractor states of a dynamical system<sup>23,33</sup>. Stable synchronization patterns at certain relative phase relationships are attractors in a dynamical system, and instabilities signify switches in perceptual state. In this framework, the state of the order parameter or the collective variable  $R$  can be representative of the overall perceptual categorization and can be studied as a function of the interactions among constituent psychophysical variables, coupling strengths, the spread in the frequencies of the oscillators as well as the time delay parameter  $\tau$ . Time delay  $\tau$  can be interpreted as the temporal window over which integration of AV information takes place.

## Results

**Behavior.** The button press responses to perceptual experience collected from each participant during the experiment were analyzed offline using customized MATLAB codes. A maximum of 40 responses were expected for each AV lag condition. Tasks with less than 35 responses were rejected since estimates of perceptual categorization may be biased during computation of percentage responses. No subjects had to be rejected based on this criteria. Behavioral responses of attempted trials of each participant were converted into percentage measures for each perceptual category, /pa, /ta or “other” corresponding to AV lags over a range [−300, 450] ms. Subsequently, 34 participants who reported perceiving /ta for at least 60% of the total responses at any AV Lag are defined as “McGurk-perceivers” and simply referred to as perceivers.

We observed that perceivers report /ta maximally at an AV lag of 0 ms when the lip movement of the speaker is synchronous with the onset of auditory stimulus (Fig. 3). Conversely, /pa perception was reported the least number of times at AV lag of 0 ms. We performed an Analysis of Variance (ANOVA) over the percentage responses (without repetitions) in each category, /pa, /ta and “other” with lags and subjects treated as factors. For threshold of statistical significance set at  $p = 0.001$ , there was a significant change in /ta responses across lags ( $F(5, 165) = 24.46, p < 0.001$ ), but not across subjects ( $F(33, 165) = 1.72, p = 0.0141$ ). /pa responses also significantly varied across lags ( $F(5, 165) = 12.57, p < 0.001$ ) and also across subjects ( $F(33, 165) = 5.43, p < 0.001$ ). The “other” responses also significantly varied across lags ( $F(5, 165) = 7.11, p < 0.001$ ) and subjects ( $F(33, 165) = 4.25, p < 0.001$ ).

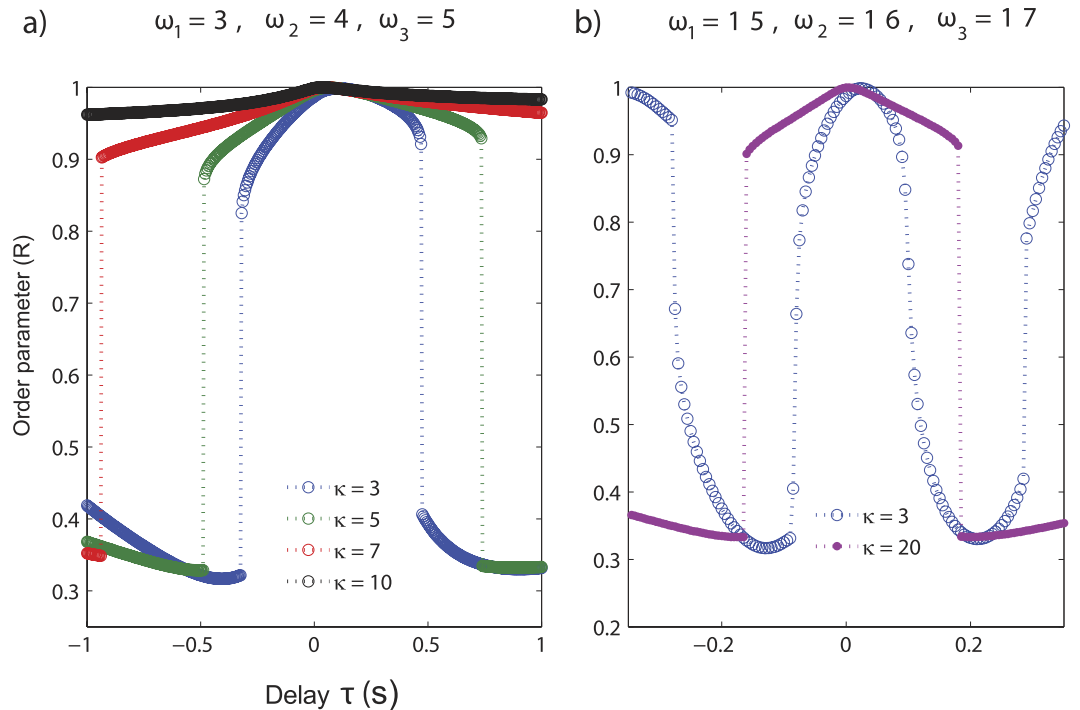


**Figure 4.** Time series of the phases  $\theta_1(t)$ ,  $\theta_2(t)$  and  $\theta_3(t)$  of the three Kuramoto oscillators having intrinsic frequencies  $\omega_1 = 3$ ,  $\omega_2 = 4$  and  $\omega_3 = 5$  respectively (a) at delay  $\tau = 0.07$  and coupling strength  $\kappa_1 = \kappa_2 = 5$ . At this value of delay, all three oscillators are almost synchronized in-phase and hence the order parameter  $R = 0.9995$ . (b) at delay  $\tau = 0.95$  and coupling strength  $\kappa_1 = \kappa_2 = 5$ . At this value of delay, auditory A and multisensory AV oscillators are oscillating almost in-phase whereas the visual oscillator V is out-of phase with the other two. Hence the order parameter  $R = 0.3326$

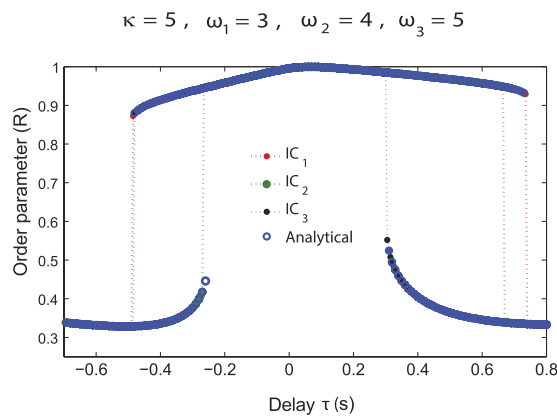
**Dynamical model of multisensory processing.** We investigated the system of equations 2 and 3 numerically and analytically to gain insights about how time-delays may influence the stability of perceptual states. We studied the model under two scenarios, one with balanced coupling and the other with unbalanced coupling. For sake of simplicity, the analytical derivations are computed for balanced coupling conditions whereas we investigate the unbalanced coupling scenarios numerically. All numerical simulations were performed using customized MATLAB (www.mathworks.com) codes and delay-differential equation solver dde23. We were interested in the collective behavior of the system of oscillators (Fig. 2, equation 2), in particular the dynamics of the order parameter that quantified the synchronized states as a function of time-delay and coupling parameters.

**Balanced coupling.** We solve the system of equations 2 and 3 for a range of  $\kappa = \kappa_1 = \kappa_2$  values from 3–20 and for different values of the time delay ( $\tau$ ). In Fig. 4 we have plotted the time series of the phases  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  at two delay values when coupling strength  $\kappa = 5$ . At delay  $\tau = 0.07$ , all three oscillators are almost synchronized in-phase and hence the order parameter  $R = 0.9995$ . However at delay  $\tau = 0.95$ , auditory A and multisensory AV oscillators are oscillating almost in-phase whereas the visual oscillator V is out-of phase and the order parameter  $R = 0.3326$ . Hence, the transition from almost complete synchronization to a lower degree of synchronization occurs at a critical value of  $\tau$ . To explore the parameter space further, we computed the order parameter for different time delays ( $\tau$ ) and coupling strengths ( $\kappa$ ) (Fig. 5). The right halves of the curves correspond to the “positive” audio-visual lags (where the visual stimulus precedes the auditory stimulus) and are obtained from the numerical integration of equation 2. The left halves of the curves correspond to “negative” audio-visual lags (where the auditory stimulus precedes the visual stimulus) and are obtained numerically from equation 3. In Fig. 5a, the initial intrinsic frequencies were set at  $\omega_1 = 3$ ,  $\omega_2 = 4$ ,  $\omega_3 = 5$ . This was roughly in  $\theta$  frequency range (3–5 Hz) which was earlier shown to be relevant for the existence of perceptual states<sup>17</sup>. The initial conditions for simulations were randomized to span all possible relative phase states. For weak coupling  $\kappa < 1$ , the existence of synchronization states is not possible consistently and  $R$  oscillates with time. For  $\kappa \geq 1$ , a clear synchronization state appears at least for a range of  $\tau$  values. Two stable states of order parameter emerge, one synchronization state around  $R \approx 1$ , and the other around  $R \approx 0.33$ . Furthermore, we also observed that the critical value of  $\tau$  increases with increase in  $\kappa$ . In other words the island of synchronization expands with increasing  $\tau$ . For example for  $\kappa = 10$  a transition happens for a high value of  $\tau$  (not shown in Fig. 4), that is not behaviourally relevant. On the other hand, we also observed that higher intrinsic frequencies of the oscillators can recede the island of synchronization. For example, if the intrinsic frequencies are set to  $\omega_1 = 15$ ,  $\omega_2 = 16$ ,  $\omega_3 = 17$  (beta band frequency) we see the transition from synchronization state occurs at lower values of  $\tau$ , and the extent of synchronization regime decreases (Fig. 5b). An important point to note here is that the critical transition value of  $\tau$  may not be same for “positive” and “negative” audio-visual lags. This stems from the disparity in the intrinsic frequencies of the oscillators.

**Multistability and analytical solutions.** Multistable states of a dynamical system can be relevant for describing perception and action, in particular the inter-trial and inter-subject variability. We investigated the existence of such states in our dynamical model. The presence of a time delayed coupling provides for a rich collective dynamics of the system including the existence of multiple collective frequencies and hence the possibility of coexistence of some of these states. We have looked at this possibility by theoretical and numerical investigation of the model



**Figure 5.** Variation of the order parameter ( $R$ ) of the system of oscillators with delay ( $\tau$ ) for the case of balanced coupling  $\kappa_1 = \kappa_2 = \kappa$  when the intrinsic frequencies are a)  $\omega_1 = 3, \omega_2 = 4$  and  $\omega_3 = 5$  and b)  $\omega_1 = 15, \omega_2 = 16$  and  $\omega_3 = 17$ . We have obtained the above curves for various  $\kappa$  values for the same set of initial phases.



**Figure 6.** Order parameter ( $R$ ) as a function of time delay ( $\tau$ ) for  $\omega_1 = 3, \omega_2 = 4, \omega_3 = 5$  and coupling strengths  $\kappa_1 = \kappa_2 = \kappa = 5$  computed using the analytical expression Eq. 15 (blue circles). Red curve is obtained by numerically integrating the model equations (Eqs 2 and 3) to get the individual phases and then using the order parameter expression (Eq. 15) for one set of initial conditions  $IC_1$ . The green and black dash curves are obtained numerically for different sets of initial phases  $IC_2$  and  $IC_3$  respectively. We see that the system is multistable, i.e., the order parameter can take different values for same  $\tau$  on changing initial conditions. The analytical results are in great agreement with the numerical values of the order parameter.

equations over a wide range of parameter values and randomized initial conditions in Figs 5 and 6. The system is multi-stable, meaning order parameter can take different values at same  $\tau$  based on initial conditions (Fig. 6). The multistability seen in this model system may be related to the multistability seen in speech perception in terms of responses registered as /pa and “others” that occur 50% and approximately 10% of time respectively (Fig. 3).

The synchronized state (where all the oscillators oscillate with same frequency  $\Omega$ ) is given by  $\theta_i(t) = \Omega t + \psi_i$ . Let the phase shift between auditory and AV oscillator be  $\theta_3 - \theta_1 = \psi_3 - \psi_1 = \phi_1$  and the phase shift between visual and AV oscillator is  $\theta_3 - \theta_2 = \psi_3 - \psi_2 = \phi_2$ . This gives  $\theta_2(t - \tau) - \theta_3(t) = -\phi_2 - \Omega\tau$  and  $\theta_3(t - \tau) - \theta_2(t) = \phi_2 - \Omega\tau$ . Using these expressions, system of equations 2 can then be written as

$$\begin{aligned} \Omega &= \omega_1 + \kappa \sin \phi_1, \\ \Omega &= \omega_2 + \kappa \sin(\phi_2 - \Omega\tau), \\ \Omega &= \omega_3 - \kappa \sin \phi_1 - \kappa \sin(\phi_2 + \Omega\tau). \end{aligned} \tag{5}$$

Summing across the sub-equations of Eq. 5, we obtain

$$\Omega = \bar{\omega} - \frac{2\kappa}{3} \cos \phi_2 \sin(\Omega\tau), \tag{6}$$

where  $\bar{\omega} = \frac{\omega_1 + \omega_2 + \omega_3}{3}$ . Substituting  $\Omega$  from Eq. 6 in Eq. 5, we obtain  $\phi_1 = \sin^{-1}\left(\frac{\Omega - \omega_1}{\kappa}\right)$  and  $\phi_2 = \Omega\tau + \sin^{-1}\left(\frac{\Omega - \omega_2}{\kappa}\right)$ . Eq. 6 gives the collective frequency  $\Omega$  as a function of delay  $\tau$  when we have “positive” audio-visual lags. Since it is a transcendental equation, therefore we get multiple values of synchronization frequency  $\Omega$  as the solution of Eq. 6 for a given value of delay. To determine which of these corresponds to a stable synchronization state, one has to do a stability analysis.

To obtain the expression of collective frequencies for the “negative” audio-visual lags, one has to do the same analysis as above using Eq. 3. For “negative” audio-visual lags, we have

$$\Omega = \bar{\omega} - \frac{2\kappa}{3} \cos \phi_1 \sin(\Omega\tau), \tag{7}$$

where  $\bar{\omega}$  is same as above but  $\phi_1 = \Omega\tau + \sin^{-1}\left(\frac{\Omega - \omega_1}{\kappa}\right)$  and  $\phi_2 = \sin^{-1}\left(\frac{\Omega - \omega_2}{\kappa}\right)$ .

*Derivation of stability condition.* We perform a linear stability analysis to determine the local stability of the synchronization state by adding a small perturbation

$$\theta_i(t) = \Omega t + \psi_i + \varepsilon \xi_i(t), \tag{8}$$

where  $0 < \varepsilon < 1$ . Taking  $\psi_3 - \psi_1 = \phi_1$  and  $\psi_3 - \psi_2 = \phi_2$  gives

$$\begin{aligned} \theta_3(t) - \theta_1(t) &= \varepsilon [\xi_3(t) - \xi_1(t)] + \phi_1, \\ \theta_3(t - \tau) - \theta_2(t) &= \varepsilon [\xi_3(t - \tau) - \xi_2(t)] + \phi_2 - \Omega\tau, \\ \theta_2(t - \tau) - \theta_3(t) &= \varepsilon [\xi_2(t - \tau) - \xi_3(t)] - \phi_2 - \Omega\tau. \end{aligned} \tag{9}$$

Substituting these in equations 2 we obtain

$$\begin{aligned} \Omega + \varepsilon \dot{\xi}_1(t) &= \omega_1 + \kappa \sin[\varepsilon(\xi_3(t) - \xi_1(t)) + \phi_1], \\ \Omega + \varepsilon \dot{\xi}_2(t) &= \omega_2 + \kappa \sin[\varepsilon(\xi_3(t - \tau) - \xi_2(t)) + \phi_2 - \Omega\tau], \\ \Omega + \varepsilon \dot{\xi}_3(t) &= \omega_3 - \kappa \sin[\varepsilon(\xi_3(t) - \xi_1(t)) + \phi_1] - \kappa \\ &\quad \sin[\varepsilon(\xi_3(t) - \xi_2(t - \tau)) + \phi_2 + \Omega\tau]. \end{aligned} \tag{10}$$

Linearizing the set of equations 10 and using 5 we obtain

$$\begin{aligned} \dot{\xi}_1(t) &= \kappa [\xi_3(t) - \xi_1(t)] \cos \phi_1, \\ \dot{\xi}_2(t) &= \kappa [\xi_3(t - \tau) - \xi_2(t)] \cos(\phi_2 - \Omega\tau), \\ \dot{\xi}_3(t) &= -\kappa [\xi_3(t) - \xi_1(t)] \cos \phi_1 - \kappa [\xi_3(t) - \xi_2(t - \tau)] \cos(\phi_2 + \Omega\tau). \end{aligned} \tag{11}$$

Taking  $\xi_i(t) = \nu_i e^{\lambda t}$  and substituting in the system of linear equations 11, where  $\lambda$  is the set of eigenvalues and  $\nu_i$  are the corresponding eigenvectors, we obtain the characteristic polynomial

$$\begin{vmatrix} -A - \lambda & 0 & A \\ 0 & -B - \lambda & B e^{-\lambda\tau} \\ A & C e^{-\lambda\tau} & -A - C - \lambda \end{vmatrix} = 0, \tag{12}$$

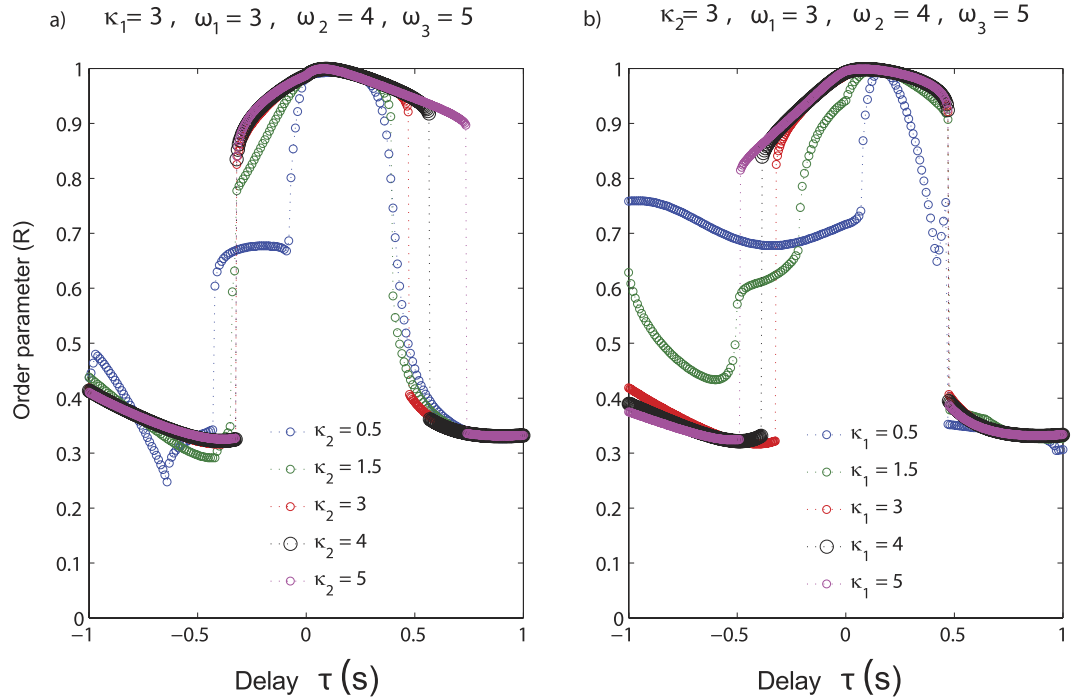
where  $A = \kappa \cos \phi_1$ ,  $B = \kappa \cos(\phi_2 - \Omega\tau)$  and  $C = \kappa \cos(\phi_2 + \Omega\tau)$ . The stability of the synchronization frequency  $\Omega$  for a given value of  $\kappa$  and  $\tau$  can be checked using the above equation by looking at the signs of the eigenvalues  $\lambda$ . The frequency synchronized state with frequency  $\Omega$  is stable if none of the eigenvalues have positive real parts.

*Analytical expression for order parameter.* The order parameter of our system is defined as

$$R e^{i\Phi} = \frac{1}{3} \sum_{j=1}^3 e^{i\theta_j}. \tag{13}$$

After a little algebra, we can express  $R$  in the form





**Figure 7.** Variation of the order parameter ( $R$ ) with delay ( $\tau$ ) of the system of oscillators having intrinsic frequencies  $\omega_1 = 3, \omega_2 = 4$  and  $\omega_3 = 5$  (a) when the coupling strength between A and AV, ( $\kappa_1$ ) is kept constant at 3 and the coupling between V and AV, ( $\kappa_2$ ) is varied from 0.5 to 5. (b) the coupling strength between V and AV, ( $\kappa_2$ ) is kept constant at 3 and the coupling between A and AV, ( $\kappa_1$ ) is varied from 0.5 to 5.

$$R = \frac{1}{3} \sqrt{[3 + 2 \cos(\theta_1 - \theta_2) + 2 \cos(\theta_3 - \theta_2) + 2 \cos(\theta_3 - \theta_1)]}. \quad (14)$$

For a frequency synchronized state,  $\theta_i(t) = \Omega t + \psi_i$  and we have taken  $\theta_3 - \theta_1 = \psi_3 - \psi_1 = \phi_1$  and  $\theta_3 - \theta_2 = \psi_3 - \psi_2 = \phi_2$ . Therefore the analytical expression for the variation of the order parameter  $R$  is given by

$$R = \frac{1}{3} \sqrt{[3 + 2[\cos(\phi_1 - \phi_2) + \cos(\phi_2) + \cos(\phi_1)]]}. \quad (15)$$

For an in-phase synchronized state  $\psi_1 = \psi_2 = \psi_3$ . Therefore,  $\phi_1 = \psi_3 - \psi_1 = 0$  and  $\phi_2 = \psi_3 - \psi_2 = 0$  and hence  $R = 1$ . We already have the analytical expressions for  $\phi_1$  and  $\phi_2$ :  $\phi_1 = \sin^{-1}\left(\frac{\Omega - \omega_1}{\kappa}\right)$  and  $\phi_2 = \Omega\tau + \sin^{-1}\left(\frac{\Omega - \omega_2}{\kappa}\right)$ . For a given set of  $\kappa, \tau$  and intrinsic frequencies, the synchronization frequency  $\Omega$  is obtained from the transcendental equation Eq. 6 by a numerical root finding method and the stability of the corresponding state is then checked using Eq. 12.  $\Omega$  values corresponding to the stable states are substituted in the expressions of  $\phi_1$  and  $\phi_2$  and the values of the order parameter  $R$  are then obtained from Eq. 15. For “negative” audio-visual lags,  $R$  follows the same Eq. 15 but here  $\Omega$  is given by Eq. 7,  $\phi_1 = \Omega\tau + \sin^{-1}\left(\frac{\Omega - \omega_1}{\kappa}\right)$  and  $\phi_2 = \sin^{-1}\left(\frac{\Omega - \omega_2}{\kappa}\right)$ .

In Fig. 6, we have plotted the analytical as well as the numerical values of the order parameter ( $R$ ) at corresponding values of time delay ( $\tau$ ) for the system of oscillators having intrinsic frequencies  $\omega_1 = 3, \omega_2 = 4, \omega_3 = 5$  and coupling strength  $\kappa_1 = \kappa_2 = \kappa = 5$  for different initial conditions. The order parameter  $R$  can have more than one value at a given value of delay giving rise to multistability and hysteresis. Figure 6 shows that the analytical results are in excellent agreement with the numerical values of the order parameter. The analytical expression for  $R$  shown in Eq. 15 holds when all the oscillators are frequency synchronized and oscillating with the synchronization frequency  $\Omega$ .

**Unbalanced coupling.** We have investigated the unbalanced coupling scenarios following two routes. First, we have kept  $\kappa_1$  fixed at the value 3 and varied  $\kappa_2$  between  $[0.5, 5]$  (Fig. 7a).  $\kappa_2 < \kappa_1$ , is equivalent to the situation where ambiguity is introduced in visual stream because of which the coupling changes, e.g. in the Nath and Beauchamp study<sup>15</sup>.  $\kappa_2 > \kappa_1$  is the situation where visual coupling is increased, auditory being unreliable. In both cases existence of a synchronization state and non-synchronous state is possible as long as  $\kappa_i \geq 1$ . When  $\kappa_1$  is fixed the island of synchronization increases with increasing values of  $\kappa_2$  indicating, that time-delayed interactions with the visual-multimodal system is crucial for perception. For negative AV lags, e.g., when auditory precedes the visual stimulus no drastic change in critical delay value occurs when  $\kappa_1$  is unchanged.

Second, we vary  $\kappa_1$  while keeping  $\kappa_2$  constant. Here we see that the critical delay value for transition from phase-synchronous to non-synchronous states shifts such that the regime of synchronization increases for negative AV lags corresponding to the scenario where auditory stimulus precedes the visual stimulus (Fig. 7b). Multistable states are possible and the possibility of existence of two states, one synchronous and one non-synchronous is critically dependent on  $\kappa_1$  (Fig. 7b).

Overall we find time delays and coupling strength facilitate the existence of synchronous states and the intrinsic frequency of oscillations reduces the temporal window over which AV signals can integrate. Thus, the synchronization window can extend over an entire range of AV lags. This is in qualitative agreement with the trend seen for the McGurk stimuli (Fig. 3). Existence of multistability is demonstrated in Fig. 6 explicitly but also observed for the coupling scenarios discussed in Fig. 5 and Fig. 7. Small amounts of perturbations from numerical instabilities was sufficient to generate transitions between the stable states of the order parameter.

## Discussion

The main purpose of this work is to conceptualize the observations from an experimental paradigm, that has been over the years a bedrock to study multisensory information processing, with a simple dynamical model to illustrate the role of environmental variables and connectivity topologies between neural subsystems in shaping of perceptual states. The Kuramoto oscillator framework provides the added advantage of linking the perceptual states to simultaneously observed neuronal oscillations from EEG/ MEG data. The perceptual dynamics observed for /ta response by Munhall and colleagues<sup>16</sup> and our behavioral recordings (Fig. 3) are in complete agreement with the order parameter dynamics observed from our model (Figs 5,6 and 7). We see that the illusory perception is reported maximally within a range of lags [−150, 300] ms. Beyond this range, the auditory /pa response dominates, for both positive and negative AV lags. We have considered a system of three phase oscillators, each representing the auditory, visual and multisensory systems. Superior colliculus (SC) and posterior superior temporal sulcus (pSTS) are two structures at the sub-cortical and cortical level respectively that can get representation in the multisensory system. SC is known to have bimodal neurons that can receive inputs from both the auditory and visual systems<sup>55</sup>. Similarly, pSTS has been established by several researchers as the locus of cortical processing of multisensory integration<sup>15,56–59</sup>. Thus a network comprising of auditory, visual areas and pSTS (with contributions from SC) can form the most elemental network for multisensory processing. Experimental validation about the presence of such networks have been provided by earlier studies such as Nath and Beauchamp<sup>15</sup> using a fMRI paradigm. An important parameter in our model is the coupling between AV - A and AV - V oscillators. Careful design of such coupling parameters motivated from experimental studies can capture how the biophysically realistic symmetry that exists in the underlying neural system can influence behavior<sup>48</sup>. Broadly, the multisensory functional unit can represent qualitatively the composite contributions of all structures such as superior colliculus, pSTS and other areas. Detailed modelling of each region along with the coupling observed in experimental fMRI data is possible, but the results from such descriptions are hard to generalize and fail to capture the overarching symmetry. Due to similar reasons we chose to ignore the possible coupling between A and V resulting from direct anatomical connections between A1 and V1 areas<sup>54,60,61</sup> in our model. In our model the signals from A and V are conveyed to the multisensory oscillator (AV) where information from the two modalities are integrated. The feedback connections from the pSTS to the unimodal areas are directly considered in the form of a bidirectional coupling between pSTS and auditory and visual areas. Nevertheless, the weighted contributions of direct coupling between A1-V1 and their interplay with the coupling between multisensory to unisensory areas can be studied in the dynamical framework. In future, a corresponding behavioral paradigm needs to be developed to test the predictions of such a model.

The coupling parameter in our model helps in two ways, first in the explanation of the existence and destabilization of perceptual states upon variation of time delays and second to provide a link between observed perceptual dynamics and neuronal oscillations. It is well-known that environmental demands modulate the effective connectivity dynamics among individual sensory systems and multisensory system<sup>15</sup>. The information from the more reliable modality is given a stronger weight<sup>62</sup>. For the McGurk paradigm, Nath and colleagues<sup>15</sup> found that the STS was connected more strongly to a sensory cortex when the corresponding sensory modality was reliable (less noisy). In our model,  $\kappa_1 > \kappa_2$  may correspond to a scenario of increased functional connectivity between the STS and the auditory cortex when the auditory modality is more reliable and  $\kappa_2 > \kappa_1$  can be realized as increased functional connectivity between the STS and visual cortex when the visual modality is more reliable. Since, there are both top-down and bottom-up connections throughout the cortical processing hierarchy<sup>63,64</sup>, we have incorporated bidirectional connections into our dynamical model. Importantly, for unbalanced coupling scenarios also the synchronized states exist indicating illusory experience is possible in presence of noisy stimuli. Electrophysiological studies indicate that a pre-stimulus theta band activity primes the network for multisensory information processing towards illusory perceptual categorization<sup>17</sup>. Our results indicate how the intrinsic frequencies of oscillators set at theta regime can support the perceptual dynamics as observed by varying the AV lag in the multisensory stimulus (compare Fig. 5 to Fig. 3). We predict that tuning of oscillators at higher frequency band may shorten the window of temporal integration. In other words narrower window of temporal integration will require phase synchrony at higher frequency such as beta and gamma. Interestingly our results suggest that coupling strength has to be relatively high if temporal integration is facilitated by beta band synchrony to produce the cross-modal percepts for a wider value of AV lags (Fig. 5b). This is in-line with the proposition of Luo and Poeppel<sup>41</sup>, who proposed a low and high frequency segregation of auditory processing. From our study we can predict that low frequency processing (e.g. syllabic speech) may be carried out by relatively weakly coupled unisensory and multisensory systems whereas high frequency processing e.g., diphonic speech/ tones require more stronger connectivity strengths among sensory systems. In summary, the minimal model of multisensory integration is geared towards linking the symmetries in neuronal connectivity and dynamics with environmental demands and behavior.

The existing models of multisensory integration are mostly based on Bayesian framework and neural networks. These studies consider different kinds of interactions among multisensory and unisensory areas. One category of models considers that the unisensory stimuli are processed separately in the primary cortices, without a significant cross-modal interaction, and multisensory integration take place in higher associative cortical areas such as SC via feedforward convergence from multiple unimodal areas<sup>65</sup>. Second category of models assumes only direct lateral connections between the two unimodal areas<sup>29,30</sup> and excludes the involvement of multisensory regions. They argue that the direct connections among early processing areas (modality-specific areas such as visual and auditory) play a pivotal role in multisensory integration. These models are based on recent anatomical tracing studies in monkeys (macaque) and human subjects that have shown direct connections between auditory and visual areas including primary cortices (V1 and A1)<sup>54,60</sup>. Similar studies have also reported projections to/from somatosensory cortex from/to auditory and visual areas<sup>66</sup>. The heteromodal connections between the primary somatosensory cortex and the primary auditory cortex were also reported in the gerbils<sup>67</sup>. In marmosets, projections from the retroinsular area of the somatosensory cortex to the caudomedial belt auditory area were also observed<sup>63</sup> in line with a similar observation in the Old World monkey<sup>68</sup>. Third category of models excludes direct connections between the unimodal areas but considers both a feedforward connection to a multisensory area, and a feedback from the multisensory area to the unisensory ones<sup>28</sup>. Finally, the fourth category of models incorporates all previous connections together, i.e., the feedback and feedforward connections between unimodal and multimodal systems along with the direct connections between the unimodal systems<sup>11,69</sup>. Most of these studies do not consider the processing of temporal relationships among the unimodal stimulus components, e.g., AV lags<sup>16</sup>, gap duration in speech perceptual categorization<sup>33</sup>. Our dynamical model (which comes under the third category) comprising of three Kuramoto oscillators attempts to do so in a minimalist manner. An earlier attempt in this direction was made by Dhamala and colleagues<sup>23</sup> who modelled the rhythmic multisensory paradigm in the audio-visual domain by considering the phase dynamics of two interacting periodic oscillators and investigated the behavioral effects of relative timings of different sensory signals. In their rhythmic paradigm, temporally congruent multisensory stimuli were expected to cause a percept of synchrony. On the other hand, incongruent stimuli could cause a percept of asynchrony or another possible state, the non-phase locked state (drift or neutral percept) because it represented failure in multisensory integration. The non-phase locked state was qualitatively different from the percept of asynchrony. Dhamala and colleagues then used fMRI and simultaneous behavioral recordings to confirm the involvement of a distributed brain network for the multisensory processing of periodic auditory-visual stimuli. The study proposed that the solution space of a hypothetical system of two coupled oscillators corresponded to the perceptual solution space of human multisensory integration and to selected activation of brain regions. Though the authors propose that their results indicate definite involvement of superior colliculus in the perception of synchrony, they have not captured the dynamics of multisensory system as a functional unit directly in the theoretical model, rather emphasizing on the periodicity as the source of categorization. Nonetheless, the rhythmic paradigm is somewhat unrealistic in the context of multisensory speech stimuli. Hence, a broader dynamical framework for multisensory perception is warranted. A key difference between our model and that of Dhamala *et al.*<sup>23</sup> is that they have not incorporated the time lag or asynchrony between the audio and visual stimuli explicitly in their theoretical model. In our model we capture the AV lag by introducing a time delay  $\tau$  between the multisensory (AV) and unisensory (V/A) oscillators and study the variation in perceptual stability as a function of this parameter. Subsequently, we propose that the order parameter or the collective variable  $R$  of the oscillator system can be representative of overall perceptual categorization.

Multistability is a key aspect of perceptual behavior and in general biological systems<sup>70</sup>. In the McGurk paradigm, presence of responses registered as /ka (the visual stimuli) or fused percepts such as /pa-ka in addition to illusory /ta or auditory /pa can be conceptualized within the framework of multistability. We propose that the different responses registered as /pa, /ta and “others” in our behavioral data can be related to the multistability in the final dynamical state of our coupled oscillator system. The advantage of a dynamical systems approach is that the presence of multistability can be very elegantly explained<sup>71</sup>. Detailed analysis of multistability can be extended to psychophysics of speech perception studies<sup>33</sup> in future as well as with other paradigms. As we have shown in the Results section, the presence of time delay makes our dynamical system highly multi-stable. For different choices of initial conditions, the system can go to different stable synchronization states. We propose that the multistability in the final state of the dynamical system can be related to the variation observed in the responses of the participants to same stimuli.

The significant achievement of our model is that it captures the key features of complex multisensory integration processes at the level of behavior and links the structural and functional constraints in the underlying neural systems to ongoing behavioral states. Our results show that our model successfully simulates the temporal constraints on the McGurk effect and the variation in the responses to the same stimuli. Any multisensory behavior can be broken into combinations of individual unisensory and multisensory systems. One can then map the constraints posed by the environmental variables onto the parameter space of a dynamical system and study the emerging perceptual states. For example, recent studies question the idea of multisensory awareness stemming from a mere integration of unisensory systems<sup>13,72</sup>. This is a challenge to the traditional way of looking at a multisensory task from the perspective of serial or parallel unisensory processes. Dynamical systems offer an attractive approach to test the environmental constraints on a hypothesized multisensory pathway using modelling, behavior and brain mapping tools. Here, the integration is not identified as a separate process but rather an outcome of the dynamical interactions among unisensory and multisensory neuron populations. In fact the role of factors like attention can be incorporated in the dynamical model by introducing more mathematical complexity, either by coupling parameter or introducing a separate module, and presents a scope of future work. Experimentally this will require recording and analyzing eye-tracking data to monitor and parametrize attentional factors. Nonetheless, new experimental paradigms have to be developed to eventually test the predicted outcomes from modelling approaches such as ours.

## References

- Gibson, J. J. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin, (1966).
- Calvert, G. A., Spence, C. & Stein, B. E. *The Handbook of Multisensory Processes*. Cambridge: MA: MIT Press, (2004).
- van Wassenhove, V., Grant, K. W. & Poeppel, D. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. USA* **102**, 1181–1186 (2005).
- Hershenson, M. Reaction time as a measure of intersensory facilitation. *J. Exp. Psychol.* **63**, 289–293 (1962).
- Stein, B. E., Meredith, M. A., Huneycutt, W. S. & McDade, L. Behavioural indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J. Cogn. Neurosci.* **1**, 12–24 (1989).
- Sumbly, W. & Pollack, I. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* **26**, 212–215 (1954).
- MacLeod, A. & Summerfield, Q. Quantifying the contribution of vision to speech perception in noise. *Br. J. Audiol.* **21**, 131–141 (1987).
- Calvert, G. A., Campbell, R. & Brammer, M. J. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* **10**, 649–657 (2000).
- Callan, D. E. *et al.* Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport* **14**, 2213–2218 (2003).
- Stein, B. E., Jiang, W., Wallace, M. T. & Stanford, T. R. Nonvisual influences on visual-information processing in the superior colliculus. *Prog. Brain Res.* **134**, 143–156 (2001).
- Guipponi, O., Cléry, J., Odouard, S., Wardak, C. & Ben Hamed, S. Whole brain mapping of visual and tactile convergence in the macaque monkey. *NeuroImage* **117**, 93–102 (2015).
- Schulz, M., Ross, B. & Pantev, C. Evidence for training-induced crossmodal reorganization of cortical functions in trumpet players. *SchulzReport* **14**, 157–161 (2003).
- Deroy, O., Chen, Y. & Spence, C. Multisensory constraints on awareness. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130207 (2014).
- McGurk, H. & Macdonald, J. Hearing lips and seeing voices. *Nature* **264**, 746–748 (1976).
- Nath, A. R. & Beauchamp, M. S. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J. Neurosci.* **31**, 1704–1714 (2011).
- Munhall, K., Gribble, P., Sacco, L. & Ward, M. Temporal constraints on the McGurk effect. *Atten., Percept., and Psychophys.* **58**, 351–362 (1996).
- Keil, J., Müller, N., Ihssen, N. & Weisz, N. On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. *Cereb. Cortex* **22**, 221–231 (2012).
- de Gelder, B. & Bertelson, P. Multisensory integration, perception and ecological validity. *Trends Cogn. Sci.* **7**, 460–467 (2003).
- Macaluso, E., Frith, C. D. & Driver, J. Modulation of human visual cortex by crossmodal spatial attention. *Science* **289**, 1206–1208 (2000).
- Kopinska, A. & Harris, L. R. Simultaneity constancy. *Perception* **33**, 1049–1060 (2004).
- Massaro, D. W. & Cohen, M. M. Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *J. Acoust. Soc. Am.* **13**, 127–134 (1993).
- Massaro, D. W. & Cohen, M. M. Perception of asynchronous and conflicting visual and auditory speech. *Speech Commun.* **100**, 1777–1786 (1996).
- Dhamala, M., Assisi, C. G., Jirsa, V. K., Steinberg, F. L. & Kelso, J. A. Multisensory integration for timing engages different brain networks. *NeuroImage* **34**, 764–773 (2007).
- Körding, K. P. *et al.* Causal inference in multisensory perception. *PLoS One* **2**, e943 (2007).
- Macaluso, E. Spatial Constraints in Multisensory Attention. In: Murray, M. M., Wallace, M. T., (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton (FL): CRC Press/Taylor & Francis; (2012). Chapter 25.
- Olasagasti, I., Bouton, S. & Giraud, A. Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex* **68**, 61–75 (2015).
- Cuppini, C., Magosso, E., Rowland, B., Stein, B. & Ursino, M. Hebbian mechanisms help explain development of multisensory integration in the superior colliculus: a neural network model. *Biol. Cybern.* **106**, 691–713 (2012).
- Magosso, E., Cuppini, C., Serino, A., Pellegrino, G. D. & Ursino, M. A theoretical study of multisensory integration in the superior colliculus by a neural network model. *Neural Networks* **21**, 817–829 (2008).
- Magosso, E., Cuppini, C. & Ursino, M. A Neural Network Model of Ventriiloquism Effect and Aftereffect. *PLoS One* **7**, e42503 (2012).
- Cuppini, C., Magosso, E., Bolognini, N., Vallar, G. & Ursino, M. A neurocomputational analysis of the sound-induced flash illusion. *NeuroImage* **92**, 248–266 (2014).
- Meredith, M. A. & Stein, B. E. Interactions among converging sensory inputs in the superior colliculus. *Science* **221**, 389–391 (1983).
- Ursino, M., Cuppini, C. & Magosso, E. Neurocomputational approaches to modelling multisensory integration in the brain: A review. *Neural Networks* **60**, 141–165 (2014).
- Tuller, B., Case, P., Ding, M. & Kelso, J. A. S. The non-linear dynamics of speech categorization. *J of Exp psychol: Human percept. & performance* **60**(1) 3–16 (1994).
- Senkowski, D., Schneider, T. R., Foxe, J. J. & Engel, A. E. Crossmodal binding through neural coherence: Implications for multisensory processing. *Trends Neurosci.* **31**, 401–409 (2008).
- Hipp, J. F., Engel, A., K. & Siegel, M. Oscillatory synchronization in largescale cortical networks predicts perception. *Neuron* **69**(2), 387–396 (2011).
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S. & Puce, A. Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* **12**, 106–113 (2008).
- Giraud, A. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517 (2012).
- von Stein, A., Rappelsberge, r. P., Sarnthein, J. & Petsche, H. Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cereb. Cortex* **9**, 137–150 (1999).
- Sakowitz, O. W., Quiroga R., Schurmann, M. & Basar, E. Spatiotemporal frequency characteristics of intersensory components in audiovisually evoked potentials. *Brain Res Cogn Brain Res* **23**, 316–326 (2005).
- Doesburg, S. M., Emberson, L. L., Rahi, A., Cameron, D. & Ward, L. M. Asynchrony from synchrony: long-range gamma-band neural synchrony accompanies perception of audiovisual speech asynchrony. *Exp. Brain Res.* **185**, 11–20 (2008).
- Luo, H. & Poeppel, D. Cortical oscillations in auditory Perception and speech: Evidence for two temporal windows in human auditory cortex. *Front Psychol* **3**, 170 (2012).
- Park, H., Kayser, C., Thut, G. & Gross, J. Cortical lip movements entrain the observer's low-frequency brain oscillations to facilitate speech intelligibility *eLife* **5**, e14521 (2016).
- Ganesh, A. C., Berthommier, F., Vilain, C., Sato, M. & Schwartz, J. L. A possible neurophysiological correlate of audiovisual binding and unbinding in speech perception *Front. Psychol.* **5**, 1340 (2014).
- Kuramoto Y. *Chemical Oscillations, Waves, and Turbulence*. New York, NY: Springer-Verlag, (1984).
- Breakspear, M., Heitmann, S. & Daffertshofer, A. Generative Models of Cortical Oscillations: Neurobiological Implications of the Kuramoto model. *Front Hum Neurosci* **4**, 1–14 (2010).

46. Cabral, J., Hugues, E., Sporns, O. & Deco, G. Role of local network oscillations in resting-state functional connectivity. *NeuroImage* **4**, 1–11 (2011).
47. Schmidt, H., Petkov, G., Richardson, M. P. & Terry, J. R. Dynamics on networks: the role of local dynamics and global networks on the emergence of hypersynchronous neural activity. *PLoS Comput Biol.* **10**, e1003947 (2014).
48. Banerjee, A. & Jirsa, V. K. How do neural connectivity and time delays influence bimanual coordination? *Biol. Cybern.* **96**, 265–278 (2007).
49. Cappe, C., Thut, G., Romei, V. & Murray, M. M. Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.* **30**, 12572–12580 (2010).
50. Jirsa, V. K. & Ding, M. Will a large complex system with time delays be stable? *Phys. Rev. Lett.* **93**, 070602 (2004).
51. Meredith, M. A. & Stein, B. E. Descending efferents from the superior colliculus relay integrated multisensory information. *Science* **227**, 657–659 (1985).
52. Keil, J., Müller, N., Hartmann, T. & Weisz, N. Prestimulus beta power and phase synchrony influence the sound-induced flash illusion. *Cereb. Cortex* **24**, 1278–1288 (2014).
53. Acebrón, J. A., Bonilla, L. L., Pérez Vicente, C. J., Ritort, F. & Spigler, R. The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* **77**, 137–185 (2005).
54. Falchier, A., Clavagnier, S., Barone, P. & Kennedy, H. Anatomical Evidence of Multimodal Integration in Primate Striate Cortex. *The Journal of Neuroscience* **22**, 5749–5759 (2002).
55. Peck, C. K. Visual-auditory integration in cat superior colliculus: implications for neuronal control of the orienting response. *Prog. Brain Res.* **112**, 167–177 (1996).
56. Powers, A. R., Hevey, M. A. & Wallace, M. T. Neural correlates of multisensory perceptual learning. *J. Neurosci.* **32**, 6263–6274 (2012).
57. Kassuba, T., Menz, M. M., Röder, B. & Siebner, H. R. Multisensory interactions between auditory and haptic object recognition. *Cereb. Cortex.* **23**, 1097–1107 (2013).
58. Okada, K., Venezia, J. H., Matchin, W., Saberi, K. & Hickok, G. An fMRI Study of Audiovisual Speech Perception Reveals Multisensory Interactions in Auditory Cortex. *PLoS One* **8**, e68959 (2013).
59. Watson, R. *et al.* Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. *J. Neurosci.* **34**, 6813–6821 (2014).
60. Rockland, K. S. & Ojima, H. Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology* **50**, 19–26 (2003).
61. Wallace, M. T., Ramachandran, R. & Stein, B. E. A revised view of sensory cortical parcellation. *PNAS* **101**, 2167–2172 (2004).
62. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
63. de la Mothe, L. A., Blumell, S., Kajikawa, Y. & Hackett, T. A. Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J. Comp. Neurol.* **496**, 27–71 (2006).
64. van Atteveldt, N., Roebroek, A. & Goebel, R. Interaction of speech and script in human auditory cortex: insights from neuroimaging and effective connectivity. *Hear Res.* **258**, 152–164 (2009).
65. Rowland, B. A., Stanford, T. R. & Stein, B. E. A model of the neural mechanisms underlying multisensory integration in the superior colliculus. *Perception* **36**, 1431–1443 (2007).
66. Cappe, C. & Barone, P. Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience* **22**, 2886–2902 (2005).
67. Budinger, E., Heil, P., Hess, A. & Scheich, H. Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience* **143**, 1065–1083 (2006).
68. Smiley, J. F. *et al.* Multisensory convergence in auditory cortex, I. cortical connections of the caudal superior temporal plane in macaque monkeys. *The Journal of Comparative neurology* **502**, 894–923 (2007).
69. Hoshino, O. Neuronal responses below firing threshold for subthreshold cross-modal enhancement. *Neural Computation* **23**, 958–983 (2011).
70. Pastukhov, A. *et al.* Multi-stable perception balances stability and sensitivity. *Front. Comput. Neurosci.* **7**, 17 (2013).
71. Freyer, F., Roberts, J. A., Ritter, P. & Breakspear, M. A canonical model of multistability and scale-invariance in biological systems. *PLoS Comput. Biol.* **8**, e1002634 (2012).
72. Lakatos, P., Chen, C., O’Connell, M. N., Mills, A. & Schroeder, C. E. Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron* **53**, 279–292 (2007).

## Acknowledgements

The study was supported by IPR and NBRC Core support staff and infrastructure. Ramalingaswami fellowship, DBT (BT/RLF/Re-entry/31/2011) and Innovative Young Bio-technologist Award (IYBA), DBT (BT/07/IYBA/2013) from the Department of Biotechnology (DBT), Ministry of Science and Technology Government of India to Arpan Banerjee (AB). AB also acknowledges the support of Centre of Excellence in Epilepsy and MEG (BT/01/COE/09/08/2011) from DBT and Dr. Dipanjan Roy (IIIT-Hyderabad) for helpful comments on the manuscript.

## Author Contributions

B.T., A.S. and A.B. conceived the study; A.M. and A.B. collected and analysed the behavioral data; B.T., A.S. and A.B. did the modelling work; B.T., A.M., A.S. and A.B. prepared the figures; B.T. and A.B. wrote the manuscript; B.T., A.M., A.S. and A.B. read and commented on the manuscript.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Thakur, B. *et al.* A dynamical framework to relate perceptual variability with multisensory information processing. *Sci. Rep.* **6**, 31280; doi: 10.1038/srep31280 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016