frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 15 May 2020
doi: 10.3389/fnbot.2020.00029

From Near-Optimal Bayesian Integration to Neuromorphic Hardware: A Neural Network Model of Multisensory Integration

Timo Oess[1]*, *Maximilian P. R. Löhr*[2], *Daniel Schmid*[2], *Marc O. Ernst*[1] *and Heiko Neumann*[2]

[1] *Applied Cognitive Psychology, Institute of Psychology and Education, Ulm University, Ulm, Germany,* [2] *Vision and Perception Science Lab, Institute of Neural Information Processing, Ulm University, Ulm, Germany*

While interacting with the world our senses and nervous system are constantly challenged to identify the origin and coherence of sensory input signals of various intensities. This problem becomes apparent when stimuli from different modalities need to be combined, e.g., to find out whether an auditory stimulus and a visual stimulus belong to the same object. To cope with this problem, humans and most other animal species are equipped with complex neural circuits to enable fast and reliable combination of signals from various sensory organs. This multisensory integration starts in the brain stem to facilitate unconscious reflexes and continues on ascending pathways to cortical areas for further processing. To investigate the underlying mechanisms in detail, we developed a canonical neural network model for multisensory integration that resembles neurophysiological findings. For example, the model comprises multisensory integration neurons that receive excitatory and inhibitory inputs from unimodal auditory and visual neurons, respectively, as well as feedback from cortex. Such feedback projections facilitate multisensory response enhancement and lead to the commonly observed inverse effectiveness of neural activity in multisensory neurons. Two versions of the model are implemented, a rate-based neural network model for qualitative analysis and a variant that employs spiking neurons for deployment on a neuromorphic processing. This dual approach allows to create an evaluation environment with the ability to test model performances with real world inputs. As a platform for deployment we chose IBM's neurosynaptic chip TrueNorth. Behavioral studies in humans indicate that temporal and spatial offsets as well as reliability of stimuli are critical parameters for integrating signals from different modalities. The model reproduces such behavior in experiments with different sets of stimuli. In particular, model performance for stimuli with varying spatial offset is tested. In addition, we demonstrate that due to the emergent properties of network dynamics model performance is close to optimal Bayesian inference for integration of multimodal sensory signals. Furthermore, the implementation of the model on a neuromorphic processing chip enables a complete neuromorphic processing cascade from sensory perception to multisensory integration and the evaluation of model performance for real world inputs.

Keywords: multisensory integration, spiking neural network, neural network, neuromorphic processing, Bayesian inference, audio-visual integration, computational modeling

OPEN ACCESS

Edited by:
Jun Tani,
Okinawa Institute of Science and Technology Graduate University, Japan

Reviewed by:
Fernando Perez-Peña, University of Cádiz, Spain
Christian Tetzlaff, University of Göttingen, Germany

***Correspondence:**
Timo Oess
timo.oess@uni-ulm.de

Received: *07 February 2020*
Accepted: *22 April 2020*
Published: *15 May 2020*

Citation:
Oess T, Löhr MPR, Schmid D, Ernst MO and Neumann H (2020) From Near-Optimal Bayesian Integration to Neuromorphic Hardware: A Neural Network Model of Multisensory Integration. Front. Neurorobot. 14:29. doi: 10.3389/fnbot.2020.00029

1. INTRODUCTION

While interacting with the world our senses are exposed to a rich and constant flow of information. Making sense of this vast of information is one of the most important task of our brain and crucial for survival. It does this by combing complementary information about the same event from different senses into a single percept. This integration process leads to an enhancement of the combined signal, thus supports the detection of events or objects of interest, improves disambiguation and allows for faster and more accurate processing than could be derived by a mere linear combination of unimodal information streams (Stein and Stanford, 2008).

Humans and other mammals are equipped with complex neural circuits to ensure fast, reliable and optimal combination of signals from various sensory organs (Marrocco and Li, 1977; Edwards et al., 1979; Cadusseau and Roger, 1985). This multisensory integration (MSI) process can be found already in the superior colliculus (SC) of the brain stem where auditory, visual and vestibular signals are combined to facilitate fast reflexive eye movements (Stein et al., 1983). This integration process is refined on ascending cortical pathways for higher level processing and decision making.

The SC is a melting pot of information from various sensory modalities and neurons in the SC are the first multimodal processing units in ascending sensory pathways (Meredith and Stein, 1983; Wallace and Stein, 1997) with spatially aligned receptive fields to these modalities (Meredith and Stein, 1996). The superficial layers of the SC receive mainly retinotopic inputs from the visual system and respond only to visual signals (Wallace et al., 1998). However, neurons in deeper layers of the SC gradually receive ascending inputs from other modalities and exhibit receptive fields for these modalities. In addition, their responses are multi-modal, i.e., receiving input from two different modalities leads to response characteristics that are different than responses to uni-modal signals (Stein and Stanford, 2008). Inputs to neurons in deep layers come from a diverse set of sensory systems and range from auditory signals from the inferior colliculus to proprioceptive signals from the vestibular system. To create a common frame of references for these different signals and, thus, spatially align them the retinotopic visual input is used as a guidance signal. This has been demonstrated in neurophysiological studies as well as modeling investigations (Rees, 1996; Wallace et al., 2004; Oess et al., 2020a).

Despite of all the ascending sensory signals in the SC, neurophysiological studies in cats indicate that there are several descending projections from the association areas (AES) of the cortex. Unimodal cortical projections from anterior ectosylvian visual area (AEv) and the auditory field of the anterior ectosylvian region (FAEs) are observed (Meredith and Clemo, 1989; Wallace et al., 1993; Wallace and Stein, 1994). These projections seem to play an essential role for the integration ability of SC neurons. Studies demonstrate that when these projections are deactivated, the neurons in the SC loose all their multisensory response characteristics (Alvarado et al., 2007a). These characteristics of SC neurons are complex and are the result of not just descending cortical projections but also neural circuitry and dynamics in the SC as we will describe later.

One of such a response characteristic is the so called *multisensory enhancement* which describes an enhanced activity for multisensory input signals that is higher than the linear combination of all unisensory inputs (Stein and Stanford, 2008). Such multisensory enhancement changes with the intensities of the input signals and creates the commonly observed and described *inverse effectiveness* for multi-modal signals of MSI neurons (Perrault et al., 2003; Stein and Stanford, 2008). That is, low intensity multimodal stimuli in spatial and temporal register lead to an enhanced response of MSI neurons which is greater than the summed responses for separately presented unimodal stimuli (super-additivity). In contrast, for high intensity multimodal stimuli, responses tend to be smaller than the sum of unimodal responses (sub-additivity). As a consequence, the probability of detecting low intensity events registered by two or more senses is increased.

Another important response characteristic of MSI neurons is the suppression for bimodal signals outside the receptive field of the neuron (Meredith and Stein, 1996). That is, the otherwise strong activity of SC neurons is suppressed by input signals of another modality with spatial or temporal offsets (spatial and temporal principle of multisensory integration). This suppression leads to a sub-additive combination of the two stimuli and thus can be seen as a means to prevent fusion of input stimuli that do not belong to the same event.

All these response characteristics can only be observed for active descending cortical feedback from association areas to MSI neurons in the SC (Jiang et al., 2001, 2007; Jiang and Stein, 2003; Alvarado et al., 2007a, 2009). When cortical projections or corresponding cortical areas are deactivated, multimodal response characteristics vanish (Rowland et al., 2014; Yu et al., 2016). Thereby, AES cortical feedback projections mediate multisensory integration abilities in SC neurons.

The aim of such a complicated integration process is to infer a percept that is more reliable and robust than unimodal perceptions. In fact, studies pointed out that humans integrate signals from different modalities in a statistically optimal fashion (Ernst and Banks, 2002), so called *Bayes optimal*. That is, they weight each signal based on its reliability before linearly combining them. Thereby, increasing the certainty of the combined signal. In addition, it has been suggested that in order to integrate signals in such an optimal way, neural populations need to be able to encode and integrate sensory signals Bayes optimally (Deneve et al., 2001; Ma et al., 2006). Hence, one of the challenges in computational modeling of multisensory integration is to demonstrate that a model integrates its input signals in a Bayes optimal or at least near-optimal way and to explain how the variety of response characteristics can emerge from population dynamics.

We introduce a neural network model of multisensory integration of audio-visual signals that exhibits such near-optimal Bayesian behavior, incorporates cortical feedback and demonstrates typical multi-sensory response characteristics. The contribution of this work is several-fold: First, we introduce

a neural network model of conductance-based neurons in the superior colliculus that incorporates neurophysiological plausible cortical feedback connections. We investigate how this feedback alters the responses of multisensory neurons and enables them to integrate signals from multiple sensory streams. In addition, we examine what enables this process to integrate multimodal signals in a Bayesian optimal fashion and demonstrate that the introduced model does near-optimal Bayesian inference. This finding links the algorithmic mechanisms and representations to functionality. In a second part, we incorporate a spike-based output encoding of the model and deploy it on IBM TrueNorth neurosynaptic system (Cassidy et al., 2014), a neuromorphic processing chip with connections to neuromorphic sensory systems. Evaluations with this neuromorphic model demonstrate the performance for real world input data. This is a novel approach of testing a biological inspired architecture since it enables a complete neuromorphic processing cascade from sensory perception to multisensory integration and the evaluation of model performance for real world inputs.

2. MATERIALS AND METHODS

In this section we introduce the architecture of the neural network model which interactions and components are based on physiological findings. The first part describes a rate-based model implementation with first-order ordinary differential equations defining the change of a neuron's voltage based activation. In a second part we introduce a spiking neural network model implementation and describe how it is realized on the TrueNorth neurosynaptic chip.

The overall architecture of the model is inspired by the SC of mammals (**Figure 1**). That is, SC neuron populations (r) receive two modality specific excitatory inputs from a visual (S^v) and auditory (S^a) input population, respectively. Divisive inhibition of model neurons simulated by pool neurons (p^{pool}) and feedforward inhibition of the sensory inputs ensure a normalized level of activity. This is realized via a coincidence detection mechanism of the feedforward inhibitory neurons (p^{sen}). Thereby, these neurons provide inhibitory input to SC neurons only when bimodal sensory inputs are present but remains inactive for unimodal sensory inputs. Modulatory input from the cortex (green box CTX in **Figure 1**) facilitates multisensory integration abilities of the neurons by feedback top-down signals. We simulate this by a population of modulatory interneurons q^m. Interactions in the cortex (gray box in **Figure 1**) are modeled such that they produce such signals only when both modalities (C^v and C^a) receive inputs. This is achieved via a feedforward cross-inhibition in the feedback path of the cortical projections (gray box in **Figure 1**) between auditory and visual

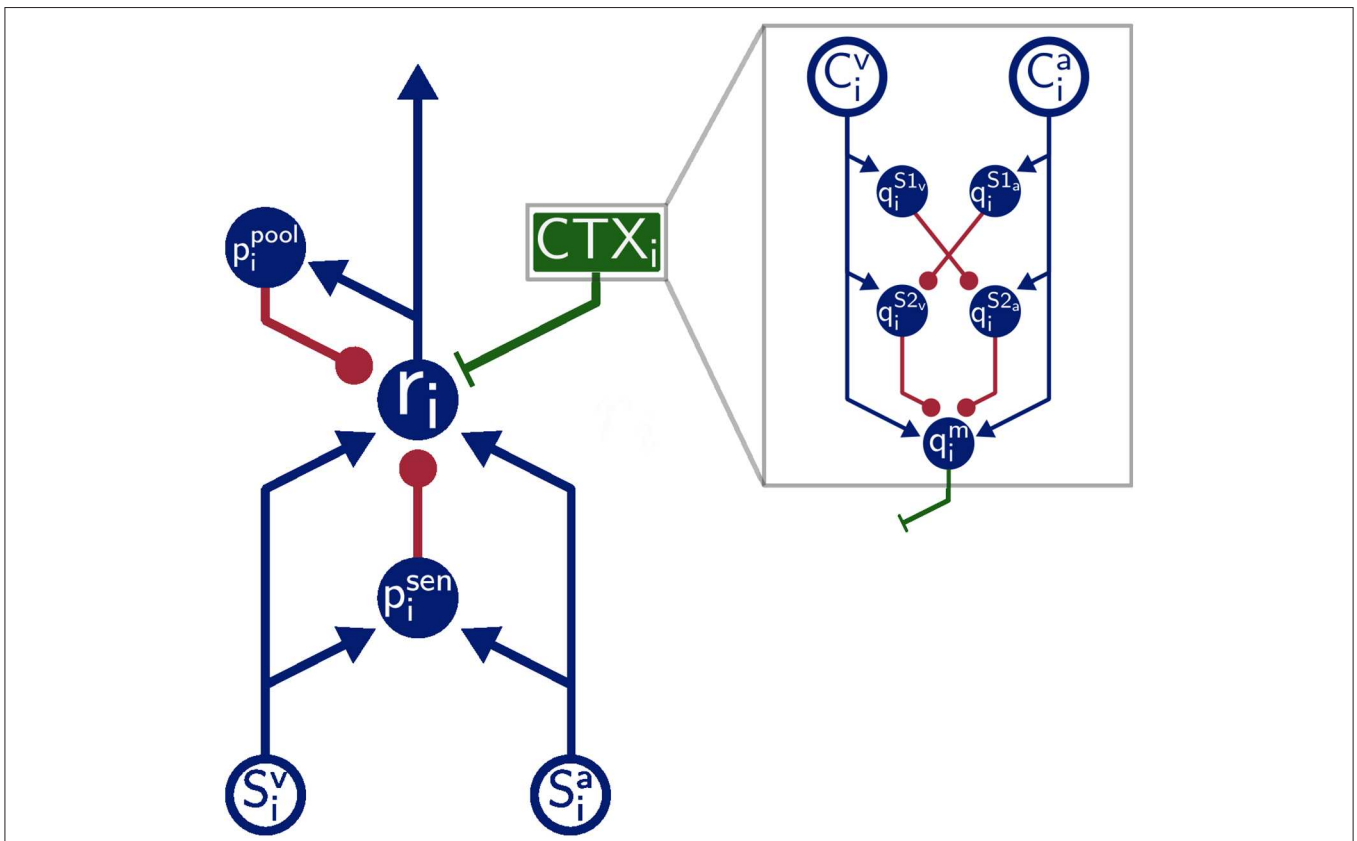

FIGURE 1 | Model Architecture. Overall architecture of the multisensory integration model. Blue lines indicate excitatory connections, green lines indicate modulatory connections, and red lines indicate inhibitory connections. Green box is modulatory cortical signal that is describe in more detail in gray box. Filled circles represent model neurons, hollow circles indicate inputs to the model. Letters indicate the name of neurons and inputs.

sensory areas (here q_a^{S1}, q_v^{S1}, q_a^{S2}, and q_v^{S2}). We will explain this in more detail in the next subsection.

2.1. Rate-Based Model

The rate-based model comprises several populations of neurons with dynamical interactions that together facilitate multisensory integration characteristics of SC neurons. Each of these populations comprises an array of $N = 20$ neurons, each one selective to a specific spatial location i in azimuthal direction (the center of the receptive field of the neuron). Such a count of neurons in a population is sufficient to achieve a satisfying resolution of the input space, which is arbitrarily chosen to be between 0 and 20 for the rate-based model. The number of neurons is not crucial and the model works well with larger neuron populations (>40) as well as smaller (<10) The membrane potential (as described in Equation 2) is governed by conductance-based integration of the neuron's excitatory and inhibitory inputs. Synapses are not modeled individually, but only represented in weight kernels which collect activities from pools of neurons. The inputs to such a neuron are described by S_i^a and S_i^v, for auditory and visual inputs, respectively. They represent the activity of tonotopical visual neurons in superficial layers of the SC and auditory neurons in the external nucleus of the inferior colliculus (ICx) (Oliver and Huerta, 1992) that have presumably spatially ordered connections to the SC (Hyde and Knudsen, 2000, 2002; Knudsen, 2002). All inputs are assumed to be spatially aligned so that for a combined event of audio and visual signals at location i, SC neuron at location i receives the most activity from inputs S_i^a and S_i^v. In addition, cortical activity of the AEv and FAEs is described with C_i^a and C_i^v and have the same activity characteristics as S_i^a and S_i^v, respectively.

Inputs are assumed to be of Gaussian shape with uncertainty σ_z and intensity I_z (where z is s_a sensory auditory input, s_v sensory visual input, c_a cortical auditory signal or c_v cortical visual signal). Thus, auditory and visual inputs at location i can be described by

$$y_i(x_t) = \exp\left(\frac{-(i - x_t)^2}{2 \cdot \sigma_z^2}\right) \cdot I_z, \tag{1}$$

where x_t is the location of a stimulus at time t.

The core of the model is a population of multisensory SC neurons that integrates excitatory, inhibitory and modulatory inputs. The change of membrane potential r_i of an SC neuron selective to location i is described by

$$\tau_d \dot{r}_i = -\alpha_d r_i + (\beta_d - r_i) \cdot EX_i \cdot (1 + \lambda \cdot MOD_i) - \kappa_r \cdot r_i \cdot INH_i, \tag{2}$$

where parameter τ_d defines membrane time constant, α_d is a passive membrane leakage rate, β_d describes a saturation level of excitatory inputs and κ_r defines the strength of divisive inhibition. Parameter values are given in **Table 1**. The parameter λ defines the influence of the modulatory input, thus the multisensory enhancement strength of the model neuron. The firing rate of an SC neuron is calculated by a sigmoidal activation function h of its membrane potential r_i

$$h(r_i) = \frac{2}{(1 + exp(-(r_i \cdot 3.4)^2))} - 1. \tag{3}$$

TABLE 1 | Model parameters.

General parameters			
N (# neurons)	20		
τ_d	1.0	α_d	1.0
β_d	1.0		
σ^m	3.0	σ	1.0
SC neuron			
κ_r	0.25	λ	0.4
l	3.6		
Modulatory neuron			
κ_m	1	γ_m	5.0
β_m	2.0		
S2 cortical neuron			
κ_{S2}	1.0	γ_{S2}	5.0

Excitatory inputs at location i are summarized in the term EX_i, modulatory inputs from cortical projections are described by the term MOD_i and inhibitory inputs are summarized in the term INH_i. We will describe each of these terms separately in the following.

The excitatory input to SC neurons directly arises from visual and auditory neurons in the outer layers of the SC and ICx, respectively with spatially aligned receptive fields

$$EX_i(t) = S_i^a(t) + S_i^v(t). \tag{4}$$

The modulatory input to SC neurons originates in the cortex and is defined by

$$MOD_i(t) = \sum_j \Lambda_{ij}^m \cdot g(q_j^m(t)), \tag{5}$$

where Λ_{ij}^m defines the interaction kernel of modulatory cortical projections to SC neurons and $g(q_j^M(t))$ defines the activity of model neuron $q_j^m(t)$ at location j and time t. We define this neuron later in this section.

The inhibitory input to SC neurons comprises a signal from feedforward inhibitory neurons of sensory inputs and a self-inhibition neuron that is fed by a pool of integration neurons. It is defined by

$$INH_i = \sum_j \Lambda_{ij} \cdot g(p_j^{pool}) + \sum_j \Lambda_{ij} \cdot g^{sen}(p_j^{sen}) \tag{6}$$

where $g(p_j^{pool})$ is the activity of a modeled self-inhibitory neuron p_j^{pool} and $g(p_j^{sen})$ the activity of an inhibitory neuron p_j^{sen} of feedforward inputs.

The feedforward inhibition of sensory inputs ensures similar intensity levels for unimodal and multimodal inputs by a

coincidence integration of both inputs. That is, the feedforward inhibition inhibits SC neurons for simultaneously active bimodal inputs but not for unimodal inputs. Each SC neuron has a feedforward inhibitory neuron driven by spatially aligned auditory and visual inputs. The activation of the membrane potential for such a neuron is defined by

$$\tau_{sen}\dot{p}_i^{sen} = -\alpha_{sen}p_i^{sen} + (\beta_d - p_i^{sen}) \cdot S_i^a \cdot S_i^v \tag{7}$$

Due to the multiplication of the neuron inputs, the feedforward inhibition neuron is active only if there are spatially aligned inputs of both modalities. Thus, the neuron behaves like a coincidence detector of its inputs.

A population of interneurons is modeled to realize pool normalization of SC neurons. A pool neuron is driven by neighboring SC neurons and feeds back on those SC neurons via inhibitory connections. The membrane potential of pool neurons is described by

$$\tau_d\dot{p}_i^{pool} = -\alpha_d p_i^{pool} + (\beta_d - p_i^{pool}) \cdot \sum_j \Lambda_{ij} \cdot h(r_j(t)). \tag{8}$$

Together with the feedforward inhibition, the pool inhibition of model SC neurons serves as a normalization mechanism to ensure a normalized energy level over different input intensity levels.

Modulatory inputs that facilitate multisensory response characteristics have their origin in the cortex, namely in the association areas AEv and FAEs.

The membrane potential of such neurons is modeled by

$$\begin{aligned}\tau_d\dot{q}_i^m = &-\alpha_d q_i^m + (\beta_m - q_i^m) \cdot (C_i^a + C_i^v) - (\gamma_m + \kappa_m \cdot q_i^m) \\ &\cdot \sum_j \Lambda_{ij} \cdot \left(g(q_j^{S2_v}) + g(q_j^{S2_a})\right),\end{aligned} \tag{9}$$

where parameter γ_m defines the subtractive influence of the inhibitory inputs $q_j^{S2_a}$ and $q_j^{S2_v}$ originating from a cross-modal inhibition circuit. This circuit (upper right part of gray box in **Figure 1**) ensures that only when both cortical inputs are present a modulatory signal is generated and fed back to SC neurons. If only one modality input is present the circuit is activated and generates strong inhibitory inputs for the q^m population resulting in a suppressed response of it. The circuit comprises four populations of neurons q^{S2_a}, q^{S2_v}, q^{S1_a}, and q^{S1_v} with connections as shown in **Figure 1** and membrane state equations for the auditory modality

$$\begin{aligned}\tau_d\dot{q}_i^{S2_a} &= -\alpha_d \cdot q_i^{S2_a} + (\beta_d - q_i^{S2_a}) \\ &\quad \cdot C_i^a - (\gamma_{S2} + q_i^{S2_a}) \cdot \sum_j \Lambda_{ij} \cdot g(q_j^{S1_v}), \\ \tau_d\dot{q}_i^{S1_a} &= -\alpha_d \cdot q_i^{S1_a} + (\beta_d - q_i^{S1_a}) \cdot C_i^a,\end{aligned} \tag{10}$$

and visual modality

$$\begin{aligned}\tau_d\dot{q}_i^{S2_v} &= -\alpha_d \cdot q_i^{S2_v} + (\beta_d - q_i^{S2_v}) \\ &\quad \cdot C_i^v - (\gamma_{S2} + q_i^{S2_v}) \cdot \sum_j \Lambda_{ij} \cdot g(q_j^{S1_a}), \\ \tau_d\dot{q}_i^{S1_v} &= -\alpha_d \cdot q_i^{S1_v} + (\beta_d - q_i^{S1_v}) \cdot C_i^v,\end{aligned} \tag{11}$$

Together these neurons synthesize the feedforward cross-modal inhibition circuit of cortical modulatory feedback.

The interaction kernel Λ between neuron populations is Gaussian and defined by :

$$\Lambda_{ij} = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left(-0.5 \cdot \left(\frac{i-j}{\sigma}\right)^2\right) \tag{12}$$

with $\sigma = \sigma^m = 3$ for Λ^m (the modulatory connection) and $\sigma = 1$ for Λ for all other connections.

The activation function $g()$ for all neurons in the model except for neuron r_i to generate a firing rate from its membrane potential is a linear rectified function with saturation level of 1

$$g(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x > 1, \\ x \cdot k, & \text{else,} \end{cases} \tag{13}$$

with $k = 2$ for p_{sen} input and $k = 1$ otherwise.

2.2. Spike-Based Model

We also implemented the proposed MSI model on the IBM TrueNorth Neurosynaptic System. TrueNorth is a highly efficient, spiking, neuromorphic hardware (Merolla et al., 2014) that provides a million neurons and 256 million synapses organized into 4096 cores (Cassidy et al., 2014). This platform has been demonstrated in numerous real-time applications ranging from speech recognition (Tsai et al., 2017) over probabilistic inference (Ahmed et al., 2016) to motion detection (Haessig et al., 2018).

When transferring any rate-based model to a spike-based architecture one must choose a representation of real valued rates. Common options range from the spike-rate of single or groups of neurons, population codes, order of spike times (Trappenberg, 2010; Kasabov, 2019) to the (inverse) time between spikes (Haessig et al., 2018). The proposed rate-based model needs to evaluate several products of variables (e.g., in Equation 2), so we choose spike-rate as representation of real valued activation. Here, multiplication is realized simply by a logical `AND` operation, each of which can be done by a single hardware neuron on the TrueNorth.

The implementation follows our approach in Löhr et al. (2020): Any model equation is split into elementary operations, each of which can be handled by a hardware neuron. Such operations range from sums and products to non-linear functions. Neuronal activation is encoded directly as spike-rate, such that only values in [0,1] can be represented. Any exceeding value will be clipped, thus we must ensure that operational regimes of all dynamics lie in this range. Neurons

with larger activation range are scaled down accordingly, and the weights of all their post-synaptic neurons are increased, respectively. Eventually, any differential equation in Equations (2, 7–11) is represented by a sub-graph of hardware neurons where the root neuron's activation follows said equation. The composition of these sub-graphs, forming the proposed model, is shown in **Figure 2**. The elementary functions to be performed by hardware neurons can be grouped into unary, binary and complex operations. *Unary* operations involve a constant and a variable, such as sum ($\oplus$) and difference ($\ominus$) or multiplication ($\odot$) by a factor. *Binary* operations involve two variables and examples are the weighted sum and product ($+$, $\bullet$). *Complex* operations subsume all remaining neurons: Convolution over channels is done by weighted sum neurons ($*$), sharing afferent axons on the same core. Root neurons evaluating ODEs ($\bullet$) and, finally, the sigmoidal (σ) activation function of Equation (3). The decomposition of equations into these elementary operations is shown in **Table 2**. For instructions on how to compute the TrueNorth neuron parameters of the above operations, please refer to the detailed Table 1 in Löhr et al. (2020).

As TrueNorth cores have a capacity of 256 neurons each, the proposed architecture must be split over several cores if more than six feature channels are to be used. At the same time, any hardware neuron's axon can only be routed to a single core. If its response is required on different cores, splitter neurons must be inserted which duplicate the neuron's response to provide additional axons. To keep the diagram simple in **Figure 2** these splitters are connected to the neurons they duplicate via dashed arrows. However, they actually share the exact same input connections and internal parameters as their originals and thus produce a perfect copy of their spike patterns.

Thus we divided the spike-based model into six functional blocks of similar neuron count, each assigned to a respective core. To reduce the amount of splitter neurons, no sub-graph of a root neuron was split over different cores. Likewise, neurons realizing a convolution were placed onto a common core, so they can share presynaptic axons. The final hardware implementation of the

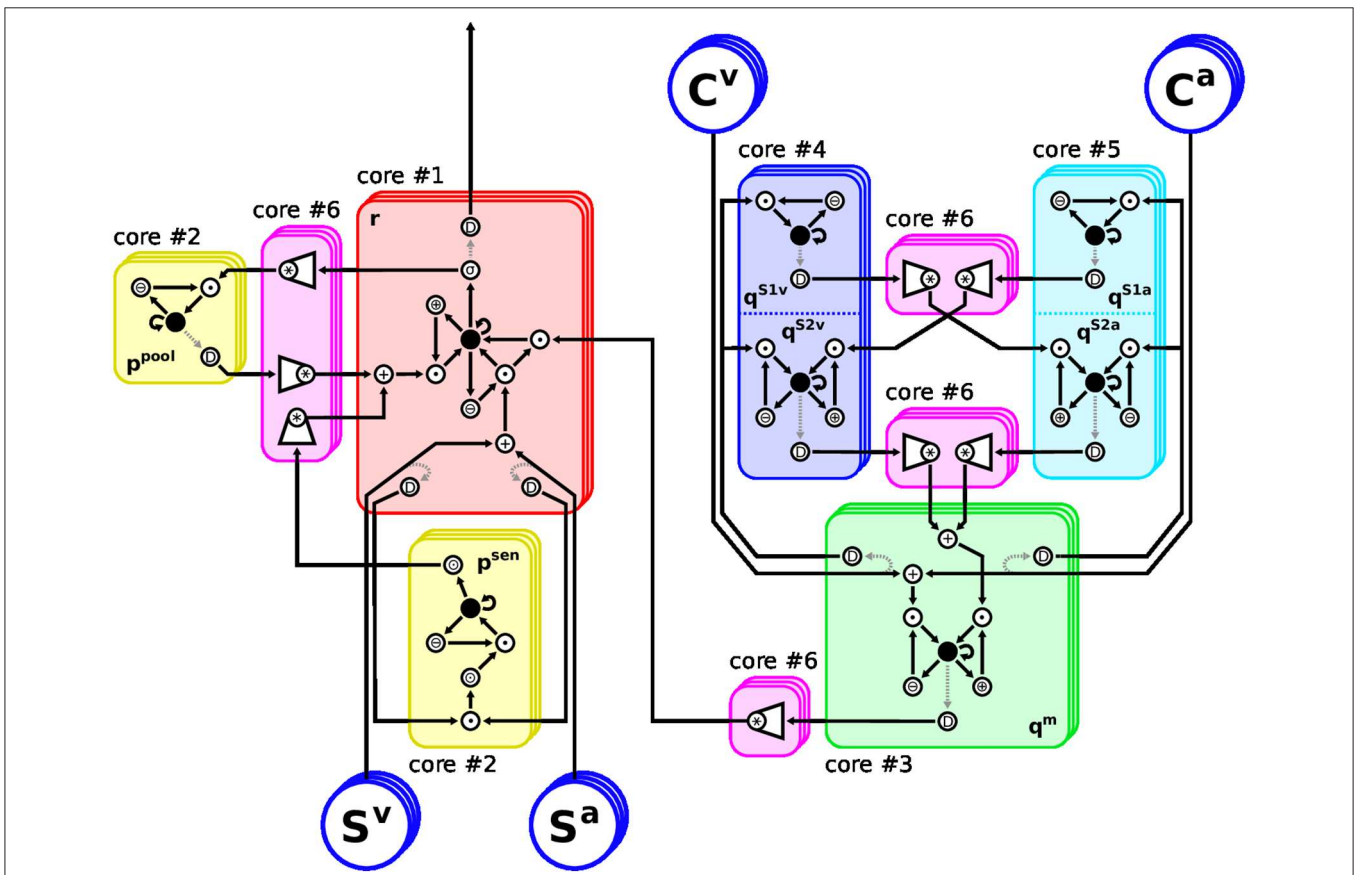

FIGURE 2 | Arrangement of the Model Architecture on the TrueNorth neurosynpatic chip. Restrictions of the hardware require careful placement of neurons onto cores: The eight differential equations (ODEs) in Equations (2, 7–11) must be split into elementary operations, because they cannot be evaluated by a single neuron each. Also, when any neuron is required to deliver spikes to different cores, a splitter neuron must be inserted to duplicate their axon. Therefore, each feature channel consists of 60 hardware neurons instead of eight rate-based ones. The chosen layout realizes up to 21 feature channels of the proposed MSI model using six of the 4096 cores of TrueNorth. Stacked frames indicate channels. Neuron: $\bullet$, root neurons evaluating ODEs; $*$, convolution over channels; $+/\bullet$, weighted sum/product of two variables; $\oplus/\ominus/\odot$, adding to/subtracting from/multiplying by a constant; D, splitter neurons; σ, sigmoidal transfer function. Rectified linear transfer functions are implemented using $\odot$ and clipping.

TABLE 2 | Elementary operations realized by hardware neurons.

Operation	Usage
$\oplus$(const + var)	Inhib. terms ($\gamma_m + \kappa_m \cdot q_i^m$), ($\gamma_{S2} + q_i^{S2_a}$), ($\gamma_{S2} + q_i^{S2_v}$) in Equations (9–11)
	Feedback term ($1 + \lambda \cdot MOD_i$) in Equation (2)
$\ominus$(const − var)	Conductance terms ($\beta_d - r_i$), ($\beta_m - q_i^m$) in Equations (2, 9)
	and ($\beta_d - p_i^{sen}$), ($\beta_d - p_i^{pool}$) in Equations (7, 8)
	and ($\beta_d - q_i^{S1_a}$), ($\beta_d - q_i^{S2_a}$), ($\beta_d - q_i^{S1_v}$), ($\beta_d - q_i^{S2_v}$) in Equations (10, 11)
$\odot$(const • var)	Transfer func. $g^{sen}()$ used in Equation (6) and scaling of $S_i^a \cdot S_i^v$ in Equation (7)
+(var + var)	Components EX_i and INH_i of neuron r in Equations (4, 6)
	Excit. ($C_i^a + C_i^v$) and inhib. ($g(q_j^{S2v}) + g(q_j^{S2a})$) inputs of q_m in Equation (9)
•(var • var)	Excitatory input $S_i^a \cdot S_i^v$ of p^{sen} in Equation (7)
	Products with conduct. terms $(\dots) \cdot EX_i$, $(\dots) \cdot (C_i^a + C_i^v)$ in Equations (2, 9)
	and $(\dots) \cdot (S_i^a + S_i^v)$, $(\dots) \cdot \sum_j \dots$ in Equations (7, 8)
	and $(\dots) \cdot C_i^a$, $(\dots) \cdot C_i^v$ in Equations (10, 11)
	Products with inhib. terms $(\dots) \cdot INH_i$, $(\dots) \cdot (\sum_j \dots)$ in Equations (2, 9)
	and $(\dots) \cdot \sum_j \dots$ in Equations (10, 11)
	Product with feedback term $EX_i \cdot (1 + \lambda \cdot MOD_i)$ in Equation (2)
*convolution	Weighted sums $\sum_j \Lambda_{ij} \dots$ in Equations (5, 6, 8–11)
●root neurons	ODE terms $\dot{r}_i = -\alpha_d r_i + (\dots) - (\dots)$, $\dot{q_i}^m = -\alpha_d q_i^m + (\dots) - (\dots)$
	and $\dot{p_i}^{sen} = -\alpha_{sen} p_i^{sen} + (\dots)$, $\dot{p_i}^{pool} = -\alpha_d p_i^{pool} + (\dots)$
	and $\dot{q_i}^{S1_a} = -\alpha_d q_i^{S1_a} + (\dots)$, $\dot{q_i}^{S2_a} = -\alpha_d q_i^{S2_a} + (\dots) - (\dots)$
	and $\dot{q_i}^{S1_v} = -\alpha_d q_i^{S1_v} + (\dots)$, $\dot{q_i}^{S2_v} = -\alpha_d q_i^{S2_v} + (\dots) - (\dots)$
σ sigmoid	$h(r_i)$ in Equation (3)

proposed MSI model consists of 60 neurons per feature channel. Restricted to six of the 4096 cores this scheme allows to synthesize the MSI model with up to 21 feature channels. If more cores are used, the amount of feature channels can readily be increased to 256; a limitation due to the convolution operation. Convolutions over larger numbers of axons would need to be split over multiple cores, however, this would require additional splitter neurons as some input axons would be needed on different cores.

3. RESULTS

Simulation results of the rate-based and spike-based model implementations for multisensory integration demonstrate characteristic multisensory integration properties and indicate near-optimal Bayesian inference of bimodal inputs.

In a first section, functional properties of the rate-based model and the multisensory integration are examined. Multisensory neurons are defined by their typical response behavior for bimodal and unimodal inputs (as described in section 1). Hence, in a first experiment we demonstrate the rate-base model's response behavior to multi- and unimodal inputs and investigate how inputs with spatial offsets are integrated. In particular, we test the response behavior for different stimulus intensities.

It has been shown that humans integrate signals from various modalities in an optimal fashion (Ernst and Banks, 2002). Where this integration exactly takes place is still under investigation. Model results demonstrate that already sub-cortical regions like the SC could integrate signals in a near optimal way when provided with a cortical control signal. Therefore, after presenting the characteristic response properties of multisensory integration neurons, we test the model's ability to integrate bimodal signals in a Bayes optimal fashion.

Having demonstrated that the rate-based model integrates multisensory signals near-optimally, in the second section we investigate the spike-based model implemented on IBM's neurosynaptic chip TrueNorth. In a first experiment, we reproduce the multisensory integration response characteristics to validate the correct functioning of the spike-based model. In a last experiment, we present real world data to the model recorded with neuromorphic hardware and evaluate its ability to integrate these signals.

3.1. Rate-Based Model Simulations

All experiments in this section are conducted with the rate-based model of multisensory integration. Simulation results in the following are computed from responses of model SC neurons after presenting the stimuli for 4,000 time steps. This duration is sufficient for each neuron in the neuron population to dynamically converge to its equilibrium membrane potential of numerical integration of the state equations. We chose *Euler's method* with a step size of $\delta t = 0.001$ for numerical integration. Model parameters for following simulations are chosen to fit a variety of neurophysiological experiments (Meredith and Stein, 1996; Stein and Stanford, 2008). In particular, we focused on the response characteristics of multisensory neurons, e.g., inverse effectiveness, spatial principle, and sampled the parameter space manually to achieve qualitatively similar results.

In all, we test 6 distinct stimulus conditions (see **Figure 3**): To demonstrate the importance of cortical feedback for multisensory integration, in the first condition (solid black line in following figures) all cortical inputs are absent ($C^a = 0, C^v = 0$), whereas both sensory inputs (S^a, S^v) are active according to Equation 1. The second condition (solid orange line) is the multisensory response for simultaneously active cortical and sensory inputs. In the third and fourth conditions (solid pink and purple line, respectively), both sensory inputs are present but only a single unimodal cortical input is given to demonstrate that both cortical unimodal signals are needed to facilitate multisensory integration. In the fifth and sixth conditions (solid blue and red lines, respectively), only unimodal sensory and unimodal cortical inputs (either visual or audio) are present as a control to show that the typical response characteristics emerge only for bimodal inputs which is in line with neurophysiological findings (Meredith and Stein, 1983; Meredith et al., 1992; Stein and Stanford, 2008).

3.1.1. Inverse Effectiveness

An essential property of responses of SC neurons to multimodal inputs is the *inverse effectiveness* of stimulus intensity. That is, weak multimodal inputs create strong multisensory enhancement, whereas strong multimodal inputs only produce weak or no multisensory enhancement. To examine whether our model exhibits this response property we present spatially

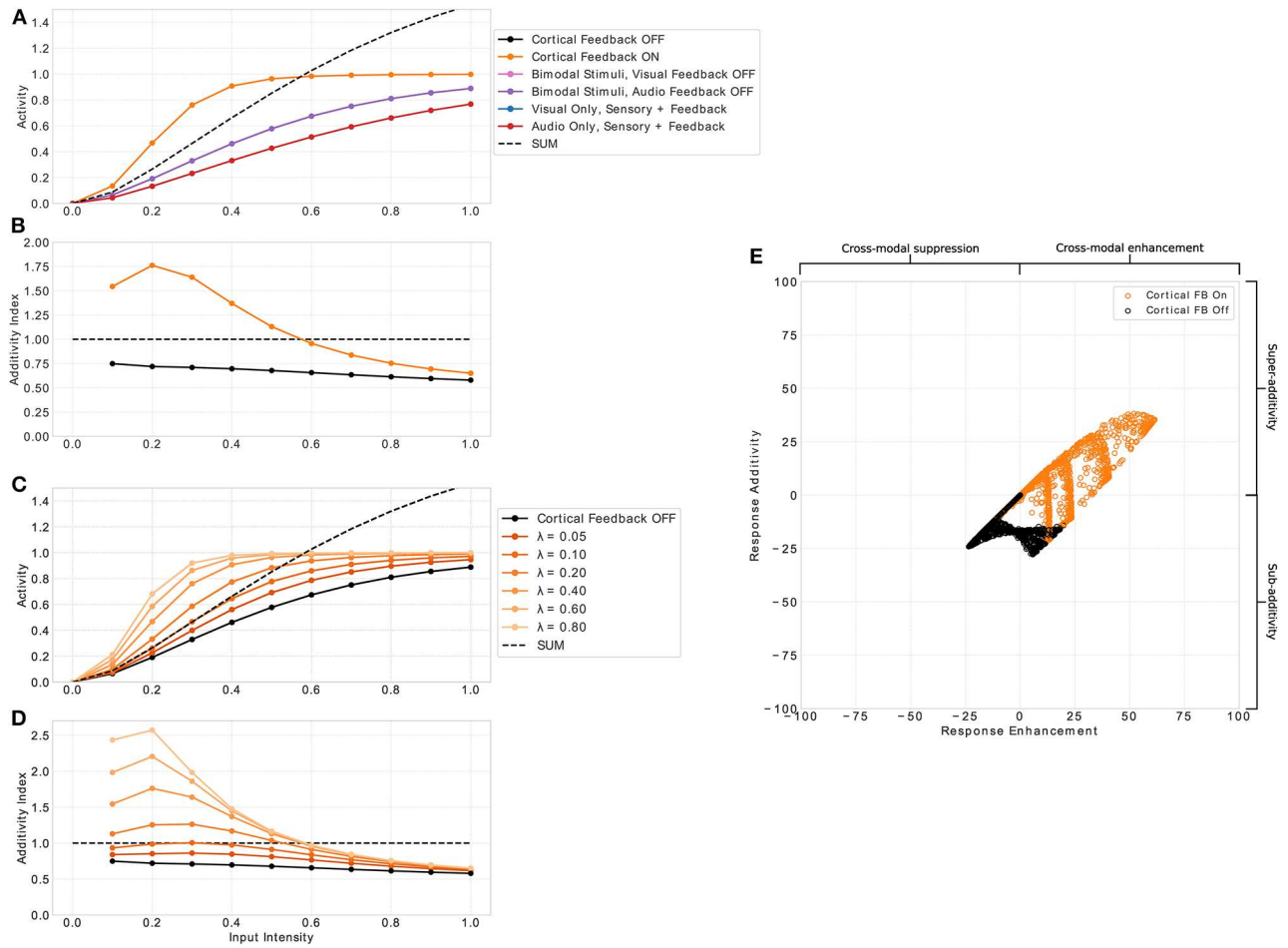

FIGURE 3 | Multisensory enhancement of MSI neurons. **(A)** Displays neuron activity over input intensities. Black and orange lines indicate presence of both input modalities for sensory and cortical inputs (bimodal response). Orange lines indicates that cortical feedback is active whereas black lines shows responses when the feedback is turned off. Pink and purple lines indicate both sensory inputs present with cortical visual and audio input off, respectively. Blue and red lines show unimodal inputs that is only visual and auditory sensory and cortical inputs, respectively. Black dashed line is the sum of the unimodal inputs (red and blue lines). **(B)** Displays the additivity index over input intensities. It is calculated by the bimodal response divided by the sum of the two unimodal response strengths. Orange and black lines are the same conditions as in **(A)**. Plot **(C)** shows the additivity index over input intensities for several λ parameter values (fading orange lines) and cortical feedback projections off (black). **(D)** Displays the additivity index for the different values of λ. Right panel **(E)** displays a summary of responses taken from model neuron $i = 8$ of the spatial principle experiment (**Figures 4**, **7**). The x-axis shows response enhancement of model neuron's responses. Positive values represent cross-modal enhancement whereas negative values indicate cross-modal suppression. The y-axis depicts the response additivity of the model neuron. Positive and negative values represent super- and sub-additivity, respectively. Orange dots indicate active cortical feedback projections. Black dots denote cortical feedback deactivated. Each dot corresponds to a certain input intensity, spatial offset value and randomly chosen value of σ_{S_a} and σ_{S_v} of input uncertainty in range [0.5, 5].

aligned auditory and visual inputs at location i to the model and measure the response of a representative neuron with receptive field centered at location i. Input intensities (I_z in Equation 1) are varied and a total of 11 intensities equally spaced in range $[0, 1]$ are tested. The responses of the neuron as a function over intensities are depicted in **Figure 3A**. For input intensities lower than 0.55 the multisensory response is stronger than the sum of the responses to unimodal inputs. In contrast, for input intensities higher than 0.55 the neuron response is weaker than the sum. This property only emerges for conditions where both cortical inputs are present. If one or both of them are not present, the response of the model neuron is constantly below the sum of responses to unimodal sensory stimulation (purple and pink line). Thus, no multisensory enhancement takes place. The neuron parameter λ in Equation (2) controls the effect of the cortical projections and thus directly influences the multisensory enhancement of the neuron (see **Figures 3C,D**). This effect can be quantified by the *additivity index* which is defined as the ratio of the bimodal response to the sum of responses for unimodal inputs ($\frac{M}{V+A}$, where M is the multisensory response for active cortial projections, V unimodal visual response and A unimodal auditory response). An additivity index of 1 means the response for bimodal inputs is exactly as strong as the sum of both unimodal responses (see Meredith and Clemo, 1989 for details). Index values above 1 indicate super-additivity whereas index values below 1 indicate sub-additivity. The model neuron exhibits super-additivity for low input intensities and sub-additivity for high input intensities (see **Figure 3B**). Thereby, it exhibits inverse effectiveness response characteristics of multisensory neurons.

3.1.2. Spatial Principle

The spatial principle of multisensory integration is commonly described by the inhibition of a stimulus in one modality by a stimulus of the other modality outside the receptive field of the neuron. Such a spatially separated stimulus combination not just leads to a reduction in the multisensory enhancement of the neuron but even to a suppression of its response. This suppression is usually ascribed to the Mexican hat shape of the receptive field. Our network model shows similar properties that emerge merely from the network dynamics (see **Figure 4**). The suppression of responses for spatially separated bimodal stimuli is facilitated by the feedforward inhibition of inputs in the network. In particular, the coupling of the absence of a spatial convolution kernel for excitatory inputs to SC neurons and the presence of such a spatial convolution for feedforward inhibitory inputs lead to a reduced activity of SC neurons outside the receptive field. The inhibition imposed by the spatially offset unimodal input still effects neighboring neurons whereas the direct excitatory does not. For bimodal stimuli with no spatial offset the network response is equal to the one shown in the previous experiment. However, for increasing spatial offset values (measured in σ of input Gaussian) the multisensory enhancement effect decreases (decreasing additivity index). For a stimulus with 3σ offset, the multisensory response is suppressed and lower than the unimodal response. This can also be seen in the additivity index curve that is below 1 at this offset. For offset values larger 3σ the suppression vanishes and the multisensory response is equal to the unimodal response, thus having an additivity index of 1.

3.1.3. Interactions Among Within-Modality Inputs

Neurophysiological studies have shown that multisensory neurons exhibit multisensory enhancement and inverse effectiveness only for bimodal inputs but not for multiple unimodal inputs (Alvarado et al., 2007b). In addition, the authors demonstrated in another study that the cortical projections only modulate multisensory but not unisensory

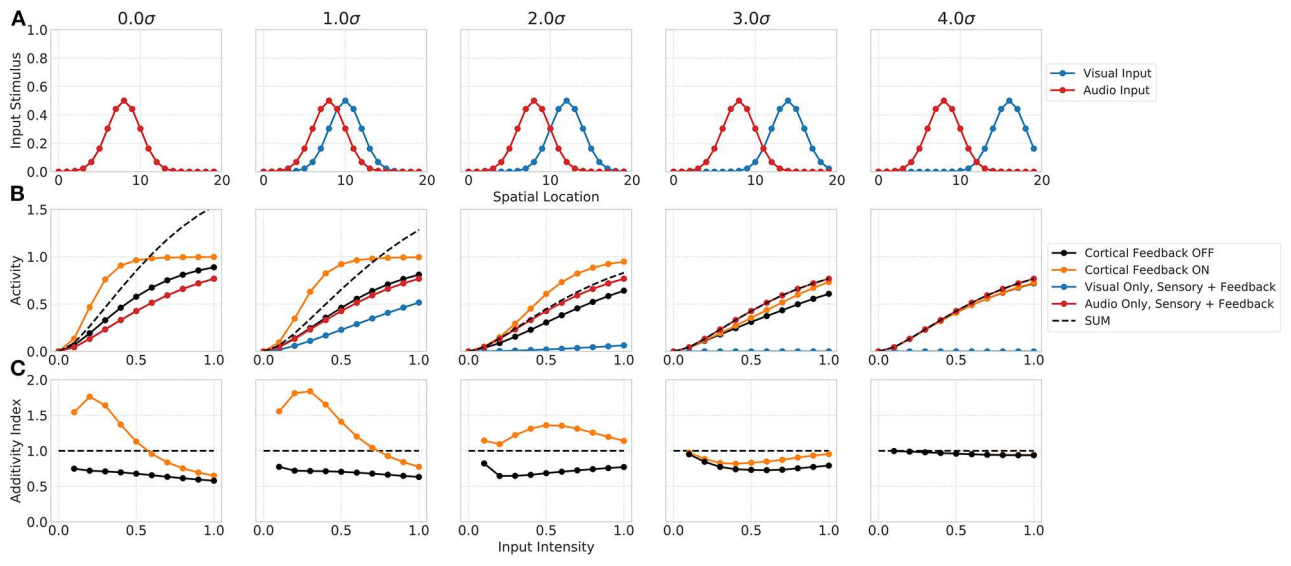

FIGURE 4 | Spatial principle of MSI neurons. Upper row **(A)** displays the audio (red) and visual (blue) inputs to the model over spatial locations with increasing offset values. Middle row **(B)** shows the responses of rate-based model neuron $i = 8$ over input intensities for corresponding audio and visual inputs as given in the first row. Color code of lines is the same as in **Figure 3**. Bottom row **(C)** depicts the additivity index of the same model neuron $i = 8$ over input intensities.

integration (Alvarado et al., 2007a). Thus, in a third experiment we investigate model responses for two unimodal (auditory) inputs. In this simulation, we assume model input S^v to represent a second auditory stimulus (Audio Input B) and the visual cortical input C^v to be 0. We simulate two auditory inputs with activated auditory cortical projections but without visual cortical activity. Model responses for a combination of within-modal stimuli are higher than for a single unimodal input but show almost no super-additivity, expect for low intensity inputs (**Figure 5**). When one of the stimulus is moved outside the receptive field of the neuron, the combined response activity becomes lower than the response for a single unimodal input. This can be ascribed to a within-modal suppression (Alvarado et al., 2007b).

3.1.4. Cortical Modulatory Projections

Cortical projections from association areas FAEs and AEv to the SC seem to have a crucial role in the multisensory integration behavior of MSI neurons and their response properties. For example, it has been shown that when these connections are removed or the corresponding cortical areas are deactivated, multisensory response properties in the SC vanish (Jiang et al., 2001; Alvarado et al., 2007a, 2009). We model these connections with modulatory input to the SC neuron that originates from a cross-modal inhibition circuit located in the cortex. That circuit ensures that only when a cross-modal stimulus is present the SC neuron receives modulatory inputs. Without such an input multisensory enhancement of SC neurons vanishes (see **Figure 3B** black line) and their response is very similar to responses for unisensory inputs. To investigate the role of cortical inputs in more detail we calculate the response enhancement $(M - max(V, A))/(M + max(V, A))$ that defines cross-modal enhancement and suppression for positive and negative values, respectively, and the response additivity $(M - (V + A))/(M + (V + A)) \cdot 100$ that indicates super-additivity and sub-additivity for positive and negative values, respectively (see Avillac et al., 2007 for details). Only if the cortical inputs are active, multimodal enhancement and super-additivity can be observed (see **Figure 3E**).

3.1.5. Multisensory Inference

Several studies have shown that when multimodal sensory cues are simultaneously available, humans integrate these cues based on the reliability of each cue (Ernst and Banks, 2002). Thereby, human observers perform a weighted linear combination of cues from different sensory perceptions to maximize the certainty of the fused signal. The weight associated to a cue is proportional to the relative reliability of the perception of the corresponding cue. For example, estimating the size of an object by a combination visual and haptic sensory perceptions is usually based on the visual input. However, once visual input is blurred, thus the reliability of the perception is decreased, humans rely more on their haptic estimate (Ernst and Banks, 2002). By taking the reliability of the sensory perception as a weight in the

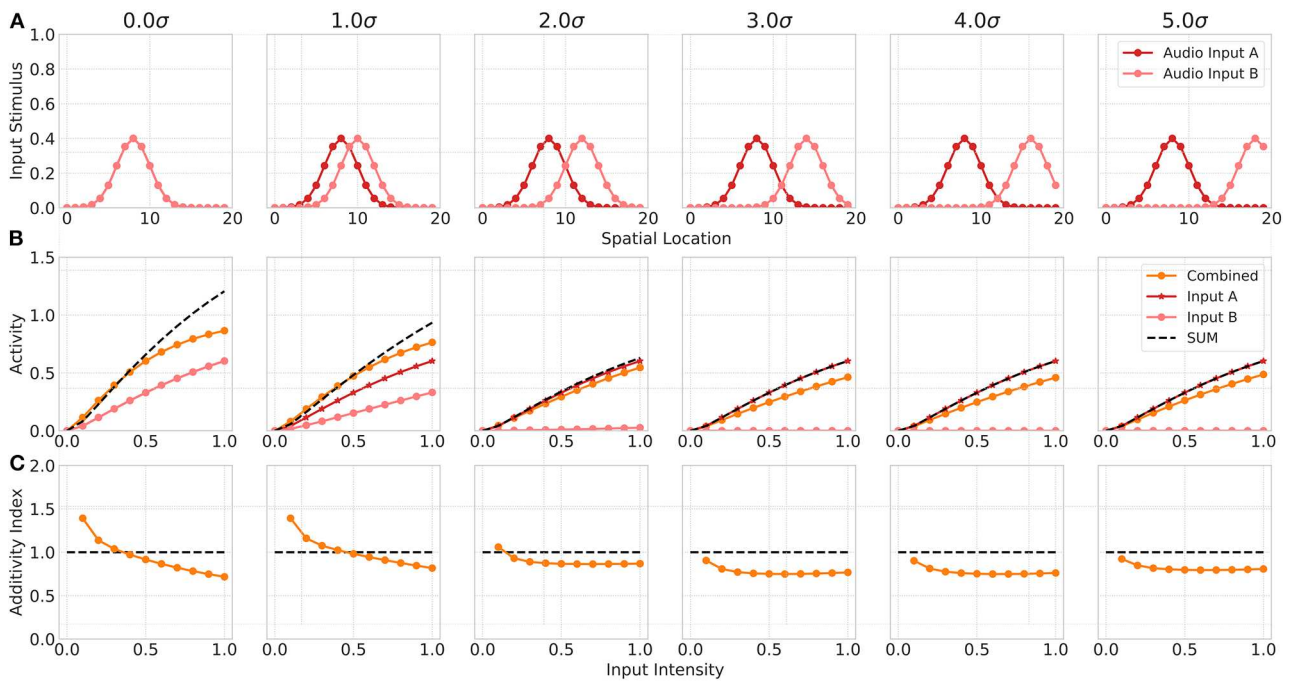

FIGURE 5 | Within modality responses. Upper row **(A)** displays the audio A (red) and audio B (light red) inputs to the model over spatial locations with increasing offset values. Middle row **(B)** shows the responses of model neuron $i = 8$ over input intensities for corresponding unimodal inputs as given in the first row. Orange lines indicates combined response. Red line shows the response for audio input A, light red line for audio input B. Bottom row **(C)** depicts the additivity index of the same model neuron $i = 8$ over input intensities.

integration process, humans perform optimal Bayesian inference for multisensory stimuli.

On which level in the processing hierarchy this inference takes place is not fully understood yet. Some researches argue that it is a rather high level cognitive process located in and between cortical areas (Kayser and Shams, 2015; Rohe and Noppeney, 2015). However, model simulations and neurophysiological recordings indicate that already on a level of two neural populations, Bayesian inference can take place (Ma and Pouget, 2008; Beck et al., 2012; Pouget et al., 2013). In our model we assume a combination of low level subcortical dynamics that provide a basis for cue integration together with high level cortical integration processes that facilitate subcortical near optimal multisensory integration.

To demonstrate that this network structure allows our model to perform for near-optimal Bayesian inference we conduct a multisensory integration simulation experiment of auditory and visual inputs. For that, we define for each modality a normal distribution of stimuli location with a specific mean (auditory stimulus distribution's mean: $\mu_a = [8, 14.5]$, visual stimulus distribution's mean: $\mu_v = [5, 8]$) and variance (auditory stimulus distribution's variance: $\sigma_a = [0.5, 3.0]$, visual stimulus distribution's mean: $\sigma_a = [0.5, 3.0]$) and draw a sample location of a visual and auditory stimulus, respectively, from these distributions (**Figure 6A**). This location of audio and visual stimuli is then applied as input to the model, i.e., defines location i for a visual and auditory stimulus independently (**Figure 6B**). We independently draw a visual and auditory stimulus, respectively, 200 times from two distributions that have different mean and variance values. For each draw we present the two stimuli to the model, compute its responses and calculate the maximum of this response. We use only the maximum of the distribution to model a maximum-likelihood approach of choosing the stimulus location which has been previously observed in humans (Ernst and Banks, 2002). Taking together all maximum values over draws in a histogram results in a posterior distribution with a mean and variance value of model responses for a given stimuli set (see **Figures 6C,E**). The real posterior (see **Figure 6D**) of the combination of the two given distributions can be determined analytically with mean $\mu_{CP} = \mu_a \cdot w_a + \mu_v \cdot w_v$, with $w_a = \frac{\frac{1}{\sigma_a^2}}{\frac{1}{\sigma_a^2}+\frac{1}{\sigma_v^2}}$ and $w_v = \frac{\frac{1}{\sigma_v^2}}{\frac{1}{\sigma_a^2}+\frac{1}{\sigma_v^2}}$, as well as variance $\sigma_{CP} = \frac{\sigma_v^2 \cdot \sigma_a^2}{\sigma_v^2+\sigma_a^2}$.

We show the model's inference capability by comparing it's response with calculated mean and variance of the stimulus input (see **Figures 6G,H**). Two conditions are tested, cortical feedback on and off. For a fair comparison only model responses that show a single peak (fused responses) are considered. Model and analytical calculated mean are similar under both conditions. However, the cortical feedback allows for a more precise computation of the inferred variance. The variances of

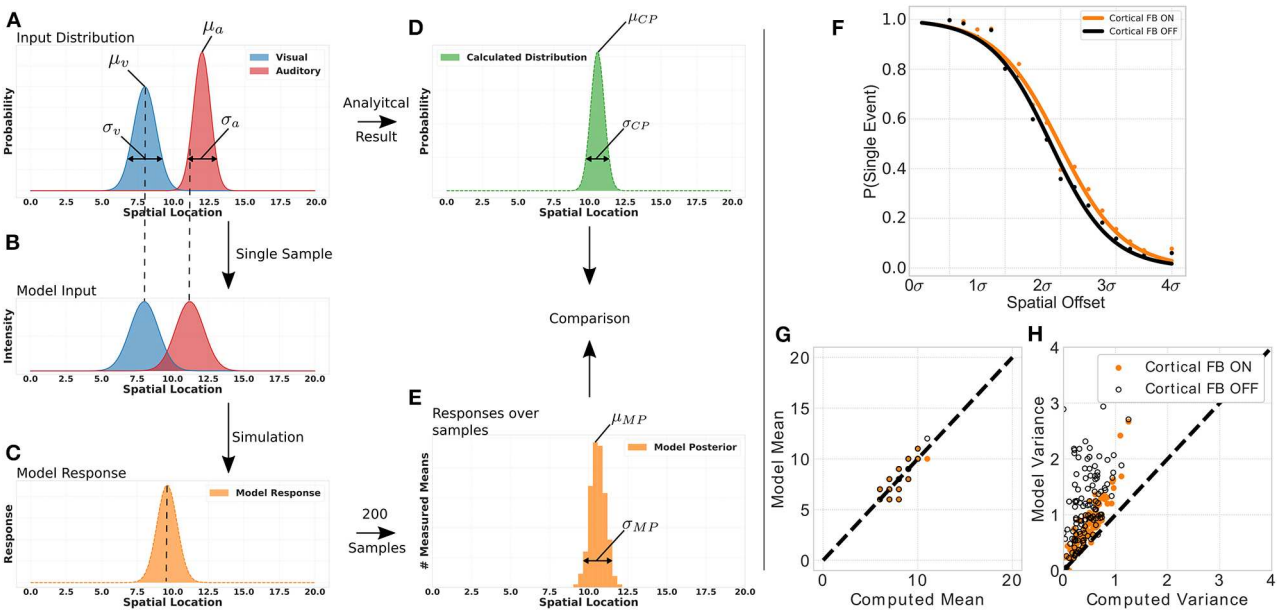

FIGURE 6 | Bayesian inference experiment. Plot **(A)** displays the stimuli distributions for visual (blue) and audio (red) inputs, respectively, from which stimuli are sampled. Plot **(B)** represents an example of a single sample with stimuli for audio and visual inputs to the MSI neuron. Plot **(C)** depicts the model response for this single sample. The histogram in plot **(D)** is calculated analytically from the input distributions in plot **(A)**. The histogram in plot **(E)** is calculated from the model response over all samples. The probability of event fusion (single peak in model response) is given over spatial offset values in **(F)**. Dots represent mean values over several simulations with cortical feedback on (orange dots) and cortical feedback off (black dots). Sigmoid functions are fitted to the dots for better visualization. Active cortical feedback enables higher probability of event fusion over spatial offsets compared to inactive cortical feedback. Plot **(G)** compares the means of model responses vs. the analytically calculated means of the samples drawn as described in left panel. **(H)** Displays model variance values vs. computed variance values. Orange dots indicate active cortical feedback whereas black dots indicate cortical feedback off. Dashed line indicates perfect model behavior.

model responses without cortical feedback has a higher offset and increases dramatically for large input variances. In contrast, with cortical feedback this increment is smaller and there is almost no offset.

The nervous system can combine two events from different modalities to form a combined single percept of the event. Under what conditions such a *mandatory cue fusion* takes place has been investigated thoroughly (Hillis et al., 2002). One crucial factor for event fusion is the spatial discrepancy between the different modalities. If an auditory event origins from the same or similar location as a visual event, it is likely to be fused to a single event. However, when the spatial offset between the two events increases the likelihood of perceiving a single event decreases (Andersen et al., 2004; Stevenson and Wallace, 2013). We investigate this behavior for our model by calculating the percentage of samples where a single event (one peak in response, mandatory fusion) has been detected in contrast to samples where two events (multiple peaks in response, no fusion) are present (see **Figure 6F**). The probability of event fusion is constantly higher for activate cortical projections than without.

3.2. Spike-Based Model Simulations

Simulation experiments in this section are conducted with the spike-based model implementation on the TrueNorth neurosynaptic chip. Since implementation of the model on this chip is fundamentally more complex than the rate-based variant, we first demonstrate similar behavior of the two implementations by presenting the same stimulus set of increasing input intensities as for the rate-based model in section 3.1.1. Typical response characteristics of MSI model neurons can be observed for the spiking model implementation (**Figure 7**). For low intensity inputs the bimodal response is greater than the sum of the two unimodal inputs (super-additivity). For increasing intensities this enhancement is reduced until the bimodal response is lower than the sum of the two bimodal inputs (sub-additivity). Thereby, the spike-based model demonstrates inverse effectiveness of MSI neurons.

We conduct the following simulation experiment to demonstrate the model's ability to cope with real world event-based sensory input data.

Inputs to the model are generated by a neuromorphic vision sensor (DVS) (Lichtsteiner et al., 2006) and an artificial neural implementation of functions of a cochlea. Sounds are recorded from 19 locations equally spaced in azimuthal range $[-90, 90]^\circ$ via two microphones placed inside the right and left ear canal of human-like shaped ears on a dummy head as depicted in **Figure 8A**. The device can turn around its longitudinal axis to create a relative displacement of the sound source location in the horizontal plane. The distance to the speaker (standard speaker) remains constant (1m) during movement (**Figure 8B**). We choose the sound of a vacuum cleaner for real world recordings. The presented sound type is a monaurally recorded sound of a vacuum cleaner. This sound was presented for azimuthal head directions of $[-90, -70, -50, -30, -10, 10, 30, 50, 70, 90]^\circ$ and 0° elevation. It was played back from the speaker and recorded in stereo with the two in-ear microphones for the duration of the sound. All recordings were done in a sound attenuated room. Subsequently, a bank of gammatone-filters is separately applied to these recordings of the right and left ear, thus creating spectrograms with 64 frequency channels resembling the output of the cochlea. Each spectro-temporal bin in the spectrogram is converted to a spike train and fed to a spiking neural network model of the lateral superior olivary (LSO) complex for computation of interaural level differences (ILD). Output of LSO model neurons is averaged over frequency bands. The weight channel of the left hemispherical output with maximum activity is subtracted from the weight channel of the right hemispherical output with maximum activity. This leads to a combined response of the left and right hemisphere over the entire range of perceived ILD values. Subsequently, this signal is converted to a one-dimensional estimate of spatial activations of sound sources by a set of 19 radial basis functions (RBFs). These functions are tuned to a specific response rate of LSO neurons, thus encoding a unique spatial location in range $[-90^\circ, +90^\circ]$ in 10° steps from the rate of LSO neurons (see Oess et al., 2020b for a detailed description of this preprocessing).

Videos are recorded of a stationary tea cup placed at evenly spaced positions (range $[-27, +27]cm$ in $3cm$ steps) in front

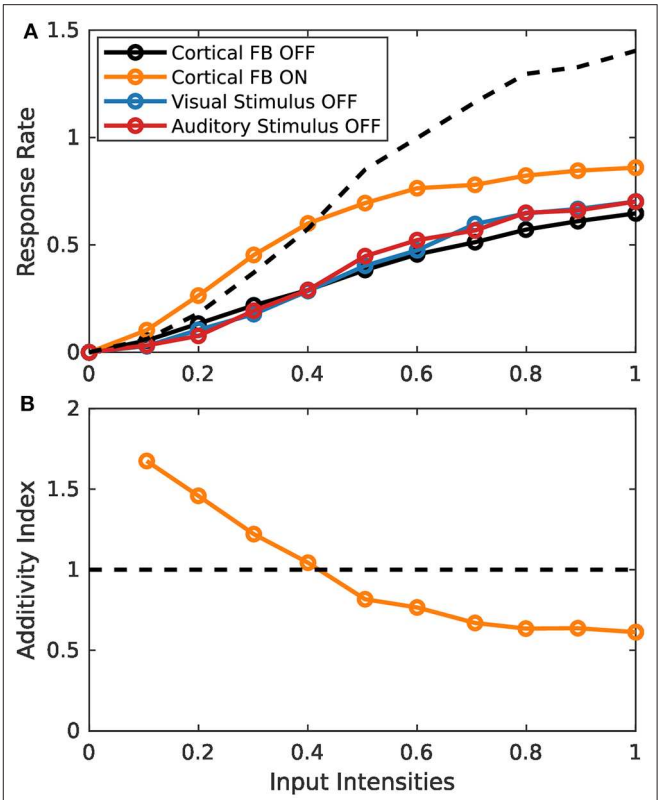

FIGURE 7 | Inverse effectiveness of spiking MSI neuron implementation. **(A)** Displays spike-based neuron activity over input intensities. Black and orange lines indicate presence of both input modalities for sensory and cortical inputs (bimodal response). Color code as in **Figure 3**. **(B)** Displays the activity index over input intensities. It is calculated by the bimodal response divided by the sum of the two unimodal response strengths. Orange and black lines are the same conditions as in **(A)**.

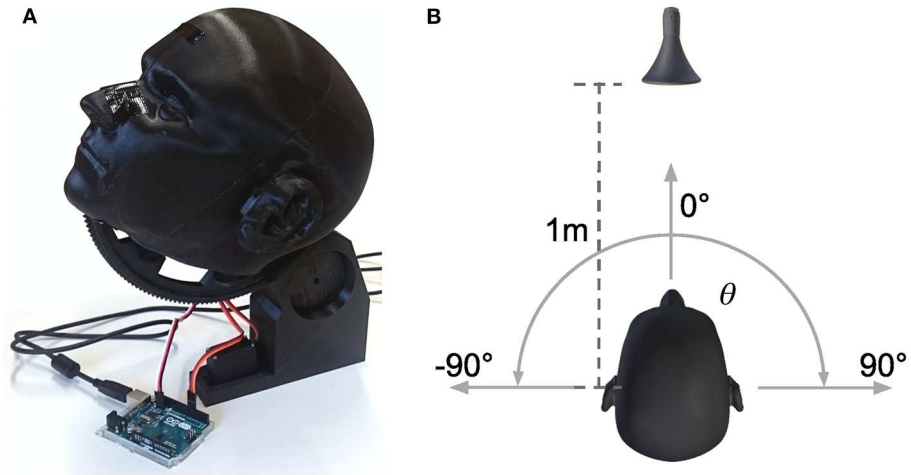

FIGURE 8 | Device and setup for real world sound recordings. **(A)** Shows the 3D printed, anatomical correct replication of a human head and ears. In the ears, ∼2 cm inside the ear canal, standard microphones are placed for recording. The head is controlled by two servo motors, one for turning around the longitudinal axis (azimuth), the other for turning around the frontal axis (elevation). All sounds are recorded for 0° elevation (head facing the speaker) and different azimuth directions. **(B)** Shows the setup for the recordings. The head is centered 1*m* in front of a standard speaker and can rotate around its longitudinal axis from $\theta = -90°$ to $\theta = 90°$ with a minimal step increment of 1°.

of the camera (distance 177*cm*). The cup's positions in all videos combined span the complete horizontal field of view of the camera. To ensure that stationary contrasts are detected the mirror setup of Löhr and Neumann (2018) is used which adds random tremor to the DVS's optical axis. A Gaussian subsampling scheme reduces the visual input of 128 × 128 pixels to 1 × 19 neurons, which relate to azimuthal direction. These neurons resemble those in superficial layers of superior colliculus and directly serve as visual sensory inputs to the MSI model.

Visual and auditory real world stimuli with increasing spatial offsets are presented to the model as sensory inputs. Audio and visual cortical inputs are simulated as described in Equation (1) and follow their corresponding sensory counterpart over spatial offsets. That is, their mean is set to the location of the maximum sensory response of the real world input. For small spatial offsets the two stimuli are integrated and a single peak in the model response is present (**Figure 9A**). We take this as an indication that multisensory fusion takes place and a single event is perceived. For increasing spatial offsets (>12 cm) the model response without feedback shows two separate peaks. This indicates that no integration takes place anymore and two separate events are perceived. The offset value for which this change of perception takes place changes with active cortical feedback projections (>18 cm). This demonstrates that cortical feedback facilitates larger offsets for which two stimuli are fused to a single percept. For an offset of >30 cm multisensory enhancement vanishes and responses for activated and deactivated cortical feedback projections are similar.

To test whether this reduction in multisensory enhancement is due to the spatial offset of sensory inputs or cortical inputs, the location of the cortical projections is fixed at the location of the auditory input (−9 cm) over all offset values. Thereby, only sensory inputs at this location receive modulatory cortical feedback (**Figure 9C**). The visual stimulus is shifted away from the auditory stimulus which leads to the perception of two different events for offset values larger than 12 cm (cortical feedback projections inactive) or 18 cm (cortical feedback projections active). This is similar to responses in **Figure 9B**. However, multisensory enhancement is maintained even when the visual input is shifted further away from this location. For offset value of 30 cm multisensory enhancement is still present at the location of the auditory stimulus whereas for the visual stimulus such an enhancement does not take place, as can be seen in **Figure 9D**.

4. DISCUSSION

We introduced two implementations of a neural model simulating functions of SC neurons for integration of audio-visual signals. The model incorporates modulatory cortical feedback connections to facilitate enhancement of multisensory signals. The rate-based implementation of the model and its responses were evaluated in various simulation experiments and we demonstrated the importance of cortical feedback projections for near-optimal integration of signals. Furthermore, the spike-based model implementation on neuromorphic hardware showed its capability of integrating real world spike inputs from neuromorphic sensors.

4.1. Multisensory Integration

Typical multisensory neurons show response enhancement for multimodal stimuli that arrive in temporal and spatial coincidence (Meredith and Stein, 1996). Previous studies report that this property only arise for enabled cortical feedback

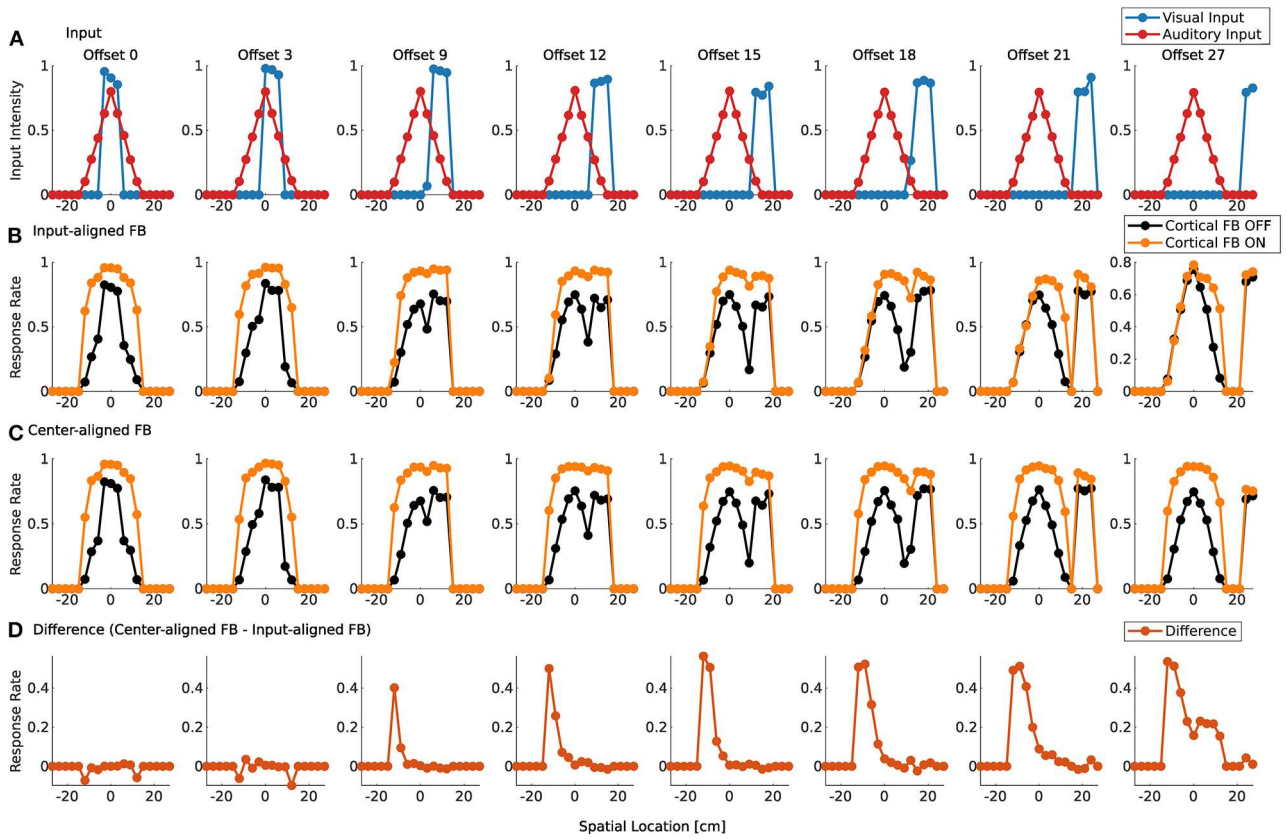

FIGURE 9 | Spatial fusion for spiking MSI neurons. Upper row **(A)** displays the audio (red) and visual (blue) real world inputs to the model over spatial locations with increasing offset values. The inputs originate from LSO model and camera recordings. Offset is given in evenly spaced displacement of stimuli in cm. Row **(B)** shows multisensory response rates of spike-based model neurons over spatial locations for corresponding audio and visual inputs as given in the first row. Orange lines indicate cortical feedback. Black lines indicate cortical feedback off. Spatial locations are given in cm. Note that the offsets are not equally distributed in the figure due to spatial constraints. We choose to display only offsets for responses that are substantially different. Row **(C)** shows multisensory response rates of spike-based model neurons as in row **(B)**. However, in contrast to **(B)**, visual and auditory cortical feedback locations are fixed to the location of the auditory sensory stimulus (0 cm). The lower row **(D)** displays the difference between response rates in **(B,C)** for active cortical feedback projections.

projections (Stein et al., 1983; Wallace and Stein, 1994; Jiang et al., 2001; Alvarado et al., 2007b). Our model results replicate such observations and show that response enhancement can vary with the gain of the modulatory cortical feedback projections controlled by λ parameter (Equation 2) in the model (see **Figures 3D,E**). Thus, the gain of how neurons integrate modulatory feedback could explain the observed variety of multisensory enhancement in responses of SC neurons as has been observed previously (Kadunce et al., 2001). Without cortical projections the response to multisensory input remains sub-additive (see **Figure 3B**) even for high input intensities. Such cortical projections are only activated when both modality specific cortical signals are active. If only one cortical region is active multisensory response properties vanish. This is in line with findings in cats, where multisensory integration disappears for deactivated cortical areas (Meredith and Clemo, 1989; Alvarado et al., 2009). In our model, this is achieved with a specially designed cortical cross-modal forward inhibition circuit in the feedback projections (see Equation 9).

Furthermore, the model follows the previously described spatial principle of MSI neurons (Meredith and Stein, 1996) (see section 1 for definition) by suppressing responses for bimodal stimuli with large spatial offsets. We would like to point out that this suppression is achieve merely by dynamic interactions between the pool normalization, the feedforward inhibition and excitation of sensory neurons, thus implicitly creating a center surround receptive field of MSI neurons.

4.1.1. Bayesian Inference

Several investigations show that afferent connections from cortical regions to the SC are necessary for multisensory integration (Alvarado et al., 2007a, 2008, 2009). However, the functional purpose of such feedback projections is still unclear. Our Simulation experiments show that multisensory integration of two input signals in a near-optimal Bayesian way appears only when cortical feedback projections are active. The variance of the integrated signal is substantially similar to the computed, optimal value for active projections than compared to responses without

these projections (see **Figure 6G**). This is especially true for larger spatial offsets of the two input stimuli (see **Figure 6F**). Thus, we hypothesize that one purpose of cortico-collicular feedback is to facilitate optimal integration of multimodal signals and that such an optimal integration might already happen on the level of the SC. Presented model variance values (**Figure 6H**) exhibit an offset for higher variance values which we assume to result from the static size of the receptive field of model neurons and could be compensated with a dynamically changing receptive field depending on sensory certainty.

4.2. Neuromorphic Implementation

We demonstrated that the proposed model architecture is suitable for robotic applications by implementing it on a real-time neuromorphic processing chip. Preliminary results for real world spike recordings obtained by neuromorphic sensory hardware suggest that the model is robust and capable of integrating real world multisensory signals. It was shown that the model's ability to fuse two modalities into a single percept changes with cortical feedback projections. This supports the hypothesis that cortex plays a crucial role in determining whether two stimuli belong to the same event or if they represent two separate events. This is further investigated in a last experiment in which the cortical feedback signals are fixed to the location of the auditory sensory input while the visual sensory input is spatially shifted. The response enhancement remains at the auditory location even if the sensory visual input is not present anymore. This can be interpreted as an increased cortical focus for this specific location. Thus, cortical projections might be controlling the mandatory fusion range of multisensory neurons and in addition serve as a spatial attention signal, as has been suggested by McDonald et al. (2001), Mozolic et al. (2008), and Talsma et al. (2010).

In future experiments, we are planning to implement such a spatial attention mechanism in order to selectively choose which multisensory signals should be enhanced. We believe that this could be accomplished by a more sophisticated cortical feedback signal with spatial properties different than the perceived sensory inputs.

4.3. Comparison to Other Models for Multisensory Integration

Several models that account for multisensory integration in the colliculus of different granularity and focus have been suggested over the years. Some of them try to explain the various response properties of MSI neurons (Anastasio and Patton, 2003; Ursino et al., 2017) whereas others focus more on the biological detailed architecture (Cuppini et al., 2011, 2017; Casey et al., 2012). In the following, we will describe two of them and point out their strengths and weaknesses compared to our presented model.

In Rowland et al. (2007), the authors presented an algebraic and compartmental model of multisensory integration that incorporate cortico-collicular projections and try to explain the existence of AMPA and NMDA receptors in MSI neurons. Their goal was to reproduce a variety of physiological findings without paying much attention to the underlying biological anatomy and structure. Like our model, the authors are able to reproduce several MSI characteristics like multisensory enhancement, inverse effectiveness and super- and sub-additivity. In addition to our presented results, they also demonstrate the MSI neuron dependence on NMDA receptors and the temporal window of integration of their model. However, they did not present any results that indicates a Bayesian optimal integration of the signals.

Another approach is taken by Ohshiro et al. (2011) and their normalization model in which they show that many of the MSI response characteristics can be achieved by a pool normalization of the neuron output. Their model assumes MSI neurons that integrate signals according to a linear weighted sum with different input weights across modalities and neurons. In addition to the replication of MSI characteristics, the authors performed a virtual experiment of vestibular-visual integration task with their model and provided data that closely resembles findings in monkeys. Despite the profound analysis of their model and resemblance of experimental data, the authors neglect cortical projections to MSI neurons entirely.

4.4. Limitations of the Model

As we have shown, the two proposed model implementations using rate-based and spike-based encoding are both able to replicate several physiological findings, predict the purpose of cortical modulatory projections and are capable of reliably processing real world spiking data. One of the drawbacks of the current implementations is the lack of any learning mechanism in the process. The model assumes that all connections are already established and inputs are spatially aligned, even though, studies show that multisensory integration emerges during maturation of the nervous system by a constant exposure to multimodal signals (Wallace and Stein, 1997). This long term exposure influences how and to what extent multisensory integration takes place. This limitation in our model could be tackled by incorporating a Hebbian correlation learning mechanism between the cortical feedback projections and MSI neurons as well as the inputs of the model. The current assumption that the two modalities are spatially aligned is a strong constraint and simplifies the model architecture but is not biologically plausible. We are confident that this can be overcome with a previously proposed architecture of spatial map alignment of visual and auditory inputs (Oess et al., 2020a).

4.5. Outlook

The proposed model implementations of MSI neurons set a solid basis for future investigations. One important question we are planning to investigate is the role of the cortical feedback. One plausible hypothesis is that the feedback projections can be controlled by an attention mechanism to set special focus on a particular region and thereby enhances signals at that spatial location. This is an essential mechanism when conflicting events are present. In addition, the spike-based implementation of the model on neuromorphic hardware is an important step toward a real-time capable robotic platform. This platform will be equipped with audio and visual sensory hardware which directly communicates with the neuromorphic processing chips via spike trains, thereby creating a complete neuromorphic

system from the sensory perception to decision making and action execution.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

TO contributed in rate-based model conception and design, experiment conduction, analysis and interpretation of data, and drafting manuscript. ML contributed in spike-based model conception and realization and drafting the manuscript. DS contributed in spike-based model conception and realization and data interpretation and analysis. ME contributed in analysis and interpretation of data, and critical review. HN contributed in model conception and design, analysis and interpretation of data, and critical review.

FUNDING

This research has been conducted as part of the VA-MORPH project financed by the Baden-Württemberg foundation in the Neurorobotik program (project no. NEU012).

ACKNOWLEDGMENTS

Simulations have been performed with hardware and software granted by a field test agreement with IBM Research Almaden and we are grateful to TrueNorth team for their support. In addition, the authors like to express their gratitude to the reviewers' suggestions that helped to improve the manuscript.

REFERENCES

Ahmed, K., Shrestha, A., Qiu, Q., and Wu, Q. (2016). "Probabilistic inference using stochastic spiking neural networks on a neurosynaptic processor," in *2016 International Joint Conference on Neural Networks (IJCNN)* (Vancouver, BC), 4286–4293. doi: 10.1109/IJCNN.2016.7727759

Alvarado, J. C., Rowland, B. A., Stanford, T. R., and Stein, B. E. (2008). A neural network model of multisensory integration also accounts for unisensory integration in superior colliculus. *Brain Res.* 1242, 13–23. doi: 10.1016/j.brainres.2008.03.074

Alvarado, J. C., Stanford, T. R., Rowland, B. A., Vaughan, J. W., and Stein, B. E. (2009). Multisensory integration in the superior colliculus requires synergy among corticocollicular inputs. *J. Neurosci.* 29, 6580–6592. doi: 10.1523/JNEUROSCI.0525-09.2009

Alvarado, J. C., Stanford, T. R., Vaughan, J. W., and Stein, B. E. (2007a). Cortex mediates multisensory but not unisensory integration in superior colliculus. *J. Neurosci.* 27, 12775–12786. doi: 10.1523/JNEUROSCI.3524-07.2007

Alvarado, J. C., Vaughan, J. W., Stanford, T. R., and Stein, B. E. (2007b). Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *J. Neurophysiol.* 97, 3193–3205. doi: 10.1152/jn.00018.2007

Anastasio, T. J., and Patton, P. E. (2003). A two-stage unsupervised learning algorithm reproduces multisensory enhancement in a neural network model of the corticotectal system. *J. Neurosci.* 23, 6713–6727. doi: 10.1523/JNEUROSCI.23-17-06713.2003

Andersen, T. S., Tiippana, K., and Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cogn. Brain Res.* 21, 301–308. doi: 10.1016/j.cogbrainres.2004.06.004

Avillac, M., Ben Hamed, S., and Duhamel, J.-R. (2007). Multisensory integration in the ventral intraparietal area of the Macaque monkey. *J. Neurosci.* 27, 1922–1932. doi: 10.1523/JNEUROSCI.2646-06.2007

Beck, J., Pouget, A., and Heller, K. A. (2012). "Complex inference in neural circuits with probabilistic population codes and topic models," in *Advances in Neural Information Processing Systems 25*, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Lake Tahoe, CA: Curran Associates, Inc.), 3059–3067.

Cadusseau, J., and Roger, M. (1985). Afferent projections to the superior colliculus in the rat, with special attention to the deep layers. *J. Hirnforsch.* 26, 667–681.

Casey, M. C., Pavlou, A., and Timotheou, A. (2012). Audio-visual localization with hierarchical topographic maps: modeling the superior colliculus. *Neurocomputing* 97, 344–356. doi: 10.1016/j.neucom.2012.05.015

Cassidy, A. S., Alvarez-Icaza, R., Akopyan, F., Sawada, J., Arthur, J. V., Merolla, P. A., et al. (2014). "Real-time scalable cortical computing at 46 giga-synaptic ops/watt with 100× speedup in time-to-solution and 100,000× reduction in energy-to-solution," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '14* (New Orleans, LA: IEEE Press), 27–38. doi: 10.1109/SC.2014.8

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

Cuppini, C., Stein, B. E., Rowland, B. A., Magosso, E., and Ursino, M. (2011). A computational study of multisensory maturation in the superior colliculus (SC). *Exp. Brain Res.* 213, 341–349. doi: 10.1007/s00221-011-2714-z

Deneve, S., Latham, P. E., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.* 4, 826–831. doi: 10.1038/90541

Edwards, S. B., Ginsburgh, C. L., Henkel, C. K., and Stein, B. E. (1979). Sources of subcortical projections to the superior colliculus in the cat. *J. Comp. Neurol.* 184, 309–329. doi: 10.1002/cne.901840207

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a

Haessig, G., Cassidy, A., Alvarez, R., Benosman, R., and Orchard, G. (2018). Spiking optical flow for event-based sensors using IBM's TrueNorth neurosynaptic system. *IEEE Trans. Biomed. Circuits Syst.* 12, 860–870. doi: 10.1109/TBCAS.2018.2834558

Hillis, J. M., Ernst, M. O., Banks, M. S., and Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science* 298, 1627–1630. doi: 10.1126/science.1075396

Hyde, P. S., and Knudsen, E. I. (2000). Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl. *J. Comp. Neurol.* 421, 146–160. doi: 10.1002/(SICI)1096-9861(20000529)421:2<146::AID-CNE2>3.0.CO;2-5

Hyde, P. S., and Knudsen, E. I. (2002). The optic tectum controls visually guided adaptive plasticity in the owl's auditory space map. *Nature* 415, 73–76. doi: 10.1038/415073a

Jiang, W., Jiang, H., Rowland, B. A., and Stein, B. E. (2007). Multisensory orientation behavior is disrupted by neonatal cortical ablation. *J. Neurophysiol.* 97, 557–562. doi: 10.1152/jn.00591.2006

Jiang, W., and Stein, B. E. (2003). Cortex controls multisensory depression in superior colliculus. *J. Neurophysiol.* 90, 2123–2135. doi: 10.1152/jn.00369.2003

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., and Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *J. Neurophysiol.* 85, 506–522. doi: 10.1152/jn.2001.85.2.506

Kadunce, D., Vaughan, W., Wallace, M., and Stein, B. (2001). The influence of visual and auditory receptive field organization on multisensory integration in the superior colliculus. *Exp. Brain Res.* 139, 303–310. doi: 10.1007/s002210100772

Kasabov, N. K. (2019). *Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence. Springer Series on Bio- and Neurosystems. 1st Edn.*, Heidelberg: Springer.

Kayser, C., and Shams, L. (2015). Multisensory causal inference in the brain. *PLoS Biol*. 13:e1002075. doi: 10.1371/journal.pbio.1002075

Knudsen, E. I. (2002). Instructed learning in the auditory localization pathway of the barn owl. *Nature* 417:322. doi: 10.1038/417322a

Lichtsteiner, P., Posch, C., and Delbruck, T. (2006). "A 128 × 128 120db 30mw asynchronous vision sensor that responds to relative intensity change," in *2006 IEEE International Solid State Circuits Conference-Digest of Technical Papers* (San Francisco, CA: IEEE), 2060–2069. doi: 10.1109/ISSCC.2006.1696265

Löhr, M. P. R., Jarvers, C., and Neumann, H. (2020). "Complex neuron dynamics on the IBM TrueNorth neurosynaptic system," in *IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS 2020* (Genova), doi: 10.1109/AICAS48895.2020.9073903

Löhr, M. P. R., and Neumann, H. (2018). "Contrast detection in event-streams from dynamic vision sensors with fixational eye movements," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)* (Florence), 1–5. doi: 10.1109/ISCAS.2018.8351084

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci*. 9, 1432–1438. doi: 10.1038/nn1790

Ma, W. J., and Pouget, A. (2008). Linking neurons to behavior in multisensory perception: a computational review. *Brain Res*. 1242, 4–12. doi: 10.1016/j.brainres.2008.04.082

Marrocco, R., and Li, R. (1977). Monkey superior colliculus: properties of single cells and their afferent inputs. *J. Neurophysiol*. 40, 844–860. doi: 10.1152/jn.1977.40.4.844

McDonald, J. J., Teder-Sälejärvi, W. A., and Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science* 292, 1791–1791. doi: 10.1126/science.292.5523.1791a

Meredith, M. A., and Clemo, H. R. (1989). Auditory cortical projection from the anterior ectosylvian sulcus (Field AES) to the superior colliculus in the cat: an anatomical and electrophysiological study. *J. Comp. Neurol*. 289, 687–707. doi: 10.1002/cne.902890412

Meredith, M. A., and Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science* 221, 389–391. doi: 10.1126/science.6867718

Meredith, M. A., and Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *J. Neurophysiol*. 75, 1843–1857. doi: 10.1152/jn.1996.75.5.1843

Meredith, M. A., Wallace, M. T., and Stein, B. E. (1992). Visual, auditory and somatosensory convergence in output neurons of the cat superior colliculus: multisensory properties of the tecto-reticulo-spinal projection. *Exp. Brain Res*. 88, 181–186. doi: 10.1007/BF02259139

Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., et al. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345, 668–673. doi: 10.1126/science.1254642

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., and Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Exp. Brain Res*. 184, 39–52. doi: 10.1007/s00221-007-1080-3

Oess, T., Ernst, M. O., and Neumann, H. (2020a). Computational investigation of visually guided learning of spatially aligned auditory maps in the colliculus. *bioRxiv*. doi: 10.1101/2020.02.03.931642

Oess, T., Löhr, M. P. R., Jarvers, C., Schmid, D., and Neumann, H. (2020b). "A bio-inspired model of sound source localization on neuromorphic hardware," in *IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS 2020* (Genova), doi: 10.1109/AICAS48895.2020.9073935

Ohshiro, T., Angelaki, D. E., and DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nat. Neurosci*. 14, 775–782. doi: 10.1038/nn.2815

Oliver, D. L., and Huerta, M. F. (1992). "Inferior and superior colliculi," in *The Mammalian Auditory Pathway: Neuroanatomy, Springer Handbook of Auditory Research*, eds D. B. Webster, A. N. Popper, and R. R. Fay (New York, NY: Springer New York), 168–221. doi: 10.1007/978-1-4612-4416-5_5

Perrault, T. J., Vaughan, J. W., Stein, B. E., and Wallace, M. T. (2003). Neuron-specific response characteristics predict the magnitude of multisensory integration. *J. Neurophysiol*. 90, 4022–4026. doi: 10.1152/jn.00494.2003

Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nat. Neurosci*. 16, 1170–1178. doi: 10.1038/nn.3495

Rees, A. (1996). Sensory maps: aligning maps of visual and auditory space. *Curr. Biol*. 6, 955–958. doi: 10.1016/S0960-9822(02)00637-1

Rohe, T., and Noppeney, U. (2015). Cortical hierarchies perform bayesian causal inference in multisensory perception. *PLoS Biol*. 13:e1002073. doi: 10.1371/journal.pbio.1002073

Rowland, B. A., Jiang, W., and Stein, B. E. (2014). Brief cortical deactivation early in life has long-lasting effects on multisensory behavior. *J. Neurosci*. 34, 7198–7202. doi: 10.1523/JNEUROSCI.3782-13.2014

Rowland, B. A., Stanford, T. R., and Stein, B. E. (2007). A model of the neural mechanisms underlying multisensory integration in the superior colliculus. *Perception* 36, 1431–1443. doi: 10.1068/p5842

Stein, B. E., Spencer, R. F., and Edwards, S. B. (1983). Corticotectal and corticothalamic efferent projections of SIV somatosensory cortex in cat. *J. Neurophysiol*. 50, 896–909. doi: 10.1152/jn.1983.50.4.896

Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci*. 9:406. doi: 10.1038/nrn2331

Stevenson, R. A., and Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Exp. Brain Res*. 227, 249–261. doi: 10.1007/s00221-013-3507-3

Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci*. 14, 400–410. doi: 10.1016/j.tics.2010.06.008

Trappenberg, T. P. (2010). *Fundamentals of Computational Neuroscience. 2nd Edn*. Oxford: Oxford University Press.

Tsai, W., Barch, D. R., Cassidy, A. S., DeBole, M. V., Andreopoulos, A., Jackson, B. L., et al. (2017). Always-on speech recognition using truenorth, a reconfigurable, neurosynaptic processor. *IEEE Trans. Comput*. 66, 996–1007. doi: 10.1109/TC.2016.2630683

Ursino, M., Crisafulli, A., di Pellegrino, G., Magosso, E., and Cuppini, C. (2017). Development of a bayesian estimator for audio-visual integration: a neurocomputational study. *Front. Comput. Neurosci*. 11:89. doi: 10.3389/fncom.2017.00089

Wallace, M. T., Meredith, M. A., and Stein, B. E. (1993). Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J. Neurophysiol*. 69, 1797–1809. doi: 10.1152/jn.1993.69.6.1797

Wallace, M. T., Meredith, M. A., and Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *J. Neurophysiol*. 80, 1006–1010. doi: 10.1152/jn.1998.80.2.1006

Wallace, M. T., Perrault, T. J., Hairston, W. D., and Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration. *J. Neurosci*. 24, 9580–9584. doi: 10.1523/JNEUROSCI.2535-04.2004

Wallace, M. T., and Stein, B. E. (1994). Cross-modal synthesis in the midbrain depends on input from cortex. *J. Neurophysiol*. 71, 429–432. doi: 10.1152/jn.1994.71.1.429

Wallace, M. T., and Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *J. Neurosci*. 17, 2429–2444. doi: 10.1523/JNEUROSCI.17-07-02429.1997

Yu, L., Xu, J., Rowland, B. A., and Stein, B. E. (2016). Multisensory plasticity in superior colliculus neurons is mediated by association cortex. *Cereb. Cortex* 26, 1130–1137. doi: 10.1093/cercor/bhu295

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.