

PRIMER

Multisensory Causal Inference in the Brain

Christoph Kayser^{1*}, Ladan Shams²

1 Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom,

2 Departments of Psychology and BioEngineering and the Interdepartmental Neuroscience Program, University of California, Los Angeles, Los Angeles, California, United States of America

* christoph.kayser@glasgow.ac.uk

Abstract

At any given moment, our brain processes multiple inputs from its different sensory modalities (vision, hearing, touch, etc.). In deciphering this array of sensory information, the brain has to solve two problems: (1) which of the inputs originate from the same object and should be integrated and (2) for the sensations originating from the same object, how best to integrate them. Recent behavioural studies suggest that the human brain solves these problems using optimal probabilistic inference, known as Bayesian causal inference. However, how and where the underlying computations are carried out in the brain have remained unknown. By combining neuroimaging-based decoding techniques and computational modeling of behavioural data, a new study now sheds light on how multisensory causal inference maps onto specific brain areas. The results suggest that the complexity of neural computations increases along the visual hierarchy and link specific components of the causal inference process with specific visual and parietal regions.

Introduction

Our brain is continuously faced with a plethora of sensory inputs impinging on our senses. At any moment we see, hear, touch, and smell, and only the coordinated interplay of our senses allows us to properly interact with the environment [1]. How the brain organizes all these sensory inputs into a coherent percept remains unclear. As shown in a new study by Rohe and Noppeney, important insights can be obtained by combining computational models with carefully crafted analysis of brain activity [2].

The brain needs to solve several computational problems to make sense of its environment. Besides the analysis of specific sensory attributes (for example, to segment a scene into its constituent objects), two critical problems involve the inputs to different senses. One is the “multisensory integration problem”: how information is synthesized (or fused) across the senses. For instance, speech perception usually relies on the integration of auditory and visual information (listening to somebody’s voice while seeing his or her lips move). This integration problem can be challenging, as each sense only provides a noisy and possibly biased estimate of the respective attribute [3,4].

In addition, the brain needs to solve the “causal inference problem” [5–7]: it has to decide which sensory inputs likely originate from the same object and hence provide complementary evidence about this and which inputs originate from distinct objects and hence should be processed separately. One example is at a cocktail party, where many faces and voices can make it a challenge to know who called our name. Another example is at a ventriloquist’s performance,



OPEN ACCESS

Citation: Kayser C, Shams L (2015) Multisensory Causal Inference in the Brain. *PLoS Biol* 13(2): e1002075. doi:10.1371/journal.pbio.1002075

Published: February 24, 2015

Copyright: © 2015 Kayser, Shams. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: CK is supported by BBSRC grant BB/L027534/1, LS by NSF grant 1057969. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

where we attribute the voice to the puppet rather than the actor. In practice, the tasks of inferring the causal structure and of obtaining precise estimates of sensory attributes are highly intertwined, as causal inference depends on the similarity of different sensory features, while the estimate of each attribute depends on the inferred causal origin. For example, the association of a face and a voice depends on both the perceived location of each as well as the match between semantic, social, or physical attributes derived from faces and voices. Hence, solving the causal inference problem has to rely on a number of factors such as spatial, temporal, and structural congruency, prior knowledge, and expectations.

In the 19th century, von Helmholtz already noted that perception requires solving multiple inference problems [8]. Yet, laboratory studies on multisensory integration often avoid the causal inference problem by crafting multisensory stimuli that leave little doubt as to whether they provide evidence about the same sensory object. However, the brain mechanisms underlying multisensory perception in everyday tasks can probably only be understood by considering both the integration and causal inference problems [9]. Fortunately, the field of Bayesian perception has provided a conceptual framework in which both can be cast in unified statistical terms [10].

Bayesian Approaches to Multisensory Perception

Bayesian statistics describes sensory representations in probabilistic terms, attributing likelihoods to each possible encoding of a sensory attribute [11]. Moreover, it describes how different variables interact in determining the outcome, such as how prior knowledge affects perceptual estimates or how inputs from two senses combine. As shown in Fig. 1A, when considered independently, each sensory modality can be conceptualized as providing a noisy (probabilistic) estimate of the same attribute. Yet, under the assumption of a common source, Bayesian inference predicts the multisensory estimate arising from the combination of both senses by weighing each input in proportion to its reliability (Fig. 1B).

This approach has provided invaluable insights about various aspects of perceptual multisensory integration [3,4,12,13] and helped to identify the sensory computations likely to be carried out by the underlying neural processes [14,15]. For example, studies on visual-vestibular integration have shown that the mathematical rules by which neural populations combine visual-vestibular information follow Bayesian predictions and that the relative weights attributed to each modality in the neural code scale with sensory reliability analogous to the perceptual weights [16].

Probabilistic Models for Causal Inference

The Bayesian approach can be extended to model the causal inference problem by including inference about the environment's causal structure (Fig. 1C). Depending on the task that the nervous system has to solve, different perceptual decision-making strategies can be used to derive estimates of sensory attributes based on the probabilities of each possible causal structure [17]. For example, when trying to minimize the error in the perceptual estimate, e.g., to precisely localize a speaker, the optimal estimate is the nonlinear weighted average of two terms: one estimate derived under the assumption that two inputs originate from a single source (fusion) and one derived under the assumption that they have separate sources (segregation), with each estimate weighted by the probability of the respective causal structure. This strategy is known as “model averaging” (Fig. 1C).

In testing whether Bayesian causal inference can account for human perception, localization tasks have been proved useful (Fig. 2) [6,17–19]. When visual and acoustic stimuli are presented with a small spatial disparity (Fig. 2A), they are likely perceived as originating from the same

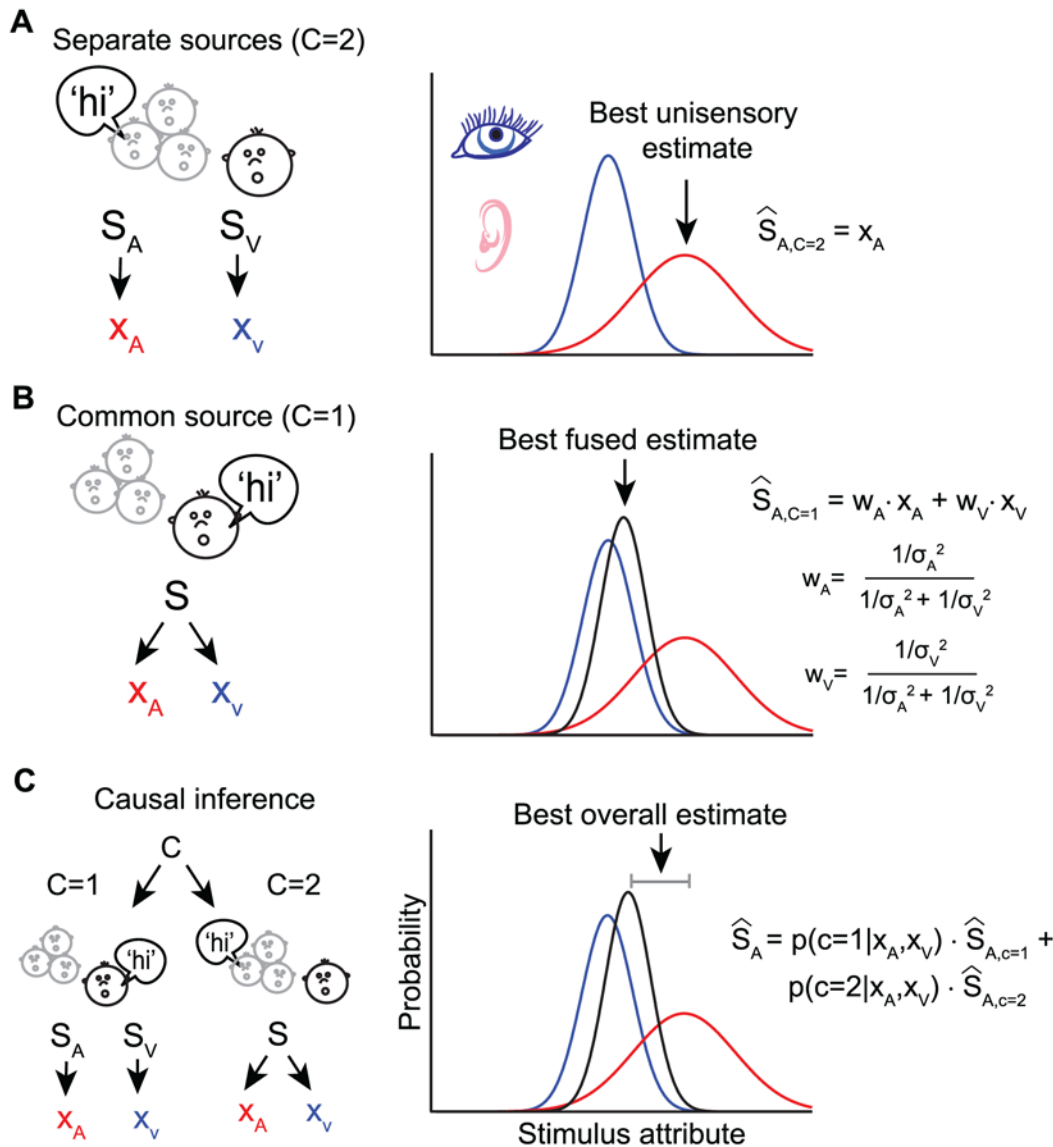


Fig 1. Bayesian models of multisensory integration. Schematic of different causal structures in the environment giving rise to visual and acoustic inputs (e.g., seeing a face and hearing a voice) that may or may not originate from the same speaker. The left panels display the inferred statistical causal structure, with S_A , S_V , and S denoting sources for acoustic, visual, or multisensory stimuli and X_A and X_V indicating the respective sensory representations (e.g., location). The right panels display the probability distributions of these sensory representations and the optimal estimate of stimulus attribute (e.g., location) derived from the Bayesian model under different assumptions about the environment. For the sake of simplicity of illustration, it is assumed that the prior probability of the stimulus attribute is uniform (and therefore not shown in the equations and figures). (A) Assuming separate sources ($C = 2$) leads to independent acoustic and visual estimates of stimulus location, with the optimal value matching the most likely unisensory location. (B) Assuming a common source ($C = 1$) leads to integration (fusion). The optimal Bayesian estimate is the combination of visual and acoustic estimates, each weighted by its relative reliability (with σ_A and σ_V denoting the inverse reliability of each sense). (C) In Bayesian causal inference (assuming a model-averaging decision strategy), the two different hypotheses about the causal structure (e.g., one or two sources) are combined, each weighted by its inferred probability given the visual and acoustic sensations. The optimal stimulus estimate is a mixture of the unisensory and fused estimates.

doi:10.1371/journal.pbio.1002075.g001

source, hence their evidence about the spatial location should be fused. In contrast, if their spatial disparity is large (Fig. 2B), the two inputs likely arise from distinct sources and should be segregated. Interestingly, the model also predicts a more surprising behaviour in conditions in which spatial disparity is moderate (Fig. 2C). In this case, the two senses get partially integrated, weighted by the relative likelihood of one or two origins [17]—assuming that the brain adopts a model

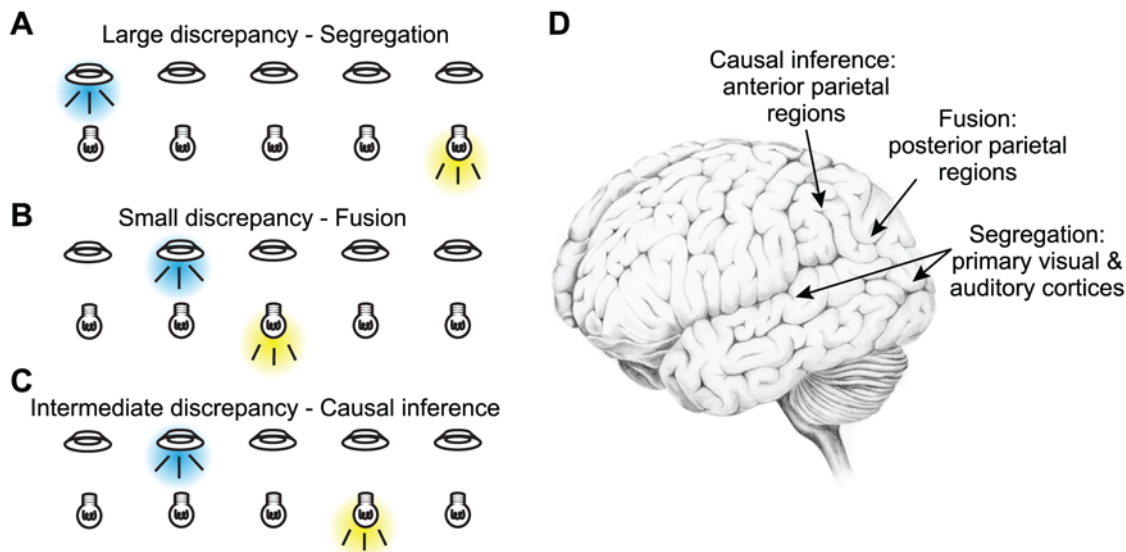


Fig 2. Causal inference about stimulus location. (A–C) Schematized spatial paradigm employed by several studies on audio-visual causal inference. Brief and simple visual (flashes) and auditory (noise bursts) stimuli are presented at varying locations along azimuth and varying degrees of discrepancy across trials. When stimuli are presented with large spatial discrepancy (panel A), they are typically perceived as independent events and are processed separately. When they are presented with no or little spatial discrepancy (panel B), they are typically perceived as originating from the same source and their spatial evidence is integrated (fused). Finally, when the spatial discrepancy is intermediate (panel C), causal inference can result in partial integration: the perceived locations of the two stimuli are pulled towards each other but do not converge. Please note that the probability distributions corresponding to each panel are shown in the respective panels in Fig. 1. (D) Schematized summary of the results by Rohe and Noppeney. Early sensory areas mostly reflect the unisensory evidence corresponding to segregated representations, posterior parietal regions reflect the fused spatial estimate, and more anterior parietal regions reflect the overall causal inference estimate. This distributed pattern of sensory representations demonstrates the progression of causal inference computations along the cortical hierarchy.

doi:10.1371/journal.pbio.1002075.g002

averaging strategy, which seems to be the case in many observers [19]. Therefore, Bayesian causal inference successfully explains perceptual judgements across the range of discrepancies, spanning a continuum, from fusion to partial integration to segregation.

Again, its success in describing human perception suggests that this Bayesian model could also provide a framework to map the underlying neural processes onto distinct sensory computations. For example, an important question is whether the same or distinct brain regions reflect the integration process and the causal inference computation. This is precisely what the study by Rohe and Noppeney addressed.

Mapping Causal Inference onto Sensory Pathways

In their study, participants localized audiovisual signals that varied in spatial discrepancy and visual reliability while their brain activity was measured using functional magnetic resonance imaging (fMRI). The authors first fit the causal inference model to the perceptual data, which enabled them to investigate the mapping between brain activity and the different spatial estimates predicted by the model; the estimates were predicted by either unisensory input (corresponding to the distinct causal origins hypothesis), by the fusion of the two sensations (corresponding to the single causal origin hypothesis), or by the causal inference model (the weighted combination of fusion and segregation). Addressing this question required an additional step of data analysis: linking the selectivity to spatial information reflected in distributed patterns of fMRI activity to the spatial estimates predicted by each model component. Luckily, methods of decoding analysis provide a means to establish such a link [20] and allowed the authors to associate each brain region of interest with the best-matching sensory estimate predicted by the inference model.

As may be expected, some regions (early visual and auditory cortices) predominantly reflected the unisensory inputs and hence were only a little affected by any multisensory computation (see Fig. 2D). Other regions, e.g., those involved in establishing spatial maps (posterior regions in the intraparietal sulcus), reflected the fused estimate. Thus, in these regions, automatic integration processes seem to occur that merge the spatial evidence provided by different modalities, weighted by their reliability, but regardless of how likely it is that these relate to the same object. And finally, regions more anterior in the intraparietal sulcus encoded the spatial estimate as predicted by the causal inference model, hence adapting their sensory representation based on the likely causal origin.

Overall, the new results show that different neural processes along the sensory pathways reflect distinct estimates about the localization of sensory events. Some estimates seem to arise mostly in a simple unisensory manner, while others exhibit the computationally complex nature required for causal inference. In addition, they suggest that sensory fusion and causal inference, at least in the context of spatial perception, are distributed processes not necessarily occurring in the same regions. And finally, they reveal a gradual emergence of multisensory computations along “visual” pathways. The data support both the traditional notion that multisensory perception is mostly implemented by higher-level association regions and the more recent notion that early sensory regions also participate in multisensory encoding [21,22]. Most importantly, however, they show how model-driven neuroimaging studies allow us to map complex sensory operations such as causal inference onto the sensory hierarchies.

One conclusion from this and previous studies is that multisensory perception does not result from a single and localized process—that would fit the often and sometimes abused term “multisensory integration” [23]. Rather, multisensory perception arises from the interplay of many processes and a network of interacting regions that implement these, each possibly relying on a different assumption about the causal structure of the environment and implementing a different sensory computation. Ultimately, it may be impossible to fully understand localized multisensory processes without considering them in the big picture of a possibly hierarchical but certainly distributed organization.

Conclusions

As with any major step forward, the results pose many new questions. For example, judging the environment’s causal structure relies on prior knowledge and experience [7,12], but we don’t know whether the processes of causal inference and incorporating prior information are implemented by the same neural processes. It will also be important to see whether there are brain regions generically involved in multisensory inference and not specific to spatial attributes. Furthermore, it seems natural to look for similar gradual changes in multisensory computations along other sensory pathways. For example, our understanding of auditory pathways may benefit from such model-based decoding studies [24]. Finally, the roles of attention and task relevance for multisensory perception remain controversial. Attentional selection modulates multisensory integration, and multisensory coincidences attract attention and amplify perception [25]. It remains unclear how attentional state or task relevance influence which sensory variables are represented in any brain region, and recent studies reveal complex patterns of mixed selectivity to task- and sensory-related variables in higher association regions [26]. Disentangling the impact of attention and task nature on multisensory encoding and what can actually be measured using neuroimaging signals remains a challenge for the future.

Neuroimaging studies provide critical insights into the large-scale organization of perception, but the underlying local mechanisms of neural population coding remain to be confirmed. Signatures of multisensory encoding at the single neuron level can be subtle [27], and the

mixed selectivity of higher-level sensory regions can render the link between neural populations and neuroimaging ambiguous [28]. Again model-driven approaches may help, for example, by providing testable hypothesis about large-scale population codes that can be extracted from electrophysiological recordings or neuroimaging [14].

On a methodological side, recent work has shown how combining fMRI with probabilistic models of cognition can be a very powerful tool for understanding brain function [29,30]. In line with this, Rohe and Noppeney show that the combination of statistical models of perception and brain decoding has the power to enlighten our understanding of perception far beyond more descriptive approaches. Yet, studies such as this require carefully crafted models and efficient paradigms to overcome the poor signal-to-noise ratio sometimes offered by neuroimaging. As a result, further advances in both perceptual models and signal understanding and analysis are required to eventually uncover why we sometimes benefit from seeing a speaker in a noisy environment and why we get fooled by the ventriloquist's puppet.

References

1. Stein BE, editor (2012) *The New Handbook of Multisensory Processing*. Cambridge: MIT Press.
2. Rohe T, Noppeney U (2015) Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biol* 13: e1002073. doi: [10.1371/journal.pbio.1002055](https://doi.org/10.1371/journal.pbio.1002055) PMID: [25575020](https://pubmed.ncbi.nlm.nih.gov/25575020/)
3. Angelaki DE, Gu Y, DeAngelis GC (2009) Multisensory integration: psychophysics, neurophysiology, and computation. *Curr Opin Neurobiol* 19: 452–458. doi: [10.1016/j.conb.2009.06.008](https://doi.org/10.1016/j.conb.2009.06.008) PMID: [19616425](https://pubmed.ncbi.nlm.nih.gov/19616425/)
4. Ernst MO, Bulthoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8: 162–169. PMID: [15050512](https://pubmed.ncbi.nlm.nih.gov/15050512/)
5. Shams L, Beierholm UR (2010) Causal inference in perception. *Trends Cogn Sci* 14: 425–432. doi: [10.1016/j.tics.2010.07.001](https://doi.org/10.1016/j.tics.2010.07.001) PMID: [20705502](https://pubmed.ncbi.nlm.nih.gov/20705502/)
6. Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, et al. (2004) Unifying multisensory signals across time and space. *Exp Brain Res* 158: 252–258. PMID: [15112119](https://pubmed.ncbi.nlm.nih.gov/15112119/)
7. Roach NW, Heron J, McGraw PV (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc Biol Sci* 273: 2159–2168. PMID: [16901835](https://pubmed.ncbi.nlm.nih.gov/16901835/)
8. Hatfield G (1990) *The natural and the normative theories of spatial perception from Kant to Helmholtz*. Cambridge, Mass: MIT Press.
9. Shams L, Ma WJ, Beierholm U (2005) Sound-induced flash illusion as an optimal percept. *Neuroreport* 16: 1923–1927. PMID: [16272880](https://pubmed.ncbi.nlm.nih.gov/16272880/)
10. Shams L, Beierholm U (2011) Humans' multisensory perception, from integration to segregation, follows Bayesian inference. In: Trommershauser J, Kording K, Landy MS, editors. *Sensory Cue Integration*. Oxford University Press. pp. 251–262.
11. Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. *Annu Rev Psychol* 55: 271–304. PMID: [14744217](https://pubmed.ncbi.nlm.nih.gov/14744217/)
12. Ernst MO (2007) Learning to integrate arbitrary signals from vision and touch. *J Vis* 7: 7 1–14.
13. Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429–433. PMID: [11807554](https://pubmed.ncbi.nlm.nih.gov/11807554/)
14. Ma WJ, Rahmati M (2013) Towards a neural implementation of causal inference in cue combination. *Multisens Res* 26: 159–176. PMID: [23713204](https://pubmed.ncbi.nlm.nih.gov/23713204/)
15. Pouget A, Deneve S, Duhamel JR (2002) A computational perspective on the neural basis of multisensory spatial representations. *Nat Rev Neurosci* 3: 741–747. PMID: [12209122](https://pubmed.ncbi.nlm.nih.gov/12209122/)
16. Fetsch CR, DeAngelis GC, Angelaki DE (2013) Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nat Rev Neurosci* 14: 429–442. doi: [10.1038/nrn3503](https://doi.org/10.1038/nrn3503) PMID: [23686172](https://pubmed.ncbi.nlm.nih.gov/23686172/)
17. Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, et al. (2007) Causal inference in multisensory perception. *PLoS ONE* 2: e943. PMID: [17895984](https://pubmed.ncbi.nlm.nih.gov/17895984/)
18. Beierholm UR, Quartz SR, Shams L (2009) Bayesian priors are encoded independently from likelihoods in human multisensory perception. *J Vis* 9: 23.21–29. doi: [10.1167/9.12.23](https://doi.org/10.1167/9.12.23) PMID: [20053114](https://pubmed.ncbi.nlm.nih.gov/20053114/)
19. Wozny DR, Beierholm UR, Shams L (2010) Probability matching as a computational strategy used in perception. *PLoS Comput Biol* 6: e1000871. doi: [10.1371/journal.pcbi.1000871](https://doi.org/10.1371/journal.pcbi.1000871) PMID: [20700493](https://pubmed.ncbi.nlm.nih.gov/20700493/)

20. Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience* 37: 435–456. doi: [10.1146/annurev-neuro-062012-170325](https://doi.org/10.1146/annurev-neuro-062012-170325) PMID: [25002277](https://pubmed.ncbi.nlm.nih.gov/25002277/)
21. Kayser C, Logothetis NK, Panzeri S (2010) Visual enhancement of the information representation in auditory cortex. *Curr Biol* 20: 19–24. doi: [10.1016/j.cub.2009.10.068](https://doi.org/10.1016/j.cub.2009.10.068) PMID: [20036538](https://pubmed.ncbi.nlm.nih.gov/20036538/)
22. Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32: 9–18. doi: [10.1016/j.tins.2008.09.012](https://doi.org/10.1016/j.tins.2008.09.012) PMID: [19012975](https://pubmed.ncbi.nlm.nih.gov/19012975/)
23. Stein BE, Burr D, Constantinidis C, Laurienti PJ, Alex Meredith M, et al. (2010) Semantic confusion regarding the development of multisensory integration: a practical solution. *Eur J Neurosci* 31: 1713–1720. doi: [10.1111/j.1460-9568.2010.07206.x](https://doi.org/10.1111/j.1460-9568.2010.07206.x) PMID: [20584174](https://pubmed.ncbi.nlm.nih.gov/20584174/)
24. Shamma S, Fritz J (2014) Adaptive auditory computations. *Curr Opin Neurobiol* 25: 164–168. doi: [10.1016/j.conb.2014.01.011](https://doi.org/10.1016/j.conb.2014.01.011) PMID: [24525107](https://pubmed.ncbi.nlm.nih.gov/24525107/)
25. Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci* 14: 400–410. doi: [10.1016/j.tics.2010.06.008](https://doi.org/10.1016/j.tics.2010.06.008) PMID: [20675182](https://pubmed.ncbi.nlm.nih.gov/20675182/)
26. Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, et al. (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497: 585–590. doi: [10.1038/nature12160](https://doi.org/10.1038/nature12160) PMID: [23685452](https://pubmed.ncbi.nlm.nih.gov/23685452/)
27. Kayser C, Lakatos P, Meredith MA (2012) Cellular Physiology of Cortical Multisensory Processing. In: Stein BE, editor. *The New Handbook of Multisensory Processing*. Cambridge, MA: MIT Press. pp. 115–134.
28. Dahl C, Logothetis N, Kayser C (2009) Spatial Organization of Multisensory Responses in Temporal Association Cortex. *J Neurosci* 29: 11924–11932. doi: [10.1523/JNEUROSCI.3437-09.2009](https://doi.org/10.1523/JNEUROSCI.3437-09.2009) PMID: [19776278](https://pubmed.ncbi.nlm.nih.gov/19776278/)
29. Lee SW, Shimojo S, O'Doherty JP (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81: 687–699. doi: [10.1016/j.neuron.2013.11.028](https://doi.org/10.1016/j.neuron.2013.11.028) PMID: [24507199](https://pubmed.ncbi.nlm.nih.gov/24507199/)
30. Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP (2013) The neural representation of unexpected uncertainty during value-based decision making. *Neuron* 79: 191–201. doi: [10.1016/j.neuron.2013.04.037](https://doi.org/10.1016/j.neuron.2013.04.037) PMID: [23849203](https://pubmed.ncbi.nlm.nih.gov/23849203/)