Deconstructing multisensory 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-Faraco1,5

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

Submitted 7 May 2014; accepted in final form 17 December 2014

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 detection. As is the case with other multisensory ensembles, evidence of early sensory interactions from audiotactile behavioral studies is somewhat mixed, often leading to discrepant conclusions. Such discrepant conclusions can arise for various reasons. For example, regarding audiotactile detection tasks, comparisons of the results of yes-no detection tasks (Schnupp et al. 2005) with two-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 include: 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); and, in multisensory studies, interpreting attentional cueing effects as evidence of early bottom-up integration (for audiovisual Lovelace et al. 2003; and for audiotactile: Gillmeister and Eimer 2007; Ro et al. 2009); an interesting approach to this issue can be found in Lippert et al. (2007).

Another factor of discrepancy between studies is in regard to differences in the statistical models. 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 multisensory research are the Pythagorean model (PM) and the 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 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 (1974) proposed a family of functions 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 multisensory 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; or Gescheider et al. 1974); also, 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 PM (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, multisensory enhancements in detection, particularly audition-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 et al. used the PM and they reported detection performances above the PM’s criterion in an audition-tactile detection task. In their experiments, the task was to detect an audition-tactile stimulus. They used vibrotactile stimulus frequencies ranging from 100 to 1,000 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 the dataset of Schnupp, we demonstrate that the discrepancy between Wilson et al. (2009) and Schnupp et al. (2005) is a consequence of the misuse of the PM’s optimality in two-alternative (2IFC) paradigms, that is, the optimal strategy in yes-no detection tasks, does not (always) apply to 2IFC tasks, rather, it depends on the strategy the subjects adopt to solve the 2IFC task. The most common strategies can be divided in two groups: 1) the observer compares information from the 2 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 concern is relevant to all experimental studies focused on multisensory detection tasks or multicomponent detection tasks.

We then put forward a simple model that 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 from 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 in the PM, but the input to the decision stage (output from the detection stage) is a continuous variable, as is the case with the LSM and the PSM. Finally, we also put forward a straightforward mechanistic implementation of the MM in an 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).

To reinforce the conclusions of our formal analysis, we present three separate experiments, designed to reproduce the violation of classic statistical models under conditions impossible for early integration through empirical tests and to illustrate a correct measurement of interactive effects from multisensory stimuli in audition-tactile detection tasks. In experiment 1, we set up the experimental protocols and established whether the apparent contradiction between the results of the studies of Schnupp et al. (2005) and Wilson et al. (2009) 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’s 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; Bender and Wang 2007 for audio), where the neuronal firing rates in early sensory areas (S1, S2, and A1) are modulated only during the stimulus presentation. We therefore hypothesized that if any (rate-based) multimodal interaction takes place at an 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 learning or historic effects). Experiment 2 thus explored audition-tactile detection across a range of stimulus onset asynchronies (SOAs up to 750 ms), with the aim of introducing experimental conditions that would clearly impede early sensory interactions. If the PM’s 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 unsensory compound stimuli (500- and 1,100-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). 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; Mac-
Millan 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 or she perceived a stimulus (yes) or not (no). Generally, stimulus is present in only 50% of the trials. Therefore, the observer has to discriminate the signal (stimulus present) from the noise (stimulus absent). In the multisensory yes-no task analyzed in this article, the stimulus is composed of two modalities and the observer is requested to report yes when he or she perceived at least one of the two modalities. In visual literature this kind of task is called a summation experiment (Graham 2001). We used the description of SDT put forward by Green (1958) and 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 multisensory research, further assumptions are:

As2. In each trial, the observer records a measure of sensory activity, s, and compares it with threshold, \( \lambda \). When s overcomes \( \lambda \), the observer’s answer is yes.

As3. \( s \) is typically assumed having normal probability distribution, with mean \( \mu \) and deviation \( \sigma \). 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 probability of false alarm (observer saying yes without stimulus). Indeed, from the assumptions I–4, the probability to correctly detect a stimulus is given by:

\[
P_{hit} = \Phi(d' - \lambda) \tag{1}
\]

where \( \Phi \) is the cumulative distribution function of the standard normal distribution and the relationship between the threshold \( \lambda \) and \( P_{hit} \) is:

\[
\lambda = -Z(P_{hit}) = \sqrt{2erf^{-1}(2P_{hit} - 1)} \tag{2}
\]

Therefore, we can write \( d' \) as:

\[
d' = Z(P_{hit}) - Z(P_{fa}) = \sqrt{2erf^{-1}(2P_{hit} - 1)} - erf^{-1}(2P_{fa} - 1) \tag{3}
\]

where the \( Z \) function is the probit function, which is the quantile function associated with the standard normal distribution, and \( erf^{-1} \) 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, ch. 10). 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 or 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 multisensory literature; second, this model has never been used as benchmark to address sensory interactions.

PROBABILISTIC SUMMATION MODEL. 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 information coming from the two separate analyses 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-looking strategy, but in the multisensory literature is commonly called PSM. For the PSM the signals from different modalities are processed and detected separately. Each detection stage has a binary output (detection vs. no detection), and the decision is an inclusive OR disjunctive 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 multisensory research.

To formalize the PSM’s hypothesis, that is, to obtain the probability of a correct answer for a yes-no (yn) task involving multisensory 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 (Luco 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 the empirical data of Schnupp et al. (2005), the condition without any stimulation registered ~3%.

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

\[
P^\text{hit}(A, T) = 1 - (1 - P^\text{A}_T)(1 - P^\text{T}_\text{Nd})
\]

where the subindexes 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 Eq. B4 of Marks et al. 2011: \( P^\text{hit}_{\text{PSM}}(A, T) = [P^\text{hit}(A, T) - fa]\frac{1}{1 - fa} \)). As we will see later, the PSM’s predicted hit probabilities are already too low when compared with the observed dataset.

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

\[
P^\text{hit}(T) = \Phi(x_T, b_T - \lambda)
\]

where \( \Phi \) 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 \( \lambda \) 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:

\[
P^\text{hit}(T, A) = 1 - (1 - \Phi(x_T, b_T - \lambda))(1 - \Phi(x_A, b_A - \lambda))
\]

as in Schnupp et al. (2005) we used a single threshold parameter \( \lambda \) 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 \( \Phi \), while Quick used \( 1 - e^{-\frac{x^2}{2\sigma^2}} \); 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. (2005) as it is more feasible to compare it with the approach to the multisensory 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 \( P^\text{T}_\text{Nd} \) can

condition \( P^\text{T}_\text{Nd} \) and vice versa. In this way we are dealing with a nonindependent PSM.

NONINDEPENDENT PSM FOR YES-NO TASKS. For this model the detection of each modality depends on the detection of the other modality. For this model we can keep calling \( P^\text{A}_T \) 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 \( P^\text{A}_T \). Similarly the probability to detect the audio stimulus (A) conditioned on a failed detection of the tactile stimulus (T) is given by: \( P^\text{T}_\text{Nd} \). The hit probability for compound stimuli (A + T) then becomes:

\[
P^\text{hit}(A, T) = 1 - (1 - P^\text{A}_T)(1 - P^\text{T}_\text{Nd})
\]

From this equation it is clear that the probability to detect the compound stimulus is higher when \( P^\text{A}_T < P^\text{T}_\text{Nd} \).

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

PYTHAGOREAN MODEL. Under this strategy, information, i.e., sensory activity (\( s_T \) and \( s_A \)), is integrated in a single element and then compared with 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 \( \lambda \), using \( d' \) of each modality as weight:

\[
S_Td_T + S_Ad_A > \lambda.
\]

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), \( d'_{\text{PM}}(A, T) = \sqrt{d'_A + d'_T} \); this equation is the actual reason why this model is called PM. The threshold \( \lambda \) is then a function of the various \( d' \)s:

\[
\lambda_{\text{PM}} = d'_{\text{PM}}(A, T)/2 - \text{logit}(P_s)/d'_{\text{PM}}(A, T)
\]

where logit is the inverse of the sigmoidal logistic function and \( P_s \) is the probability to have signal; \( \text{logit}(P_s) = 0 \) for \( P_s = 50\% \). This last term takes into account the possibility that the probability of the signal and the noise will not be the same. The PM’s hit probability is given by Eq. 1 using the detectability \( d'_{\text{PM}} \) and the threshold \( \lambda_{\text{PM}} \).

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 either the threshold or the weight for the modalities before stimulus appearance. The strategy described below, the LSM, does not suffer from this problem as weight and threshold are fixed.

LINEAR SUMMATION MODEL. 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:

\[
S_T + S_A > \lambda.
\]

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:

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

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 LSM. Note that, when the stimuli with single and double modality are interleaved, the \( p^\text{hit} \) for the single modality (originally given by Eq. 1) has to be replaced with:

\[
P^\text{hit}(T) = \Phi((d'_T - \lambda)/\sqrt{2})
\]
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, LSM, PM, and OM) 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 one-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 one-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 → 0), the PM’s decision bound coincides with the one-look model.

For the reanalysis of the dataset of Schnupp et al. with the LSM we again used Eq. 5, and in this case the product b_xk corresponds to the d’. The hit probability with two stimuli (A + T), Eq. 11 is:

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

ORTHOGONAL MODEL. The OM is only one of a family of functions that Schnupp et al. used to fit the probability of hit of their dataset for the double-modality stimuli, this probability is:

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

As Schnupp et al. indicated, this latter function can indeed be interpreted within the SDT framework, interpreting the products \( \sqrt{x_T^2 b_T^2 + x_A^2 b_A^2} \) as the discriminability and \( \lambda \) as the threshold (although, as we argue below, these functions are adequate for fitting but not for modeling). 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 \( k \)-value that we analyzed here. For the OM, the multisensory detection probability is given by:

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

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, p. 185).

However, only one of these functions, the one with \( k = 1 \), is effectively a mechanistic model, while 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 \( \Phi \) 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 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, \( \lambda \) 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 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’s 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 \( d' \), and \( d'_k \), 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 \( \Phi \) loses its role, so how can we interpret the OM (and all the models based on Eq. 14 with \( p \neq 1 \))? There are two possible answers: 1) OM is just an useful formula that can describe succinctly the results of Schnupp et al.; and 2) OM can be seen 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 \( \Phi \)) 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 the OM of Schnupp et al.) 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 it is not effectively a mechanistic model but only a useful formula to fit the data. We now push 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. 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 that depends on the activation state, high/low, and the stimulus intensity (as in the PM).

The MM is composed of three stages: a sensory stage, a detection stage and a decision stage (see Fig. 4). The MM has an effective detection stage whose output can take a continuous value, \( v \). 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 \( f_h(v) \) and \( f_l(v) \), respectively. We used the subscripts 1 and 2 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 \( f_h(v) \) a delta function around one; the value of \( v \) 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 \( \lambda_v \) and \( v \), when \( v > \lambda_v \), 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: \( v_T + v_A > \lambda_v \).

ATTRACTOR-BASED NEURAL NETWORK MODEL. We propose that the behavioral results of detection tasks with multimodal stimuli can
be interpreted within the framework of 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, N-methyl-D-aspartate 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 aims 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 layer 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 were set to $w = 0.07$ and between detection and decisional stage were set to $w = 0.11$. Once the parameters of the sensory, detection, and decision modules had been set, we chose the values of the interlayers connections 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 stimulusonly activated (corresponding to no detection) or high activity (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, also causing 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 is the source 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, $\sim 1$ 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 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 multisensory 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 information 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, which represents the PSM as an example of the first strategy and the difference model (DM; as an example of the 2nd 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 DM and its pictorial representation is shown in 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:

\[ As1. \text{ Subject compares the sensory activity of the first interval, } s_1, \text{ to the one of the second interval, } s_2, \text{ and calculates the difference between them, } s_1 - s_2. \text{ The sign of the difference determines the answer: first (or second) interval when it is positive (or negative).} \]

Apart from this, the other assumptions are commonly used in multisensory literature to characterize this model and we therefore adopt them here (note that the subindexes denote 1st and 2nd interval of the 2IFC):

\[ As2. s_1 \text{ and } s_2 \text{ are Gaussian distributed with unitary deviance.} \]

\[ As3. s_1 \text{ and } s_2 \text{ have unitary deviance.} \]

\[ As4. s_1 \text{ and } s_2 \text{ are statistically independent.} \]

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 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, $P_{\text{correct}}$, is given by:

\[ P_{\text{correct}} = \sqrt{2 \Phi(d')} \] (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).
Table 1. ANN model summary and details

<table>
<thead>
<tr>
<th>ANN Model Summary and Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. ANN Model Summary</strong></td>
</tr>
<tr>
<td><strong>Populations</strong></td>
</tr>
<tr>
<td><strong>Name</strong></td>
</tr>
<tr>
<td>A, T (sensory A, T)</td>
</tr>
<tr>
<td>Det A, Det T (detection A, T)</td>
</tr>
<tr>
<td>Dec (decision)</td>
</tr>
</tbody>
</table>

**C. Neuron and Synapse Model**

**Type**
Leaky integrate-and-fire, conductance-based synapses

**Subthreshold dynamics**
\[ C_m \frac{dV(t)}{dt} = -g_L(V(t) - V_r) - I_{AMPA,ext}(t) - I_{AMPA,rec}(t) - I_{NMDA}(t) - I_{GABA}(t) \]

**Synaptic currents**
\[ I_{AMPA,ext}(t) = g_{AMPA,ext}(V(t) - V_E) \]
\[ I_{AMPA,rec}(t) = g_{AMPA,rec}(V(t) - V_E) \sum_j w_{j,k} \]
\[ I_{NMDA}(t) = g_{NMDA}(V(t) - V_E) \frac{1}{1 + [Mg^{2+}]_0 \exp(-0.062V(t)/3.57)} \sum_j w_{j,k} \]
\[ I_{GABA}(t) = g_{GABA}(V(t) - V_E) \sum_j w_{j,k} \]

**Fraction of open channels**
\[ \frac{d_j^{AMPA,ext}(t)}{dt} = \frac{j^{AMPA,ext}(t)}{TAMPA} + \sum_i \delta(t-t_i^j) \]
\[ \frac{d_j^{AMPA,rec}(t)}{dt} = \frac{j^{AMPA,rec}(t)}{TAMPA} + \sum_i \delta(t-t_i^j) \]
\[ \frac{d_j^{NMDA}(t)}{dt} = \frac{j^{NMDA}(t)}{TAMPA,decay} + \delta(t-t_i^j) \]
\[ \frac{d_j^{NMDA}(t)}{dt} = \frac{j^{NMDA}(t)}{TAMPA,rise} + \sum_i \delta(t-t_i^j) \]
\[ \frac{d_j^{GABA}(t)}{dt} = \frac{j^{GABA}(t)}{GABA} + \sum_i \delta(t-t_i^j) \]

**Spiking**
If \( V(t) \geq V_s \), then:
1. \( t^* = t \)
2. Emit spike at time \( t^* \)
3. \( V(t) = V_{reset} \)

**D. Input**
Type: Poisson generator
Description: fixed rate, \( N_{ext} \) Poisson generators per neuron, each one projects to one neuron

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

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 PM.

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

\[ d_x^A s_{A,1} + d_x^A s_{A,2} > d_x^A s_{A,1} + d_x^A s_{A,2} \]

As5. The sensory activities from each modality are linearly summed.

As6. The weights of this linear sum are the respective \( d_x^A \) of each modality (Wickens 2001).

The decision rule in this case would be:

\[ d_x^A s_{A,1} + d_x^A s_{A,2} > d_x^A s_{A,1} + d_x^A s_{A,2} \]
Table 2. Parameters used in the simulations for the ANN

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_{ex} (excitatory)</td>
<td>0.5 nF</td>
</tr>
<tr>
<td>C_{inh} (inhibitory)</td>
<td>0.2 nF</td>
</tr>
<tr>
<td>f</td>
<td>0.15</td>
</tr>
<tr>
<td>$\text{AMP}_{ax}$ (excitatory)</td>
<td>2.08 nS</td>
</tr>
<tr>
<td>$\text{AMP}_{ax}$ (inhibitory)</td>
<td>1.62 nS</td>
</tr>
<tr>
<td>$\text{AMP}_{ax}$ (excitatory)</td>
<td>0.104 nS</td>
</tr>
<tr>
<td>$\text{AMP}_{ax}$ (inhibitory)</td>
<td>0.081 nS</td>
</tr>
<tr>
<td>$\text{GABA}_{ax}$ (excitatory)</td>
<td>1.287 nS</td>
</tr>
<tr>
<td>$\text{GABA}_{ax}$ (inhibitory)</td>
<td>1.002 nS</td>
</tr>
<tr>
<td>$\text{NMDA}_{ax}$ (excitatory)</td>
<td>0.327 nS</td>
</tr>
<tr>
<td>$\text{NMDA}_{ax}$ (inhibitory)</td>
<td>0.258 nS</td>
</tr>
<tr>
<td>N_{E}</td>
<td>800</td>
</tr>
<tr>
<td>N_{I}</td>
<td>200</td>
</tr>
<tr>
<td>N_{ext}</td>
<td>800</td>
</tr>
<tr>
<td>V_{E}</td>
<td>0 mV</td>
</tr>
<tr>
<td>V_{I}</td>
<td>-70 mV</td>
</tr>
<tr>
<td>V_{L}</td>
<td>-70 mV</td>
</tr>
<tr>
<td>V_{reset}</td>
<td>-55 mV</td>
</tr>
<tr>
<td>V_{p}</td>
<td>-50 mV</td>
</tr>
<tr>
<td>$w_{p}$ (decision-making network)</td>
<td>1.8</td>
</tr>
<tr>
<td>$w_{c}$ (confidence network)</td>
<td>1.7</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.5 ms^{-1}</td>
</tr>
<tr>
<td>$\lambda_{ref}$</td>
<td>40 Hz</td>
</tr>
<tr>
<td>D_{I}</td>
<td>2.4 kHz</td>
</tr>
<tr>
<td>D_{I}</td>
<td>[30] Hz</td>
</tr>
<tr>
<td>$\tau_{AMP}$</td>
<td>2 ms</td>
</tr>
<tr>
<td>$\tau_{GABA}$</td>
<td>10 ms</td>
</tr>
<tr>
<td>$\tau_{NMDA}_{aux}$</td>
<td>100 ms</td>
</tr>
<tr>
<td>$\tau_{NMDA}_{res}$</td>
<td>2 ms</td>
</tr>
</tbody>
</table>

See Abeles (1991) for definitions of parameters.

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):

\[
d'_{PM}(A, T) = \sqrt{d'_{A}^2 + d'_{T}^2} \tag{17}
\]

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 PSM.

First, consider a 2IFC for a single modality, and let us call \(P_{T}^{IFC}\) 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. Therefore, let us call \(P_{T}^{IFC}\) and \(P_{T}^{IFC}\) the probabilities of this two outcomes, respectively, whose sum is equal to 1: \(P_{T}^{IFC} + P_{T}^{IFC} = 1\).

As2. The probability of false alarm, \(P_{T}^{IFC}\), is zero (but see below for a model with \(P_{T}^{IFC} > 0\)).

As3. When the observer fails to detect the stimulus, with probability \(P_{T}^{IFC}\), 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:

\[
P_{T}^{IFC} = P_{T}^{IFC} + P_{T}^{IFC}/2 = (1 + P_{T}^{IFC})/2 \tag{18}
\]

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

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

Therefore, we have four possible outcomes: full perception of tactile stimulus, but not the audio one (\(P_{T}^{IFC} \times 1 - P_{T}^{IFC}\)), full perception of audio, but not tactile (\(P_{T}^{IFC} \times 1 - P_{T}^{IFC}\)), full perception of both audio and tactile (\(P_{T}^{IFC} \times P_{T}^{IFC}\)), and no perception of either one (\(1 - P_{T}^{IFC} \times 1 - P_{T}^{IFC}\)). Therefore, the probability of detection for the case of a bimodal stimulus in a 2IFC becomes:

\[
P_{T}^{IFC}(A, T) = (1 + P_{T}^{IFC} + P_{T}^{IFC} - P_{T}^{IFC} P_{T}^{IFC})/2 \tag{19}
\]

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 \(P_{T}^{IFC}\) can condition \(P_{T}^{IFC}\), and vice versa.

NONINDEPENDENT PSM FOR 2IFC TASKS. We adopted the same nomenclature of the PSM for nonindependent modalities of the yes-no task. The probability to correct answer for compound stimuli \((A + T)\) then becomes:

\[
P_{T}^{IFC}(A, T) = 1 + (1 - P_{T}^{IFC})(1 - P_{T}^{IFC})/2 \tag{20}
\]

From this equation it is clear that the probability to detect the compound stimulus is higher when \(P_{T}^{IFC} < P_{T}^{IFC}\).

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 NONZERO FALSE ALARMS. For the PSM with nonzero 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 \((A + T)\) or not, \(P_{T}^{IFC}\).

As2. During the interval without the stimulus the observer can perceive (hallucinate) it with a probability \(P_{T}^{IFC}\).

As3. The observer will answer correctly in trials with correct detections and without false alarms, or will respond randomly in trials with correct detections and false alarms, or, finally, will respond 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:

\[
P_{T}^{IFC}(A, T) = (1 + P_{T}^{IFC} - P_{T}^{IFC})/2 \tag{21}
\]

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

\[
P_{T}^{IFC}(A, T) = (1 + P_{T}^{IFC} + P_{T}^{IFC} - P_{T}^{IFC} P_{T}^{IFC} - P_{T}^{IFC} P_{T}^{IFC} - P_{T}^{IFC} P_{T}^{IFC})/2 \tag{22}
\]

Thus we see how for the PM the adaptation to a 2IFC implies a competition of the sensory activity and as a consequence an increases in 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 Theoretical Analysis 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 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.

*J Neurophysiol* • doi:10.1152/jn.00341.2014 • www.jn.org
Fig. 2. Schematic representation of 2 models in a 2-interval forced choice task (2IFC) for a single modality (A) and audiotactile (B) detection tasks, showing the distributions of the sensory activities ($S_1$ and $S_2$) for the noise and for the stimulus (in B both modalities of the multisensory compound are represented, $S^T_1$, $S^A_1$, $S^T_2$, and $S^A_2$). We arbitrarily represented the noise stimulus as being presented first. For the PSM strategy (A and B, top row); after each interval the observer is in 1 of 2 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 audiotactile 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 (A) or 4 (B) states are not in contradiction and correct (i.e., stimulus in 2nd interval); error answers are given only when the 2 (A) or 4 (B) states are not in contradiction, but incorrect (i.e., stimulus in 1st interval); in all the other cases, where at least one of the decision is incoherent with the others, the observer has to guess. According to the difference model (DM; A, bottom row) in the single modality the decision is made upon the comparison of the 2 sensory activities $S_1$ and $S_2$, so the sensory activity of the 1st interval, $S_1$, has to be memorized until after the 2nd interval. According to the PM strategy (B, top row) the decision is made upon the comparison of the weighted sum of the 2 sensory activities ($S^T_1$, $S^T_2$). The weighted sum of the audiotactile sensory activity of the 1st interval, $S^T_1$, has to be memorized until after the 2nd interval. According to the PM strategy (B, top row) the decision is made upon the comparison of the weighted sum of the 2 sensory activities ($S^T_1$, $S^T_2$). To generate these distributions we simulated the PMs for 104 trials by extracting pseudorandom Gaussian values for the sensory activities.
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 before their inclusion in the study and reported normal women, age 21–25) and received monetary reward for their participation.

Apparatus. Participants sat ~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 1-mm ø probe moved by a custom-built vibratory device mounted on a table (Dancer Design, Vibrotactile Laboratory System, Liverpool, UK). The stimulation device consisted of a high-precision vibrotactile actuator (amplitude error was about \( \pm 3\% \) of the measured value). We used a custom made finger pad 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 1,100 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 5,000 Hz) of 4-ms length modulated by Gaussian envelopes. These stimuli were modeled after Liang et al. (2002). During the stimulation interval, background white noise was delivered at a 70-dB intensity 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 1-mm 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 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 and 1,100 ms. Participants underwent multiple experimental sessions (between 3 and 6) of 2 h 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 two sessions. After this phase (45 min on average), the session consisted of several test blocks of 75 trials each (~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 1,100 Hz), participants performed only the combined condition (that is, with compounded auditory stimuli of 500 and 1,100 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 et al. 1979). 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). Finally, once the mentioned voltage increase was observed, the probe was displaced 500 \( \mu \text{m} \) 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 micrometers. To establish a tolerance criterion for variations in the force exerted over the probe, the voltage necessary for a 31-Hz stimulus with a suprathreshold amplitude (100 \( \mu \text{m} \) ) 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, and 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 the studies of Schnupp (2005) and Wilson (2009, 2010a,b) because the results from these studies are in apparent contradiction. Namely, Wilson et al. reported superadditive 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 1 no stimulation condition. To better understand the results of Schnupp et al., 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 multisensory combinations based on the single modality detection data. For the test phase (see Fig. 3) against multisensory 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). To better understand the results of Schnupp et al., 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 multisensory 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 Eqs. 6, 13, and 15 in MATERIALS AND METHODS. 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, A–C. As illustrated in Fig. 3, the best fit to the empirical results is achieved by the OM (Fig. 3A): the differences between the hit probability predicted from the OM and the experimental values tend to zero. The PSM (Fig. 3B) has a tendency to underestimate the detection probabilities (reddish color). Finally, the LSM overestimates the bimodal detection probabilities (bluish color of Fig. 3C) in the majority of the cells.

To help interpreting these results we presented in Fig. 3D the bimodal detection probabilities (bimodal 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 bimodal 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 Eqs. 11 and 15, respectively (values of \( b_{1(A)} \) and \( \lambda \) were obtained from the additional hypotheses that the false alarm rate was \( \sim 3\% \), that is in accordance with the experimental data reported by Schnupp et al.). Thus, ordering the models in terms of bimodal detection probability, we can see that \( P_{\text{LSM}} > P_{\text{OM}} > P_{\text{PSM}} \), which, together with the empirical results shown in Fig. 3, A–C, clearly indicates that the experimental data lie in an intermediate level between the PSM’s 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 (underestimation) for the PSM.

As was the case with the other benchmark models, 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 to achieve two objectives: 1) having performances ranging between the PSMs and LSMs; and 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. 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 nondetected cases, are \( f_{h}(v) \) and \( f_{l}(v) \), respectively, where the output variable is \( v \) and the subscripts \( h \) and \( l \) indicate high and low state, respectively.
when two stimuli are delivered at the same time, like auditory probability to detect it is: MATERIALS AND METHODS). This neural network, akin to the MM, and a detailed description of the ANN model is reported in we implemented it in an ANN Wang (2002) model with of the OM but with completely different assumptions. D

Fig. 3 precisely what allows this model to outperform the PSM. In On the contrary for the MM this term is not zero, which is For the PSM values only at higher when the answer is Yes then when it is No. (detection) then when it is not (no-detection). The p.d.f. for the Dec pool is depicted for the 2 possible outcome of the whole net: No and Yes. Its activity is usually

The probabilities to have a low and a high state are

The p.d.f.s of the activity of Det A and Det T are plotted for the 2 possible observer internal states. Their activity is usually higher when the stimulus is detected 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 2 possible observer internal states. Their activity 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 2 possible outcome of the whole net: No and Yes. Its activity is usually higher when the answer is Yes then when it is No.

The probabilities to have a low and a high state are $P_l$ and $P_h$ (for more details, see MATERIALS AND METHODS). Therefore, when two stimuli are delivered at the same time, like auditory (A) and tactile (T), we can simply use the sum of the two: $v_A + v_T$. Therefore, under presentation of the stimulus the probability to detect it is:

$$P_{Det}^{A+T} = \int \int (v_A + v_T > \lambda) P_{h}^{A} f_{i}^{A}(v^{A}) P_{f}^{T} f_{i}^{T}(v^{T}) + P_{h}^{T} f_{i}^{T}(v^{T}) P_{f}^{A} f_{i}^{A}(v^{A}) + P_{h}^{A} f_{i}^{A}(v^{A}) P_{f}^{T} f_{i}^{T}(v^{T}) + P_{h}^{T} f_{i}^{T}(v^{T}) P_{f}^{A} f_{i}^{A}(v^{A}) d v^{A} d v^{T} \tag{2.2}$$

For the PSM $f_{i}^{A}(v)$ and $f_{h}^{A}(v)$ are delta functions with nonzero values only at $v_A$ and $v_T$, and $v_A + v_T > \lambda$. As a consequence, the last term in the above formula equals zero:

$$\int \int (v_A + v_T > \lambda) P_{h}^{A} f_{i}^{A}(v^{A}) P_{f}^{T} f_{i}^{T}(v^{T}) d v^{A} d v^{T} = 0$$

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 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 before 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 mim-

\[ J \text{ Neurophysiol} \cdot \text{doi:10.1152/jn.00341.2014} \cdot \text{www.jn.org} \]
ics 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 yes–no 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 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. The results of Schnupp et al. (2005) indicate that the ability to integrate the multisensory signals is below the LSM and therefore below the PM’s prediction. In a more recent set of studies, with a different paradigm, but again with a multisensory (audiotactile) detection task, Wilson and colleagues (2009, 2010a,b) reported superadditive (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 to 1,000 Hz. 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 the results of Schnupp et al. and Wilson et al. 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 with 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 with 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. (2005) belongs to the first of the two situations we just described: it is a yes–no multisensory 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). However, for the paradigm used by Wilson et al. (2009), 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. Also, 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 with 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 ~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 MATERIALS AND METHODS), and therefore, PSM’s predicted d’-ratio (~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 around 3% (as in the dataset of Schnupp et al. 2005), the PSM predicts about 81% performance, and its d’-ratio is still ~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 the studies of Schnupp et al. (2005) and Wilson et al. (2009) stems from differences in stimulus predictability between paradigms.

**Experiment 1: Does Audiotactile Interaction Lead to an Enhancement of Stimulus Detection?**

In this experiment we analyzed detection performance for audiotactile and unimodal stimuli using three audio (A) frequencies (13, 31, and 49 Hz) and one tactile (T) frequency (31 Hz), namely unimodal A13, A31, 768 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. Figure 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, A and 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). Figure 6, D and H, shows, respectively, the accuracy and d’-ratio averaged over observers for the different auditory frequencies.

The values found here for the d’-ratio (d’exp/d’PM) 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 conclude 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 laboratory 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). 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 the Wilson’s results, the effect is not frequency dependent; indeed, we did not find any statistical difference among 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 multisensory 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, while ANN has to assume memory of the detection. Accordingly, in experiment 2, we used a bimodal audiotactile detection task introducing a range...
of SOAs between the two sensory events forming the bimodal stimulus as long as 1s.

Experiment 2: Audiotactile 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 multimodal 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 milliseconds) or instead infer the presence of late-decision interactions (for time windows amounting to several hundreds of milliseconds). In experiment 2, the interest was thus on determining whether interactive effects, as signaled by a violation of the Pythagorean criterion, were observed for audiotactile stimuli at long SOA values. 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 1,100 Hz, as used in the seminal work of Marill (1956). We are aware, that contrary to what was claimed by Marill, more recent articles 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 one-hit model. However, this debate, interesting as it may be, is completely out of the goal of the present article. We only chose 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 +
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 SOA 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. 8B). One can see again that the average value of the $d'$-ratio, for the condition with SOA = 750 ms, is similar to the one reported in our two previous experiments (bootstrapping from our distribution a dataset with 1,000 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 = 750 ms, detection of the compound surpasses the PM’s criterion. On the contrary the average value for the $d'$-ratio, for the condition with SOA = 0 ms, is $\sim 0.7$. That is, at SOA = 0 ms, 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 (audiotactile or audioaudio). Taken together, the results of experiment 3 suggest that the $d'$-ratio $\sim 1.3$ 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 multisensory (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 the dataset of Schnupp, 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 the observer when detecting unimodal stimuli, in the 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 the PM’s criterion) and Schnupp et al. (2005) (where observed data fell below LSM’s 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 to capitalize on the OM’s simplicity and power to establish a good noninteraction benchmark while at the same time having a direct translation to models based on well-known neurophysiological processes. The MM shares characteristics of both the PM and the PSM. The main objective of this model is to show the
inadequacy of concluding for early sensory interactions based on multisensory results that surpass the PM or 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 quasibinary detection stage (whereas 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 nonzero 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 quasibinary 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 multisensory 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. 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. From this description we cannot calculate quantitatively the prediction for the hit probability, but we can only indicate whether the hit probability 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 is 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 anticorrelated (see mathematical description in MATERIALS AND METHODS). Similarly, the probability of a correct answer to the compound stimuli is higher when the probabilities of both modalities are anticorrelated. Even though our description of the nonindependent 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 multisensory 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 nonindependence 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, it 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 (Otto and Mamassian 2012; Otto et al. 2013) 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 measurement of interactive effects from multisensory stimuli in audiotactile detection tasks, we presented three separate experiments. Experiment 1 demonstrated integrative effects with audiotactile stimuli in the flutter range in a paradigm akin to the one used by 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 audiotactile) were very similar to those reported by Wilson and colleagues, who 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. Experiment 2 sought to further test the accuracy of the prediction made by the PM by introducing a bimodal audiotactile detection task across a range of stimulus SOAs, that is long intervals (−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 multisensory interactions at the sensory level. Importantly, the relevance of the results of experiment 3, where a compound of two audio stimuli was used, cannot be understated. Indeed, the results from experiment 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

---

1 We are grateful to an anonymous referee that suggested to us to explore this possibility.
at which it takes place is not completely clear (Kiang and Sachs 1968). Experiment 3 demonstrated that, given the right conditions (2 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 multisensory 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 multisensory interactions in detection tasks and we do not make claims about other cognitive processes such as spatial representation (i.e., see theoretical 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 to measure such interactions one has to use an appropriate baseline and that past studies have often used baselines that tended to overestimate integration.

ACKNOWLEDGMENTS

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

GRANTS

M. Panunzi and G. Deco were funded by the Consolider-Ingenio 2010 CDS 2007-00012, European Research Council (ERC) Advanced Grant DY-STRUCTURE (no. 295129), the European Commission Seventh Research Framework Programme (FP7)-Future and Emerging Technologies (FET) FlagSHIP Human Brain Project (no. 604102), the FP7-Information and Communication Technologies (ICT) BrainScaleS (no. 269921), and Plan Estatal de Fomento de la Investigación Científica y Técnica de Excelencia (PSI2013-42091-P). S. Soto-Faraco, A. Pérez-Bellido, and J. López-Moliner were funded by the Ministerio de Economía y Competitividad (PSI2013-42626-P), Agència de Gestió d’Ajuts Universitaris i de Recerca (AGAUR) Generalitat de Catalunya (2014SGR856), and the ERC (SG10-2013-263145).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS


REFERENCES


