Articles in PresS. J Neurophysiol (December 17, 2014). doi:10.1152/jn.00341.2014

Deconstructing multi-sensory enhancement in detection

Mario Pannunzi[1] ✉, **Alexis Pérez-Bellido**[2], **Alexandre Pereda-Baños**[3], **Joan López-Moliner**[2,4], **Gustavo Deco**[1,5] and **Salvador Soto-Faraco**[1,5]

[1]Universitat Pompeu Fabra, Barcelona (Spain),
[2]Universitat de Barcelona (UB), Barcelona (Spain),
[3]Barcelona Media, Barcelona (Spain),
[4]Institute for Brain, Cognition and Behaviour (IR3C), Barcelona (Spain),
[5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona (Spain),

✉DTIC-Universitat Pompeu Fabra (55.120) C/ Roc Boronat, 138, 08018 Barcelona, Spain; mario.pannunzi@gmail.com

Tuesday 16th December, 2014

Abstract

The mechanisms responsible for the integration of sensory information from different modalities have become a topic of intense interest in psychophysics and neuroscience. Many authors now claim that early, sensory-based cross-modal convergence improves performance in detection tasks. An important strand of supporting evidence for this claim is based on statistical models such as the Pythagorean model or the probabilistic summation model. These models establish statistical benchmarks representing the best predicted performance under the assumption that there are no interactions between the two processing paths. Following this logic, when observed detection performances surpass the predictions of these models, it is often inferred that such improvement indicates early cross-modal convergence. We present a theoretical analyses scrutinizing some of these models and the statistical criteria most frequently used to infer early cross-modal interactions during detection tasks. Our current analysis shows how common misinterpretations of these models lead to their inadequate use and, in turn, to contradictory results and misleading conclusions. To further illustrate the latter point, we introduce a model that accounts for detection performances in multimodal detection tasks, but for which surpassing of the Pythagorean or probabilistic summation benchmark can be explained without resorting to early sensory interactions. Finally, we report three experiments that put our theoretical interpretation to the test, and further clarify how to adequately measure multimodal interactions in audio-tactile detection tasks.

Introduction

Understanding how humans detect and react to everyday life complex events must include an account of how cues in different sensory modalities are integrated in the brain. However, the rules that govern these interactions and the underlying brain mechanisms (for a review see Alais et al. (2010); Fetsch et al. (2013); van Atteveldt et al. (2014)) are still far from agreed upon.

Many studies report that cross and within-modality convergence at sensory processing levels produce benefits in detectability. Behavioral (e.g., Frassinetti et al. (2002); Gillmeister and Eimer (2007); Pérez-Bellido et al. (2013)) and neuro-physiological studies (e.g., Murray et al. (2005); Kayser et al. (2005); Lemus et al. (2010)) designed to provide evidence of such early multi-sensory interactions abound in the literature, even though their interpretation is not always straightforward (Alais et al. (2010); Driver and Noesselt (2008)). We focus on audio-tactile interactions during detection tasks, these sensory interactions can putatively take place at various levels of processing, including the earliest ones at the peripheral nervous system (Ghazanfar and Schroeder (2006); Lakatos et al. (2007); Lemus et al. (2010)), which makes them an ideal candidate to test for early sensory interactions. As is the case with other multi-sensory ensembles, evidence of early sensory interactions from audio-tactile behavioral studies is somewhat mixed, often leading to discrepant conclusions. Such discrepant conclusions arise for various reasons. For example, regarding audio-tactile detection tasks, comparisons of the results of yes-no tasks (Schnupp et al. (2005)) with 2 interval forced choice tasks, 2IFC, (Wilson et al. (2009)) are not straightforward (see for example Yeshurun et al. (2008)), and this can easily lead to erroneous interpretations of the observed improvements in performance. Further reasons are: Confounding detection with discrimination (e.g., Ro et al. (2009), both for frequency and for amplitude discrimination Soto-Faraco and Deco (2009)), interpreting measures that do not reflect changes in sensitivity as if they did (compare Schürmann et al. (2004); Soto-Faraco and Deco (2009) with Yarrow et al. (2008)) or, in multi-sensory studies, interpreting attentional cueing effects as evidence of early integration (for audio-visual Lovelace et al. (2003); and for audio-tactile: Gillmeister and Eimer (2007); Ro et al. (2009)); an interesting approach to this issue can be found in Lippert et al. (2007).

The present study focuses on some of the most used benchmark models that have been taken as criteria to decide on the presence of early sensory interactions (Wuerger et al. (2003); Meyer et al. (2005); Sperdin et al. (2009, 2010); Wilson et al. (2009, 2010a,b); Arnold et al. (2010); Marks et al. (2011)). Two popular models in multi-sensory research are the Pythagorean model (PM) and linear summation model (LSM), both implementations of the Signal Detection Theory (SDT, e.g., Wickens (2001); MacMillan and Creelman (2005)). Another common approach is the probabilistic summation model (PSM, see e.g., Green (1958)). Briefly, the PM and the LSM represent the the signal or the noise with a continuous value (sensory activities), and assume a linear summation of these continuous sensory activities in a stage preceding the decision. The PSM on the other hand assumes that the final decision about the presence of the stimulus

is made, in a probabilistic fashion, upon the independent decisions made on each individual modality, therefore considering a few finite states (detection, no detection). Within the PSM framework, Quick (Quick (1974)) proposed a family of functions in order to describe how the different decisions about each modality can be integrated. These mathematical tools have been used principally in vision research, but they have become relevant in multi-sensory research as well, and in particular for detection tasks (Dalton et al. (2000); Wuerger et al. (2003); Alais and Burr (2004); Meyer et al. (2005)).

Several issues can arise from incorrect interpretations of these models. For example, neglecting/forgetting changes in the decision criterion, can lead to an overestimation of detectability improvements (compare e.g. report given by Alais et al. (2010); Soto-Faraco and Deco (2009) of Gescheider et al. (1974)); or confusion and mix-up of methods based on the SDT and the ones based on Quick Pooling models can lead to ambiguous interpretations of experimental results (see Introduction in Arnold et al. (2010), or Alais and Burr (2004)); the lack of a uniform interpretation of the different mathematical frameworks utilized to model psychophysical detection data impedes a straightforward comparison of the results. For example for the 2IFC tasks, some studies apply the approach based on the SDT (e.g., Wilson et al. (2009); Ernst and Banks (2002)) while others apply the one based on the PSM (e.g., Alais and Burr (2004); Meyer et al. (2005); Wuerger et al. (2003)). However, even though these issues can help us to put in perspective our study, coping with all of them is beyond its scope, and we will analyze in detail only those related to our main focus, that is, multi-sensory enhancements in detection, particularly audio-tactile interactions. To this aim we carry out a reanalysis of previous results in the literature (Schnupp et al. (2005); Wilson et al. (2009)) and run three further experiments to provide empirical evidence for our theoretical conclusions.

In a thorough series of studies (Wilson et al. (2009, 2010a,b)), Wilson and colleagues used the PM and they reported detection performances above the PM's criterion in an audio-tactile detection task. In their experiments, the task was to detect an audio-tactile stimulus. They used vibrotactile stimulus frequencies ranging from 100 Hz to 1000 Hz (Wilson et al. (2009)) and found frequency-specific interactions. These results are in contradiction with the conclusions of a previous study by Schnupp et al. (2005), where a very similar detection paradigm was used. Indeed the detection data reported by Schnupp *et al.* fell below the PM's criterion, and were well fitted instead by their Orthogonal Model (OM), which assumes no sensory interactions at early levels of processing.

Based on our own theoretical analysis of these models and a reanalysis of Schnupp's dataset, we demonstrate that the discrepancy between Wilson et al. (2009) and Schnupp et al. (2005) is a consequence of the misuse of the PMs optimality in two interval two alternative forced choice paradigms (2IFC), that is, the optimal strategy in yes/no tasks, does not (always) apply to 2IFC tasks rather, it depends on the strategy the subjects adopt to solve the task. The most common strategies can be divided in two groups: (1) The observer compares informations

from the two intervals and then makes a unique decision or (2) the observer makes a decision at each interval and then integrates these decisions. The models therefore need to be adjusted in accordance to which of the latter two strategies is adopted by each participant. Note that this warning is relevant to all experimental studies focused on multi-sensory detection tasks, or multi-component detection tasks.

We then put forward a simple model which assumes cross-modal sensory independence (no interactions until the decision stage), and whose behavior is in agreement with previous experimental data (Schnupp et al. (2005); Wilson et al. (2009)). The intention in putting this model forward is not to propose a further benchmark model, but to exemplify how surpassing the PM's benchmark in a 2IFC does not necessarily imply early interactions. The model we propose has features extrapolated both from the PM, the LSM and the PM, therefore we referred to it as a mixed-model (MM). In the MM the detection stage takes place separately for each modality as the PSM, but the input to the decision stage (output from the detection stage) is a continuos variable, as is the case with the LSM and the PSM. Finally, we also put forwar a straightforward mechanistic implementation of the MM in a attractor neural network (ANN, (Wang (2002)). The models based on ANN are adequate for modeling detection decision making (Deco et al. (2007)), and in addition, are able to qualitatively reproduce neurophysiological results (e.g., De Lafuente and Romo (2005)).

In order to reinforce the conclusions of our formal analysis, we present three separate experiments, designed demonstrate the lack of early integration through empirical tests, and to illustrate a correct measurement of interactive effects from multi-sensory stimuli in audio-tactile detection tasks. In experiment 1, we set up the experimental protocols and established whether the apparent contradiction between the results of Schnupp et al. and Wilson *et al.*'s studies was caused by the paradigm , i.e., stimulus predictability. Two additional experiments are then presented, both based on the simple idea that if the interaction that generates the violation of the PM criterion takes place in sensory areas, then its strength will depend heavily on the temporal overlap between the two stimuli. This dependence on the temporal overlap is based on neurophysiological findings (De Lafuente and Romo (2005); de Lafuente and Romo (2006) for somatosensory; Bendor and Wang (2007) for audio), where the firing rates of the early-sensory areas (S1, S2 and A1) are modulated only during the stimulus presentation. We therefore hypothesized that if any (rate-based) multi-modal interaction takes place at early-sensory processing stage the detection performance will depend heavily on the temporal overlap between the two stimuli (we do not make claims about other possible modulations not based on actual activity, such as long term learnign or historic effects). Experiment 2 thus explored audio-tactile detection across a range of stimulus onset asynchronies (SOAs up to 750ms), with the aim of introducing experimental conditions that would clearly impede early sensory interactions. If the PM prediction is surpassed in these conditions then it is evidently not an effective benchmark. Finally, in experiment 3, we reinforce the rationale of experiment 2 by illustrating how

long SOAs do indeed suppress well established interactions at early-sensory stages (Fastl and Zwicker (2001)) in detectability (Marill (1956)) using unisensory compound stimuli (500 Hz and 1100 Hz pure tones).

Materials and methods

This section is divided in two parts. In the first part we provide an in-depth mathematical description of the models that will be the object of our subsequent scrutiny. These models are the ones traditionally employed to tell apart early from late multisensory interactions (Wuerger et al. (2003); Meyer et al. (2005); Sperdin et al. (2009, 2010); Wilson et al. (2009, 2010a,b); Arnold et al. (2010)). In order to have a more detailed understanding of the principles underlying these models, readers may refer to the corresponding chapters of the more relevant books (Green and Swets (1966); Wickens (2001); Graham (2001); MacMillan and Creelman (2005)). In the second part we describe the materials and methods used for the psychophysical experiments right before the experimental results.

Theoretical methods

Mathematical models of detection behavior for yes-no tasks

In the yes-no paradigm the observer is requested to report after each trial whether he perceived a stimulus (yes) or not (no). Generally, stimulus is presented in only 50% of the trials. Therefore, the observer has to discriminate the signal (stimulus present) from the noise (stimulus absent). In the multi-sensory yes-no task analyzed in this paper, the stimulus is composed of two modalities and the observer is requested to report yes when she perceived at least one of the two modalities. In visual literature this kind of task is called summation experiment (Graham (2001)).
We used the description of signal detection theory (SDT) put forward by Green and Swets (Green (1958); Green and Swets (1966)) as a starting point for explaining the considered models. The main assumption is that participants sample a continuous distribution of signal values, and that the noise distribution overlaps with the signal's distribution (Assumption 1). For the simplest version, usually adopted in multi-sensory research, further assumptions are:

As2 In each trial, the observer records a measure of sensory activity, s, and compares it with threshold, λ. When s overcomes λ, the observer's answer is yes.

As3. s is typically assumed having normal probability distribution, with mean μ and deviation σ. The signal's mean is defined as d' and the noise's mean is equal to 0.

As4. In both cases, signal and noise, the deviation is, in the standard description, equal to 1.

With these assumptions, there exists a simple analytic relationship between d', the single-modality detectability, P^{hit}, the probability to correctly detect a stimulus, and P^{fa}, the proba-

bility of false alarm (observer saying yes without stimulus). Indeed, from the assumptions [1-4], the probability to correctly detect a stimulus is given by:

$$P^{hit} = \Phi(d' - \lambda) \tag{1}$$

where Φ is the cumulative distribution function of the standard normal distribution and the relationship between the threshold λ and P^{fa} is:

$$\lambda = -\mathrm{Z}(P^{fa}) = \sqrt{2}\,\left(\mathrm{erf}^{-1}(2\,P^{fa} - 1)\right) \tag{2}$$

Therefore we can write d' as:

$$d' = \mathrm{Z}(P^{hit}) - \mathrm{Z}(P^{fa}) = \sqrt{2}\,\left(\mathrm{erf}^{-1}(2\,P^{hit} - 1) - \mathrm{erf}^{-1}(2\,P^{fa} - 1)\right) \tag{3}$$

where the Z function is the probit function, that is the quantile function associated with the standard normal distribution, erf^{-1} is inverse of the error function. When the stimulus is a compound of two modalities there are many ways to separate observations that come from the signal(s) distribution(s) from those coming from the noise(s) distribution(s), and "even an incomplete catalog of the possibilities is large enough to be confusing" (Wickens (2001), ch10). We will only consider those that are commonly taken into account in the multisensory literature. As a guide, we refer the reader to fig.1A, where we depicted a schematic representation of two categories of processing models in a yes-no task. For the model shown in the top row, the audio and tactile signals are processed and detected separately. Only after this stage the two paths intersect, and a not-specified sum takes place, after which the decision is made. In the model depicted in the bottom row of fig.1A, the interaction between audio and tactile information takes place before the detection. These two models do not assume any early-sensory interactions, nor top-down influences from higher decisional areas, and the reason, as we already anticipated and as we will discuss below, is precisely that they have been used as benchmarks to test early-sensory interactions.

One-look strategy

A plausible, and maybe the simplest, detection strategy is for the observer to use only the best of the two modalities. Wickens (2001) named it one-look strategy for obvious reasons. This strategy would of course lead to a low performance, as a combination of the two signals should have better performance than either of both signals alone. We discarded analyzing this option for two reasons: First, to our knowledge there is no evidence in favor of it in the multi-sensory literature; second, this model has never been used as benchmark to address sensory interactions.

Probabilistic summation model (PSM)

An a priori slightly better strategy assumes that the observer analyzes each component separately, and then she combines the output in a single decision (see top row of fig.1A). The informations coming from the two separate analysis can be characterized with a finite-state, so this strategy is based on similar hypotheses to those formulated by the High-Threshold Theory

(HTT). Wickens (2001), referred to it as dual-look strategy, but in the multi-sensory literature is commonly called Probabilistic summation model (PSM). For the PSM the signals from different modalities are processed and detected separately. Each detection stage has a binary output (detection vs not-detection), and the decision is an inclusive OR function of these individual modality outputs (see top row of fig.1A). Despite strong theoretical and experimental criticisms (see e.g., Laming (1997); Tyler and Chen (2000)), the PSM remains a very popular approach in multi-sensory research.

To formalize the PSM's hypothesis, that is, to obtain the probability of a correct answer for a yes/no (yn) task involving multi-sensory events, some simple assumptions about the participants strategy are commonly adopted:

As1. Detection is a binary process, that is, it can either happen or not (Luce (1963), Norman (1964));

As2. Information processing leading to detection proceeds in a completely independent way for each modality.

As3. The observer has zero false alarm (fa) rate, that is, there are no detections of any signal in the interval without signal. This assumption is of course an oversimplification. In Schnupp *et al.*'s empirical data, the condition without any stimulation registered around 3%.

Let us call P_T^d the probability to correctly detect the stimulus (a tactile stimulus, for instance), and P_T^{Nd} the probability to fail to detect it (with $P_T^d + P_T^{Nd} = 1$). The probability of a correct detection with a compound of two stimuli (A+T) then becomes:

$$P^{hit}(A,T) = 1 - (1 - P_T^d)(1 - P_A^d) \tag{4}$$

Where the subindices A and T indicate the two modalities, i.e. audio and tactile. MacMillan and Creelman (2005) and Marks et al. (2011) have done a similar analysis with the same results, and show how the PSM's prediction would change when relaxing assumption As3. However we did not report here that analysis because the predicted detection performance in that case, is even lower than the one obtained with eq.4 (see equation B4 of Marks et al. (2011): $P_{fa}^{hit}(A,T) = (P^{hit}(A,T) - fa)/(1 - fa)$). As we will see later, the PSM's predicted hit probabilities are already too low when compared with the observed dataset.

For the re-analysis of the Schnupp *et al.* dataset, we adopted the following psychophysical curve of probability of detection as a function of stimulus intensity (as used in Schnupp et al. (2005)):

$$P^{hit}(T) = \Phi(x_T\, b_T - \lambda) \tag{5}$$

where Φ is the cumulative normal function, x_T is the difference of the stimulus intensity between the value of interest and a baseline, b_T is a proportional factor dependent on the modality and,

λ is a decision stage threshold parameter. The subindex indicates the stimulus modality (here T stands for tactile, but of course the same equation is valid for the other modality, A=audio).

For the PSM the probability of a correct detection with two stimuli (A+T), eq.4, can be rewritten as:

$$P^{hit}(T,A) = 1 - (1 - \Phi(x_T\, b_T - \lambda))\,(1 - \Phi(x_A\, b_A - \lambda)) \tag{6}$$

as in Schnupp et al. (2005) we used a single threshold parameter λ to facilitate the comparison between different models. This approach is very similar to the one proposed in Quick (1974), differing only in the function used to fit the psychophysical curve. Schnupp and colleagues used the Φ, while Quick used $1 - 2^{-(b_T\, x_T)^k}$, in fact other functions can and have been used (e.g., Wuerger et al. (2003)). In our analysis we preferred to stick to the function used by Schnupp et al. as it is more feasible to compare it with the approach to the multi-sensory SDT. Indeed the PSM is an example of a model where sensory interaction takes place after the detection stage. However if we relax the second of the PSM's assumptions in the yes-no task, then P^d_T can condition P^d_A, and viceversa. In this way we are dealing with a non-independent PSM.

Non-indpendent PSM for yes-no tasks

For this model the detection of each modality dependends on the detection of the other modality. For this model we can keep calling P^d_T the probability to correctly detect the tactile stimulus, but now the probability to correctly detect the audio stimulus is conditioned on correctly detecting the tactile stimulus and we can call it $P^d_{A/T}$. Similarly the probability to detect the audio stimulus (A) conditioned on a failed detection of the tactile stimulus ($\bar{T}$) is given by: $P^d_{A/\bar{T}}$. The hit probability for compound stimuli (A+T) then becomes:

$$P^{hit}(A,T) = 1 - (1 - P^d_T)\,(1 - P^d_{A/\bar{T}}) \tag{7}$$

From this equation it is clear that the probability to detect the compound stimulus is higher when $P^d_A < P^d_{A/\bar{T}}$.

In the next two models (represented in fig.1A), this interaction takes place before the detection stage.

Pythagorean model (PM)

Under this strategy, information, i.e. sensory activities (s_T and s_A), is integrated in a single element and then compared to a threshold (see bottom row of fig.1B). The decision rule of the optimal model (whose prediction coincides with the likelihood-ratio estimate (LRE)) is based on the comparison of the weighted linear sum of the sensory activities to the threshold λ, using d' of each modality as weight:

$$s_T\, d'_T + s_A\, d'_A > \lambda. \tag{8}$$

According to this proposal the detectability of the compound stimuli equals the root square of the squared sum of all individual detectability scores (Wickens (2001)), $d'_{PM}(A,T) = \sqrt{d'^2_A + d'^2_T}$; this equation is the actual reason why this model is called Pythagorean model (PM). The

threshold λ is then a function of the various d's:

$$\lambda_{PM} = d'_{PM}(A,T)/2 - \text{logit}(Ps)/d'_{PM}(A,T) \tag{9}$$

where logit is the inverse of the sigmoidal logistic function and P_s is the probability to have signal; $\text{logit}(Ps) = 0$ for $P_s = 50\%$. This last term takes into account the possibility that the probability of the signal and the noise won't be the same.

The PM's hit probability is be given by eq.1 using the detectability d'_{PM} and the threshold λ_{PM}.

It is worth noting that this strategy is not directly applicable in experimental protocols where different modalities and/or amplitudes of the stimuli appear in an interleaved fashion (like in Schnupp et al. (2005)), because the observer cannot set properly neither the threshold nor the weight for the modalities prior to stimulus appearance. The strategy described below, the linear summation model (LSM), does not suffer from this problem as weight and threshold are fixed.

Linear summation model (LSM)

A simpler option with respect to the previous PM is a decision rule based on a linear sum with fixed values of the weights and the threshold:

$$s_T + s_A > \lambda. \tag{10}$$

This model is also represented by the bottom row of fig.1A, as the interaction takes place before the detection stage. The hit probability is then:

$$P^{hit}(A,T) = \Phi\left(\frac{d'_A + d'_T - \lambda}{\sqrt{\sigma_A^2 + \sigma_T^2}}\right) \tag{11}$$

Where $\sigma_A^2 + \sigma_T^2 = 2$, as the two modalities are assumed to have standard deviation equal to 1. We refer to this model here and thereafter as the linear summation model (LSM). Note that, when the stimuli with single and double modality are interleaved, the P^{hit} for the single modality (originally given by eq.1) has to be replaced with:

$$P^{hit}(T) = \Phi((d'_T - \lambda)/\sqrt{2}) \tag{12}$$

Indeed even in single modality trials the fluctuations of the two modalities are relevant for the decision.

To help interpreting and comparing these models (1-look model, PSM, PM and LSM) we represented in fig.1B their decision bounds, that is how their decision rule divides the space of the sensory activity into yes or no answer regions. In this plane the axes (X_T, X_A) represent the stimuli amplitudes for the two modalities. There are two key observations to be made in this graph: First, the LSM, the PM and the 1-look model are similar in that their decision bound (green, purple and blue lines) is a single straight line, while the PSM has decision bound formed by two lines (red lines). Indeed LSM and PM belong to the type of model shown in the bottom

row of fig.1A, while the PSM belongs to the one depicted in the top row of fig.1A. Of course as the 1-look model lacks an actual second path, it can be cataloged in both groups. Second, the PM's decision bound is tilted to the right of the LSM's decision bound, because the PM assumes that the observer's strategy is to give more relevance (weight) to the modality having higher d' (in this case we arbitrarily represent $d'_A > d'_T$). For the limiting case where one of the two modalities is completely irrelevant (for example $d'_A/d'_T \to 0$) the PM's decision bound coincides with the 1-look model's.

For the re-analysis of the Schnupp *et al.* dataset with the LSM we again used eq.5, and in this case the product $b_T\, x_T$ corresponds to the d'. The hit probability with two stimuli (A+T), eq.11 is:

$$P^{hit}(T, A) = \Phi(x_T\, b_T + x_A\, b_A - \lambda) \tag{13}$$

Orthogonal-Model (OM)

The Orthogonal-Model (OM) is only one of a family of functions that Schnupp *et al.* used in order to fit the probability of hit of their dataset for the double-modality stimuli, this probability is:

$$P^{hit}(T, A) = \Phi\left(\sqrt[k]{(x_T\, b_T)^k + (x_A\, b_A)^k} - \lambda\right) \tag{14}$$

As Schnupp and colleagues indicated, this latter function can indeed be interpreted within the SDT framework, interpreting the products $\sqrt[k]{(x_T\, b_T)^k + (x_A\, b_A)^k}$ as the discriminability and λ as the threshold (although, as we argue below, these functions are adequate for fitting but not for modelling). Schnupp et al. (2005) showed that the best fit is achieved when the free parameter k characterizing the functions through the k-norm sum is close to 2, and therefore close to the OM's k-value, that we analyzed here. For the OM, the multi-sensory detection probability is given by:

$$P^{hit}(T, A) = \Phi(\sqrt{(x_T\, b_T)^2 + (x_A\, b_A)^2} - \lambda) \tag{15}$$

Schnupp et al. hypothesized that the nervous system uses a 'sensory metric space' to register and compare stimuli of various types. In their words: "In the orthogonal sensory metric space envisaged here, different sensory modalities are thought to occupy separate and mutually orthogonal dimensions. Stimuli are represented as points in this space, and these points may move as the stimulus intensity in each modality/dimension changes." (Schnupp et al. (2005), pg.185).

However only one of these functions, the one with $k = 1$, is effectively a mechanistic model, whilst the others are simply useful formulas to fit the data.

To illustrate this point it is worth remembering the interpretation underlying these formulas, in particular the formula on which the linear sum is based. Indeed the Φ function, for the SDT framework is not just an arbitrary choice useful to obtain a good fit of the sigmoidal psychophysical curve, rather it is a consequence of the assumptions of the Gaussian distribution of the sensory activity (see Green and Swets (1966); Wickens (2001)).

For the compound stimuli, eq.14 gives the prediction for the LSM setting $k = 1$ (compare with

eq.11 of the Materials and Methods). When setting $k = 2$, eq.14 can be confounded with the PM's hit prediction (see Discussion in Schnupp et al. (2005)), but they are not the same. Indeed the PM's threshold (eq.9 in the Materials and Methods) is dependent on the d', while for the OM, and in general for all the functions of eq.14, λ is a constant. It is indeed important not to confound the optimality of the Pythagorean sum of the detectability d' of the linear sum model, and the Pythagorean sum of the sensory activity of the OM (see Discussion of Schnupp et al. (2005)). Even more important is to rule out the possibility of the PM implementation for this task is the PM's decision rule, eq.8, that is based on the idea that the observer knows her detectability for the two modalities in each trial. This knowledge is clearly impossible, when the modalities and amplitude are interleaved during the experiment, as in Schnupp et al. (2005).
As for the case with $k = 2$, eq.14 for all $k \neq 1$ cannot be interpreted in terms of the decision model described by the SDT. Indeed if we interpret d'_T and d'_A as the mean of a Gaussian distribution for the stimulus signal, then the k-norm sum of the d's is not the mean of the k-norm sum of the Gaussian variables. Therefore for these cases the function Φ loses its role.
So, how can we interpret the OM, (and all the models based on eq.14 with $p \neq 1$)? There are 2 possible answers: 1. The OM is just an useful formula that can describe succinctly the results of Schnupp et al.; 2. The OM can be see as a deterministic model, that is x_T and x_A are not random Gaussian variables, and the whole stochasticity of the behavior as well as the shape of the psychophysical curve (well fitted by the probability function ϕ) has to be attributed to actual characteristics of the decision process. While the first answer lacks any mechanistic explanation for the results, the second answer is clearly implausible for its assumption of lack of noise in the sensory activity.

So far we have described some of the most widely used statistical models (plus Schnupp et al.'s OM) for benchmarking sensory interactions in yes/no multimodal detection tasks. We have seen how the main difference between them is their assumption regarding the order of the detection phase and 'sum' phase (see fig.1). Importantly, we have seen that even though the OM seems to provide the better fit to the data is not effectively a mechanistic model, but only a useful formula to fit the data. We now put forward a model, that, as said in the introduction, is not intended as the model that should replace those described so far, but as an example that it is possible to have a model able to describe the results of Schnupp *et al.* without hypothesizing early-sensory interaction between modalities at an early stage.

Mixed-Model (MM)

The model proposed here can be seen as an intermediate step between the PM and the PSM. The MM assumes a separate intermediate detection stage for each modality, similar to the PSM. The 'detected/not-detected' states are encoded into the activity of pools of neurons as high/low firing rates. We assume that, in the brain, these states are not encoded by a binary code (as in the PSM), but they are encoded in a continuous value which depends on the activation state - high/low - and the stimulus intensity (as in the PM).

The MM is composed of 3 stages: a sensory stage, a detection stage and a decision stage (see Fig.4). The MM has an effective detection stage whose output can takes a continuous value, ν. The probabilities of having a high or a low activation state, determined within the detection stage, are P_h and $P_l = 1 - P_h$, respectively and P_h is given by eq.1. The probability distribution functions of the detection stage output, for both the detected and non-detected cases, are $f_l(\nu)$ and $f_h(\nu)$ respectively. We used the subscripts l and h to indicate high and low activation states respectively (in terms of neural activity activation state amounts to firing rates). Here for the sake of simplicity, we chose for $f_h(\nu)$ a delta function around one; the value of ν for the low state is proportional to the difference of the stimulus intensity between the value of interest and a baseline value (x as defined above), as in the PM. The final decision is a comparison between a threshold value λ_ν and ν; when $\nu > \lambda_\nu$ the detection answer is yes. When two stimuli are delivered at the same time, like audio (A) and tactile (T), we can simply use the sum of the two: $\nu_T + \nu_A > \lambda_\nu$.

Attractor-based neural network model (ANN)

We propose that the behavioral results of detection tasks with multi-modal stimuli can be interpreted within the framework of attractor-based neural networks (ANN). The neural network adopted here can be described with the three distinct stages used for the previous models (see Fig.1 of the main text); the first two stages (sensory and detection) are composed of two separate modules each, and the third (decision) stage has a single module (see Fig.4). Each module is composed of one excitatory and one inhibitory neuron population of single compartment leaky integrate-and-fire neurons that incorporate biophysically realistic parameters (Abeles (1991)). Neurons in this network are connected by three types of receptors that mediate the synaptic currents flowing into them: AMPA, NMDA glutamate, and GABA receptors. A detailed description of the neural and synapse model dynamics adopted can be found in the following sections (its most relevant features are also reported in Table 1 generated following the prescriptions of Nordlie et al. (2009)).

The main assumption made by this model is that the stimuli are processed within separate channels during the sensory and detection stages, and that these signals are integrated, in a nonlinear fashion, only in a final, decisional stage.

The first modules, or sensory stage mimics the behavior of neurons in sensory areas and aim to reproduce the activity patterns described in De Lafuente and Romo (2005) for the somatosensory module, and in Bendor and Wang (2007) for the auditory module. The detection modules in the second stage represent an intermediate step between the sensory stage and high level brain areas described in neurophysiological experiments as the seat of decision processes (e.g. de Lafuente and Romo (2006)), and theoretically analyzed in Deco et al. (2007). One detection module was implemented for each modality. Finally, the third, decisional stage, encodes the output/decision of the detection task.

All parameters were set according to Abeles (1991), with the exception of the recurrent synaptic weights of the selective excitatory populations and the synaptic weights between selective excitatory populations of different layers. Sensory layers' excitatory recurrent synaptic weights were set to $w = 1.1$, for the detection stage layers were set to $w = 1.54$, and for the decisional stage layers were set to $w = 1.52$. Synaptic weights between the sensory and detection stages where set to $w = 0.07$, and between detection and decisional stage where set to $w = 0.11$. Once the parameters of the sensory, detection and decision modules had been set, we chose the values of the inter-layers connections in order to emulate the neurophysiological results observed in De Lafuente and Romo (2005); Bendor and Wang (2007). All the parameters of the network are reported in Table 2.

This model assumes that the strength of the input impinging on the excitatory population of the sensory module is proportional to the strength of the stimulus (just like, for example, these types of stimuli are encoded in S1, i.e., the input to MPC is transmitted from S1). The detection and decisional stages have each a neural network implementing a decision-making process (Wang (2002)). The behavior of the modules in the detection and decision stages is bistable as a function of the activation state of the excitatory population, that is, either lowly activated (corresponding to no detection), or highly activated (corresponding to the detection). The excitatory pools of neurons encode the detection, and its neurons have strong excitatory recurrent connections. When external input is delivered the activity of the neurons in the corresponding pool increases, causing also a subsequent enhancement of the inhibitory connections.

We propose that the perceptual response results from neurodynamical bistability. In this framework, each of the stable states corresponds to one possible perceptual response: 'stimulus detected' or 'stimulus not detected'. In fact, the probabilistic character of the system results from the stochastic nature of the networks. The impinging random spike train together with finite-size effects are the sources of this stochasticity. Thus, for weak external inputs the network has one stable state, and the excitatory pool fires at a weak level (spontaneous state, around 1 Hz). This spontaneous state encodes the 'stimulus not detected' state. For stronger external input a state corresponding to the strong activation of the excitatory pool emerges. We call this excited state encoding the 'stimulus detected' state.

Spiking simulations were implemented in custom C++ programs. For the spiking simulations we used a second-order Runge-Kutta routine with a time step of 0.02 ms to perform numerical integration of the coupled differential equations that describe the dynamics of all cells and synapses. The population firing rates were calculated by performing a spike count over a 50 ms window moved with a time step of 5 ms. This sum was then divided by the number of neurons in the population and by the window size.

As mentioned above, one of the main sources of discrepancies in the interpretation of these models is the confusion regarding their application to different paradigms, and how the change in paradigm affects the assumptions regarding the optimal strategy for detection. In this sense,

similar interpretations have been given for the yes/no and the 2IFC paradigms, and we think that it is important to briefly explain how the models proposed so far need to be adapted to correctly model the detection process in 2IFC paradigms.

Mathematical models of detection behavior for 2IFC tasks

In the two-interval forced choice (2IFC) paradigm each trial consists of two temporal intervals and, for the detection task, the observer is requested to indicate in which of the two intervals is the stimulus present. For different reasons, discussed below, the 2IFC is widely used in psychophysics research to measure the sensitivity of the observer. In this sense multi-sensory research is not an exception.

As mentioned before, when the stimulus is a compound of two modalities, there are many ways to separate signal(s) from noise(s), and with the 2IFC the task acquires a new degree of freedom as the observer can choose between different strategies to solve the task. The strategies commonly used can be divided in two groups: (1) The observer compares informations from the two intervals and then makes a unique decision or (2) the observer makes a decision at each interval and then integrates these decisions. A schematic representation of these strategies is illustrated in fig.2A, that represents the probabilistic sum model (PSM) as an example of the first strategy and the difference model (DM, as an example of the second strategy) in a 2IFC.

Pythagorean model (in a 2IFC)

An example of the first group of strategies (i.e., independent decisions and then integrate) is the so-called Difference Model (DM) and its pictorial representation is shown fig.2A. This is just an illustration assuming single modality events to exemplify how PSM and DM work in a 2IFC. The main assumption characterizing this model is:

As1. Subject compares the sensory activity of the first interval, s_1, to the one of the second interval, s_2, and calculates the difference between them, $s_1 - s_2$. The sign of the difference determines the answer: First (or second) interval when it is positive (or negative);

Apart from this, the other assumptions are commonly used in multi-sensory literature to characterize this model and we therefore adopt here (note that the subindices denote first and second interval of the 2IFC):

As2. s_1 and s_2 are Gaussian distributed with unitary deviance;

As3. s_1 and s_2 have unitary deviance;

As4. s_1 and s_2 are statistically independent.

Note that even relaxing these three assumptions, the DM could be represented by the pictorial description of fig.2A. While the first two assumptions are helpful to generate the optimal strategy for independent signals, the third assumption has been used in order to use the model as a

benchmark for early-sensory interactions (see below and the Discussion). One of the two sensory activities is the signal with mean d' and the other is the noise with mean 0. In this case the relationship between detectability and the probability of correct answer, P^{cor}, is given by:

$$P^{cor} = \sqrt{2}\,\Phi(d') \tag{16}$$

The symmetry between the first and the second interval is part of the assumptions, even though it is not undisputed (Yeshurun et al. (2008)).

In order to model the behavior with compound stimuli, further assumptions are necessary, and we will describe one of the possible implementation of this extra assumptions, the Pythagorean model (PM).

The PM is chosen on the basis that its performance is optimal and equivalent to the likelihood-ratio estimate (LRE). To characterize the PM in a 2IFC the following assumptions are necessary:

As5. The sensory activities from each modality are linearly summed

As6. The weights of this linear sum are the respective d' of each modality (Wickens (2001))

The decision rule in this case would be:

$$d'_T\, s_{T,1} + d'_A\, s_{A,1} > d'_T\, s_{T,2} + d'_A\, s_{A,2}$$

As in the yes-no task discussed earlier, the detectability parameter d' of the compound stimuli is the root square of the squared sum of each individual detectability score (Green and Swets (1966)):

$$d'_{PM}(A,T) = \sqrt{d'^2_A + d'^2_T} \tag{17}$$

A schematic representation of the PM for the 2IFC is presented in fig.2B. The same figure shows a schematic representation of the PSM, described in the following paragraph, to help the comparison of these two models.

Probabilistic summation model (in a 2IFC)

Let us now analyze the most commonly used model belonging to the second kind of strategy described above, that a separate decision is made at each interval, and the final decision is then selected upon the two decisions through a probabilistic choice. This strategy does not make any restriction on the way the separate decisions are made on each interval (for example PSM, MM, PM or OM seen for the yes-no task). Even for this case, the model is referred to as 'Probabilistic summation model' (PSM).

First, consider a 2IFC for a single-modality, and let us call P_T^{2IFC} the probability to correctly individuate the interval when, for instance, a tactile stimulus (T), has been presented. This model is commonly characterized (e.g., Green (1958); Wickens (2001); Marks et al. (2011)) with the following assumptions:

As1. The detection process is binary, that is, either the observer perceives the stimulus or she does not. So, let us call P_T^d and P_T^{Nd} the probabilities of this two outcomes, respectively, whose sum is equal to 1: $P_T^d + P_T^{Nd} = 1$.

As2. The probability of false alarm, P^{fa}, is zero (but see below for a model with $P^{fa} > 0$).

As3. When the observer fails to detect the stimulus, with probability P_T^{Nd}, she must guess and, she will give the correct answer on half of the trials, if this guess is unbiased.

With these assumptions the probability of a correct answer for a 2IFC is then:

$$P_T^{2IFC} = P_T^d + P_T^{Nd}/2 = (1 + P_T^d)/2 \tag{18}$$

Let us now analyze the case of a compound audio-tactile stimulus, completely correlated (they are presented in the same interval) and synchronized. To this aim, a further assumtion is needed:

As4. Two stimuli are detected in a completely independent way.

Therefore we have 4 possible outcomes: full perception of tactile stimulus, but not the audio one (P_T^d and $1 - P_A^d$), full perception of audio, but not tactile (P_A^d and $1 - P_T^d$), full perception of both audio and tactile (P_A^d and P_T^d), and no perception of either one ($1 - P_A^d$ and $1 - P_T^d$). So the probability of detection for the case of a bimodal stimulus in a 2IFC becomes:

$$P^{2IFC}(A,T) = (1 + P_T^d + P_A^d - P_T^d\, P_A^d)/2 \tag{19}$$

As for the yes-no task we can modify the PSM for the 2IFC by relaxing the fourth assumption regarding the independence of the two modalities, assuming that P_T^d can condition P_A^d, and viceversa.

Non-indpendent PSM for 2IFC tasks

We adopted the same nomenclature of the PSM for non-independent modalities of the yes-no task. The probability to correct answer for compound stimuli (A+T) then becomes:

$$P^{2IFC}(A,T) = 1 + (1 - P_T^{2IFC})\,(1 - P_{A/\bar{T}}^d)/2 \tag{20}$$

From this equation it is clear that the probability to detect the compound stimulus is higher when $P_A^d < P_{A/\bar{T}}^d$.

In the description of the PSM for 2IFC we have assumed (As2) that the false alarm probability was zero. However, when the possibility of false alarms is introduced in the probability summation model, eq.19 needs to be adapted.

PSM for 2IFC tasks with non-zero false alarms

For the PSM with non-zero false alarms, the new assumptions can be stated in this way:

As1. The detection process is still a bistable process when a stimulus is effectively presented, that is the observer perceives the stimulus P_T^d or not, P_T^{Nd}.

As2. During the interval without the stimulus the observer can perceive (hallucinate) it with a probability P_T^{fa}.

As3. The observer will answer correctly in trials with correct detections and without false-alarm, or will respond randomly in trials with correct detections and false-alarms, or finally randomly as well in trials without both correct detections and false-alarms.

The probability of a correct answer for a 2IFC for single modality is then given by this formula:

$$P_{A+T}^{2IFC} = (1 + P_T^d - P_T^{fa})/2 \tag{21}$$

With the same assumptions, with two modalities and applying algebra we can obtain:

$$P_{A+T}^{2IFC} = (1 + P_T^d + P_A^d - P_T^d\,P_A^d + P_A^{fa}\,P_T^{fa} - P_T^{fa} - P_A^{fa})/2 \tag{22}$$

Thus we see how for the PM the adaptation to a 2IFC implies a comparison of the sensory activity and as a consequence an increases of the d' respect to the yes/no paradigm. On the other hand hypothesizing a different state (or partial decision) at each interval implies for the observer to adopt a strategy based on the PSM. We will see in the 'Theoretical analysis' section how the PM and the PSM are related; more importantly we will show the relevance of knowing whether the observer is comparing the sensory activity or is making a partial decision at each interval in order to argue that the PM is the optimal strategy in 2IFC detection tasks.

Experimental methods

Ethics Statement

The experimental protocols were approved by the local Ethical Review Committee (CEIC Parc de Mar), and conform to the ethical standards laid down in the 1964 Convention of Helsinki.

Participants

Six participants took part in the experiments (5 women, age 21-25) and received monetary reward for their participation (10 euros per hour). All of the participants gave their informed consent prior to their inclusion in the study and reported normal hearing. Six in experiment 1, five in experiment 2 and 3.

Apparatus

Participants sat approximately 50 cm away from a computer monitor in a sound attenuated, dimly lit room. Audio stimuli were delivered through headphones. Tactile stimulation was

delivered on the second phalanx of the middle finger of the right hand using a probe with a 1mm ø probe moved by a custom built vibratory device mounted on a table (Dancer Design, Vibrotactile Laboratory System, VLS; Liverpool, UK). The stimulation device consisted of a high precision vibrotactile actuator (amplitude error was about +/- 3% of the measured value). We used a custom made finger-pad in order to reduce non desired/uncontrolled finger movements, that were further controlled by establishing a threshold criterion for baseline probe displacement (see 'Experimental Procedures').

Stimuli

Auditory stimuli with frequencies above the flutter range (250, 500 and 1100 Hz) were pure tones. Auditory stimuli within the flutter range (13, 31 and 49 Hz) were generated with pulse trains. These pulses were sinusoidal fragments (at a characteristic frequency of 5000 Hz) of 4 ms length modulated by Gaussian envelopes. These stimuli were modelled after Liang et al. (2002). During the stimulation interval, background white noise was delivered at 70 dB intensity in order to mask any possible noise produced by the stimulator device. We briefly tested (20 / 30 trials) that the participants were at chance level when the audio stimuli volume were at its minimal values. Tactile stimuli were delivered through a 1mm ø rounded tip metal probe, in contact with the second phalanx of the middle finger of the right hand. The probe moved following a sinusoidal waveform generated with Matlab 9.1 software and delivered through a Texas Instruments motherboard that controlled the operation of the tactile stimulator.

Experimental procedures

All three experiments involved a two interval two-alternative forced-choice (2IFC) paradigm. Each trial was composed of two intervals, the stimulus was presented only in one of them and observers had to report, by pressing a specified key on the keyboard, through a keyboard in which interval they perceived the stimulus. The stimulus lasted 500 ms, centered around an interval between 700-1100 ms. Participants underwent multiple experimental sessions (between 3 and 6) of two hours each, on different days. Subjects were trained during the first three sessions and these data were discarded from the analysis. Depending on the experiment, there were a minimum of two and a maximum of three such experimental sessions per condition. In experiments 1 and 2 both audio (A) and tactile (T) stimuli were presented, whereas experiment 3 involved compounds of auditory tones at different frequencies.

Each experimental session began by determining the stimulation amplitude threshold at which A and T stimuli were detected with 70% accuracy level in the 2IFC detection task (estimated using the Quest adaptive procedure). The sensory thresholds were considered stable only when the Quest procedure gave the same values for more than 2 sessions. After this phase (45 min on average), the session consisted of several test blocks of 75 trials each (approximately 4/5

blocks, depending on time available). For experiments 1 and 2, the modality to be presented in each trial was made unpredictable by randomly and equiprobably selecting A, T or A+T stimuli on a trial by trial basis. Participants were instructed to be attentive to both modalities. For experiment 3, after identifying the amplitude producing 70% performance for each sound frequency (500 and 1100 Hz), participants performed only the combined condition (that is, with compound auditory stimuli of 500 Hz and 1100 Hz).

Tactile thresholds are very sensitive to variations due to training, fatigue, temperature, lack of sleep, changes of mood, position, etc (see e.g., Green and Stevens (1979)). In order to obtain a measure as stable as possible, the voltage that was necessary to displace the probe, reported by the tactile stimulation device, was used to guide the indenting of the probe in the skin, and to discard sessions where the participant had changed the force exerted over the probe above a predefined tolerance criterion (see below). In particular, the following procedure was employed every time the observer needed to move her hand and to set up the initial position of the probe: First, the voltage necessary to generate a tactile vibration (31 Hz) without contacting any surface was obtained. Second, the probe was progressively moved into the skin until observing an abrupt increase in the voltage necessary to generate the 31 Hz stimulus (about a 30% voltage increase). And finally, once the mentioned voltage increase was observed, the probe was displaced 500 μm away from the contact point. This procedure was repeated up to three times, until the error range was not larger than a few tens of microns. In order to establish a tolerance criterion for variations in the force exerted over the probe, the voltage necessary for a 31 Hz stimulus with a supra-threshold amplitude (100 μm) was obtained at the beginning of each session, and this measurement was then repeated every 15 trials throughout the whole block. When the difference between the voltage exceeded 4% from test to test the last 15 trials were discarded and repeated.

Although we have not compared how the control of this variability affects our results, it was evident from our interaction with the participants that thresholds were very sensible to environmental conditions (mainly temperature) which required some time to be stabilized, given the observed adaptation of the thresholds in time, it is undeniable that results would have been affected.

Results

Theoretical analysis

We report now a reanalysis of part of the results reported by Schnupp's and Wilson's studies (Schnupp et al. (2005); Wilson et al. (2009, 2010a,b)) because the results from these studies are in apparent contradiction. Namely Wilson and colleagues reported super-additive performance for a 2IFC detection task above the PM's criterion, while Schnupp *et al.*, using a yes-no paradigm, reported detection performance above the PSM's criterion, but well below the PM's criterion.

Schnupp et al. (2005) measured detection performance for paired ensembles of auditory, visual and tactile stimuli, with different amplitude values in a yes-no task. Stimuli were selected from a set of 64 (8-by-8) combinations arising from crossing 8 amplitude values (including the zero value) per each modality; therefore there are 49 actual bimodal stimuli, a subset of 14 single modality stimuli, and one no stimulation condition. In order to better understand Schnupp *et al.*'s results, we further analyzed their data by comparing the three models, LSM, PSM and OM, on the basis of their capacity to fit the detection probabilities for the 49 possible multi-sensory combinations based on the single modality detection data. For the test phase (see fig.3) against multi-sensory data we excluded the aforementioned single modality combinations, which we used only to find the best parameters for the models.

As reported in Schnupp et al. (2005), for the LSM, in most cases (12 out of 17) the observed data fell short of the model's prediction ($p > 0.05$). They also reported that both the PSM and the OM (their own proposed model) provided a good fit of the observed data, and in all the observers the deviance statistic for both models was below what is expected by chance ($p > 0.05$). However the authors showed that PSM and OM produced quite different fits: the PSM produced worst fits than the OM (higher deviance in 14 over 17 observers, $p = 0.0148$, Wilcoxon sign rank test). In order to better understand Schnupp *et al.*'s results, we further analyzed their data by comparing the three models, LSM, PSM and OM, on the basis of their capacity to fit the hit probabilities for the 49 possible multi-sensory combinations (as said above we excluded the 15 combinations yielding unisensory or no-stimulus conditions, which were used only to adjust the model's parameters). The equations used for the PSM, LSM and OM are eq.6, 13 and 15 in the Materials and Methods section. For this subset of conditions we calculated the differences in detection probabilities between the prediction of each of the three models and the experimental values, see Fig.3 panels a-c. As illustrated in Fig.3 the best fit to the empirical results is achieved by the OM (panel a): The differences between the hit probability predicted from the OM and the experimental values tend to zero. The PSM (panel b) has a tendency to under-estimate the detection probabilities (reddish color). Finally, the LSM over-estimates the double modality detection probabilities (blueish color of panel c) in the majority of the cells.

To help interpreting these results we presented in Fig.3d the bimodal detection probabilities (bi-modal vs unimodal) for a yes-no task generated with numerical simulations for the three statistical models LSM, PSM and OM, for different values of probability of unimodal correct detections. Our objective was to compare the bi-modal detection probability of the three models given the same unimodal detection probabilities. For the sake of simplicity, we only take into consideration the cases where the two modalities have identical unimodal detection probabilities. To obtain the value for PSM's bimodal detection probability, given the unimodal detection rates, eq.4 can be used, with equal detection probabilities for the two modalities. For the LSM and OM we have used eq.11 and eq.15 respectively. The values of $b_{T(A)}$ and λ were obtained from the additional hypotheses that the false alarm rate was around 3%, that is in accordance

with the experimental data reported by Schnupp *et al.*. Thus, ordering the models in terms of detection probability we can see that $P_{LSM} > P_{OM} > P_{PSM}$, which, together with the empirical results shown in panels a-c clearly indicates that the experimental data lies in an intermediate level between the PSM and LSM's predictions. In other words Fig.3 does not only reveal the model that best fits the behavioral data, but also the sign of the prediction mismatch between each model and the observed data, positive (overestimation) for the LSM and negative (understimation) for the PSM.

One interesting aspect of the OM is that it can account for detection results without requiring early interactions. However, one limiting aspect of this model, if it is to be taken as an implementation of the detection process, is that its proposed interactions are not feasible in terms of neurophysiological processes. The performance of the OM can be matched by a model grounded on different assumptions and having a straightforward mechanistic implementation, namely, the MM, which provides a similarly good statistical description of the results and, in addition, it can be implemented with ANN (see Fig.4). The MM assumes a separate intermediate detection stage for each modality, similar to the PSM. The 'detected/not-detected' states are encoded into the activity of pools of neurons as high/low firing rates. A more detailed description of the MM is provided in Materials and Methods. As mentioned there, it was named mixed model because it includes features from both the PSM and PM, we refer to this model as the MM.

The MM model has been devised in order to achive two objectives: (1) Having performances ranging between the PSM's and LSM's. (2) providing a straightforward mechanistic implementation, namely, the ANN, which provides a similarly good description of the results in a neural network. The ANN adopted for this model has the appeal that stands in its good level of biological plausibility and the good results achieved in modeling both behavioral and neurophysiological results of decision making and detection task (Wang (2002); Deco et al. (2007)).

It is worth asking, similarly to the OM, whether detection performance for the MM can be higher than the PSM's. To answer this question we reasoned as following: As mentioned above the probability distribution functions of the detection stage output, for both the detected and non-detected cases, are $f_l(\nu)$ and $f_h(\nu)$ respectively; where the output variable is ν and the subscripts l and h indicate high and low state respectively. The probabilities to have a low and a high state are P_l and P_h (for more details see Materials and Methods). So, when two stimuli are delivered at the same time, like auditory (A) and tactile (T), we can simply use the sum of the two: $\nu^A + \nu^T > \lambda_\nu$. Therefore under presentation of the stimulus the probability to detect it is:

$$\begin{aligned} P_{Det}^{A+T} = & \int\int_{(\nu_A+\nu_T)>\lambda_\nu} P_h^A f_h(\nu^A) P_l^T f_l(\nu^T) + P_l^A f_l(\nu^A) P_h^T f_h(\nu^T) + \\ + & \quad P_l^A f_l(\nu^A) P_l^T f_l(\nu^T) + P_h^A f_h(\nu^A) P_h^T f_h(\nu^T)\, d\nu^A d\nu^T \end{aligned} \tag{23}$$

For the PSM $f_l(\nu)$ and $f_h(\nu)$ are delta functions with non-zero values only at ν_l and ν_h, and

$\nu_l^A + \nu_l^T < \lambda_\nu$. As a consequence, the last term in the above formula equals zero:

$$\int\int_{(\nu_A+\nu_T)>\lambda_\nu} P_l^A f_l(\nu^A) P_l^T f_l(\nu^T)\, d\nu^A d\nu^T = 0$$

On the contrary for the MM this term is not zero which is precisely what allows this model to outperform the PSM. In Fig.3d we have shown that this model can match the results of the OM, but with completely different assumptions.

Indeed capitalizing on the working assumptions of the MM, we implemented it in an attractor neural network (ANN) Wang (2002) model with parameters derived from actual physiological data (see Fig.4 and a detailed description of the ANN model is reported in Materials and Methods). This neural network, akin to the MM, comprises three stages (see Fig.5b), and assumes that stimulation is processed along separate channels prior to the decisional stage, where information is finally merged. Again, the first stage of the model corresponds to processing in the sensory stage (corresponding to primary sensory areas), the second stage mimics the perceptual processing and detection of each modality in a separate pathway, and the final stage incorporates perceptual decision. For more details on the ANN see Materials and Methods.

To illustrate the feasibility of the ANN, we simulated a 'psychophysical curve', that is the probability of 'detection' for a given amplitude of a stimulus of each modality with a yn paradigm. This curve was then compared with those simulated using the OM and the PSM (see Fig.5) for different values of stimulus intensity. The dashed gray line is the fit for the prediction of the ANN single modality (gray circles) with a cumulative normal distribution function. For simplicity we used two modalities having identical performance. The predictions by the OM (black line) and by the PSM (gray line) are generated on the basis of this fit. The ANN detection probability for the bimodal stimulus (empty circles) is very similar to the OMs and a few points higher than the PSM. This suggests that the OM and ANN model can reproduce the detection data with similar performance.

As mentioned before, we are interested in the comparison of the detection data of yes-no tasks with 2IFC tasks, as they are in apparent contradiction. Schnupp *et al.*'s results indicate that the ability to integrate the multi-sensory signals is below the LSM's and therefore below the PM's prediction. In a more recent set of studies, with a different paradigm, but again with a multi-sensory (audio-tactile) detection task, Wilson and colleagues reported super-additive (as indicated by the violation of the PM's criterion), frequency specific interactions between audio and tactile processing in a detection task using stimulus frequencies ranging from 100 Hz to 1000 Hz (Wilson et al. (2009, 2010a,b)). We address this discrepancy on the following analysis.

We claim that the assumption that the PM's strategy is optimal for 2IFC paradigms, is the reason behind the apparent contradiction between Schnupp *et al.*'s and Wilson *et al.*'s results. It should be kept in mind that the PM represents the optimal strategy for detecting a stimulus in the following situations: (1) In a yes/no task with bimodal stimuli: The $d'-$weighted linear sum of the stimuli strength (when compared to other strategies that reach the same performance

with unimodal stimuli in a yes/no task); and (2) in a 2IFC task with unimodal stimuli, as it represents the strategy of comparing the strength values of the two intervals (signal vs noise, again compared to other strategies that reach the same performance with unimodal stimuli in a yes/no task). As we see, in both cases the key fact is that we are comparing strategies known to work equally well with unimodal stimuli in a yes/no task; or in other words models can be compared once the distance between the mean value of the stimulus distribution and the noise distribution (stimulus-noise separation) is fixed.

The paradigm analyzed by Schnupp *et al.* belongs to the first of the two situations we just described: It is a yes-no multi-sensory detection task and the various models are compared with the condition that all have equal performance in a yes-no unimodal detection task. As such, we can actually expect the PM to be the best strategy for their paradigm (as we showed in that case, even the LSM is well above the predictions of the PSM). But for the paradigm used by Wilson and colleagues, the models compared reach the same performance in a 2IFC unimodal detection task, and not in a yes-no unimodal detection stimuli. In such a situation we can have for example an observer adopting a strategy based on the PM and another one using a strategy based on the PSM (see fig.2). If these two observers have the same 2IFC performance, then their stimulus-noise separation will differ. And being the PM a better strategy than the PSM, the stimulus-noise separation of the observer using PM will be smaller than the PSM observer's one. This higher value of the stimulus-noise separation for the PSM compared to the PM will generate better performance for double modality stimuli detection.

To clarify this point, let us provide a simple numerical example: Consider individual hit rates of 70% in a 2IFC unisensory detection task. The PM predicts a performance of around 77% for bimodal stimuli. The PSM's prediction, if we assume that the false alarm rate is zero, for bimodal stimuli, is of 82% (see Material and Methods), and therefore PSM's predicted d'-ratio ($\frac{d'_{exp}}{d'_{PM}}$) is around 1.2. Despite the latter, given that both the actual false alarm rate and the way models can be influenced by it are unknown, we cannot make firm conclusions about the suitability of the PSM at this point. Indeed, considering false alarms (see Supp.Materials) around 3% (as in the Schnupp *et al.* dataset), the PSM predicts about an 81% performance, and its d'-ratio is still around 1.2.

The following experiments aim at submitting the ideas suggested by this analysis to critical empirical tests. The following experiment 1 aimed to control whether the apparent contradiction between the results of Schnupp *et al.* and Wilson *et al.*'s studies stems from differences in stimulus predictability between paradigms.

Experiment 1: Does audio-tactile interaction lead to an enhancement of stimulus detection?

In this experiment we analyzed detection performance for audio-tactile and unimodal stimuli using three audio (A) frequencies (13, 31 and 49 Hz) and one tactile (T) frequency (31 Hz),

namely unimodal A13, A31, A49 and T31. Just one audio frequency was tested per session and day, but the three stimuli (A, T and A+T) were interleaved. Once unimodal thresholds were established for A and T stimuli, interactive effects were measured using intermixed presentations of A, T and A+T trials in several successive blocks of 75 trials, as explained in 'Experimental Procedures'. Fig.6 summarizes the main results from experiment 1. The increase in detectability for the combined A+T condition is depicted as a percentage of correct answers (Fig.6, panels a-g), and as the ratio between the experimental d' (d'_{exp}) for A+T combinations and the predicted d' using the PM (d'_{PM}; see Fig.6h), which was derived from the unimodal d' empirical values according to eq.17 (see Materials and Methods). The panels d and h show respectively the accuracy and d'-ratio averaged over observers for the different auditory frequencies.

The values found here for the d'-ratio $\left(\frac{d'_{exp}}{d'_{PM}}\right)$ are clearly well above one (two-tailed signed rank Wilcoxon test, $p < 0.01$). A ratio of one implies that the observed result equals the value predicted by the benchmark, and ratios above one imply that observed result surpasses the benchmark value and would in principle lead to conlude a sensory integration. This pattern is very similar to that reported by Wilson *et al.*, and not statistically different from the results we obtained in our own lab using a 250 Hz stimulus (two-tailed Mann-Whitney-Wilcoxon, $p > 0.5$). Of course, despite these similar results, a key difference between Wilson *et al.* and our paradigm lies in the predictability of the modality of incoming trials, unpredictable in our case, but fully predictable in theirs. This may have produced different expectancy conditions (for example, observers could get ready for a particular stimulus modality in Wilson's experiment, but had to prepare for both in ours). In order to clear out any possible effects related to this difference, we replicated the experiment introducing a predictable and an unpredictable condition. We did not find any statistical difference between the performances in the two paradigms (two-tailed Mann-Whitney-Wilcoxon, $p > 0.5$). In summary, and in agreement with the results shown by Wilson *et al.*, the present results indicate that the detectability of the compound multisensory stimuli is higher than the Pythagorean sum of the individual detectability scores of each stimulus modality. Secondly, and contrary to Wilson's results, the effect is not frequency dependent, indeed we did not find any statistical difference between the three frequencies (two-tailed Mann-Whitney-Wilcoxon, $p > 0.5$). That is, a similar enhancement was observed for every frequency combination.

Our results showed that empirical detection probabilities were very similar to those reported by Wilson and colleagues, therefore excluding the paradigm as the cause of these contradictory results between Schnupp et al. (2005) and Wilson et al. (2009), and pointing to the underlying model instead.

To further test our framework, we based the following experiments on neurophysiological evidence (de Lafuente and Romo (2006) for somatosensory; Bendor and Wang (2007) for audio) that neural activity related to stimulation in sensory areas is present only when the stimulus itself is present: We hypothesized that if the interaction takes place in sensory areas its strength

will depend heavily on the temporal overlap between the two stimuli. Therefore, if a violation of the Pythagorean criterion is observed even when a long empty temporal interval separates the stimuli in different modalities, it will be difficult to maintain the claim that this criterion is a viable baseline to benchmark a genuine sensory origin for multi-sensory interactions in detection. Instead, the MM and OM models clearly allow the possibility of interaction at long SOAs, even though for slightly different reasons: The OM has to assume memory of the stimulus value, whilst ANN has to assume memory of the detection. Accordingly, in experiment 2, we used a bimodal audio-tactile detection task introducing a range of stimulus onset asynchronies (SOAs) between the two sensory events forming the bimodal stimulus as long as 1s.

Experiment 2: Audio-tactile interactions across variations in SOA

Due to the temporal course of low-level sensory processing in auditory and somatosensory sensory areas (Bendor and Wang (2007); De Lafuente and Romo (2005)), different types of multi-modal interactions take place on different time windows. Thus, depending on the width of these temporal windows, one can maintain the assumption of early-sensory (for time windows amounting to tens of ms) or instead infer the presence of late-decision interactions (for time windows amounting to several hundreds of ms). In experiment 2 the interest was thus on determining whether interactive effects, as signaled by a violation of the Pythagorean criterion, were observed for audio-tactile stimuli at long SOA values. To this end, 9 different SOAs ranging from -1 to 1s were used, that is we inserted a stimulus-free time interval ranging between 0 and 1s between the onsets the audio and tactile stimuli. If interactive effects are observed for SOA values as large as -1s and/or 1, it would be difficult to claim that such effects are attributable to interactions at early sensory areas, as neurophysiological data show that these brain regions have barely any persistence of activity in the absence of external stimulation.

The results (see Fig.7) are shown as the average over the five participants. The top panel shows the percent-correct scores across all audio-tactile SOA values plus unimodal A and T conditions. As can be seen, the enhancement on the audio-tactile stimuli detectability with respect to the unimodal conditions is stable across all the SOAs up until SOA values of ± 1 s (Friedman test, $p > 0.1$). The bottom panel plots the d'-ratio (d'_{exp}/d'_{PM}). The detection enhancement obtained for audio-tactile stimuli, as compared to unimodal conditions, was evaluated against the prediction of the PM. In general, and for each SOA value, the d' ratio is not statistically different from the results of experiment 1 (Friedman test, $p > 0.1$) and, critically, it is superior to the Pythagorean prediction (bootstrapping from our empirical distribution a dataset with 1000 datapoints of the d'-ratio: Its mean value was 1.3 and its 95% confidence of interval is [1.2; 1.4]). According to a strict adherence to the PM, and to the results of experiment 2, we would need to accept that sensory integration occurs for bimodal compounds even under conditions where stimuli were presented as far as 1s apart. However, this is rather incompatible with physiological knowledge; it is well known that brain areas supporting early sensory processing do not sustain

stimulus-driven activity after the stimulus has been physically removed. Therefore, the validity of the PM so often used as a benchmark to test integration is clearly put into question by these results.

To reinforce the latter rationale, in experiment 3, we resorted to unimodal compounds of auditory stimuli rather than bimodal ones. There are well know within-modality sensory interaction phenomena, which depend on the temporal properties of neural responses in sensory areas (or earlier, see e.g. Fastl and Zwicker (2001)). The main aim of experiment 3 was therefore to further test the assumption that early sensory interactions disappear at long SOAs, and moreover, continue to put the PM and PSM to test.

Experiment 3: Auditory multicomponent stimuli

Experiment 3 was devised to provide a further validation of the rationale in experiment 2 that the use of a long SOA condition serves as a test for early integration. In within-modality detection, sensory interactions are usually of a competitive nature, so that two sounds of different frequency tend to mask each other and they do so maximally when overlapping in time. To this end, we compared two experimental conditions involving presentation of within-modality compounds of two auditory stimuli, one of 500 Hz and one of 1100 Hz, as used in the seminal work of Marill (1956). We are aware, that contrary to what was claimed by Marill (1956), more recent papers report increases in detection with multicomponent signals with respect to single component signals (see e.g., Dubois et al. (2011)). Other studies (see e.g., Thompson et al. (2013)) report detection data that can be explained with the 1-look model. However, this debate, interesting as it may be, it is completely out of the goal of the present paper. We only choose this experimental condition, because we were certain that the detection performance was affected by an early-sensory interaction (see e.g., Fastl and Zwicker (2001)). This interaction is strong enough to imply a detection performance well below the PSM's criterion (Marill (1956); Dubois et al. (2011); Thompson et al. (2013)). In the measurement session each block was composed of 75 trials of the combined condition (500+1100), randomizing simultaneous presentation with +750 ms SOA trials where the two tones were offset in time. Here, any interactive effects should be seen when the stimuli are presented in synchrony but the long SOAs condition should strongly reduce the possibility of any interactive effects at early sensory stages.

The results of experiment 3 are summarized in Fig.8. The percent-correct scores of all but one of the participants were significantly higher for the long SOA than for the synchronous condition (Wilcoxon test $p < 0.005$). This indicates that the two auditory stimuli, when delivered in synchrony, engaged in an interaction, in this case of a competitive nature, actually hindering detectability of the compound stimulus. From these scores, we calculated the d'-ratio between the result observed for the compound stimulus, and the prediction based on the two unisensory detection rates according to the PM (Fig.8, panel on the right). One can see again that the average value of the d'-ratio, for the condition with SOA=750ms, is similar to the ones reported

in our two previous experiments (bootstrapping from our distribution a dataset with 1000 data-points of the d'-ratio: Its mean value is 1.36 and its confidence of interval is [1.10; 1.52]). That is, at SOA=750ms, detection of the compound surpasses the PM's criterion. On the contrary the average value for the d'-ratio, for the condition with SOA=0ms, is around 0.7. That is, at SOA=0ms, the compound is detected much less than predicted by the PM.

These analyses confirmed that the detectability of the compound auditory stimulus presented in synchrony decreased with respect to that of the probabilistic sum (PSM) of individual stimulus detectabilities. However, when a long delay was introduced between the two stimuli the interaction disappeared, so that the detection performance for the compound stimulus with long SOA was comparable to the levels predicted based on single-stimulus detection assuming no interaction (e.g., by the PSM).

It is interesting to note that this value of d'-ratio does not seem to depend on the particular modality pairing (audio-tactile or audio-audio). Taken together, the results of Exp.3 suggest that the d'-ratio around 1.3 is to be expected whenever we have a double-stimulus (compound) detection task without any kind of sensory interaction (be facilitative or competitive) at the sensory level. That is, this value of d' provides a baseline for testing additive interactions. More importantly, these results support the claim that the multi-sensory (experiment 2) and unisensory (experiment 3) results with long SOAs are a valid demonstration of the inadequacy of the PM as a benchmark for testing early sensory integration of bimodal stimuli, and are indeed in contradiction with the hypothesis of early interactive effects between the two stimuli (A+T or A+A).

Discussion

Based on our theoretical analysis of the statistical models that are often used to test for early integration in detection tasks and a reanalysis of Schnupps dataset, we obtained two main results: First, the PM does not always represent the optimal strategy in a 2IFC, given the probability of a correct answer in a 2IFC detection task for unimodal stimuli. This result is relevant as the PM is used under the assumption that it represents the optimal strategy in the detection of multimodal respect to unimodal stimuli under the assumption of no integration. As we have shown, for the PM, optimality depends on the strategy used by observer when detecting unimodal stimuli, in absence of any information about the observer's strategy the use of this model as benchmark can be misleading. In the literature, for example, for 2IFC detection tasks, both the DM and the PSM are used to describe observer's strategies (compare for example Wilson et al. (2009); Schnupp et al. (2005) with Alais and Burr (2004); Wuerger et al. (2003); Meyer et al. (2005). Moreover these results explain the apparent discrepancy between Wilson et al. (2009) (where the observed data go beyond PM's criterion) and Schnupp et al. (2005) (where observed data fell below LSM' criterion); Second, we put forward the MM, a simple

model of 'late' multimodal interactions. The MM, like the OM, proposed by Schnupp et al. (2005), shows the best fit for the behavioral data in a yes-no detection task. However, as we showed, the OM cannot be a mechanistic model. We put forward the MM in order to capitalize on the OM's simplicity and power to establish a good non-interaction benchmark whilst at the same time having a direct translation to models based on well-known neuro-physiological processes. The MM shares characteristics of both the PM and the PSM. The main objective of this model is to show the inadequacy of claiming for early-sensory interaction based upon which ones multi-sensory behavioral results that surpass the PM on the PSM.

The three key elements of the MM proposed here are the following: First, sensory and detection processing occur independently for each modality, and information is merged only at the decision stage (like in the PSM). Second, the MM has a quasi-binary detection stage (wheres the PSM has a fully binary detection stage), meaning that the two states of the excitatory population can be considered as on/off states, but the distributions of the activities for these states have non-zero standard deviations. Third, the output from the detection stage, despite being bimodal, is proportional to the output of the sensory stage. Elaborating on this idea, we implemented the MM in an ANN Wang (2002) model with parameters derived from actual physiological data (see Materials and Methods). Interestingly most of the features of the ANN, such as the specific neuron dynamics implemented here, are not indispensable, in the sense that similar implementations can produce the same results, but they have to share the three indispensable features reported above. For example, in the ANN, the fact that the attractor dynamics generate two fixed points, and the distributions around them are a consequence of the internal (quenched and thermal) and the external (thermal) noise, underlies the resemblance between the behavior of its detection stage and that of the MM. The existence of a quasi-binary detection stage (second element) is the feature that is more feasible to be tested with a neurophysiological study, like the de Lafuente and Romo (2006) or Lemus et al. (2009), with compound audio tactile stimuli.

Of course, the MM is not the only possible model able to describe the multi-sensory detection results without hypothesizing interaction between modalities at an early stage. Indeed the models we analyzed are commonly used to benchmark the hypothesis that the interaction between the modalities take place in an early stage of processing. As such the lack of this interaction can be seen as a reason to include the assumption of modality independence in the benchmark models. However other mechanisms can induce dependency between the modalities, such as the allocation of the attention. Indeed attention oriented to one modality could influence the processing to the other modality. In order to understand what are the potential consequences of removing the independency between the two modalities, we adopted a slight variation of the PSM described in Materials and Methods[1]. From this description we cannot calculate quantitatively the prediction for the hit probability, but we can only indicate whether the hit probability

[1]We are grateful to an anonymous referee that suggested us to explore this possibility.

prediction of this model overcomes or not the PSM when the hit probabilities of both modalities are positively (or negatively) dependent. Indeed, the hit probability of the compound stimuli are higher when the probabilities of the single modalities are negatively dependent (see Materials and Methods).

Following the hypothesis that dependency between modalities is due to how attention is deployed (for example, paying more attention to one modality could decrease the amount of attention available for the other modality), as a consequence the hit probability of the second modality conditioned on a 'hit' of the first modality decreases. Such interaction can be described as a top-down interaction where the modalities are anti-correlated (see mathematical description in the Materials and Methods). Similarly, the probability of a correct answer to the compound stimuli is higher when the probabilities of both modalities are anti-correlated.

Even though our description of the non-independent PSM, both for yes-no and 2IFC paradigms, lacks quantitative predictions, we cannot rule out that it could actually be another possibility, as good as the MM, to interpret these multi-sensory detection data without implying early-sensory interactions.

Recent results reported by Otto and colleagues (Otto and Mamassian (2012); Otto et al. (2013)) are very relevant to the issue of the non-independency of the modalities. Indeed by means of a reaction-time (RT) paradigm, Otto *et al.* reported sequential effects on the detection's RT of the different modalities: They reported a strong negative correlation between response latencies for unisensory stimuli. The authors claimed that this correlation across trials can induce the effect of overcoming the Boole's inequality (Miller (1982)). Therefore is not difficult to see that the correlation across trials between modalities automatically discards interpretations based on sensory level interactions. However, the extent to which the results of Otto and colleagues translate to similar results with RT-detection tasks (Murray et al. (2005); Sperdin et al. (2009)) or to psychophysical detection paradigms, such as the ones reported in the present work, remains to be clarified.

To submit the idea of lack of early integration to critical empirical tests and to illustrate our approach to the measurment of interactive effects from multi-sensory stimuli in audio-tactile detection tasks, we presented three separate experiments. Exp. 1 demonstrated integrative effects with audio-tactile stimuli in the flutter range, in a paradigm akin to the one used by Wilson and colleagues (Wilson et al. (2009, 2010a,b)), but in which participants could not predict the target modality for each incoming trial. Despite this difference in the paradigm, detection probabilities in the three basic conditions (audio alone, tactile alone, and audio-tactile) were very similar to those reported by Wilson and colleagues, whom interpreted this results as reflecting an early sensory integration of auditory and tactile information leading to an enhancement in the detection of bimodal respect to unimodal stimuli. Exp. 2 sought to further test the accuracy of the prediction made by the PM by introducing a bimodal audio-tactile detection task across a range of stimulus SOAs, that is long intervals (around 1 s) between stimuli. Contrary to

what an account based on low-level sensory integration would predict, it was observed that detection performance for compound stimuli clearly surpassed the Pythagorean sum prediction for both short and long SOA values, as any sensory effects should have clearly faded in the latter case. As a consequence, even though our results cannot unambiguously indicate the level (or levels) of processing at which interaction occurs, they seem to exclude a rate-based multi-sensory interactions at the sensory level. Importantly the relevance of the results of Exp. 3, where a compound of two audio stimuli was used, cannot be understated. Indeed the results from Exp. 3 are solidly based on the fact that a two auditory stimuli with two different frequencies, when presented at the same time, interfere rather than facilitate the detection process (Marill (1956); Dubois et al. (2011); Kiang and Sachs (1968); Thompson et al. (2013)). It is well known that this inference is of an early sensory nature, even if the exact level at which it takes place is not completely clear (Kiang and Sachs (1968)). Exp. 3 demonstrated that, given the right conditions (two concurrent stimuli within the same modality) sensory interactions do in fact occur, and that, when the interaction between stimuli presented is of an early sensory nature, this very interaction can be suppressed inserting a long temporal interval between the stimuli. Furthermore, when the early sensory interaction fades out (with long intervals), the resulting performance levels matches that of multi-sensory compounds in the two previous experiments.

However we have to warn that these results and the associated conclusions are restricted to statistical models used to evaluate multi-sensory interactions in detection tasks and we do not make claims about other cognitive processes such as spatial representation (i.e. see the oretical work of Pouget et al. (2002)), or size estimations (Ernst and Banks (2002)).

Finally, it is important to clarify that the conclusion arising from the present study does not preclude the possibility of early level interactions per se. Rather, the main conclusion here is that in order to measure such interactions one has to use an appropriate baseline, and that past studies have often used baselines that tended to overestimate integration.

Acknowledgements

We are very grateful to Nara Ikumi and Xavi Mayoral for their invaluable help with the experimental design and testing, to Jan W. H. Schnupp for providing his data for reanalysis, and to an anonymous reviewer of a previous version of the manuscript, Marc Ernst and Miguel Lechón for their enlightening suggestions.

References

Abeles, A. (1991). Corticonics. Cambridge University Press, New York.

Alais, D. and Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection. Cognitive Brain Research, **19**(2), 185–194.

Alais, D., Newell, F. N., and Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. Seeing and perceiving, **23**(1), 3–38.

Arnold, D. H., Tear, M., Schindel, R., and Roseboom, W. (2010). Audio-visual speech cue combination. PloS one, **5**(4), e10217.

Bendor, D. and Wang, X. (2007). Differential neural coding of acoustic flutter within primate auditory cortex. Nature Neuroscience, (6), 763–771.

Dalton, P., Doolittle, N., Nagata, H., and Breslin, P. (2000). The merging of the senses: integration of subthreshold taste and smell. Nature neuroscience, **3**(5), 431–432.

De Lafuente, V. and Romo, R. (2005). Neuronal correlates of subjective sensory experience. Nature Neuroscience, **8**(12), 1698–1703.

de Lafuente, V. and Romo, R. (2006). Neural correlate of subjective sensory experience gradually builds up across cortical areas. Proceedings of the National Academy of Sciences, **103**(39), 14266–14271.

Deco, G., Pérez-Sanagustín, M., De Lafuente, V., and Romo, R. (2007). Perceptual detection as a dynamical bistability phenomenon: A neurocomputational correlate of sensation. Proceedings of the National Academy of Sciences of the United States of America, **104**(50), 20073–20077.

Driver, J. and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron, **57**(1), 11–23.

Dubois, F., Meunier, S., Rabau, G., Poisson, F., and Guyader, G. (2011). Detection of multicomponent signals: effect of difference in level between components. Journal of the Acoustical Society of America, **130**(5), EL284.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, **415**(6870), 429–33.

Fastl, H. and Zwicker, E. (2001). Psychoacoustics: facts and models. Springer.

Fetsch, C. R., DeAngelis, G. C., and Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. Nature Reviews Neuroscience, **14**(6), 429–442.

Frassinetti, F., Bolognini, N., and Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. Experimental Brain Research, **147**(3), 332–343.

Gescheider, G., Kane, M., C, S., and Ruffolo, L. (1974). The Effect of Auditory Stimulation on Responses to Tactile Stimuli. Psychonomic Society.

Ghazanfar, A. A. and Schroeder, C. E. (2006). Is neocortex essentially multisensory? Trends in Cognitive Sciences, **10**(6), 278–285.

Gillmeister, H. and Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. Brain Research, **1160**, 58–68.

Graham, N. V. S. (2001). Visual pattern analyzers. Number 16. Oxford University Press.

Green, D. M. (1958). Detection of multiple component signals in noise. Journal of the Acoustical Society of America, **30**(10), 904–911.

Green, D. M. and Swets, J. A. (1966). Signal detection theory and psychophysics, volume 1. Wiley.

Green, BG Lederman, S. and Stevens, J. (1979). The effect of skin temperature on the perception of roughness. Sens. Processes., **3(4)**, 327–33.

Kayser, C., Petkov, C. I., Augath, M., and Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. Neuron, **48**(2), 373–384.

Kiang, N. Y. S. and Sachs, M. B. (1968). Two-tone inhibition in auditory-nerve fibers. Journal of the Acoustical Society of America, **43**(5), 1120–1128.

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., and Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron, **53**(2), 279–292.

Laming, D. (1997). A critique of a measurement-theoretic critique: Commentary on michell, quantitative science. British Journal of Psychology, **88**(3), 389.

Lemus, L., Hernández, A., and Romo, R. (2009). Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proceedings of the National Academy of Sciences, **106**(23), 9471–9476.

Lemus, L., Hernández, A., Luna, R., Zainos, A., and Romo, R. (2010). Do sensory cortices process more than one sensory modality during perceptual judgments? Neuron, **67**(2), 335–348.

Liang, L., Lu, T., and Wang, X. (2002). Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. Journal of Neurophysiology, **87**(5), 2237–2261.

Lippert, M., Logothetis, N. K., and Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. Brain Research, **1173**, 102–109.

Lovelace, C. T., Stein, B. E., and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Brain Research, **17**(2), 447–453.

Luce, D. R. (1963). A threshold theory for simple detection experiments. Psychological Review, **70**(1), 61–79.

MacMillan, N. and Creelman, C. (2005). Detection Theory: A User's Guide. Erlbaum.

Marill, T. (1956). Detection theory and psychophysics.

Marks, L. E., Veldhuizen, M. G., Shepard, T. G., and Shavit, A. Y. (2011). Detecting gustatory–olfactory flavor mixtures: Models of probability summation. Chemical senses, page bjr103.

Meyer, G., Wuerger, S., Rhrbein, F., and Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. Experimental Brain Research, **166**, 538–547.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. Cognitive psychology, **14**(2), 247–279.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., Schroeder, C. E., and Foxe, J. J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cerebral Cortex, **15**(7), 963–974.

Nordlie, E., Gewaltig, M.-O., and Plesser, H. E. (2009). Towards reproducible descriptions of neuronal network models. PLoS Comput Biol, **5**(8), e1000456.

Norman, D. A. (1964). Sensory thresholds, response biases, and the neural quantum theory. Journal of Mathematical Psychology, **1**(1), 88 – 120.

Otto, T. U. and Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. Current Biology, **22**(15), 1–6.

Otto, T. U., Dassy, B., and Mamassian, P. (2013). Principles of multisensory behavior. The Journal of Neuroscience, **33**(17), 7463–7474.

Pérez-Bellido, A., Soto-Faraco, S., and López-Móliner, J. (2013). Sound-driven enhancement of vision: disentangling detection-level from decision-level contributions. Journal of neurophysiology, **109(4)**, 1065–77.

Pouget, A., Deneve, S., and Duhamel, J.-r. (2002). Opinion: A computational perspective on the neural basis of multisensory spatial representations. Nature Reviews Neuroscience, **3**(9), 741–747.

Quick, J. R. (1974). A vector-magnitude model of contrast detection. Kybernetik, **16**(2), 65–67.

Ro, T., Hsu, J., Yasar, N. E., Elmore, L. C., and Beauchamp, M. S. (2009). Sound enhances touch perception. Experimental brain research Experimentelle Hirnforschung Exprimentation crbrale, **195**(1), 135–43.

Schnupp, J. W. H., Dawe, K. L., and Pollack, G. L. (2005). The detection of multisensory stimuli in an orthogonal sensory space. Experimental Brain Research, **162**(2), 181–190.

Schürmann, M., Caetano, G., Jousmäki, V., and Hari, R. (2004). Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. The Journal of the Acoustical Society of America, **115**(2), 830–832.

Soto-Faraco, S. and Deco, G. (2009). Multisensory contributions to the perception of vibrotactile events. Behavioural Brain Research, **196**(2), 145 – 154.

Sperdin, H. F., Cappe, C., Foxe, J. J., and Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. Frontiers in integrative neuroscience, **3**(March), 10.

Sperdin, H. F., Cappe, C., and Murray, M. M. (2010). Auditory–somatosensory multisensory interactions in humans: Dissociating detection and spatial discrimination. Neuropsychologia, **48**(13), 3696–3705.

Thompson, E. R., Iyer, N., and Simpson, B. D. (2013). Multicomponent signal detection: Tones in noise. In Proceedings of Meetings on Acoustics, volume 19, page 050030. Acoustical Society of America.

Tyler, C. W. and Chen, C. C. (2000). Signal detection theory in the 2afc paradigm: attention, channel uncertainty and probability summation. Vision Research, **40**(22), 3121–3144.

van Atteveldt, N., Murray, M. M., Thut, G., and Schroeder, C. E. (2014). Multisensory integration: Flexible use of general operations. Neuron, **81**(6), 1240–1253.

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron, **36**(5), 955–968.

Wickens, T. D. (2001). Elementary Signal Detection Theory. Oxford University Press, USA.

Wilson, E. C., Reed, C. M., and Braida, L. D. (2009). Integration of auditory and vibrotactile stimuli: effects of phase and stimulus-onset asynchrony. Journal of the Acoustical Society of America, **126**(4), 1960–1974.

Wilson, E. C., Reed, C. M., and Braida, L. D. (2010a). Integration of auditory and vibrotactile stimuli: effects of frequency. Journal of the Acoustical Society of America, **127**(5), 3044–3059.

Wilson, E. C., Braida, L. D., and Reed, C. M. (2010b). Perceptual interactions in the loudness of combined auditory and vibrotactile stimuli. Journal of the Acoustical Society of America, **127**(5), 3038–3043.

Wuerger, S. M., Hofbauer, M., and Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold. Perception And Psychophysics, **65**(8), 1188–1196.

Yarrow, K., Haggard, P., and Rothwell, J. C. (2008). Vibrotactile-auditory interactions are post-perceptual. Perception, **37**(7), 1114.

Yeshurun, Y., Carrasco, M., and Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. Vision Research, **48**(17), 1837–1851.

Figures and Tables Legends

Figure 1: A: Schematic representation of two classes of models of multimodal integration in yes/no multi-sensory detection tasks. Top row: For this type of model, auditory and tactile signals are processed and detected separately in their respective sensory areas. After detection, the outputs from the detection layers are summed, and only afterwards the decision is made. In the PSM, the sum-operation is realized with an inclusive OR; For the MM the sum involves as the addition of the neural outputs of the two detection processes. In both cases the decision is based on a comparison with the corresponding threshold. Bottom row: For this type of model auditory and tactile signals interact before the detection. Both he PM and the LSM belong to this class, and for both the sum involves a linear addition. For the PM, the addends are weighted with their respective d's, while for the LSM both weights are equal to one. B. Decision boundaries in a yes-no multi-sensory detection task for most common models used in multi-sensory literature (1-look model, PM, LSM, PSM, MM). The axes represent the sensory activity value for auditory (abscissa) and tactile (ordinate) stimuli. The space is divided by the decision boundaries in the two regions (YES/NO), depending on the decision rule implemented by each model. The LSM, the 1-look model and the PM have a straight line as decision bound (green, purple and blue lines), while the PSM's decision bound is formed by two straight lines crossing ordinate and abscissa in λ (red line). As a consequence of the latter that (arbitrarily, $d'_A > d'_T$), the PM's decision bound is tilted to the right with respect to the LSM's one. In the limit $d'_A/d'_T \to 0$, the PM's decision boundary tends to approach the 1-look boundary, as the tactile stimuli do not provide any information in this strategy.

Figure 2: Schematic representation of two models in a 2IFC for a single modality (A) and audio-tactile (B) detection task. Top row of A and B: Pictorial representation of the distribution of the sensory activities (S_1 and S_2) for the noise and for the stimulus (in B both modalities of the multi-sensory compound are represented, S_1^T, S_1^A, S_2^T and S_2^A). The noise has been arbitrarily represented in the first interval of the 2IFC trial. PSM strategy (central rows); after each interval the observer is in one of two states: Detection (Det) or not-Detection (NoDet). When the stimulus is not presented (1st interval) the observer can make a false alarm (FA) or a correct rejection. When the stimulus is presented (2nd interval) the observer can make a hit or a miss by detecting or not the stimulus, respectively. For an audio-tactile stimulus (B) there are 16 possible states, corresponding to the possible combinations of the 4 states of each interval. The final decision is made upon these states with a probabilistic rule. Correct answers are given only when the 2 (panel A) or 4 (panel B) states are not in contradiction and correct (i.e. stimulus in 2nd interval); error answers are given only when the 2 (panel A) or 4 (panel B) states are not in contradiction, but incorrect (i.e. stimulus in 2st interval); in all the other cases, where at least one of the decision is incoherent with the others, the observer has to guess. A panel, central row: According to the Difference model (DM) in the single-modality the decision is made upon the comparison of the two sensory activities S_1 and S_2, so the sensory activity of the first interval, S_1, has to be memorized until after the second interval. B panel, central row: According to the PM strategy the decision is made upon the comparison of the weighted sum of the two sensory activities ($\bar{S_1}, \bar{S_2}$). The weighted sum of the audio-tactile sensory activity of the first interval, $\bar{S_1}$, has to be memorized until after the second interval. C: Decision bounds for PM and PSM (without false alarms) in a detection task with audio-tactile stimuli. These boundaries are represented in the same stimulus space as in fig.1B, but for this paradigm is more complex to represent the decision, as it depends on the sensory activity during the interval with no stimulus. The PSM's decision bound separate the X_T, X_A space into two zones, where the answer can be correct (one of the two stimuli was detected) or a guess (both stimuli have sensory activity below the threshold). For the PM we depicted the contour plot of the probability distribution for correct (dashed lines) and error (continuous lines) answers. To generate these distributions we simulated the PM's for 10^4 trials by extracting pseudo-random Gaussian values for the sensory activities.

Figure 3: Panels A-C: The cells in the three checkerboards represent the differences in detection probabilities in a yes-no task between the experimental values from Schnupp et al. (2005) dataset and the OM (panel a), PSM (panel b) and LSM predictions (panel c). For each cell the color intensity indicates the magnitude of the deviation. White indicates a good match between the model's prediction and the observed value. While the color intensity (red or blue) indicates whether the models' prediction under- or over-estimates the experimental value respectively. Panel D: Detection probabilities, bimodal vs unimodal, for a yes-no task generated with numerical simulations for LSM, PSM, MM and OM (see text).

Figure 4: A: Schematic representation of the proposed ANN for bimodal detection. The gray rectangles represent the different modules: sensory (audio and somatosensory, labeled A and T), detection related (audio and somatosensory, labeled Det A and Det T) and decisional stage (Dec). Black circles represent the inhibitory pools of neurons; blue, red and purple circles represent the excitatory pools respectively for the audio, somatosensory and the compound decisional stage. Between the modules the synapses connect uniquely excitatory-excitatory neurons (light grey arrows). The excitatory recurrent synapses are differentiate qualitatively into strong (detection and decisional stage) and weak (sensory stage). B: Probability distribution function (p.d.f) of the activity of the excitatory pools (A, T, Det A, Det T and Dec) of the different layers. Excitatory pools A and T, as well as Det A and Det T, have the same qualitative behavior, and are collapsed in a unique graph. Pools A and T have higher activity when the stimulus is present (stimulus) then when it is not (noise). The p.d.f.s of the activity of Det A and Det T are plotted for the two possible observer internal states. Their acitivity is usually higher when the stimulus is detected (detection) then when it is not (no-detection). The p.d.f. for the Dec pool is depicted for the two possible outcome of the whole net: No and Yes. Its acitivity is usually higher when the answer is Yes then when it is No.

Figure 5: Psychophysical curves of detection probability calculated with the 3-stage ANN model (sensory, detection and decision) for a single modality (ANN Unimodal, gray circles) and for a compound stimulus (ANN Bimodal, empty circles). The dashed line is the fit with a normal cumulative distribution for the detection probability for the ANN unimodal (gray dashed line); the continuous lines are respectively the prediction for the PSM (gray), and for the OM (black) given the same unimodal detection probabilities of the ANN unimodal.

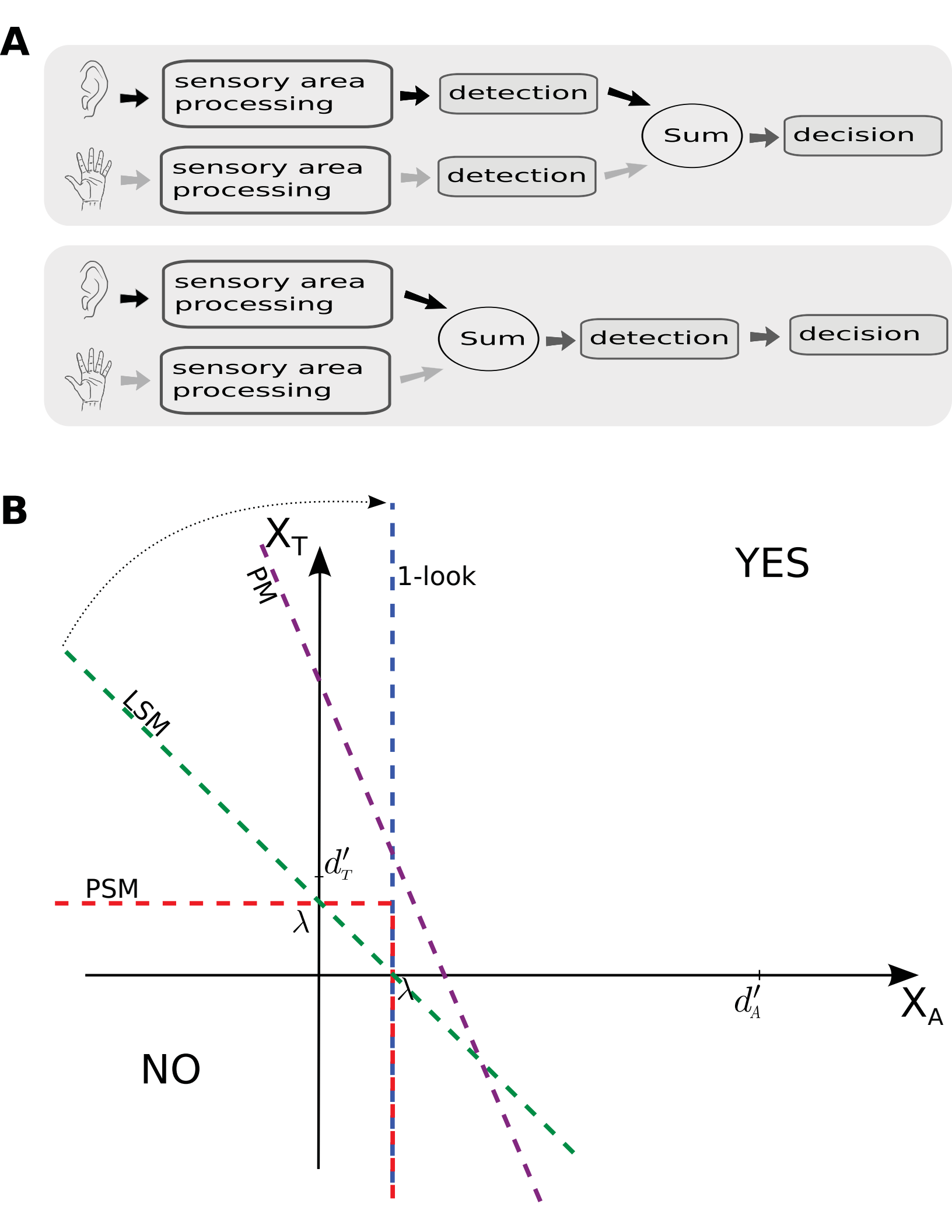
Figure 6: Results from experiment 1. The six panels on the left (A-F) show individual probabilities of correct answers for the different frequencies (13, 31 and 49 Hz) in the three conditions, T, A and A+T respectively in black, gray and white. G: Inter-observer averages of the probability of correct answers. Horizontal dashed lines for the panels A-G is the chance level for the proportion of correct. Error bars are the s.e. H: Boxplot d'-ratio of the observed (d'_{exp}) over the PM (d'_{PM}) predictions. On each box, the central mark is the median, the edges of the box are the first and third quartiles, the whiskers extend to the most extreme data-points, that is within 1.5 interquartile range below and above the lower and upper quartile respectively; the outliers are plotted individually with crosses.

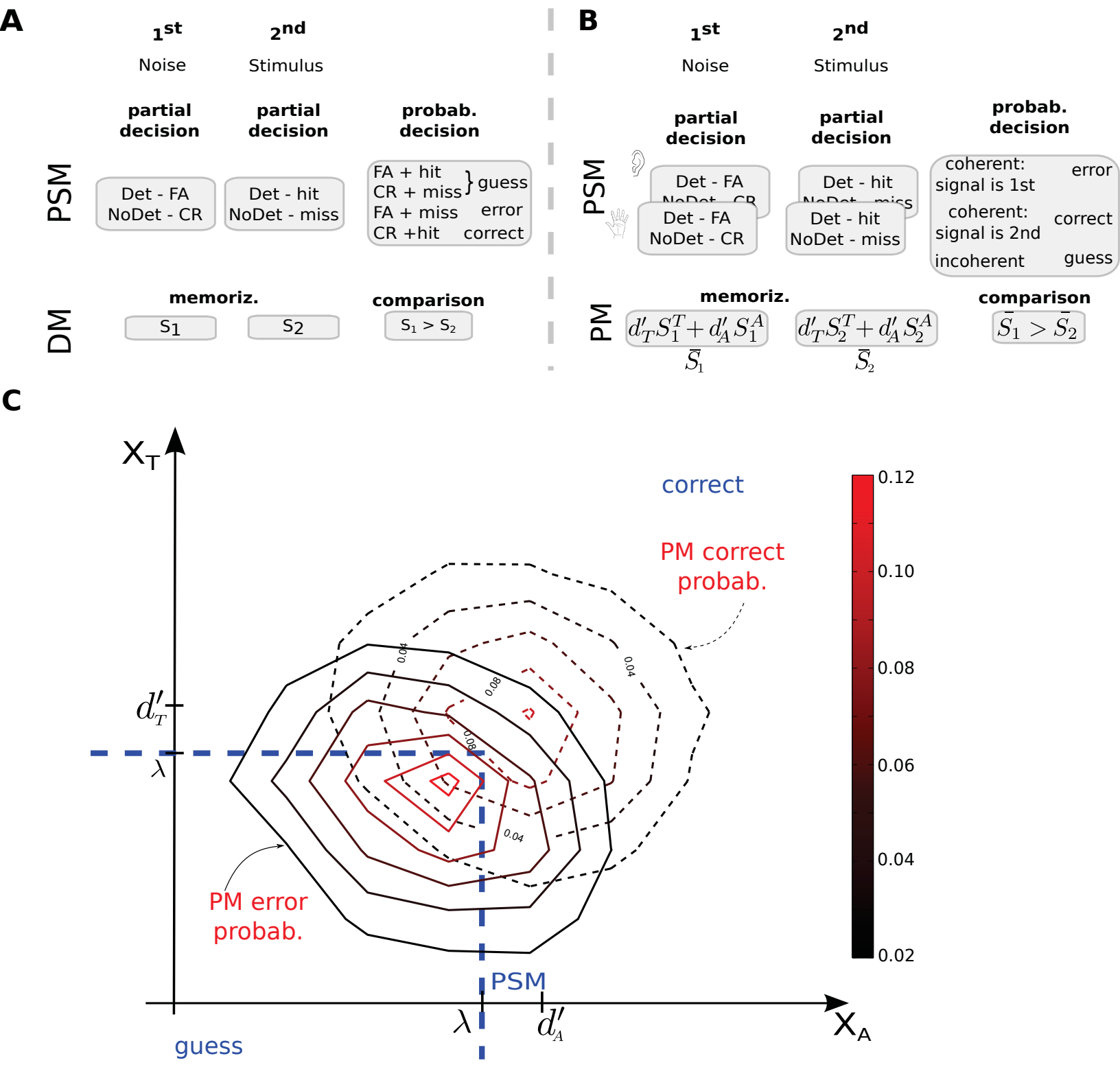
Figure 7: Results from experiment 2. A: Probability of correct answers in the A, T and AT conditions as a function of SOA averaged over the 5 participants. For SOAs greater than 0, the tactile stimulus is presented before the audio stimulus (vice-versa for negative SOAs). B: Boxplot of the d'-ratio: d'_{exp}/d'_{PM}. The red crosses represent the outliers.

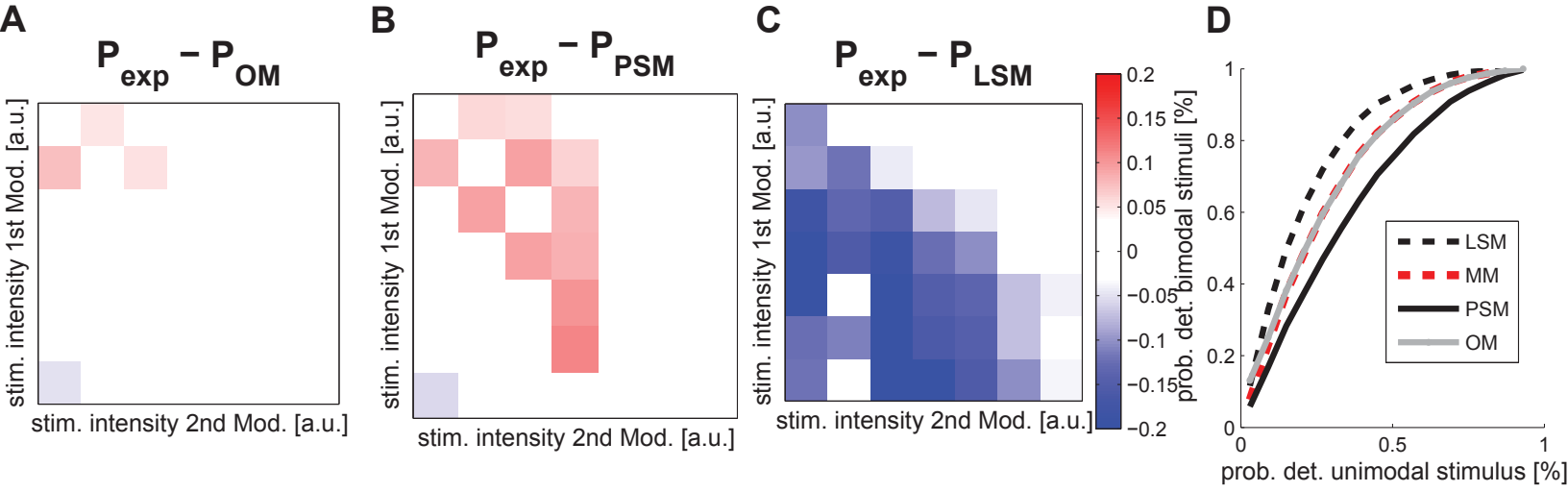
Figure 8: Results from experiment 3. A: Probabilities of correct answer for stimuli in synchrony (white bars) and long SOA (gray bars) for five participants. The rightmost bars represent the average over the five observers. The leftmost bars show the average over the five participants. B: Boxplot averages of the d'-ratio, d'_{exp}/d'_{PM}, for the two conditions long SOA and synchronous stimuli.

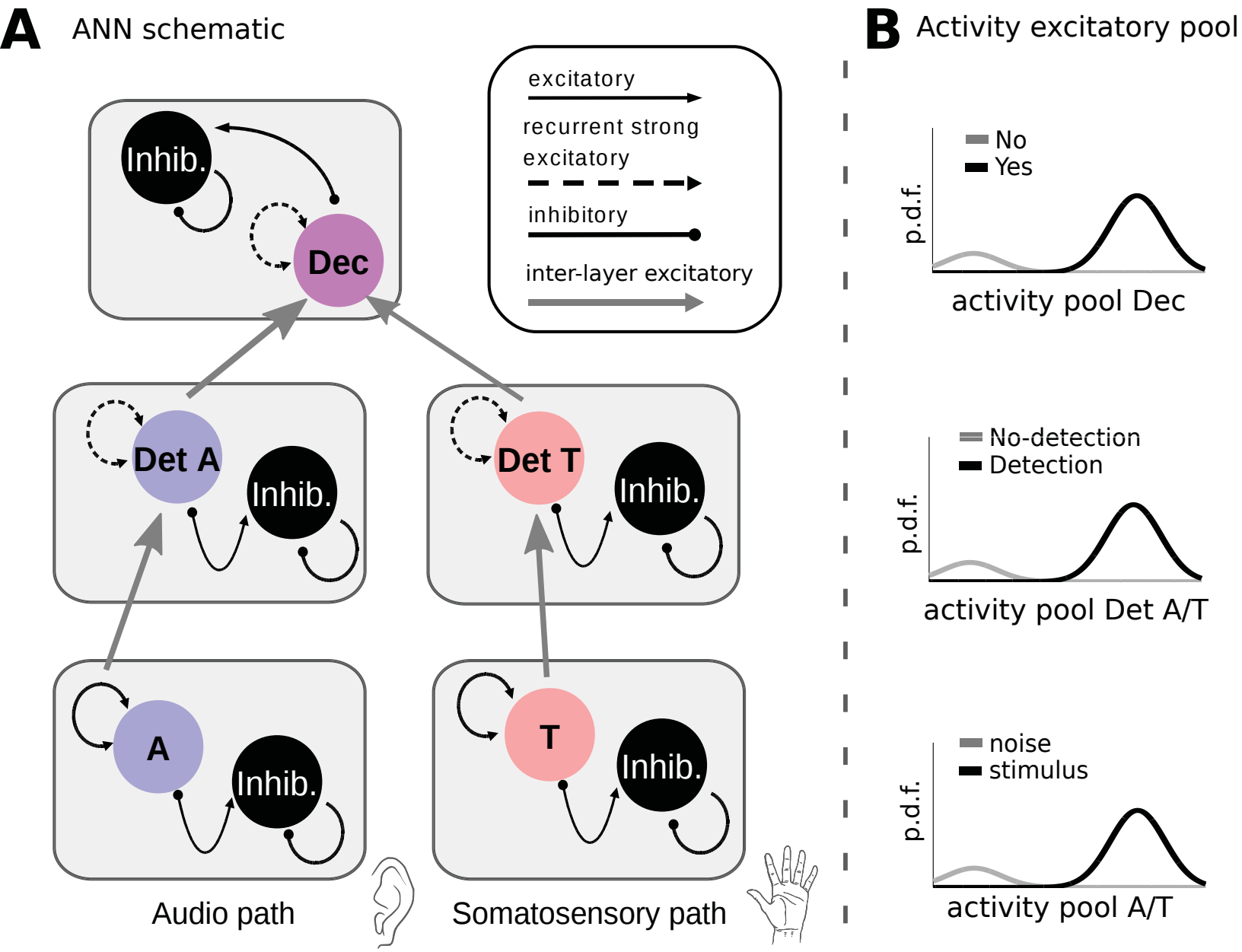
Table 1: ANN model is summarized in panel A and detailed in panels B-E. Parameters values are given in Tab.2.

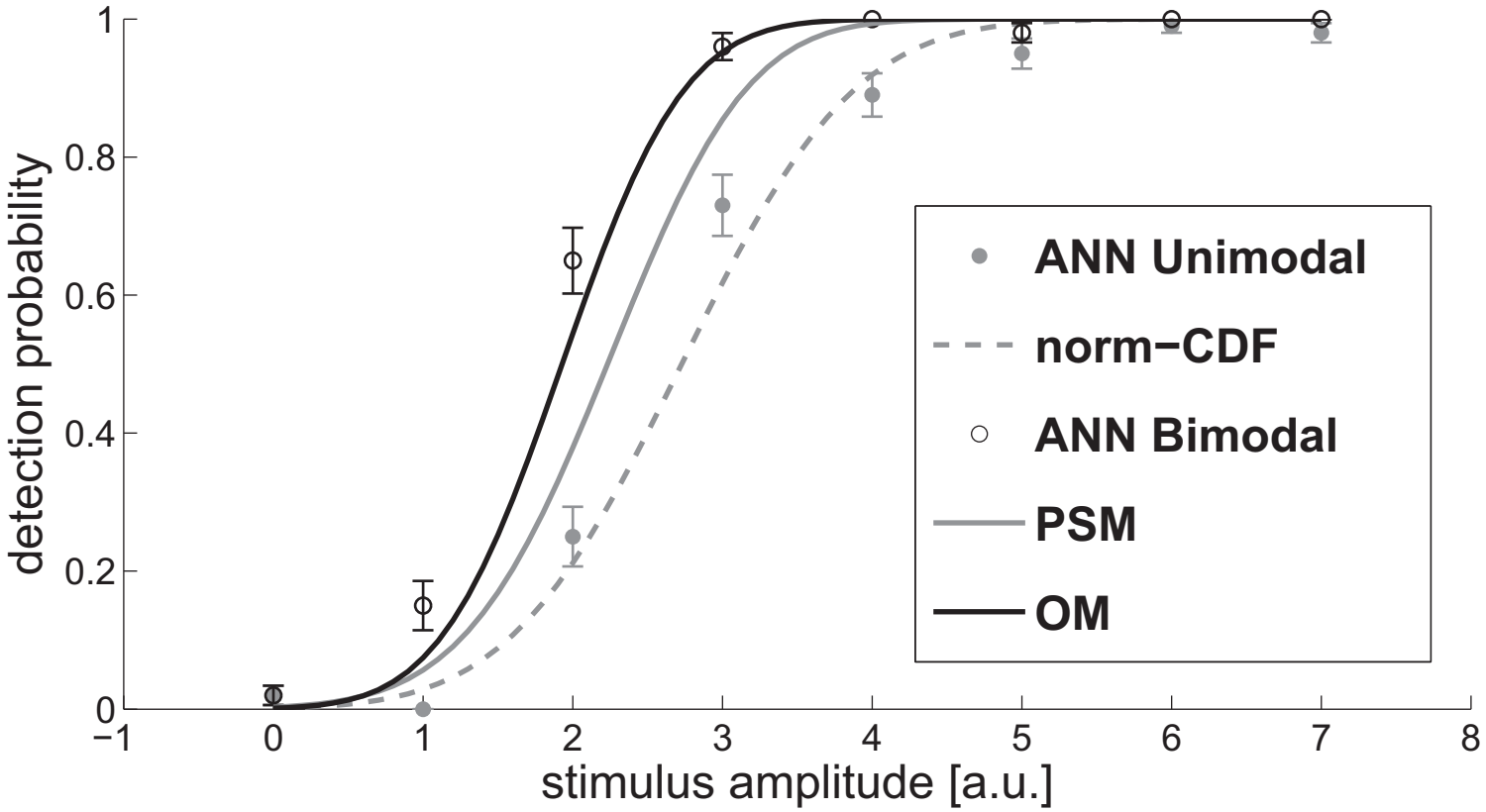
Table 2: Parameters used in the simulations for the ANN.

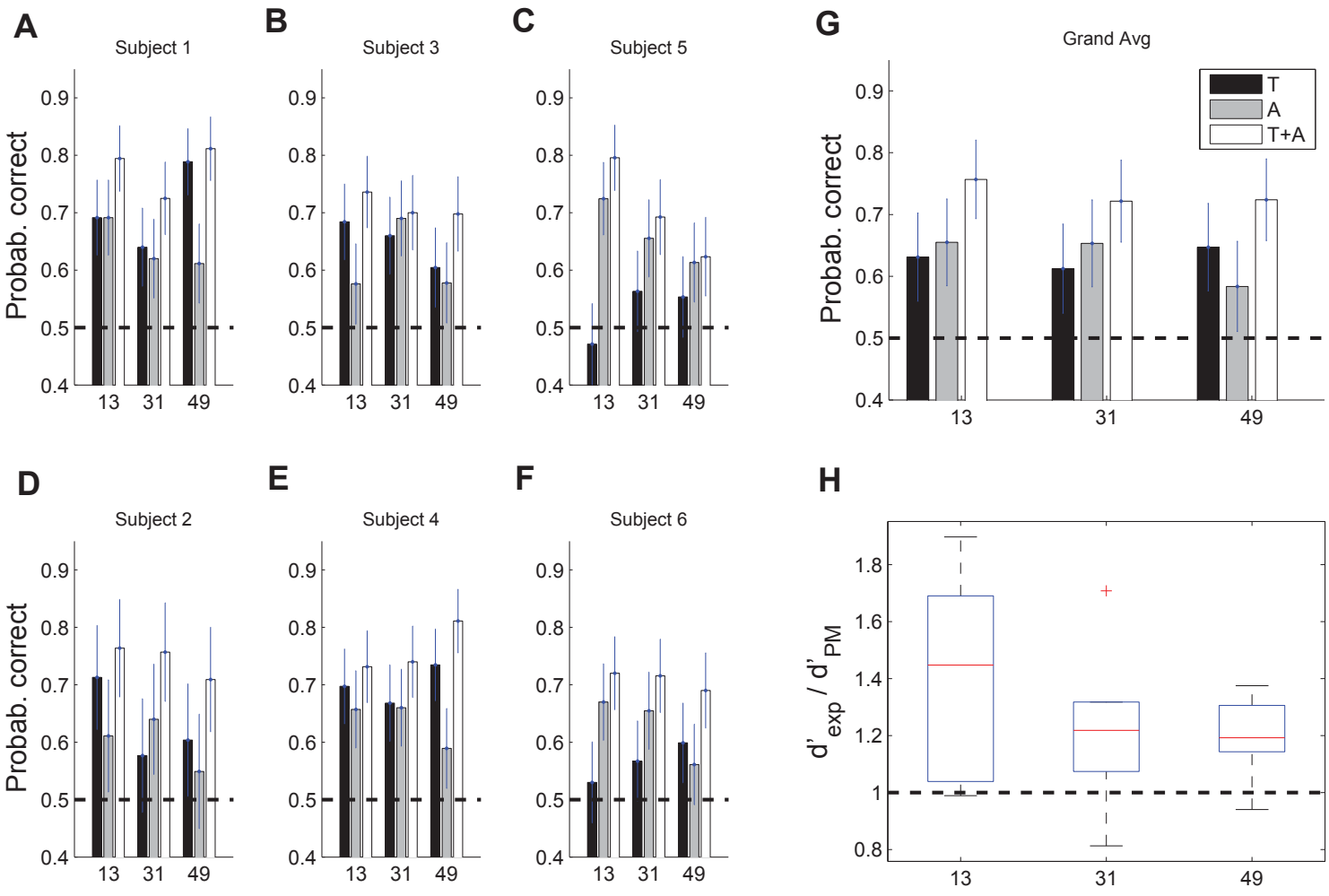

A
sensory area processing
detection
sensory area processing
detection
Sum
decision
sensory area processing
sensory area processing
Sum
detection
decision
B
X_T
PM
1-look
YES
LSM
d'_T
PSM
λ
λ
d'_A
X_A
NO

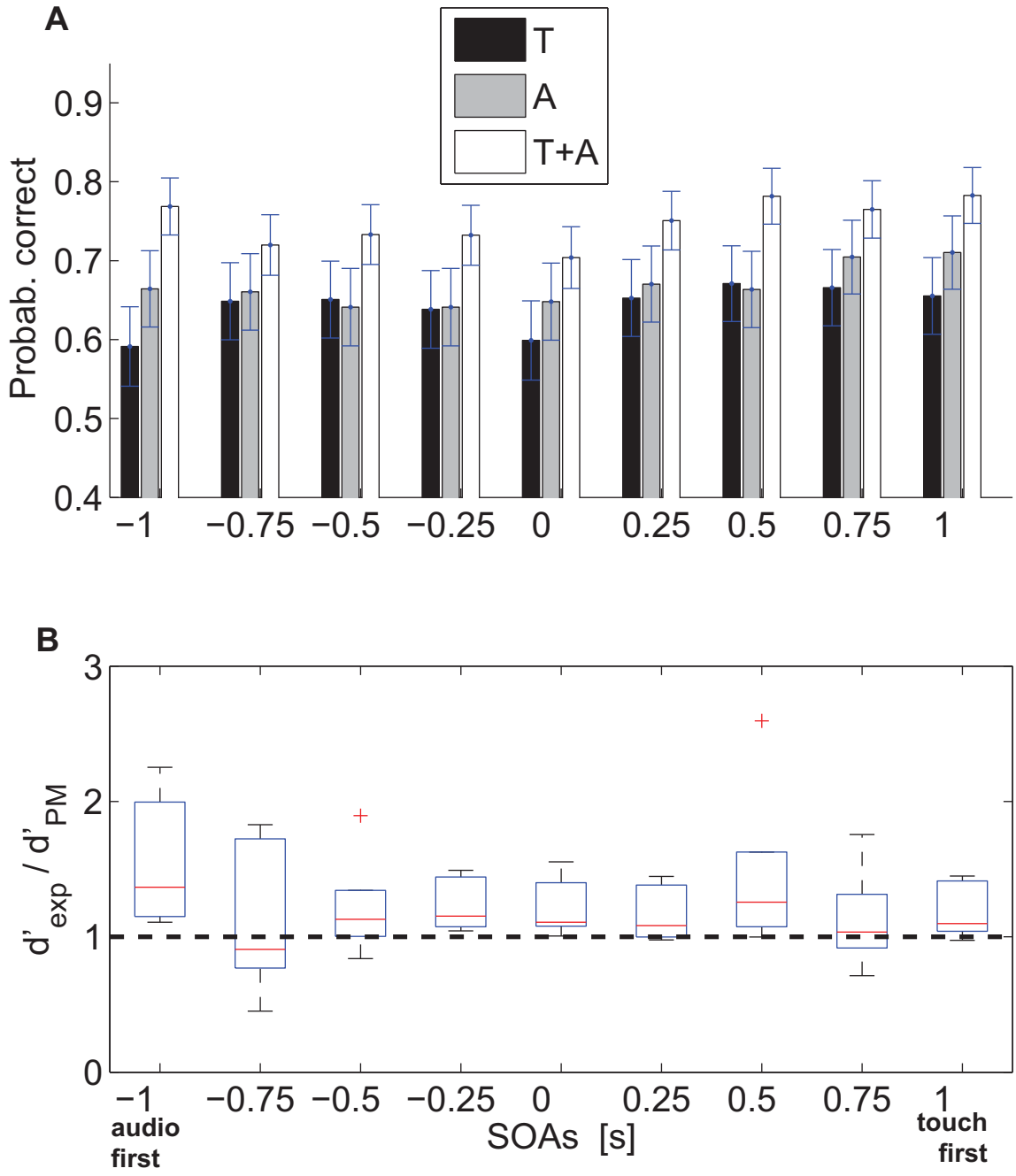

A
1st
Noise
2nd
Stimulus
partial decision
partial decision
probab. decision
PSM
Det - FA
NoDet - CR
Det - hit
NoDet - miss
FA + hit
CR + miss
guess
FA + miss
error
CR +hit
correct
DM
memoriz.
comparison
S1
S2
S1 > S2
B
1st
Noise
2nd
Stimulus
partial decision
partial decision
probab. decision
PSM
Det - FA
NoDet - CR
Det - hit
NoDet - miss
coherent: signal is 1st
error
coherent: signal is 2nd
correct
incoherent
guess
PM
memoriz.
comparison
$d'_T S_1^T + d'_A S_1^A$
$d'_T S_2^T + d'_A S_2^A$
$\bar{S}_1 > \bar{S}_2$
$\bar{S}_1$
$\bar{S}_2$
C
X_T
X_A
d'_T
λ
d'_A
correct
PM correct probab.
PM error probab.
PSM
guess
0.12
0.10
0.08
0.06
0.04
0.02

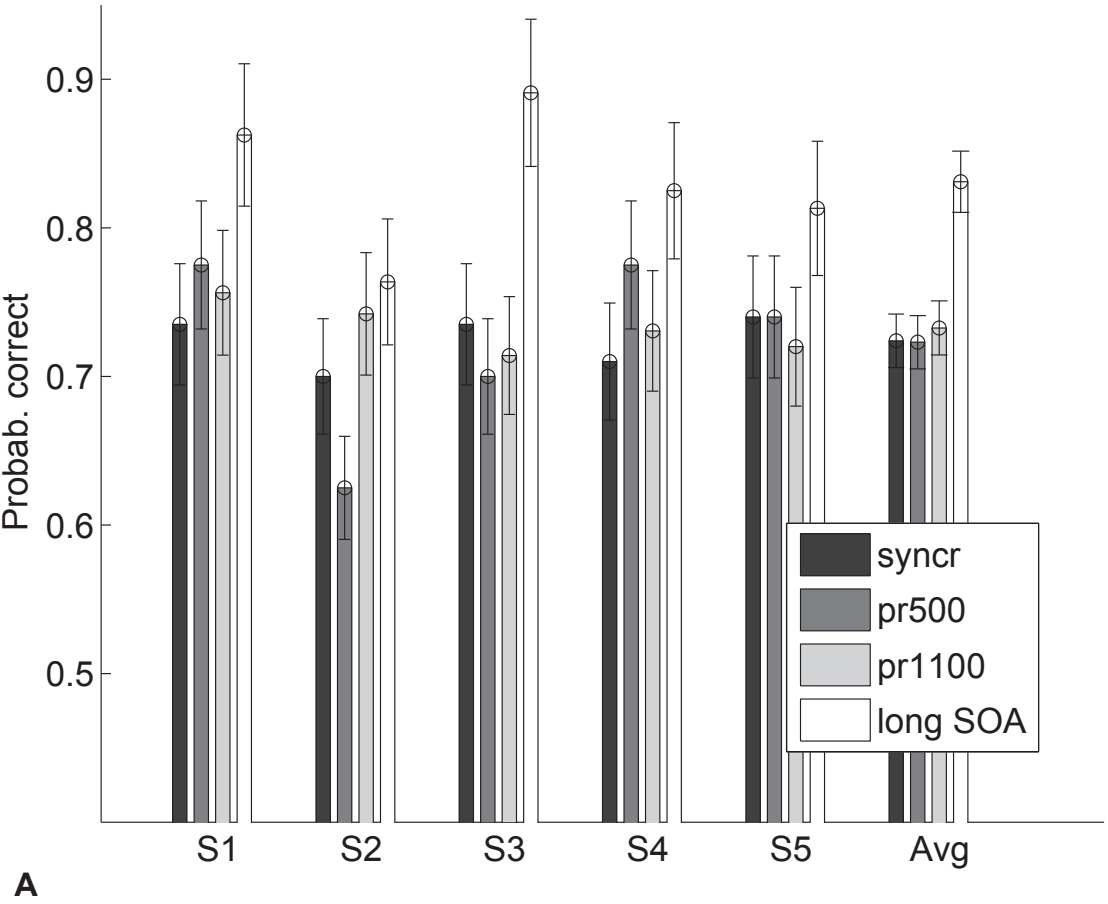

A
P exp − P OM
stim. intensity 1st Mod. [a.u.]
stim. intensity 2nd Mod. [a.u.]
B
P exp − P PSM
stim. intensity 1st Mod. [a.u.]
stim. intensity 2nd Mod. [a.u.]
C
P exp − P LSM
stim. intensity 1st Mod. [a.u.]
stim. intensity 2nd Mod. [a.u.]
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
prob. det. bimodal stimuli [%]
D
1
0.8
0.6
0.4
0.2
0
0
0.5
1
prob. det. unimodal stimulus [%]
LSM
MM
PSM
OM

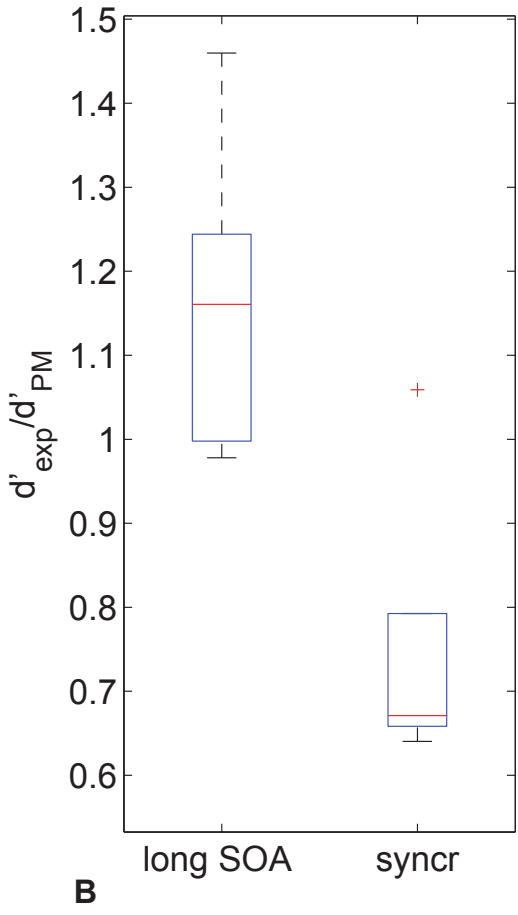

A
ANN schematic
excitatory
recurrent strong excitatory
inhibitory
inter-layer excitatory
Inhib.
Dec
Det A
Inhib.
Det T
Inhib.
A
Inhib.
T
Inhib.
Audio path
Somatosensory path
B
Activity excitatory pool
No
Yes
p.d.f.
activity pool Dec
No-detection
Detection
p.d.f.
activity pool Det A/T
noise
stimulus
p.d.f.
activity pool A/T


detection probability
stimulus amplitude [a.u.]
ANN Unimodal
norm−CDF
ANN Bimodal
PSM
OM
1
0.8
0.6
0.4
0.2
0
−1
0
1
2
3
4
5
6
7
8


A
Subject 1
B
Subject 3
C
Subject 5
G
Grand Avg
T
A
T+A
Probab. correct
0.9
0.8
0.7
0.6
0.5
0.4
13
31
49
D
Subject 2
E
Subject 4
F
Subject 6
H
d'exp / d'PM
1.8
1.6
1.4
1.2
1
0.8


A
T
A
T+A
Probab. correct
0.9
0.8
0.7
0.6
0.5
0.4
−1
−0.75
−0.5
−0.25
0
0.25
0.5
0.75
1
B
d'exp / d'PM
3
2
1
0
−1
audio first
−0.75
−0.5
−0.25
0
0.25
0.5
0.75
1
touch first
SOAs [s]


Probab. correct
0.9
0.8
0.7
0.6
0.5
S1
S2
S3
S4
S5
Avg
syncr
pr500
pr1100
long SOA
A


d'exp/d'PM
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
long SOA
syncr
B

A	ANN Model Summary
Populations	Ten
Topology	Five modules partially connected
Connectivity	full, no synaptic delay
Neuron model	Leaky Integrate-and-Fire, fixed threshold, fixed refractory time
Channel models	-
Synapse model	Instantaneous jump and exponential decay for AMPA and GABA and exponential jump and decay for NMDA receptors
Plasticity	-
Input	Independent fixed-rate poisson spike trains to all neurons
Measurements	Spike activity

B	Populations		
Total number of neurons	$N = 2000$	In each module	$N_{mod} = N/2$
Excitatory neurons in each module		$N_E = 0.8 \cdot N_{mod}$	
Inhibitory neurons in each module		$N_I = 0.2 \cdot N_{mod}$	

Name	**Size**	**Name**	**Size**
A, T (sensory A, T)	$N_{A/T} = f \cdot N_E$	Nonspecific (sensory modules)	$N_E - N_{A/T}$
		Inhibitory (sensory modules)	$0.2 \cdot N_{mod}$
Det A, Det T (detection A, T)	$N_{DetA/DetT} = f \cdot N_E$	Nonspecific (detection modules)	$N_E - N_{DetA/DetT}$
		Inhibitory (detection modules)	$0.2 \cdot N_{mod}$
Dec (decision)	$N_{Dec} = f \cdot N_E$	Nonspecific (decision module)	$N_E - N_{Dec}$
		Inhibitory (decision modules)	$0.2 \cdot N_{mod}$

C	Neuron and Synapse Model
Type	Leaky integrate-and-fire, conductance-based synapses
Subthreshold dynamics	$C_m \dot{V}(t) = -g_L(V(t) - V_L) - I_{AMPA,ext}(t) - I_{AMPA,rec}(t) - I_{NMDA}(t) - I_{GABA}(t)$
Synaptic currents	$I_{AMPA,ext}(t) = g_{AMPA,ext}(V(t) - V_E) \sum_{j=1}^{N_{ext}} s_j^{AMPA,ext}(t)$ $I_{AMPA,rec}(t) = g_{AMPA,rec}(V(t) - V_E) \sum_{j=1}^{N_E} w_j s_j^{AMPA,rec}(t)$ $I_{NMDA}(t) = \frac{g_{NMDA}(V(t)-V_E)}{1+[Mg^{2+}]exp(-0.062V(t))/3.57} \times \sum_{j=1}^{N_E} w_j s_j^{NMDA}(t)$ $I_{GABA}(t) = g_{GABA}(V(t) - V_I) \sum_{j=1}^{N_I} w_j s_j^{GABA}(t)$
Fraction of open channels	$\frac{ds_j^{AMPA,ext}(t)}{dt} = -\frac{s_j^{AMPA,ext}(t)}{\tau_{AMPA}} + \sum_k \delta(t - t_j^k)$ $\frac{ds_j^{AMPA,rec}(t)}{dt} = -\frac{s_j^{AMPA,rec}(t)}{\tau_{AMPA}} + \sum_k \delta(t - t_j^k)$ $\frac{ds_j^{NMDA}(t)}{dt} = -\frac{s_j^{NMDA}(t)}{\tau_{NMDA,decay}} + \alpha x_j(t)(1 - s_j^{NMDA}(t))$ $\frac{dx_j^{NMDA}(t)}{dt} = -\frac{x_j^{NMDA}(t)}{\tau_{NMDA,rise}} + \sum_k \delta(t - t_j^k)$ $\frac{ds_j^{GABA}(t)}{dt} = -\frac{s_j^{GABA}(t)}{\tau_{GABA}} + \sum_k \delta(t - t_j^k)$
Spiking	if $V(t) \geq V_\theta \wedge t > t^\star + \tau_{rp}$ 1. $t^\star = t$ 2. emit spike at time $t^\star$ 3. $V(t) = V_{reset}$

D	Input
Type	**Description**
Poisson generator	Fixed rate, N_{ext} poisson generators per neuron, each one projects to one neuron

E	Measurements
Spike activity, firing-rates calculated using spike count in a 50 ms time window shifted by 5 ms step	

Parameter	Value	Parameter	Value
C_m (excitatory)	0.5 nF	V_E	0 mV
C_m (inhibitory)	0.2 nF	V_I	-70 mV
f	0.15	V_L	-70 mV
$g_{AMPA,ext}$ (excitatory)	2.08 nS	V_{reset}	-55 mV
$g_{AMPA,ext}$ (inhibitory)	1.62 nS	V_θ	-50 mV
$g_{AMPA,rec}$ (excitatory)	0.104 nS	w_+(decision-making network)	1.8
$g_{AMPA,rec}$ (inhibitory)	0.081 nS	w_+(confidence network)	1.7
g_{GABA} (excitatory)	1.287 nS	α	0.5 ms^{-1}
g_{GABA} (inhibitory)	1.002 nS	$\lambda_{\text{Reference}}$	40 Hz
g_{NMDA} (excitatory)	0.327 nS	λ_{ext}	2.4 kHz
λ	45 Hz	$\Delta\lambda$	[0 30] Hz
g_{NMDA} (inhibitory)	0.258 nS	τ_{AMPA}	2 ms
N_E	800	τ_{GABA}	10 ms
N_I	200	$\tau_{NMDA,decay}$	100 ms
N_{ext}	800	$\tau_{NMDA,rise}$	2 ms