2014

Bilateral gain control; an "innate predisposition" for all sorts of things.

Wilkinson, N

http://hdl.handle.net/10026.1/9538

Bilateral gain control; an "innate predisposition" for all sorts of things

Nicholas Wilkinson[1]* and Giorgio Metta[1,2]

[1] iCub Facility, Istituto Italiano di Tecnologia, Genova, Italy
[2] Centre for Robotics and Neural Systems, University of Plymouth, Plymouth, UK

Empirical studies have revealed remarkable perceptual organization in neonates. Newborn behavioral distinctions have often been interpreted as implying functionally specific modular adaptations, and are widely cited as evidence supporting the nativist agenda. In this theoretical paper, we approach newborn perception and attention from an embodied, developmental perspective. At the mechanistic level, we argue that a generative mechanism based on mutual gain control between bilaterally corresponding points may underly a number of functionally defined "innate predispositions" related to spatial-configural perception. At the computational level, bilateral gain control implements beamforming, which enables spatial-configural tuning at the front end sampling stage. At the psychophysical level, we predict that selective attention in newborns will favor contrast energy which projects to bilaterally corresponding points on the neonate subject's sensor array. The current work extends and generalizes previous work to formalize the *bilateral correlation model* of newborn attention at a high level, and demonstrate in minimal agent-based simulations how bilateral gain control can enable a simple, robust and "social" attentional bias.

Keywords: bilateral symmetry, gain control, innate predisposition, embodiment, beamforming, attention, "like me", human–robot interaction

1. INTRODUCTION

> The empirical position is, to be sure, in agreement with the nativistic on a number of points—for example, that local signs of adjacent places on the retina are more similar than those farther apart and that the corresponding points on the two retina are more similar than those that do not correspond. Helmholtz (1968)

Psychophysical studies have demonstrated quite sophisticated spatial-configural perception in newborn babies. See for example Johnson et al. (1991); Slater and Kirby (1998); Farroni et al. (2000); and Streri et al. (2013a). Such findings have often been interpreted in terms of functionally defined "innate predispositions" (Morton and Johnson, 1991; Spelke, 1998; Streri et al., 2013b), which are seen as providing the "biological basis" of perceptual and social development (Johnson and Morton, 1991; Johnson, 2003). Other researchers have suggested that conspecifics are identified as "like me" via perceptuomotor resonance with internal representations of the self, developed prenatally through self-exploration and proprioception (Sugita, 2009; Pitti et al., 2013). Some have even proposed that an innate precursor of the mirror neuron system underlies newborn sociality (Meltzoff and Decety, 2003; Lepage and Théoret, 2007). Here we extend and generalize previous work on newborn face preference (Wilkinson et al., in press), to outline a high level formalization of a novel perspective in which *embodied sampling biases* provide "innate information" about space, configuration, and "like-me-ness."

Specifically, the sampling bias under focus in the current paper is *bilateral* sensor distribution and integration. Many authors have noted the important role of multimodal correlation in social interaction and perceptual learning (e.g., Bahrick et al., 2004; Sai, 2005). Less attention has been paid to the implications of intramodal bilateral correlations from the social and behavioral perspective, though bilateral sensory interaction has been intensively studied in its own right (see "Bilateral mutual gain control" below). Here we hypothesize a functional relationship between bilateral interaction and spatial-configural perception in newborns. A simplified formal model explains how bilateral mutual gain control can enable spatial-configural perceptual distinctions analogous to some of those observed in human newborns. Agent-based simulations address minimal analogs of visual perception of "size constancy" (Granrud, 1987; Slater et al., 1990), visual and audio-visual face perception (Goren et al., 1975; Johnson et al., 1991; Sai, 2005), and the dependency of face perception and social learning on infant directed alignment, or "direct gaze" (Farroni et al., 2002; Guellai and Streri, 2011).

In section 2, we present a case for bilateral mutual gain control as a general aspect of intersensory integration, based on the existing literature, and recruit the array signal processing formalism of beamforming (Naidu, 2001) as a way of describing spatial-configural attention and orienting. In section 3, we define a minimal formal model to describe how bilateral gain control can generate an overt attentional preference for the "like me." In section 4, we report results of simulations based on this model.

Finally, we discuss the relevance and scope of our theoretical findings, and offer some concluding remarks.

2. THE NEWBORN AS A MULTIMODAL, BILATERAL SENSOR ARRAY

Gibson (1966) argued that the sensing body should be characterized as a multimodal sensor array. In any sensing process, the front line is the sampling regime adopted; neural/computational processing can only process that which has been sampled. It is therefore impossible to meaningfully characterize sensory input to the nervous system during active behavior without considering the physical embodiment of the sensor array (Towal et al., 2011). The vertebrate sensor array is structured in a more or less bilaterally symmetric manner. A growing consensus points to mutual gain control as a fundamental feature of bilateral interactions (e.g., Li and Ebner, 2006; Ding et al., 2013; Schmidt, 2013; Wunderle et al., 2013; Xiong et al., 2013). Mutual gain control is one mode of enacting *beamforming* (e.g., Westermann, 2003; Ma, 2013), a standard technique for selective tuning in sensor array technology. In this section, we unpack this analogy between selective attention in newborn infants and selective tuning in sensor array theory. First, we review the neurobiological literature on bilateral mutual gain control and the relationship of gain control to attention. This sketches a plausible neural substrate for the current model, which we term the "bilateral correlation model" ("BCM") of newborn attention. We then briefly introduce beamforming and explain how it is related to bilateral gain control and attention.

2.1. BILATERAL MUTUAL GAIN CONTROL AND SENSORY ATTENTIONAL GATING

2.1.1. Bilateral symmetry structures vertebrate physiology

The bilateral structure of the brain and body is aligned and integrated according to symmetric correspondence at many stages of sensory and motor processing, an architecture perhaps most clearly demonstrated by the corpus callosum (Iwamura, 2000; Li and Ebner, 2006), which links corresponding bilateral points in the brain. Tactile stimulation of one hand causes bilateral cortical activation at corresponding somatotopic points (Hansson and Brismar, 1999). Binocular cross-correlation is widely thought to underly stereo vision (Banks, 2007; Filippini and Banks, 2009). Binaural cross-correlation informs orienting and looking behavior in neonates (Mendelson et al., 1976; Jiang and Tierney, 1996; Furst et al., 2004), and the foetus is capable of auditory orienting *in utero* (Voegtline et al., 2013). From an aesthetic perspective, the "sweet spot" region of binaural synchrony is manipulated by sound engineers to deliver the most enjoyable and engaging listening experience (Theile, 2000; Bauck, 2003), suggestive of a more general multimodal link between bilateral correlation, arousal and "liking."

2.1.2. Neural gain control implements selective attention

Neural *gain control* can implement multiplicative interactions *in vivo* (Rothman et al., 2009). Gain control acts like an amplifier, or "gate" for input signals. Gain control is widely thought to mediate selective attention (Hillyard et al., 1998; Salinas and Sejnowski, 2001; Aston-Jones and Cohen, 2005; Reynolds and Heeger, 2009; Feldman and Friston, 2010; Katzner et al., 2011; Sara and Bouret, 2012), and has been mechanistically linked to ascending projections from neuromodulatory hubs and the sympathetic nervous system (Aston-Jones and Cohen, 2005; Voisin et al., 2005; Sara, 2009; Fuller et al., 2011). Presynaptic sychrony can also modulate postsynaptic gain in a feedforward fashion (Huguenard and McCormick, 2007; Womelsdorf and Fries, 2007; Gotts et al., 2012). Gain control has been formally equated with the modulation of Bayesian *precision* in probabilistic generative modeling (Feldman and Friston, 2010; Moran et al., 2013).

2.1.3. Bilateral mutual gain control

There is extensive evidence that mutual gain control is an important aspect of binocular (Ding and Sperling, 2006; Meese and Baker, 2011; Ding et al., 2013), binaural (Kashino and Nishida, 1998; Ingham and McAlpine, 2005; Steinberg et al., 2013; Xiong et al., 2013), and bitactile (Hansson and Brismar, 1999; Li and Ebner, 2006) interactions in various species. Interhemispheric interactions via corpus callosum are well described by a gain control relation (Li and Ebner, 2006; Schmidt, 2013; Wunderle et al., 2013). Thus mutual gain control is the most plausible general framework for bilateral sensory interaction, though many particulars are likely to exist at a more detailed level. Given this organization, a behavioral preference for stimuli which induce strong correlations between corresponding points is practically inevitable.

2.1.4. Bilateral gain control in the neonate

Studies examining prestereoptic binocular vision in human infants have produced conflicting results, and have not included neonatal subjects (Shimojo et al., 1986; Brown and Miracle, 2003; Kavšek et al., 2013). The maturity of binocular gain control circuitry in human newborns is therefore unknown. In the rhesus macaque, considered a good model for the human visual system, binocular circuitry is quite mature in neonates (Rakic, 1976; Horton and Hocking, 1996), and responses are limited by low monocular sensitivity rather than binocular immaturity (Chino et al., 1997). Binaural integration is also functioning, if not entirely mature, in neonates (Furst et al., 2004; Litovsky, 2012). The BCM assumes that newborn bilateral integration is a qualitatively similar, if immature, version of that in adults, but does not prescribe the precise transform; many variations on mutual gain control are possible. In adults, bilateral interactions are modulated by mono and stereo normalization (Carandini et al., 1997; Moradi and Heeger, 2009; Carandini and Heeger, 2011), but the nature of normalization in the newborn human is, as far as we know, unknown.

2.2. THE THALAMUS—AN ARCHITECTURAL HUB FOR ATTENTIONAL GATING AND BILATERAL ALIGNMENT

2.2.1. Bilateral alignment in the thalamus

The thalamus embodies architectural alignment of the signals from bilaterally corresponding sensors (Jones, 1998). For example, the *lateral geniculate nucleus* (henceforth "LGN"), is comprised of six layers. Connections from contralateral nasal retina project to layers 1, 4, and 6, while ipsilateral temporal retina projects to layers 2, 3, and 5. All of these layers are precisely

in-reference with respect to retinal receptive fields. This architecture forms a structural basis for precise binocular receptive fields in efferents such as primary visual cortex. Bilateral auditory (Wrege and Starr, 1981; Fitzpatrick et al., 1997; Ingham and McAlpine, 2005; Krumbholz et al., 2005) and tactile (Mountcastle and Henneman, 1949; Emmers, 1965; Davis et al., 1998; Coghill et al., 1999) pathways also converge in the thalamus and earlier.

2.2.2. Sensory gain control and attention in thalamus

An active role for the thalamus in attention has long been theorized (Clark, 1932; Crick, 1984), and evidence supporting these suggestions is accumulating (Varela and Singer, 1987; Sillito et al., 1994; Sherman and Guillery, 2006; Sherman, 2007; Saalmann and Kastner, 2011; Ward, 2013). The thalamus is strongly associated with gain control (Saalmann and Kastner, 2009). LGN receives only about 10% of its input from retina, with the other 90% constituted in approximately equal proportions by cholinergic projections from the parabrachial nucleus of the brainstem, inhibitory control from the thalamic reticular nucleus, and feedback connectivity from layer 6 in striate cortex (Saalmann and Kastner, 2009). Recently, layer 6 in visual cortex has been shown to mediate gain control of superficial layers (Olsen et al., 2012; Vélez-Fort and Margrie, 2012). This extensive modulatory network effects gain modulation in LGN, thereby gating visual input to the cortex (Sherman, 2007; Saalmann and Kastner, 2011; Lien and Scanziani, 2013). Corticofugal feedback also modulates gain control in auditory thalamus (Grothe, 2003; He, 2003; Zhang et al., 2008).

2.3. BEAMFORMING, ORIENTING, AND MOTOR ATTENTION

Beamforming, a form of spatial filtering, is a technique for manipulating the *spatial tuning* of a sensor array (Naidu, 2001). Applications include astronomy (e.g., the SCUBA-2 project [1]), neuroimaging (Van Veen and Buckley, 1988; Siegel et al., 2008), "smart" audio technology (Kellermann et al., 2012; Sun et al., 2012), and network communication (Litva and Lo, 1996; Lakkis, 2012). The mathematical essence of beamforming is maximization of constructive interference between the signals from an array of sensors. Integrating the signals from an array of spatially distributed sensors creates a "beam," or preferred angle of arrival, for incoming signals. Mutual gain control is one possible integration function e.g., Ma (2013). When the signals from all the sensors are temporally aligned, constructive interference is maximized and the input signal is faithfully reproduced. Adding differential delays to the sensor inputs, or physically turning the array, can rotate this beam in space, so that sources at particular locations (e.g., a mobile phone) can be targeted, whilst noise from elsewhere is tuned out; a kind of technological "selective attention." See **Figure 1** for a schematic depiction.

For example, in delay-and-sum beamforming, a large set of delays is provisionally added to the signals from the array. The particular delay(s) which maximizes the constructive interference with a reference sensor (i.e., maximizes the combined signal from
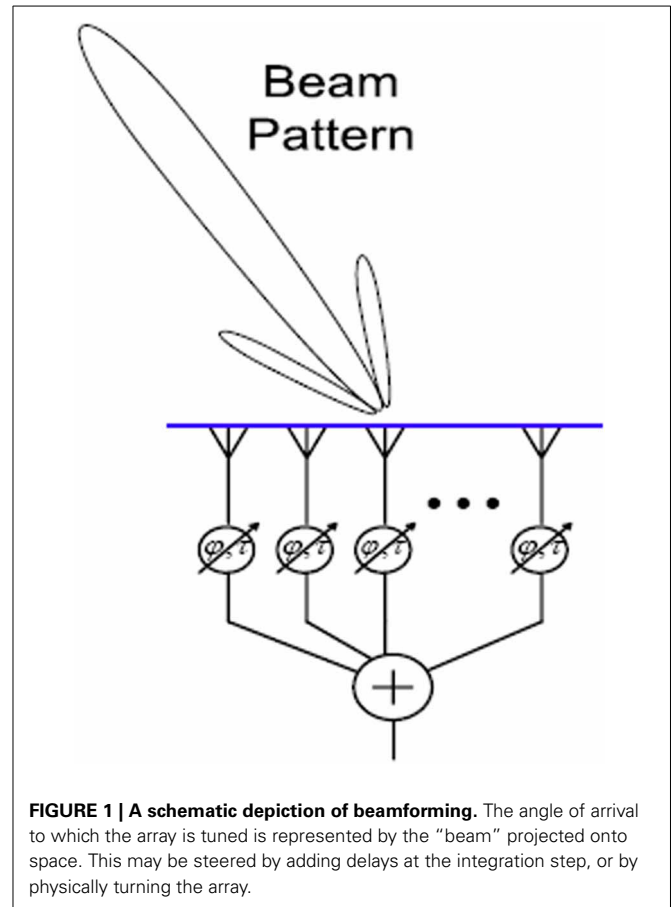
---
[1]http://www.jach.hawaii.edu/JCMT/continuum/scuba2.html

FIGURE 1 | A schematic depiction of beamforming. The angle of arrival to which the array is tuned is represented by the "beam" projected onto space. This may be steered by adding delays at the integration step, or by physically turning the array.

the whole array) is identified. This optimum delay is proportional to the angle of arrival of the signal and may be calculated as;

$$\Delta(t_i) = \frac{(i-1)d\cos\theta}{c}, i = 1, 2 \ldots n \tag{1}$$

Where d is inter-sensory distance, c is the traveling speed of the signal (e.g., the speed of sound), θ is the angle of arrival of the signal, and i represents the position of the sensor concerned relative to the reference sensor. More than two sensors can of course be used. In practive, θ is usually not known.

2.3.1. Audio Beamforming

Beamforming is mathematically similar to the widely accepted notion of sound localization based on inter-aural time difference ("ITD") (Jeffress, 1948; Fitzpatrick et al., 2000, 2002; Joris and Yin, 2007; Hartmann et al., 2013). Binaural hearing aids employ beamforming to give directional selectivity, allowing the device to focus on sound sources at the auditory midline (Greenberg and Zurek, 1992; Kompis and Dillier, 2001; Westermann, 2003; Ma, 2013). A recent study found that auditory adaptive coding mechanisms primarily target sources near the interaural midline (Maier et al., 2012), suggestive of a "special" attentional status for midline sources. Audio beamforming is fundamental to technological approaches to the "Cocktail Party Problem" (Cherry, 1953; Haykin and Chen, 2005).

2.3.2. Motor attention and beamforming

Adding the delay $\Delta(t_i)$ steers the angle of arrival to which the array is tuned, and is equivalent to physically turning the array. Physically turning the array is analogous to the psychological concept of orienting or *overt attention*. Adding delays to "virtually" orient the array is analogous to "covert attention." Covert attention and overt attention are thought to be tightly linked (Eimer et al., 2005; de Haan et al., 2008), though appear to be mediated by different cellular networks (Gregoriou et al., 2012). Physically turning the array provides the basis for the motor side of selective attention in the current model. A movement equivalent to the delay is enacted in order to physically align the array with the source. We deal only with the overt orienting case here. Thus $\Delta(t_i)$ corresponds to a motor command, rather than an internal delay. Here we reference this difference by the term "active beamforming." This may seem a little topsyturvy, as beamforming was to a large extent designed to avoid the need for slow and energy hungry mechanical orienting of the array, but we believe it conveniently expresses the underlying continuity between mechanical and "virtual" manipulation of alignment.
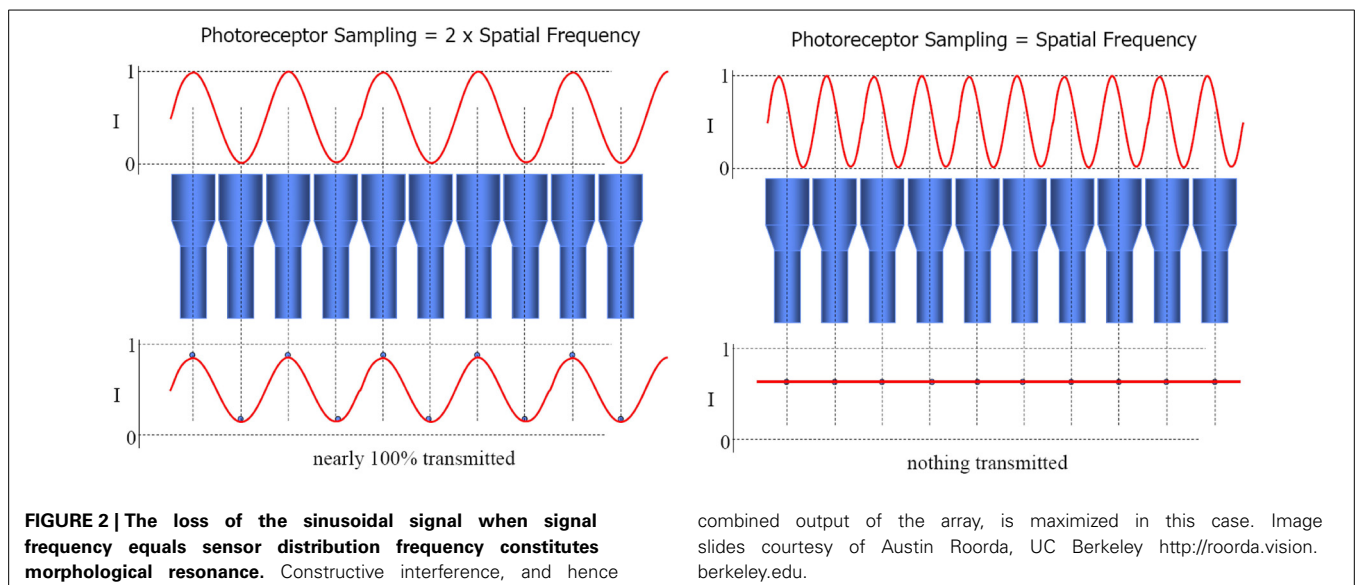
2.3.3. Visual beamforming

The speed of light renders delays virtually undetectable. Light's beamlike propagation and the design of visual optics provide direction selectivity *a priori*. Thus visual beamforming is not standard practice, and the terminology adopted here may thus be non-standard. However, the mathematical concept at the heart of beamforming—maximizing constructive interference between sensors—applies equally to the visual case in the current model. As the underlying sensory computations are the same between vision and audition, we adopt the continuous terminology "visual beamforming."

The beamforming technique provides an array tuned to two different flavors of stimulus. Firstly, *one source* that lies on the intersection of the sensors' lines of sight (as just described for audio beamforming). Secondly, *multiple sources*, one on each line of sight. This yields tuning to particular *configurations* of multiple signal sources, simply through spatial resonance between the configuration of transmitters and the configuration of receivers. In the case of sensors with parallel directional tuning, this "preferred" configuration is the *same configuration* as that of the sensor array itself. Thus it functions as a "like me" configuration detector. This is the special case of Nyquist–Shannon sampling theorem, wherein signal frequency equals sensor distribution frequency, causing maximal constructive interference between the combined signals. See **Figure 2**. Prior to the development of convergent stereopsis, newborn binocular alignment is (approximately and noisily) parallel (Thorn et al., 1994), though this depends on what the infant is looking at (Slater and Findlay, 1975).

Multiple source signals often occur in the context of problems in sensor array theory, indeed it is this sensitivity that is targeted by signal jamming techniques (Poisel, 2011). However, multiple transmitter, multiple receiver set-ups are widely used in wireless communication (termed "MIMO" multiple-input and multiple-output beamforming e.g., Raleigh and Cioffi, 1998) and may be highly relevant in the case of biological vision. Eyes are a particularly important example of a multiple visual signal source (Gliga and Csibra, 2007). Single transmitter and multiple transmitter beamforming specifies two ways in which a scene can be "like me" to a newborn infant; it can "fit" my sensor array by containing *single sources* which lie on the intersections of my bilateral lines of sight, and it can "fit" my sensor array by containing *multiple* sources which independently occupy both the lines of sight of a bilateral sensor pair. To our knowledge, this *functional relationship* between "visual beamforming," spatial-configural vision and "like-me" perception in the newborn is a novel proposal. However, the underlying mechanisms proposed are not; they are just those of standard binocular circuitry and established communications technology.

FIGURE 2 | The loss of the sinusoidal signal when signal frequency equals sensor distribution frequency constitutes morphological resonance. Constructive interference, and hence combined output of the array, is maximized in this case. Image slides courtesy of Austin Roorda, UC Berkeley http://roorda.vision.berkeley.edu.

3. A FORMAL MODEL

In this section, we define a simplified model world in which to formalize our arguments and demonstrate the resultant effects in minimal computational simulations.

3.1. A SIMPLE WORLD

Let us first define a discrete World W with two spatial dimensions P, Θ, and time T. P and Θ refer to the location (in polar coordinates) in space of arbitrary quanta each with one binary degree of freedom C. See **Figure 3**. We use lowercase letters to denote specific values of capitalized variables, and to denote specific members of capitalized sets. Specifically, W is the set of all individual discrete World points w, each of which is characterized by four state variables ρ, θ, τ, c.

$$\forall w\{w \in W\} \tag{2}$$
$$w = <\rho, \theta, \tau, c> $$

C corresponds to the presence ("1") or absence ("0") of a visual signal source. Within this World, a subset of points w have $c = 1$ (indicating a visual signal source). We denote this set $\bar{W}_C \subseteq W$;

$$\forall w\{w \in \bar{W}_C : w_c = 1\} \tag{3}$$

3.2. BUILDING AN AGENT

3.2.1. A single visual sensor

Let us introduce a binary visual sensor, which occupies one space/quanta w in the World W. It has a tight beam, perfectly
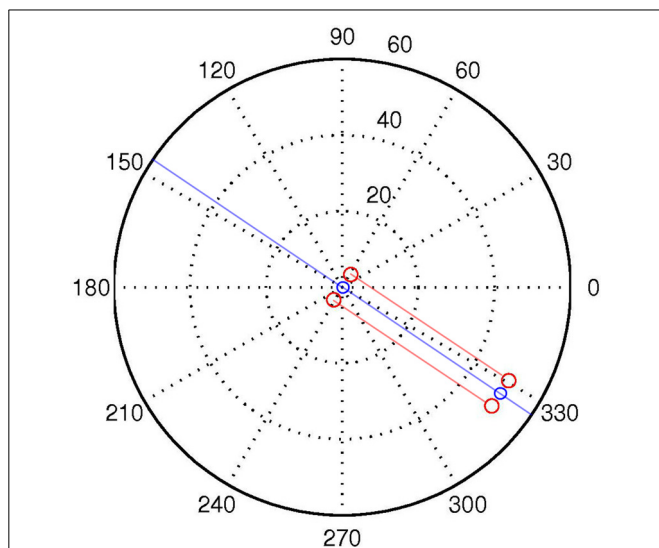

FIGURE 3 | A logpolar World, containing two "agents," one located at the origin, the other at distance of $\rho = 50$, and aligned with ("looking at") the other agent at the origin. The red dots correspond to the positions of the sensors of each agent. The semi-transparent red lines correspond to the LoS of the central visual sensors for each "eye." The semi-transparent blue line demarks the auditory midline, or line of equidistance from the two auditory sensors. This corresponds to the "beam" projected in beamforming. The blue circles represent the "mouth" of the agents.

straight line of sight. We treat the travel time of light as zero. It is convenient to place the sensor at the origin, such that for any given location in the World w that location is on the line of sight if and only if $w_\theta = \text{Sensor}_\theta$. The line of sight is then a set L containing all locations w in the World satisfying

$$\forall w\{w \in L : \theta_w = \theta_{\text{sensor}}\}. \tag{4}$$

The sensor output S_{Eye} is 1 just in case its line of sight contains at least one point w where $w_c = 1$, that is if L intersects $\bar{W}_C$, otherwise it outputs zero. This may be viewed conveniently as a logical disjunction (OR gate) with the entire line of sight as input, which we will denote $S = \vee L$. The presence of a sensor at a point w sets $w_c = 1$, i.e., the sensor is also a visual signal source.

3.2.2. A binocular agent

Let us place two visual sensors in the World, oriented in parallel and separated by an inter-sensory distance Φ. We refer to this sensor pair as "eyes" for readability. They are connected via a logical AND gate. The output of this AND operation is denoted $\dot{S}_{\text{Eyes}}$ and may be described in full as

$$\dot{S}_{\text{Eyes}} = \&(\vee L_{\text{Eye1}}, \vee L_{\text{Eye2}}). \tag{5}$$

This is a minimal analogy of mutual gain control, the AND operation being equivalent to multiplication in the binary case. It is convenient to view $\dot{S}_{\text{Eyes}}$ as a single "meta-sensor" with a double aperture and a "forked" line of sight. Let us denote this forked LoS $\dot{L}_{\text{Eyes}}$, of which L_{Eye1} and L_{Eye2} are subsets. $\dot{S}_{\text{Eyes}} = 1$ if and only if both L_{Eye1} and L_{Eye2} intersect $\bar{W}_C$.

3.2.3. Vergence

Vergence eye movement changes the distance separating the visual axes (i.e., the LoS of the eyes), here denoted $\vec{\Phi}$, over depth, according to;

$$\vec{\Phi}_\rho = \Phi - \rho \tan \theta \tag{6}$$

where ρ is depth, and θ is vergence angle in radians. Positive θ corresponds to convergence and negative to divergence. $\vec{\Phi}$ and Φ are measured in the same units as ρ. With no vergence/parallel axes (i.e., $\theta = 0$), this reduces to $\vec{\Phi} = \Phi$ at any depth. In the current model the sensors are assumed to be aligned in parallel, to keep things simple.

3.2.4. A dyad of binocular agents

In order to show how this set-up enables "like me" detection, let us introduce a second agent into the World (as in **Figure 3**). At this stage, it is useful to adopt the term "transceiver," which references a mechanism which is both a transmitter and a receiver. This terminology becomes useful in the dyadic case, as one inter-actor's sensor is the other's signal. If we place the two agents in alignment, their pairs of visual transceivers mutually satisfy one another's sensory condition $\dot{S}_{\text{Eyes}} = 1$ (**Figure 3**). However, if we translate either agent relative to the other, they no longer stimulate one another in this way. In order to see the other agent when it is not directly in front, a wider visual field of view is required.

3.2.5. Wide field of vision

Let us define an agent with numerous copies of the sensor arrangement just described, analogous to a "pin-hole camera." Each agent then has two wide angle sensors ("eyes"), while each "eye" consists of many individual light sensors ("rods") with tight beam lines of sight; a *nested* sensor array. Formally, this may be represented simply by expressing the above description in terms of pointwise vector or matrix operations. The calculations are just the same, but repeated in parallel across the visual field. With this wider field of vision ("FoV"), an agent may see things not only directly in front, but also off to the side. The *spatial* offset from center may then be used to guide orienting. Thus we define each eye as a vector of individual sensors;

$$\vec{S}_{\text{Eye1}} = [S_{\text{Eye1},1}, S_{\text{Eye1},2}, S_{\text{Eye1},i}, \ldots S_{\text{Eye2},n}] \qquad (7)$$

$$\vec{S}_{\text{Eye2}} = [S_{\text{Eye2},1}, S_{\text{Eye2},2}, S_{\text{Eye2},i} \ldots S_{\text{Eye2},n}]$$

and corresponding points ($\Leftrightarrow$) according to;

$$S_{\text{Eye1},1} \Leftrightarrow S_{\text{Eye2},1}, \qquad (8)$$

$$S_{\text{Eye1},i} \Leftrightarrow S_{\text{Eye2},i} \cdots$$

For each corresponding bilateral sensor pair, there is an integration node $\dot{S}_{\text{Eyes},i}$, yielding a vector of integration nodes $\vec{\dot{S}}_{\text{Eyes},1\ldots i}$. If any of these are 1, that is if the agent encounters binocular correlations in its field of view, then it goes in to "align" mode (see "Movement" below). The extent of the FoV for a visual sensor was arbitrarily set to be 0.1π, but other values could be used. Sampling resolution σ within this FoV was set to 0.0005 radians, but other values could be used.

However, there is connection to audition via $\ddot{S}$ only for the "foveal" sensor pair at the center of the field of vision $S_{\text{Eye1,Center}} \Leftrightarrow S_{\text{Eye2,Center}}$ (see "Inter-modal integration" below). Peripheral occurrences of $\dot{S}_{\text{Eyes},i} = 1$ are used only for orienting, which brings the signal to align with $S_{\text{Eye1,Center}} \Leftrightarrow S_{\text{Eye2,Center}}$.

3.2.6. Audio sensing and signalling

Let us add audio capacity to the agents. The "mouth" is located midway between the "eyes" of the agent, and generates an audio event at each time step. The omnidirectional "ears" are co-located with the "eyes." However, they have a different LoS. The momentary LoS L_{Ear} of an ear may be defined as a subset of W wherein the spatial distance from source to sensor equals the temporal distance from source to sensor. This takes the form of a cone in spacetime. See **Figure 4**. Assuming the sensor is at the origin and sound travels at one spatial quanta per temporal quanta, this is just;

$$L_{\text{Ear}} \subset W \qquad (9)$$

$$\forall w \{w \in L_{\text{Ear}} : \tau_w = -\rho_w\}.$$

The aural LoS are integrated in just the same way as the visual LoS;

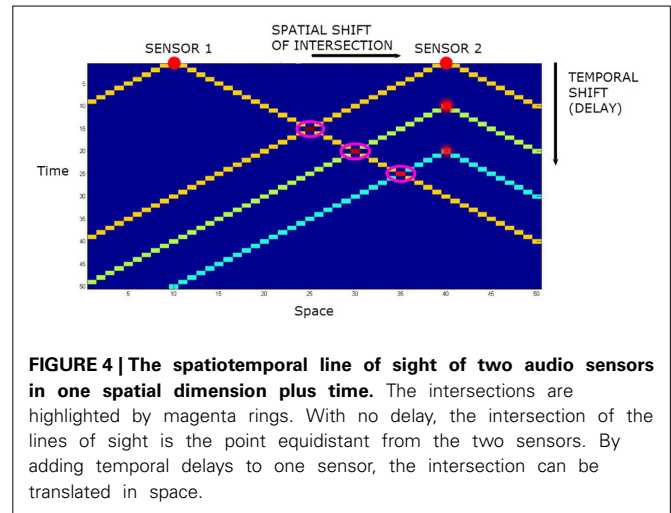$$\dot{S}_{\text{Ears}} = \&(\vee L_{\text{Ear1}}, \vee L_{\text{Ear2}}). \qquad (10)$$

FIGURE 4 | The spatiotemporal line of sight of two audio sensors in one spatial dimension plus time. The intersections are highlighted by magenta rings. With no delay, the intersection of the lines of sight is the point equidistant from the two sensors. By adding temporal delays to one sensor, the intersection can be translated in space.

Ignoring time, the intersection $L_{\text{Ear1}} \cap L_{\text{Eye2}}$ forms a straight line in space defined by equidistance from both sensors. See **Figure 3**. This is equivalent to the "beam" of directional tuning projected in beamforming (**Figure 1**). To avoid the inefficiency of simulating sound propagation, the code used for the current model implements audio beamforming by hand. For each sound source, the code directly measures its offset from the auditory midline $L_{\text{Ear1}} \cap L_{\text{Eye2}}$. This value can then be used to generate a motor command for auditory orienting as in Equation (12). Naturally, the real world case with reverberations etc. becomes much more complex, but the simplified implementation is sufficient for demonstrative purposes.

3.2.7. Intermodal integration

We now introduce an intermodal meta-sensor $\ddot{S}$, which combines the signals from the two unimodal meta-sensors $\dot{S}_{\text{Ears}}$ and $\dot{S}_{\text{Eyes,Center}}$, again with an AND gate. This operation may be described in full as;

$$\ddot{S} = \&(\&(\vee L_{\text{Ear1}}, \vee L_{\text{Ear2}}), \&(\vee L_{\text{Eye1,Center}}, \vee L_{\text{Eye2,Center}})). \quad (11)$$

$\ddot{S} = 1$ if and only if all the lines of sight of the agent's individual sensors intersect $\bar{W}_C$ (or equivalent in the appropriate modality e.g., $\bar{W}_A$ for audio events). Whether they intersect at the *same point* in $\bar{W}_C / \bar{W}_A$ or different points corresponds to the difference between location tuning and configuration tuning mentioned in "Visual beamforming" above. In terms of the momentary state of the agent's sensory apparatus, however, these different ecological situations are indistinguishable. The AND integration steps cannot decrease sparsity. Sparsity here corresponds directly to "selectiveness" in selective attention; with respect to sparsity/selectivity, $\ddot{S} \geq \dot{S}_{\text{SensorPair}} \geq S_{\text{Sensor}}$.

Thus we arrive at an increasingly exclusive, hierarchical sensory definition of "like-me-ness," in a formal sense amenable to implementation on a robot. The higher a momentary sensory sample $\breve{L}$ reaches in the bilateral, multimodal hierarchy, topped here by $\ddot{S}$, the more attention it receives.

3.2.8. Movement

We place one agent, representing the "subject" at the origin. This agent can only rotate on the spot. The other agent, representing the "stimulus," is positioned at a depth of 50 quanta and a randomized position in Θ. It does not move. The stimulus is always oriented toward the subject at the origin, except where "averted gaze" is the experimental condition. See **Figure 3**. The default "searching" behavior for the subject is to rotate in a circle. If in the current timestep, $\dot{S}_{Eyes,i} = 1$, anywhere in the visual field, the agent changes to an "align" behavior. When the subject is in align mode, it orients so as to center this stimulation by producing a motor command proportional to the offset of the correlated sensor pair from the center of the visual field. In the simulations where the ears and mouth are used, the subject also rotates to minimize the delay δ_{Audio} between the audio streams.

$$\Delta\Theta = \eta(i_{center} - i_{correlated}) + \eta\delta_{Audio} \qquad (12)$$

where η is a motor gain constant for which we used $\eta = -0.15\sigma$ (σ denotes sampling resolution). Together with the sensory bias for the center mentioned in the previous subsection, this "orienting-to-center" effectively implements an *a priori* motor assumption of special status for the midline. This "center is special" assumption probably deserves closer scrutiny in terms of its biological basis and potential justification in terms of optimal sampling strategies, but this topic is beyond the scope of the current paper.

4. AGENT-BASED SIMULATIONS

In this section, we describe a series of simulations based in the simple world just described. The simulations address minimal analogs of three newborn abilities; perception of size constancy (Granrud, 1987; Slater et al., 1990), visual face perception (Goren et al., 1975; Johnson et al., 1991) and audio–visual alignment (Guellai and Streri, 2011). In each of the simulations here, the "task" for the subject is to find and align with the stimulus, despite increasing levels of distractors which we place into the world. Distractors are randomly dispersed visual events which are permanent and immobile throughout the course of a trial. Where audio distractors are also used, each visual distractor may, with a 25% chance at each timestep, also generate an audio event. As a quantitative measure of performance, we recorded the angular distance in Θ of the subject from alignment with the stimulus at the end of the trial, and term this value the "error." Each simulation was run 100 times and results averaged over all runs.

4.1. SIMULATION 1.0—SIZE CONSTANCY

Granrud (1987) and Slater et al. (1990) found that newborns could perceive "size constancy" across changes in depth. Familiarization studies showed that the subjects could perceive a distinction between a patterned object and a double-sized (otherwise identical) object at double the distance, such that the size and form of the retinal projection were identical. The size constancy effect has been cited as some of the most convincing evidence for innate perceptual predispositions (Spelke, 1998; Streri et al., 2013b).

4.1.1. Results

This simulation did not include audio capacity or events. A stimulus of equal size to the subject was compared to a stimulus of twice the size. For the "same size" condition, the depth ρ of the stimulus was increased in increments of 1 quanta from 50 to 100 quanta. In the "double size" condition, depth increased in increments of 2 from 100 to 200 quanta; thus double the associated distance for the "same size" stimulus. **Figure 5** displays mean error against distance for the two stimulus conditions. The distance of the stimulus from the subject makes little difference to the result; it is physical size of the configuration which is targeted by visual beamforming, and translation in depth makes no difference with parallel visual axes (until the limits of resolution are reached). See **Figure 3**. The subject finds the equal-size stimulus effectively, and ignores the double-sized stimulus equally effectively.

This is of course a special case; the current model would not discriminate a double-sized from a triple-sized stimulus, as it would ignore both. More generally, otherwise identical stimuli of different physical size may project in an identical manner to one retina if depth compensates, but may project differently in terms of the spatial relations between the images on *both* retinae. Note that disparity sensitivity is not a prerequisite; interactions between corresponding points may in many cases be sufficient.

4.1.2. Discussion

Streri et al. (2013b), after Holway and Boring (1941), state that in order to recognize size constancy, infants need to combine projective size with information about viewing distance. Neither of these values are used explicitly in the current model, which nonetheless exhibits a size constancy effect. Perception of size constancy implies that the stimuli must be characterized by some difference in the way they impact the senses. If there were no detectable difference between stimuli at all, the behavioral distinction would
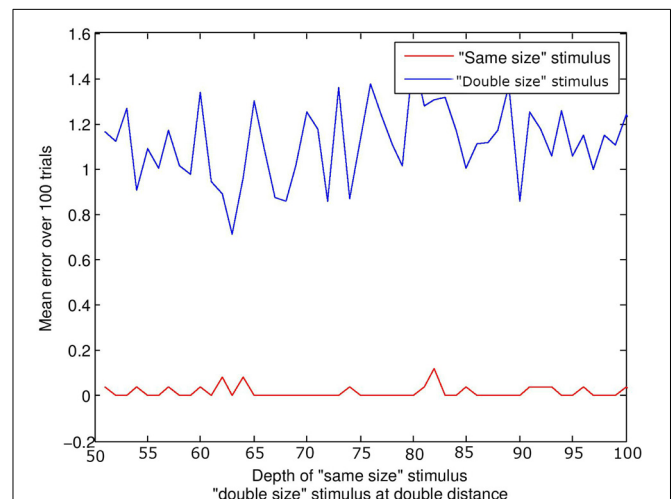

FIGURE 5 | Size constancy over depth. The lines represents the error at the end of the trial. The red line represents the "same size" condition, and the blue line the "double size" condition. For a given depth *d* of the "same size" stimulus, the "double size" stimulus was at twice the distance 2*d*. Regardless of depth, the subject always finds and aligns with the "same size" stimulus, and ignores the "double size" stimulus.

be not just innate, but supernatural. If there is a sensory difference, then a familiarization effect does not imply any specific functional adaptation. Binocular correlation patterns could provide an informational basis for the discrimation between "small" and "far away" in the newborn. Monocular motion parallax might provide another potential source of the distinction.

4.1.3. Prediction
Size constancy in newborns will be lost under monocular viewing conditions. This would distinguish the potential contributions of binocular and monocular mechanisms.

4.2. SIMULATION 2.0—FACE DETECTION
Goren et al. (1975) and Johnson et al. (1991) found evidence for an innate preference for face-like stimuli. We recently showed that these results might result from a newborn "preference" for binocular correlations (Wilkinson et al., in press). Existing theories of newborn face preference posit some kind of monocular neural filter applied to the retinal image (Morton and Johnson, 1991; Turati, 2004; Pitti et al., 2013).

4.2.1. Results—Binocular correlations versus monocular template matching
We compared a subject using bilateral gain control with a subject using a "face template" monocularly in each eye. This template consists simply in seeing two or more visual signals at the same time, analogous to applying a "two blob" template across spatial scales. If such a situation occurs, the template subject rotates to center the position of the match, averaged between the two eyes. There is no audio in this first simulation. **Figure 6** depicts average angular distance of the subject from the stimulus on the vertical axis, against increasing levels of distractors on the horizontal axis. Without distractors, finding the stimulus is a trivial matter of finding the only stimulus around, and so both BCM (red line) and template (blue line) methods perform very well. With high levels of distractors (**Figure 7**), the task is really quite difficult. As distractors are added and *selective attention* is required, the BCM subject substantially outperforms the template subject. This
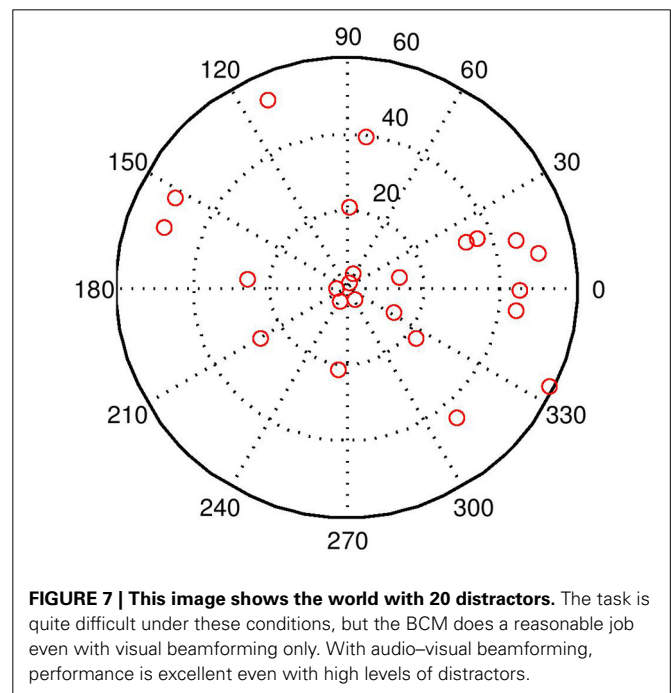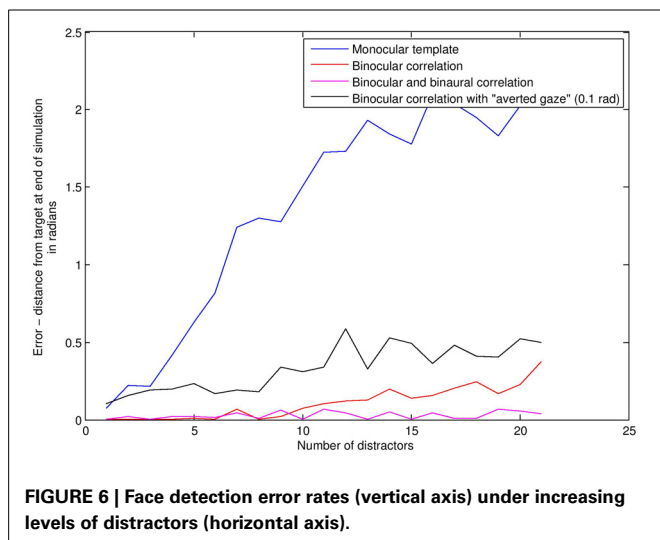
is because the BCM subject is tuned to a particular physical size of stimulus regardless of depth, whereas the monocular template subject confounds size with depth, and is therefore more prone to distractors. Confounding depth with size is a general feature of templates which match the monocular visual array.

4.2.2. Results—averted gaze
Studies have shown that good visual alignment of the stimulus with the subject ("direct gaze") plays an important role in newborn face perception (Farroni et al., 2002; Guellai and Streri, 2011). To model these findings, we alter the alignment of the stimulus, such that it no longer "looks" directly at the subject, to provide a minimal analog of what is termed in the literature "averted gaze." We examined the effect of averted gaze on face detection, using the binocular correlation based detector (black line in **Figure 6**). The result is a roughly constant increase in the error rate over all distractor levels. This arises because although the subject is often able to find the averted gaze stimulus, it is unable to align precisely with it, and so oscillates around the stimulus instead of coming to rest at perfect alignment.

4.2.3. Results—audio-visual face detection
Adding the audio capabilities described in Audio Sensing and Signalling enables improvements in face detection, because $\ddot{S}$ is more selective than $\dot{S}_{Eyes}$ (see "Intermodal integration" above). Audio distractors were used in this simulation. The stimulus was always in the "direct gaze" condition. The magenta line in **Figure 6** shows the error with audio-visual beamforming. Even with high levels of distractors (**Figure 7**), the stimulus is usually located effectively; a configuration of audio and visual distractors matching the form of the stimulus is quite rare, even with up to 20 distractors.

FIGURE 6 | Face detection error rates (vertical axis) under increasing levels of distractors (horizontal axis).

FIGURE 7 | This image shows the world with 20 distractors. The task is quite difficult under these conditions, but the BCM does a reasonable job even with visual beamforming only. With audio–visual beamforming, performance is excellent even with high levels of distractors.

4.2.4. Discussion

This minimal model demonstrates "like me" detection through audio-visual beamforming, in a manner which is highly dependent on good mutual alignment. The astute reader may note that in the real world, babies have different inter-pupillary distance ("IPD") to their adult conspecifics. This is indeed an important factor. We have previously shown that this need not be a major problem in the real world case, as the spatial extension of the signal from the eyes can compensate for some difference in IPD (Wilkinson et al., in press). In general, allometry and alignment are relevant to the behavioral outcomes of bilateral gain control, and will generate individual and inter-trial differences. To keep things simple, we do not expand on this topic in the current article.

4.2.5. Prediction

Newborns will prefer stimuli which induce strong binocular correlations over monocular contrast matched controls which do not. A certain level of "face preference" will drop out of this more general effect.

4.3. SIMULATION 3.0 FACE-VOICE INTEGRATION: THE ROLE OF DIRECT GAZE

Sai (2005) elegantly demonstrated that newborns identify their mothers face via the association of her voice, which they already recognize from the prenatal period. In recent extensions of this work, subjects showed long term recognition effects for dynamic faces only when accompanied by speech, supporting an important role for multimodal stimulation in triggering/gating learning (Guellai et al., 2011). Beyond this, the faces had to fixate the infant; averted gaze abolished recognition effects (Guellai and Streri, 2011). Here we asked whether audio-visual beamforming would produce analogous dependencies on good alignment between subject and stimulus.

In this simulation, the stimulus either aligned properly with the subject—the "direct gaze" condition—or is rotated out of alignment—the "averted gaze" condition. The amount of rotation was incrementally increased between 0.0 and 0.2 radians. We do not model learning *per se* in the current contribution, as would be required to model recognition effects completely. Instead, we constrain the discussion to selective attention, on the basis that attention defines what is learned. Learning is assumed to occur if and only if $\ddot{S} = 1$. We then measured the average (over 100 simulations) amount of time the subject spends in "learning mode" (i.e., $\ddot{S} = 1$) under each increasing misalignment of the stimulus. There were no distractors in this simulation.

4.3.1. Results

The results of this simulation are displayed in **Figure 8**. In the "direct gaze" condition (0.0 radians), the subject finds the stimulus and aligns with it effectively. As a result, it goes into "learning mode" (i.e., $\ddot{S} = 1$) for the remainder of the simulation. Small misalignments up to about 0.07 radians are pretty much undetectable at the resolution used here. Over this level, the amount of time spent learning tends quickly to zero. This is because the subject finds the stimulus, but then oscillates rapidly around it, trying but failing to align with both the audio and visual signals. As
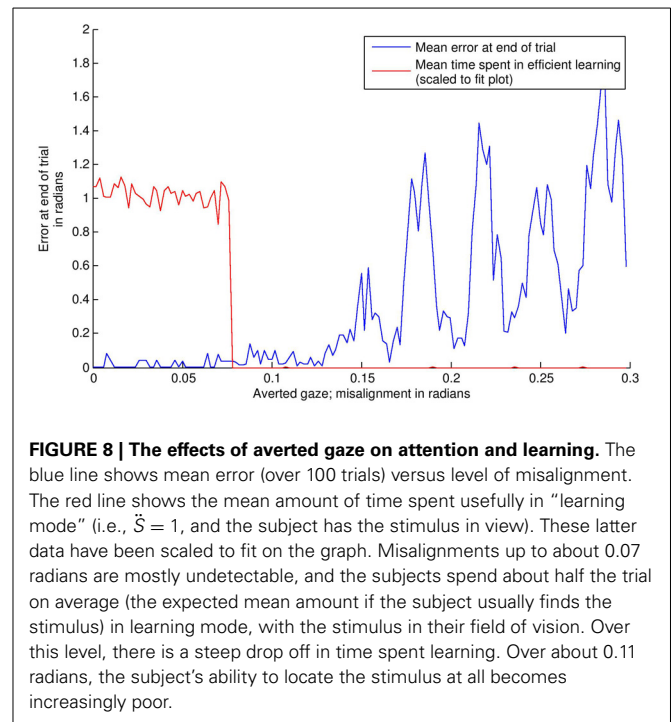

FIGURE 8 | The effects of averted gaze on attention and learning. The blue line shows mean error (over 100 trials) versus level of misalignment. The red line shows the mean amount of time spent usefully in "learning mode" (i.e., $\ddot{S} = 1$, and the subject has the stimulus in view). These latter data have been scaled to fit on the graph. Misalignments up to about 0.07 radians are mostly undetectable, and the subjects spend about half the trial on average (the expected mean amount if the subject usually finds the stimulus) in learning mode, with the stimulus in their field of vision. Over this level, there is a steep drop off in time spent learning. Over about 0.11 radians, the subject's ability to locate the stimulus at all becomes increasingly poor.
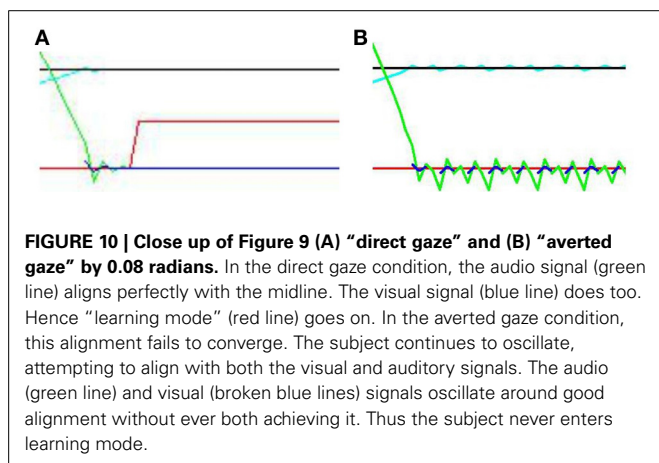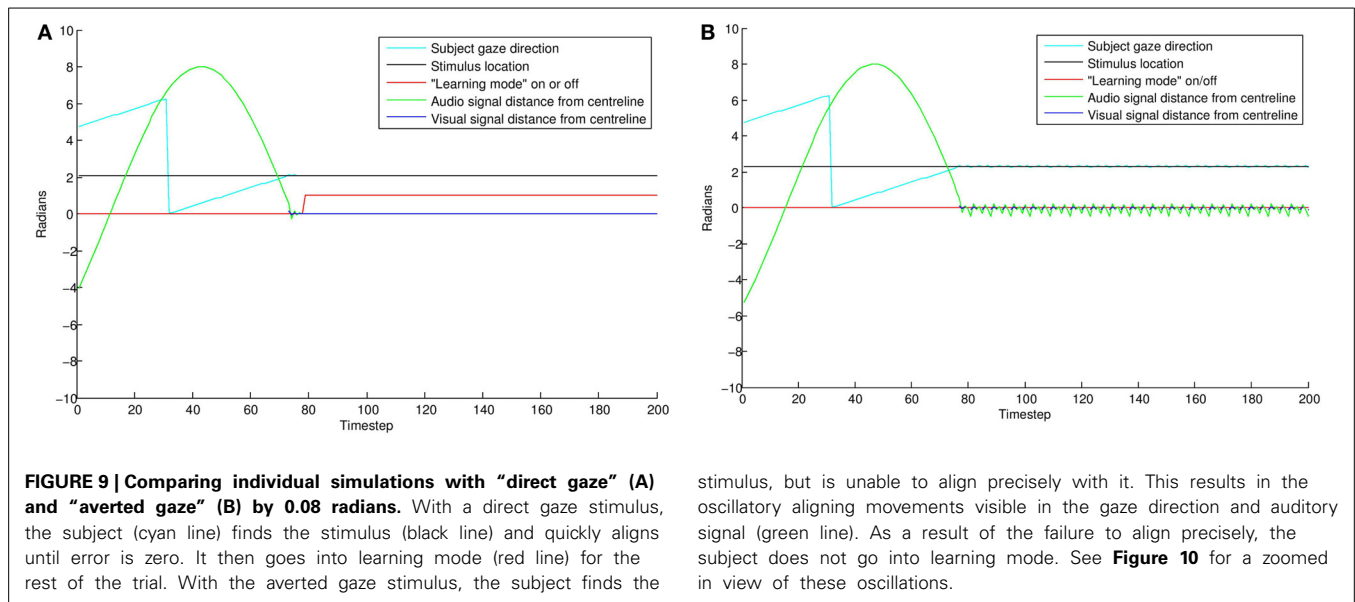
misalignment increases, the subject ceases to locate the stimulus effectively.

Figure 9 displays results from two individual trials. In the direct gaze condition **Figure 9A**, both the audio and visual transceivers of the subject quickly align perfectly with those of the stimulus. With gaze averted by 0.1 radians **Figure 9B**, the subject tries unsuccessfully to align with the stimulus, resulting in a fast oscillation around the stimulus. The green line represents the distance from synchrony of the binaural time signals. Note the fast oscillation, generated by the oscillation of the subject around the stimulus (**Figure 10**). Averted gaze means the subject cannot align properly with both the visual signal (broken blue line) and the audio signal at the same time, and so $\ddot{S} = 0$ at all times. "Learning mode" is obviously an oversimplification, but infants may be suffering from similar difficulty in aligning their sensor arrays with misaligned stimuli.

4.3.2. Discussion

These simulations show that bilateral gain control can in principle generate a strong dependence on alignment of multimodal signal sources. The beamforming technique projects the tuning of the sensor array to a particular subvolume of physical and configural space. A set of unimodal specifications combine to build up a multimodal attentional "template," which is projected onto the world as the combined multimodal "beams" of the array. This projected template describes an "ideal stimulus" in terms of what modalities of signal should be where and in what relations *in physical space*. The establishment and success of an interaction depends upon the extent and accuracy to which this set of selective conditions are met.

We are implementing a slightly more sophisticated model of the BCM on the humanoid robot iCub. The robot's visual system

FIGURE 9 | Comparing individual simulations with "direct gaze" (A) and "averted gaze" (B) by 0.08 radians. With a direct gaze stimulus, the subject (cyan line) finds the stimulus (black line) and quickly aligns until error is zero. It then goes into learning mode (red line) for the rest of the trial. With the averted gaze stimulus, the subject finds the stimulus, but is unable to align precisely with it. This results in the oscillatory aligning movements visible in the gaze direction and auditory signal (green line). As a result of the failure to align precisely, the subject does not go into learning mode. See **Figure 10** for a zoomed in view of these oscillations.

FIGURE 10 | Close up of Figure 9 (A) "direct gaze" and (B) "averted gaze" by 0.08 radians. In the direct gaze condition, the audio signal (green line) aligns perfectly with the midline. The visual signal (blue line) does too. Hence "learning mode" (red line) goes on. In the averted gaze condition, this alignment fails to converge. The subject continues to oscillate, attempting to align with both the visual and auditory signals. The audio (green line) and visual (broken blue lines) signals oscillate around good alignment without ever both achieving it. Thus the subject never enters learning mode.

consists of monocular spatiotemporal filters in a simple *predictive coding* architecture (Rao and Ballard, 1999), based on *adaptive pattern generators* (Righetti and Ijspeert, 2006). Sensory input is subjected to an *active cancelation* through negative feedback of the learned signal. Essentially this is a minimal implementation of sensory adaptation; see Lieder et al. (2013) for a more realistic example of predictive coding applied to adaptation. The remaining sensory input is then subjected to bilateral mutual gain control. Audition uses purely temporal filters, but is otherwise based on the same mechanisms. It is worth noting that the existing literature on *active cancelation* (Kuo and Morgan, 1999) might provide a useful bridge between predictive coding [in particular the influential *active inference* formulation (Friston, 2010)], and the radically embodied implications of array sensing.

Attention is expressed through saccadic eye movements and neck movements, so as to centralize the source of the strongest region of post filter energy in the visual field, using the existing iCub oculomotor system. Vision is blocked during these movements. If at any given time, a "strong" (defined by some

threshold) post-filter visual signal exists in the central region of vision, *and* is accompanied by strong binaural correlation (analogous to $\ddot{S} = 1$), then the robot enters an "attentive" state in which it records the ensuing audio-visual stream to disk. A value representing this state is incremented whenever the above conditions are met, and otherwise decays slowly. We will then analyze the recorded periods, and assess their selectivity and reliability with respect to identifying the "interaction events" we instigate during the trial. This is obviously a much simplified model of the newborn case, but may still be instructive. The "social" functionality is anticipated partly because humans form the bulk of unpredictably moving objects in the lab (and a newborn babies world), and partly because their transceiver array is the right size and shape to cause bilateral correlations in multiple modalities (in a humanoid robot). We believe that identifying and carefully characterizing such ecological factors, together with defining minimal, biologically plausible mechanisms to exploit them, is crucial both to understanding how basic, naïve social attention is apparently so *easy* for babies, and to giving robots a similar "social instinct."

4.3.3. Predictions

(1) In the psychophysical experiments detailed in Guellai and Streri (2011), "averted gaze" interfered with recognition effects. The BCM predicts that in a reproduction of this experiment, "averted voice" (or binaural misalignment of the sound source for the voice) will also impair recognition effects. Stereo speakers with delays could be used to manipulate the apparent source location of the voice in a subtle, non-intrusive way.

(2) We predict that, in the cluttered, busy laboratory environment, the robot described above will preferentially record to disk during "social interaction" events, in particular the case where an interaction partner visually aligns with, and talks to, the robot (which people in the lab will be asked to do). It will occasionally record other events which just happen to impact the senses in the right way to pass the filter. We

will also compare audio-visual beamforming against visual beamforming alone, and against mono mechanisms alone, in order to isolate the contributions of mono spatiotemporal filters and predictive coding, multi-modal coincidence, and bilateral mutual gain control.

5. GENERAL DISCUSSION

Bilaterality provides a spatially selective supramodal dimension along which to collapse the spatial and modal extent of the sensor array. Bilateral mutual gain control can enable spatial-configural tuning through beamforming. The existing evidence that bilateral gain control gates sensory flow in adults of various species practically implies a downstream effect on the assortment of observable outcomes collectively termed "attention" (Li and Ebner, 2006; Ding et al., 2013; Wunderle et al., 2013; Xiong et al., 2013). The major open questions with respect to "innate" perception, attention, and sociality are the nature of bilateral interactions in the human neonate and the implications of this physiological organization on behavioral and ontogenic timescales. The former can be addressed experimentally in neonates, while the latter can be explored via longitudinal empirical designs, mathematical modeling and developmental robotics (Morse et al., 2011).

The formal model and agent based simulations presented here are intended in a purely pedagogical sense, to specify and clarify our argument and demonstrate its consequences, and provide no evidence that newborn perception employs similar mechanisms. However, as for any useful theory, the current proposal *does* link back to the biological case via the predictions it makes regarding newborn behavior. Given knowledge of a newborn subject's bilateral sensory perspectives, a BCM based model can make real time, individual level salience predictions, with a precision that depends on the quality of perspective reproduction and the physiological detail of the model. If testing confirms these predictions, then the BCM will provide a powerful tool to understand early perceptual and social development.

The fundamental generative prediction of the bilateral correlation model of newborn attention is that the spatial-configural attentional biases of newborn babies will be a function of the extent to which multimodal contrast energy in the scene projects synchronously to bilaterally corresponding points on the subjects sensor array, given whatever mono filters are in use. Many experimental paradigms are possible, but the classical preference-on-average over a population of subjects must be replaced by a paradigm focussed at the individual level, as the BCM makes predictions which are dependent on the real time allometry and alignment of the subject's sensor array. Stereo source manipulation of the auditory "sweet spot" offers an unintrusive stimulus paradigm for assessing the role of binaural correlation in newborn attention and learning. Screen based eye-tracking could enable the use of dynamic, responsive stimuli which can manipulate binocular correlation in real time for individual subjects.

Our perspective is substantially in agreement with theories that innate sociality is based on perceptuomotor resonance and motor contagion (Meltzoff and Decety, 2003; Blakemore and Frith, 2005; Lepage and Théoret, 2007; Sugita, 2009; Pitti et al., 2013). However, the BCM differs in taking more seriously the role of the body, and resulting sensory and ecological geometry,

in which the brain and its activity are situated. Innate knowledge of the "like me" is embedded in the sampling biases implied by sensory morphology and behavioral repertoire, inviting a broader conception of the "mirror organism." The shared anatomical structure of interacting brains facilitates the interpersonal synchronization of brain activity (Dumas et al., 2012). Beamforming provides a mechanism by which shared corporal embodiment can play a similar role in mediating spatiotemporal alignment between interactors. It has been suggested that the "social brain network" and the "resting state network" may substantially overlap (Schilbach et al., 2008). The current model carries a similar message at the corporal level; bilateral sensor distribution and integration can create a sampling bias for the "like me" which frames neural and bodily development in a deeply and intrinsically social context from the earliest stages of sensory sampling and the beginnings of experience in the world. We suggest that the current proposal is best interpreted as a contribution to the recent literature developing an *enactive* theory of early social development (Varela et al., 1991; O'Regan and Noë, 2001; De Jaegher and Di Paolo, 2007; Auvray et al., 2009; De Jaegher et al., 2010; Di Paolo et al., 2010; Lenay et al., 2011; Di Paolo and De Jaegher, 2012; Froese et al., 2012; Lenay and Stewart, 2012).

The BCM is relatively easy to implement on a robot, at least in the simple "newborn" form presented here. The basic requirements are a bilaterally organized sensor array geometry, and an integration step based on mutual gain control. The nature of an audio or visual "event" in the current model is deliberately abstract. In the real world case, the choice of sensors and the mono filters applied will define what qualifies as an event. In the case of continuous valued sensor data, the binary AND used for intersensory integration may be replaced by multiplication or a more complicated gain control function, perhaps including normalization. The behavioral outcomes which emerge will depend on the form of the bilateral sensor pair's LoS and the bilateral morphology of the sensor array, shifting much of the explanatory burden for observable functional specificity to the embodiment of the agent and the ecological context. Therefore the design of the robot's sensor array becomes crucial to behavioral outcomes. Equally, this implies the natural selection can mould behavioral outcomes by moulding the phenotypic instantiation of sensor array morphology. Note for example the relatively large head and inter-pupillary distance of the human newborn (Pryor, 1969), which brings the newborn sensor array closer to good spatial tuning with that of adult conspecifics. This may be a candidate for a specific adaptation for newborn social interaction, though may also reflect other necessities such as a large brain.

The BCM is in agreement with current gain control models of bilateral integration (e.g., Ding et al., 2013) and interfaces cleanly with existing models attention based on gain modulation (Hillyard et al., 1998; Salinas and Sejnowski, 2001; Aston-Jones and Cohen, 2005; Reynolds and Heeger, 2009; Feldman and Friston, 2010; Sara and Bouret, 2012), spatiotemporal energy models of early vision (Mante and Carandini, 2005), and computational methods based on salience mapping (Itti and Koch, 2001), wherein it can simply provide another contributor to overall salience. Predictive coding (Rao and Ballard, 1999; Bastos et al., 2012) and its generalization in the free energy minimization

framework (Friston, 2010; Moran et al., 2013) are becoming increasingly influential as models of brain function. In this approach, a major factor influencing gain modulation is the *predictability* of the signal (Feldman and Friston, 2010). Models based on "artificial curiosity" or "intrinsic motivation" (Barto, 2013; Gottlieb et al., 2013) take a similar approach. Bilateral mutual gain control is more *a priori*, in that modulation is [ignoring potential interaction through bilateral normalization (Moradi and Heeger, 2009)], *independent* of the modulated signal. Bilateral mutual gain control is also (we found) awkward to formalize and justify in terms of predictive coding.

In terms of the information compression analogy, bilateral mutual gain control maps more comfortably to an embodied form of lossy *transform coding*, wherein the basis function in which a mono signal is recoded is simply (if unconventionally) the signal from its stereo twin, and vice versa. The application of lossy transform coding followed by lossless predictive coding is standard practice in data compression (Clarke, 1995). The role of transform coding is to order information in a format where threshold based mechanisms can most effectively eliminate information according to some *a priori* value function. For example, in MP3, the audio time signal is subjected to Fourier transform and components which are undetectable to a computational model of human hearing are discarded. The functional justification is reduced file size with little appreciable loss of sound quality for human listeners. In the BCM, signals which are not part of multimodal configurations "resonant" with the allometry and alignment of the array—quantified in terms of bilateral correlation—are discarded/damped. The transform consists in the large scale neural morpho-architecture collapsing the space between bilateral sensor pairs. The sharpness of the filter's cut-off may be manipulated by mono resolution and by allowing a certain extent of spatiotemporal cross-correlation as well as pure correspondence. The functional justification is selective spatial-configural tuning to the "like me," in the broadest sense of the term.

The combination of local (mono) decorrelation, followed by global (bilateral) correlation, may turn out to be justified on sparsity principles (Vinje and Gallant, 2000), and possibly on free-energy minimization principles (Feldman and Friston, 2010) as priors expressing the expectation that spatially neighboring locations in the world are likely to be correlated, whereas spatially distributed locations in the world are not. The thermodynamic unlikeliness of corporal form ensures that it is rarely approximated by random happenings, providing selectivity, while heredity validates the assumption that resonant signals are likely to be interesting, providing relevance.

The principle of beamforming through mutual gain control is not limited to the example of bilateral symmetry expanded here. An interesting and testable, if speculative, prediction is that a mutual gain control relation will exist between "corresponding points" on the fingertips of a single hand, across all age ranges. Work on this topic could help us to understand naïve and skilled use of the hand as a sensor array to apprehend the ecological geometry and dynamics (Gibson, 1961, 1966, 1977), and resulting sensorimotor contingencies (O'Regan and Noë, 2001), of interacting with different shapes, densities, and textures.

5.1. CONCLUSION

We have sketched an uncontraversial argument for bilateral mutual gain control as a basic perceptual mechanism. Bilateral mutual gain control offers an effective and efficient multimodal "like me detector" available to both biological and robotic systems. Some form of bilateral interaction in newborns is much more likely than no bilateral interaction given current evidence, and (perhaps immature) mutual gain control is the strongest candidate for integration. Perhaps more contraversially, we argue that this may explain a number of "innate predispositions" observed in newborn infants. Bilateral interactions have been largely ignored as potential causes of innate predispositions in the developmental psychology literature. Whilst we agree to a large extent with the notion of the competent newborn, the current paper is aimed to target questionable adaptationist, nativist, and internalist assumptions regarding causal structure. Empirical work is required to establish the nature of bilateral interaction in the newborn human, and its potential relevance to particular abilities observed in newborns. Though the current model is a greatly simplified one, the mechanisms involved are very general. Ongoing work is extending the model to dynamic sensor distribution through behavior, simple perceptual learning/adaptation through predictive coding, interactive scenarios, and real world embodiment on the humanoid robot iCub. Whether bilateral mutual gain control can account, fully or partially, for particular existing empirical results is a case-by-case question, and we certainly do not want to assert that the BCM can fully explain newborn social skills. However, we do wish to push the point that *ignoring* bilateral integration has led to questionable interpretations of newborn abilities. In summary, the effects of bilateral integration should be considered and controlled for in the planning, execution, and interpretation of psychophysical experiments investigating perception, attention, and sociality in newborns.

REFERENCES
Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Auvray, M., Lenay, C., and Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas Psychol.* 27, 32–47. doi: 10.1016/j.newideapsych.2007.12.002

Bahrick, L. E., Lickliter, R., and Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Curr. Dir. Psychol. Sci.* 13, 99–102. doi: 10.1111/j.0963-7214.2004.00283.x

Banks, M. (2007). The perceptual consequences of estimating disparity via correlation. *J. Vis.* 7:24. doi: 10.1167/7.15.24

Barto, A. G. (2013). "Intrinsic motivation and reinforcement learning," in *Intrinsically Motivated Learning in Natural and Artificial Systems*, eds G. Baldassarre and M. Mirolli (Berlin: Springer), 17–47. doi: 10.1007/978-3-642-32375-1_2

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Bauck, J. L. (2003). Loudspeaker array for enlarged sweet spot. US Patent 6, 633–648.

Blakemore, S.-J., and Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia* 43, 260–267. doi: 10.1016/j.neuropsychologia.2004.11.012

Brown, A. M., and Miracle, J. A. (2003). Early binocular vision in human infants: limitations on the generality of the superposition hypothesis. *Vis. Res.* 43, 1563–1574. doi: 10.1016/S0042-6989(03)00177-9

Carandini, M., and Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3136

Carandini, M., Heeger, D. J., and Movshon, J. A. (1997). Linearity and normalization in simple cells of the primary visual cortex. *J. Neurosci.* 17, 8621–8644.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975. doi: 10.1121/1.1907229

Chino, Y., Smith III, E., Hatta, S., and Cheng, H. (1997). Postnatal development of binocular disparity sensitivity in neurons of the primate visual cortex. *J. Neurosci.* 17, 296–307.

Clark, W. L. G. (1932). The structure and connections of the thalamus. *Brain* 55, 406–470. doi: 10.1093/brain/55.3.406

Clarke, R. J. (1995). *Digital Compression of Still Images and Video*. New York, NY: Academic Press, Inc.

Coghill, R. C., Sang, C. N., Maisog, J. M., and Iadarola, M. J. (1999). Pain intensity processing within the human brain: a bilateral, distributed mechanism. *J. Neurophysiol.* 82, 1934–1943.

Crick, F. (1984). Function of the thalamic reticular complex: the searchlight hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 81, 4586–4590. doi: 10.1073/pnas.81.14.4586

Davis, K. D., Kwan, C. L., Crawley, A. P., and Mikulis, D. J. (1998). Functional mri study of thalamic and cortical activations evoked by cutaneous heat, cold, and tactile stimuli. *J. Neurophysiol.* 80, 1533–1546.

de Haan, B., Morgan, P. S., and Rorden, C. (2008). Covert orienting of attention and overt eye movements activate identical brain regions. *Brain Res.* 1204, 102–111. doi: 10.1016/j.brainres.2008.01.105

De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making. *Phenomenol. Cogn. Sci.* 6, 485–507. doi: 10.1007/s11097-007-9076-9

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

Di Paolo, E., and De Jaegher, H. (2012). The interactive brain hypothesis. *Front. Hum. Neurosci.* 6:163. doi: 10.3389/fnhum.2012.00163

Di Paolo, E. A., Rohde, M., and De Jaegher, H. (2010). "Horizons for the enactive mind: values, social interaction, and play," in *Enaction: Towards a New Paradigm for Cognitive Science*, eds J. R. Stewart, O. Gapenne, and E. A. Di Paolo (Cambridge, MA: MIT Press), 33–87.

Ding, J., Klein, S. A., and Levi, D. M. (2013). Binocular combination in abnormal binocular vision. *J. Vis.* 13:14. doi: 10.1167/13.2.14

Ding, J., and Sperling, G. (2006). A gain-control theory of binocular combination. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1141–1146. doi: 10.1073/pnas.0509629103

Dumas, G., Chavez, M., Nadel, J., and Martinerie, J. (2012). Anatomical connectivity influences both intra-and inter-brain synchronizations. *PLoS ONE* 7:e36414. doi: 10.1371/journal.pone.0036414

Eimer, M., Forster, B., Velzen, J. V., and Prabhu, G. (2005). Covert manual response preparation triggers attentional shifts: erp evidence for the premotor theory of attention. *Neuropsychologia* 43, 957–966. doi: 10.1016/j.neuropsychologia.2004.08.011

Emmers, R. (1965). Organization of the first and the second somesthetic regions (si and sii) in the rat thalamus. *J. Comp. Neurol.* 124, 215–227. doi: 10.1002/cne.901240207

Farroni, T., Csibra, G., Simion, F., and Johnson, M. (2002). Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9602–9605. doi: 10.1073/pnas.152159999

Farroni, T., Valenza, E., Simion, F., and Umilta, C. (2000). Configural processing at birth: evidence for perceptual organisation. *Perception* 29, 355–372. doi: 10.1068/p2858

Feldman, H., and Friston, K. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Filippini, H. R., and Banks, M. (2009). Limits of stereopsis explained by local cross-correlation. *J. Vis.* 9, 8.1–8.18. doi: 10.1167/9.1.8

Fitzpatrick, D. C., Batra, R., Stanford, T. R., and Kuwada, S. (1997). A neuronal population code for sound localization. *Nature* 388, 871–874. doi: 10.1038/42246

Fitzpatrick, D. C., Kuwada, S., and Batra, R. (2000). Neural sensitivity to interaural time differences: beyond the jeffress model. *J. Neurosci.* 20, 1605–1615. Available online at: http://www.jneurosci.org/content/20/4/1605.full

Fitzpatrick, D. C., Kuwada, S., and Batra, R. (2002). Transformations in processing interaural time differences between the superior olivary complex and inferior colliculus: beyond the Jeffress model. *Hear. Res.* 168, 79–89. doi: 10.1016/S0378-5955(02)00359-3

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Froese, T., Lenay, C., and Ikegami, T. (2012). Imitation by social interaction? Analysis of a minimal agent-based model of the correspondence problem. *Front. Hum. Neurosci.* 6:202. doi: 10.3389/fnhum.2012.00202

Fuller, P., Sherman, D., Pedersen, N. P., Saper, C. B., and Lu, J. (2011). Reassessment of the structural basis of the ascending arousal system. *J. Comp. Neurol.* 519, 933–956. doi: 10.1002/cne.22559

Furst, M., Bresloff, I., Levine, R. A., Merlob, P. L., and Attias, J. J. (2004). Interaural time coincidence detectors are present at birth: evidence from binaural interaction. *Hear. Res.* 187, 63–72. doi: 10.1016/S0378-5955(03)00331-9

Gibson, J. (1961). Ecological optics. *Vis. Res.* 1, 253–262. doi: 10.1016/0042-6989(61)90005-0

Gibson, J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.

Gibson, J. (1977). "The concept of affordances," in *Perceiving, Acting, and Knowing*, eds R. Shaw and J. Bransford (Hillsdale, NJ: Lawrence Erlbaum), 67–82.

Gliga, T., and Csibra, G. (2007). Seeing the face through the eyes: a developmental perspective on face expertise. *Prog. Brain Res.* 164, 323–339. doi: 10.1016/S0079-6123(07)64018-7

Goren, C., Sarty, M., and Wu, P. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* 56, 544–549.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., and Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* 17, 585–593. doi: 10.1016/j.tics.2013.09.001

Gotts, S., Chow, C., and Martin, A. (2012). Repetition priming and repetition suppression: a case for enhanced efficiency through neural synchronization. *Cogn. Neurosci.* 3, 227–237. doi: 10.1080/17588928.2012.670617

Granrud, C. E. (1987). Size constancy in newborn human infants. *Invest. Ophthalmol. Visual Sci.* 28(Suppl.), 5.

Greenberg, J. E., and Zurek, P. M. (1992). Evaluation of an adaptive beamforming method for hearing aids. *J. Acoust. Soc. Am.* 91, 1662. doi: 10.1121/1.402446

Gregoriou, G. G., Gotts, S. J., and Desimone, R. (2012). Cell-type-specific synchronization of neural activity in fef with v4 during attention. *Neuron* 73, 581–594. doi: 10.1016/j.neuron.2011.12.019

Grothe, B. (2003). New roles for synaptic inhibition in sound localization. *Nat. Rev. Neurosci.* 4, 540–550. doi: 10.1038/nrn1136

Guellai, B., Coulon, M., and Streri, A. (2011). The role of motion and speech in face recognition at birth. *Vis. Cogn.* 19, 1212–1233. doi: 10.1080/13506285.2011.620578

Guellai, B., and Streri, A. (2011). Cues for early social skills: direct gaze modulates newborns' recognition of talking faces. *PLoS ONE* 6:e18610. doi: 10.1371/journal.pone.0018610

Hansson, P., and Brismar, T. (1999). Tactile stimulation of the hand causes bilateral cortical activation: a functional magnetic resonance study in humans. *Neurosci. Lett.* 271, 29–32. doi: 10.1016/S0304-3940(99)00508-X

Hartmann, W. M., Dunai, L., and Qu, T. (2013). "Interaural time difference thresholds as a function of frequency," in *Basic Aspects of Hearing*, eds B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, and H. E. Gockel (Berlin: Springer), 239–246. doi: 10.1007/978-1-4614-1590-9_27

Haykin, S., and Chen, Z. (2005). The cocktail party problem. *Neural Comput.* 17, 1875–1902. doi: 10.1162/0899766054322964

He, J. (2003). Corticofugal modulation of the auditory thalamus. *Exp. Brain Res.* 153, 579–590. doi: 10.1007/s00221-003-1680-5

Helmholtz, H. (1968). "The facts of perception (originally published 1878)," in *Helmhotz on Perception: Its Physiology and Development*, eds R. M. Warren and R. P. Warren (New york, NY: Wiley), 205–231.

Hillyard, S. A., Vogel, E. K., and Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and

neuroimaging evidence. *Phil. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 353, 1257–1270. doi: 10.1098/rstb.1998.0281

Holway, A. H., and Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *Am. J. Psychol.* 54, 21–37. doi: 10.2307/1417790

Horton, J. C., and Hocking, D. R. (1996). An adult-like pattern of ocular dominance columns in striate cortex of newborn monkeys prior to visual experience. *J. Neurosci.* 16, 1791–1807.

Huguenard, J. R., and McCormick, D. A. (2007). Thalamic synchrony and dynamic regulation of global forebrain oscillations. *Trends Neurosci.* 30, 350–356. doi: 10.1016/j.tins.2007.05.007

Ingham, N. J., and McAlpine, D. (2005). Gabaergic inhibition controls neural gain in inferior colliculus neurons sensitive to interaural time differences. *J. Neurosci.* 25, 6187–6198. doi: 10.1523/JNEUROSCI.0146-05.2005

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203. doi: 10.1038/35058500

Iwamura, Y. (2000). Bilateral receptive field neurons and callosal connections in the somatosensory cortex. *Phil. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 355, 267–273. doi: 10.1098/rstb.2000.0563

Jeffress, L. A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41, 35. doi: 10.1037/h0061495

Jiang, Z. D., and Tierney, T. S. (1996). Binaural interaction in human neonatal auditory brainstem. *Pediatr. Res.* 39, 708–714. doi: 10.1203/00006450-199604000-00024

Johnson, M. (2003). Functional brain development in infants: elements of an interactive specialization framework. *Child Dev.* 71, 75–81. doi: 10.1111/1467-8624.00120

Johnson, M., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns preferential tracking of faces and its subsequent decline. *Cognition* 40, 1–19. doi: 10.1016/0010-0277(91)90045-6

Johnson, M., and Morton, J. (1991). *Biology and Cognitive Development: The Case of Face Recognition*. Oxford: Blackwell.

Jones, E. (1998). "Chapter I the thalamus of primates," *Handbook of Chemical Neuroanatomy*, vol 14, eds F. E. Bloom, A. Björklund, and T. Hökfelt (Amsterdam: Elsevier), 1–298.

Joris, P., and Yin, T. C. (2007). A matter of time: internal delays in binaural processing. *Trends Neurosci.* 30, 70–78. doi: 10.1016/j.tins.2006.12.004

Kashino, M., and Nishida, S. (1998). Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect. *J. Acoust. Soc. Am.* 103, 3597. doi: 10.1121/1.423064

Katzner, S., Busse, L., and Carandini, M. (2011). Gabaa inhibition controls response gain in visual cortex. *J. Neurosci.* 31, 5931–5941. doi: 10.1523/JNEUROSCI.5753-10.2011

Kavšek, M. (2013). Infants' responsiveness to rivalrous gratings. *Vis. Res.* 76, 50–59. doi: 10.1016/j.visres.2012.10.011

Kellermann, P., Reindl, K., and Zheng, Y. (2012). Method and acoustic signal processing system for interference and noise suppression in binaural microphone configurations. EP Patent 2,395,506.

Kompis, M., and Dillier, N. (2001). Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise-ratio improvement. *J. Acoust. Soc. Am.* 109, 1123. doi: 10.1121/1.1338557

Krumbholz, K., Schönwiesner, M., von Cramon, D. Y., Rübsamen, R., Shah, N. J., Zilles, K., et al. (2005). Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. *Cereb. Cortex* 15, 317–324. doi: 10.1093/cercor/bhh133

Kuo, S. M., and Morgan, D. R. (1999). Active noise control: a tutorial review. *Proc. IEEE* 87, 943–973. doi: 10.1109/5.763310

Lakkis, I. (2012). Method and apparatus for creating beamforming profiles in a wireless communication network. US Patent 8,219,891.

Lenay, C., and Stewart, J. (2012). Minimalist approach to perceptual interactions. *Front. Hum. Neurosci.* 6:98. doi: 10.3389/fnhum.2012.00098

Lenay, C., Stewart, J., Rohde, M., and Amar, A. A. (2011). You never fail to surprise me: the hallmark of the other: experimental study and simulations of perceptual crossing. *Interact. Stud.* 12, 373–396. doi: 10.1075/is.12.3.01len

Lepage, J.-F., and Théoret, H. (2007). The mirror neuron system: grasping others actions from birth? *Dev. Sci.* 10, 513–523. doi: 10.1111/j.1467-7687.2007.00631.x

Li, L., and Ebner, F. F. (2006). Balancing bilateral sensory activity: callosal processing modulates sensory transmission through the contralateral thalamus

by altering the response threshold. *Exp. Brain Res.* 172, 397–415. doi: 10.1007/s00221-005-0337-y

Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J., and Stephan, K. E. (2013). Modelling trial-by-trial changes in the mismatch negativity. *PLoS Comput. Biol.* 9:e1002911. doi: 10.1371/journal.pcbi.1002911

Lien, A. D., and Scanziani, M. (2013). Tuned thalamic excitation is amplified by visual cortical circuits. *Nat. Neurosci.* 16, 1315–1323. doi: 10.1038/nn.3488

Litovsky, R. Y. (2012). "Development of binaural and spatial hearing," in *Human Auditory Development*, eds L. Werner, R. R. Fay, and A. N. Popper (Berlin: Springer), 163–195. doi: 10.1007/978-1-4614-1421-6_6

Litva, J., and Lo, T. K. (1996). *Digital Beamforming in Wireless Communications*. Boston, MA: Artech House, Inc.

Ma, G. (2013). Wireless binaural compressor. US Patent 20,130,010,973.

Maier, J. K., Hehrmann, P., Harper, N. S., Klump, G. M., Pressnitzer, D., and McAlpine, D. (2012). Adaptive coding is constrained to midline locations in a spatial listening task. *J. Neurophysiol.* 108, 1856–1868. doi: 10.1152/jn.00652.2011

Mante, V., and Carandini, M. (2005). Mapping of stimulus energy in primary visual cortex. *J. Neurophysiol.* 94, 788–798. doi: 10.1152/jn.01094.2004

Meese, T. S., and Baker, D. H. (2011). Contrast summation across eyes and space is revealed along the entire dipper function by a "swiss cheese" stimulus. *J. Vis.* 11:23. doi: 10.1167/11.1.23

Meltzoff, A. N., and Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Phil. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 358, 491–500. doi: 10.1098/rstb.2002.1261

Mendelson, M., Haith, M., and Gibson, J. (1976). The relation between audition and vision in the human newborn. *Monogr. Soc. Res. Child Dev.* 41, 1–72. doi: 10.2307/1165922

Moradi, F., and Heeger, D. J. (2009). Inter-ocular contrast normalization in human visual cortex. *J. Vis.* 9, 13.1–13.22. doi: 10.1167/9.3.13

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., and Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. *J. Neurosci.* 33, 8227–8236. doi: 10.1523/JNEUROSCI.4255-12.2013

Morse, A. F., Herrera, C., Clowes, R., Montebelli, A., and Ziemke, T. (2011). The role of robotic modelling in cognitive science. *New Ideas Psychol.* 29, 312–324. doi: 10.1016/j.newideapsych.2011.02.001 (Special Issue: Cognitive Robotics and Reevaluation of Piaget's Concept of Egocentrism).

Morton, J., and Johnson, M. (1991). Conspec and conlern: a two-process theory of infant face recognition. *Psychol. Rev.* 98, 164–181. doi: 10.1037/0033-295X.98.2.164

Mountcastle, V., and Henneman, E. (1949). Pattern of tactile representation in thalamus of cat. *J. Neurophysiol.* 12, 85–100.

Naidu, P. (2001). *Sensor Array Signal Processing*. New York, NY: CRC Press.

Olsen, S. R., Bortone, D. S., Adesnik, H., and Scanziani, M. (2012). Gain control by layer six in cortical circuits of vision. *Nature* 483, 47–52. doi: 10.1038/nature10835

O'Regan, J., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973. doi: 10.1017/S0140525X01000115

Pitti, A., Kuniyoshi, Y., Quoy, M., and Gaussier, P. (2013). Modeling the minimal newborn's intersubjective mind: the visuotopic-somatotopic alignment hypothesis in the superior colliculus. *PLoS ONE* 8:e69474. doi: 10.1371/journal.pone.0069474

Poisel, R. (2011). *Modern Communications Jamming: Principles and Techniques*. Boston, MA: Artech House.

Pryor, H. B. (1969). Objective measurement of interpupillary distance. *Pediatrics* 44, 973–977.

Rakic, P. (1976). Prenatal genesis of connections subserving ocular dominance in the rhesus monkey. *Nature* 261, 467–471. doi: 10.1038/261467a0

Raleigh, G. G., and Cioffi, J. M. (1998). Spatio-temporal coding for wireless communication. *IEEE Trans. Commun.* 46, 357–366. doi: 10.1109/26.662641

Rao, R., and Ballard, D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.2009.01.002

Righetti, L., and Ijspeert, A. J. (2006). "Programmable central pattern generators: an application to biped locomotion control," in *Proceedings 2006 IEEE International Conference on Robotics and Automation (ICRA 2006)* (Orlando, FL: IEEE), 1585–1590. doi: 10.1109/ROBOT.2006.1641933

Rothman, J., Cathala, L., Steuber, V., and Silver, R. (2009). Synaptic depression enables neuronal gain control. *Nature* 457, 1015–1018. doi: 10.1038/nature07604

Saalmann, Y. B., and Kastner, S. (2009). Gain control in the visual thalamus during perception and cognition. *Curr. Opin. Neurobiol.* 19, 408–414. doi: 10.1016/j.conb.2009.05.007

Saalmann, Y. B., and Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. *Neuron* 71, 209–223. doi: 10.1016/j.neuron.2011.06.027

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: evidence for intermodal perception at birth. *Infant Child Dev.* 14, 29–50. doi: 10.1002/icd.376

Salinas, E., and Sejnowski, T. J. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. *Neuroscientist* 7, 430–440. doi: 10.1177/107385840100700512

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev. Neurosci.* 10, 211–223. doi: 10.1038/nrn2573

Sara, S. J., and Bouret, S. (2012). Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron* 76, 130–141. doi: 10.1016/j.neuron.2012.09.011

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., and Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Conscious. Cogn.* 17, 457–467. doi: 10.1016/j.concog.2008.03.013

Schmidt, K. E. (2013). The visual callosal connection: a connection like any other? *Neural Plast.* 2013, 397176. doi: 10.1155/2013/397176

Sherman, S. M. (2007). The thalamus is more than just a relay. *Curr. Opin. Neurobiol.* 17, 417–422. doi: 10.1016/j.conb.2007.07.003

Sherman, S. M., and Guillery, R. W. (2006). *Exploring the Thalamus and its Role in Cortical Function*, 2nd edn. Cambridge, MA: MIT Press.

Shimojo, S., Jr., J. B., O'Connell, K. M., and Held, R. (1986). Pre-stereoptic binocular vision in infants. *Vis. Res.* 26, 501–510. doi: 10.1016/0042-6989(86)90193-8

Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., and Engel, A. K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. *Neuron* 60, 709–719. doi: 10.1016/j.neuron.2008.09.010

Sillito, A., Jones, H., Gerstein, G., and West, D. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature* 369, 479–482. doi: 10.1038/369479a0

Slater, A., and Kirby, R. (1998). Innate and learned perceptual abilities in the newborn infant. *Exp. Brain Res.* 123, 90–94. doi: 10.1007/s002210050548

Slater, A., Mattock, A., and Brown, E. (1990). Size constancy at birth: newborn infants' responses to retinal and real size. *J. Exp. Child Psychol.* 49, 314–322. doi: 10.1016/0022-0965(90)90061-C

Slater, A. M., and Findlay, J. M. (1975). Binocular fixation in the newborn baby. *J. Exp. Child Psychol.* 20, 248–273. doi: 10.1016/0022-0965(75)90102-2

Spelke, E. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behav. Dev.* 21, 181–200. doi: 10.1016/S0163-6383(98)90002-9

Steinberg, L. J., Fischer, B. J., and Peña, J. L. (2013). Binaural gain modulation of spectrotemporal tuning in the interaural level difference-coding pathway. *J. Neurosci.* 33, 11089–11099. doi: 10.1523/JNEUROSCI.4941-12.2013

Streri, A., Coulon, M., and Guellaï, B. (2013a). The foundations of social cognition: studies on face/voice integration in newborn infants. *Int. J. Behav. Dev.* 37, 79–83. doi: 10.1177/0165025412465361

Streri, A., Hevia, M. D. D., Izard, V., and Coubart, A. (2013b). What do we know about neonatal cognition? *Behav. Sci.* 3, 154–169. doi: 10.3390/bs3010154

Sugita, Y. (2009). Innate face processing. *Curr. Opin. Neurobiol.* 19, 39. doi: 10.1016/j.conb.2009.03.001

Sun, H., Mabande, E., Kowalczyk, K., and Kellermann, W. (2012). Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing. *J. Acoust. Soc. Am.* 131, 2828. doi: 10.1121/1.3688476

Theile, G. (2000). "Multichannel natural recording based on psychoacoustic principles," in *Audio Engineering Society Convention 108* (Paris: Audio Engineering Society).

Thorn, F., Gwiazda, J., Cruz, A. A., Bauer, J. A., and Held, R. (1994). The development of eye alignment, convergence, and sensory binocularity in young infants. *Invest. Ophthalmol. Vis. Sci.* 35, 544–553.

Towal, R., Quist, B., Gopal, V., Solomon, J., and Hartmann, M. (2011). The morphology of the rat vibrissal array: a model for quantifying spatiotemporal patterns of whisker-object contact. *PLoS Comput. Biol.* 7:e1001120. doi: 10.1371/journal.pcbi.1001120

Turati, C. (2004). Why faces are not special to newborns. *Curr. Dir. Psychol. Sci.* 13, 5–8. doi: 10.1111/j.0963-7214.2004.01301002.x

Van Veen, B., and Buckley, K. (1988). Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Mag.* 5, 4–24. doi: 10.1109/53.665

Varela, F. J., and Singer, W. (1987). Neuronal dynamics in the visual corticothalamic pathway revealed through binocular rivalry. *Exp. Brain Res.* 66, 10–20. doi: 10.1007/BF00236196.

Varela, F. J., Thompson, E. T., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.

Vélez-Fort, M., and Margrie, T. W. (2012). Cortical circuits: layer 6 is a gain changer. *Curr. Biol.* 22, R411–R413. doi: 10.1016/j.cub.2012.03.055

Vinje, W. E., and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287, 1273–1276. doi: 10.1126/science.287.5456.1273

Voegtline, K. M., Costigan, K. A., Pater, H. A., and DiPietro, J. A. (2013). Near-term fetal response to maternal spoken voice. *Infant Behav. Dev.* 36, 526–533. doi: 10.1016/j.infbeh.2013.05.002

Voisin, D., Guy, N., Chalus, M., and Dallel, R. (2005). Nociceptive stimulation activates locus coeruleus neurones projecting to the somatosensory thalamus in the rat. *J. Physiol.* 566, 929–937. doi: 10.1113/jphysiol.2005.086520

Ward, L. M. (2013). The thalamus: gateway to the mind. *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 609–622. doi: 10.1002/wcs.1256

Westermann, S. E. (2003). Binaural digital hearing aid system. US Patent 6, 549–613.

Wilkinson, N., Paikan, A., Rea, F., Metta, G., and Gredebäck, G. (in press). Staring us in the face? An embodied theory of innate face preference. *Dev. Sci.*

Womelsdorf, T., and Fries, P. (2007). The role of neuronal synchronization in selective attention. *Curr. Opin. Neurobiol.* 17, 154–160. doi: 10.1016/j.conb.2007.02.002

Wrege, K. S., and Starr, A. (1981). Binaural interaction in human auditory brainstem evoked potentials. *Arch. Neurol.* 38, 572. doi: 10.1001/archneur.1981.00510090066008

Wunderle, T., Eriksson, D., and Schmidt, K. E. (2013). Multiplicative mechanism of lateral interactions revealed by controlling interhemispheric input. *Cereb. Cortex* 23, 900–912. doi: 10.1093/cercor/bhs081

Xiong, X. R., Liang, F., Li, H., Mesik, L., Zhang, K. K., Polley, D. B., et al. (2013). Interaural level difference-dependent gain control and synaptic scaling underlying binaural computation. *Neuron* 79, 738–753. doi: 10.1016/j.neuron.2013.06.012

Zhang, Z., Liu, C.-H., Yu, Y.-Q., Fujimoto, K., Chan, Y.-S., and He, J. (2008). Corticofugal projection inhibits the auditory thalamus through the thalamic reticular nucleus. *J. Neurophysiol.* 99, 2938–2945. doi: 10.1152/jn.00002.2008