BRILL Multisensory Research (2025) DOI:10.1163/22134808-bja10150 brill.com/msr

From Exploration to Integration: 15 Years of Multisensory Research at Peking University

Lihan Chen [1,2,3,*]

[1] School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, P.R. China

[2] Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, P.R. China

[3] National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing 100871, P.R. China

[*]E-mail: CLH@pku.edu.cn

ORCID iD: 0000-0002-0337-4177

Received 8 September 2024; accepted 7 May 2025; published online 22 May 2025

Abstract

We introduce how 'the rule of thumb' of multisensory integration, which was proposed in the seminal book *The Merging of the Senses* by Stein and Meredith in 1993, inspired the empirical research work conducted at Multisensory lab, Peking University (China) for the last 15 years. We also outline the potential research trends in the multisensory research field.

Keywords

brain oscillations, ventriloquism, McGurk effect, crossmodal dynamic capture, interoception, individual differences, artificial intelligence

1. Introduction

The Merging of the Senses written by Barry E. Stein and M. Alex Meredith (Stein and Meredith, 1993), has been instrumental in laying the groundwork for the burgeoning field of multisensory studies, and stands as a foundational text that has significantly influenced my journey into the realm of multisensory research. The book introduces three key principles of multisensory integration — the Principle of Inverse Effectiveness, the Spatial Principle, and the

Published with license by Koninklijke Brill BV | DOI: 10.1163/22134808-bja10150
 | ISSN: 2213-4794 (print) 2213-4808 (online)

Temporal Principle — which provide a foundational framework for understanding how the brain integrates multisensory information, though factors like attention and context can influence these processes. Guided by these principles, our multisensory research group in Beijing has benefited from studying this book, using it to design experiments involving ventriloquism paradigms (Chen and Vroomen, 2013; Chen *et al.*, 2018), crossmodal correspondence in normal and atypical groups (Chen *et al.*, 2016; Feng *et al.*, 2023a), and recent interoception paradigms (Gong *et al.*, 2022). These principles have also helped explain our findings in behavioral, modeling, and neural oscillation studies (Chen *et al.*, 2018; Guo *et al.*, 2024; Gupta and Chen, 2016).

The book examines neural computations in the superior colliculus, focusing on concepts like 'superadditivity' and 'inverse effectiveness' crucial for understanding sensory interactions (Chapter 9). Using feline models, it explores how sensory convergence amplifies or suppresses certain stimulus combinations based on perceived common causes (Chapter 10). The authors propose that multisensory integration depends on unimodal receptive field characteristics, with optimal integration occurring through overlapping activity patterns rather than simultaneous stimulus onset. Spatially coincident stimuli enhance responses, while disparate stimuli may lead to suppression or no interaction.

2. Multisensory Contributions to the Perception of Movement: Crossmodal Dynamic Capture

My foray into the multisensory field was driven by a confluence of personal fascination and serendipitous opportunities. In 2007, as a PhD student at Ludwig Maximilian University (LMU), Munich, I initially proposed a thesis on cognitive control and the Simon effect. However, upon my arrival in Munich, I discovered that a professor I had intended to work with had retired, and meanwhile, a new multisensory lab was being established. Intrigued by this emerging field, I swiftly immersed myself in exploring potential research questions. Within a span of two months, I actively embarked on my first foray into 'multisensory experiments'. One of these was crossmodal dynamic capture.

Crossmodal dynamic capture is a multisensory phenomenon where the perception of a dynamic stimulus in one sensory modality (e.g., vision) is influenced or 'captured' by a stimulus in another modality (e.g., audition). This occurs when spatially and temporally aligned stimuli from different senses interact, often creating perceptual illusions, such as the 'sound-induced flash illusion', where a single flash paired with multiple beeps is perceived as multiple flashes (Hirst *et al.*, 2020; Shams *et al.*, 2000). This seminal concept has been noticeably realized in an influential paper by Freeman and Driver

(2008). This groundbreaking study unveiled the phenomenon of temporal ventriloquism, demonstrating that the timing of a static sound could bias the spatio-temporal processing of visual apparent motion, as induced by visual bars oscillating between opposite hemifields (Freeman and Driver, 2008).

In the spatial domain of motion, it was observed that one sense could 'dominate' the perception of motion cues in another sense, rather than causing interference. This intriguing phenomenon is known as capture. During the construction of the multisensory lab at LMU, we utilized a tactile device named the Heijobox, which was capable of delivering more potent tactile stimuli. With this device, we were able to replicate an audiotactile version of the crossmodal capture illusion, as originally described by Freeman and Driver (2008), further explored in our subsequent publication (Chen *et al.*, 2011).

The *Simon effect* occurs when reaction times are faster for spatially congruent stimulus–response pairs (same side) than incongruent ones (opposite sides), even if location is task-irrelevant. In crossmodal tasks (e.g., visual–auditory), it shows that spatial information is integrated across senses, influencing responses automatically. This effect highlights the brain's tendency to process spatial compatibility, revealing insights into attention, multisensory integration, and response selection (Cespón *et al.*, 2020).The crossmodal variant of the Simon effect, which was a part of my initial experiments, targeting on the audiotactile modalities, was later documented in a publication (Zheng and Chen, 2018) (Fig. 1). Indeed, including this study, the experimental work on spatial and temporal ventriloquism has lasted about 10 years and marked a significant milestone in my academic journey, solidifying my commitment to the multisensory research domain (Chen *et al.*, 2018).

A compelling area of inquiry within multisensory research is the debate surrounding the existence of intermodal motion. Our studies have shown that both moving and asynchronous static sounds can indeed capture intermodal apparent motion, such as visual–tactile and tactile–visual. Interestingly, while auditory directional cues exert a lesser influence on the perception of intramodal visual apparent motion compared to intramodal tactile or intermodal visual/tactile apparent motion, auditory temporal information appears to have a uniform impact across both intramodal and intermodal apparent motion scenarios. These observations suggest that intermodal apparent motion is similarly influenced by dynamic or static auditory information as intramodal visual or tactile apparent motion. The crossmodal dynamic capture effect is particularly pronounced for functionally weaker signals in spatial localization, echoing the 'inverse effectiveness principle'. Consistent with this principle, the auditory capture effect observed in Experiment 1 of Chen and Zhou (2011) was more pronounced for tactile stimuli than for visual stimuli, as was the susceptibility to visual–tactile or tactile–visual stimuli.

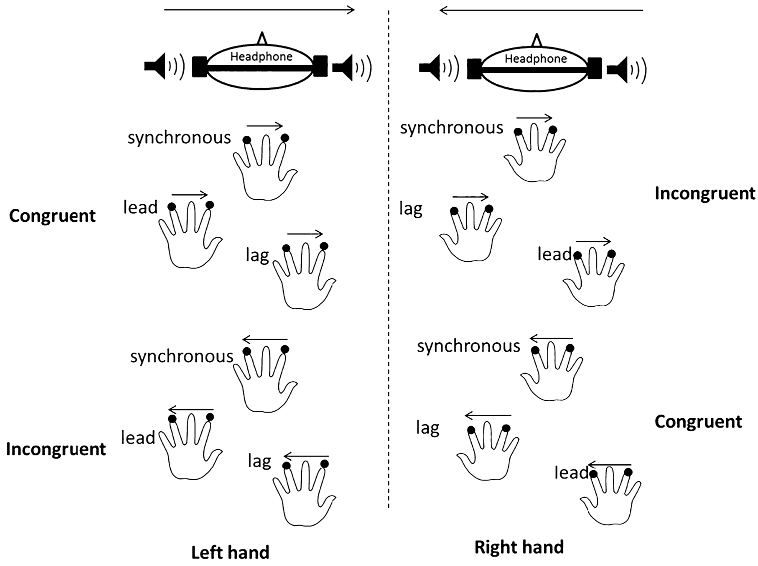

Figure 1. Tactile Simon apparent motion and crossmodal capture in Zheng and Chen (2018). We used a headset to deliver sounds. A 2 (left *vs* right hand) × 2 (leftward *vs* rightward motion) × 4 (conditions: baseline, auditory leading, auditory synchronous and auditory lagging) factorial design was adopted. Participants were required to discriminate the tactile motion, irrespective of the beeps (Zheng and Chen, 2018). Temporal course of resolving conflicts between spatial codes during attentional shifts, including attentional re-engagement, may account for the tactile Simon-like effect.

In preparation for establishing the multisensory lab in Beijing, we had to tailor specific devices to facilitate the study of crossmodal dynamic capture in spatial localization and the discrimination of crossmodal apparent motion across various combinations of visual, auditory, and tactile modalities (Chen *et al.*, 2014; Jiang and Chen, 2013). By adhering to the spatial and temporal rules, as well as considering sensory dominance (Fig. 2), we were able to build a platform (Fig. 3) and construct a 'lookup' table of sensory dominance relationships (somehow I would ask students to check before they implement the experiments). This allowed us to predict the potential outcomes of crossmodal capture in the spatial domain, providing a robust framework for our experimental design and analysis.

In a comprehensive synthesis of empirical data, Vroomen and I authored a popular review paper on ventriloquism (Chen and Vroomen, 2013). Our work expanded beyond the traditional exploration of ventriloquism to encompass the McGurk effect, with additional studies conducted by our team (Feng *et al.*, 2021, 2023a, 2023b). These two seminal paradigms (ventriloquism and McGurk effect) have been extensively utilized in the field, leading to a robust

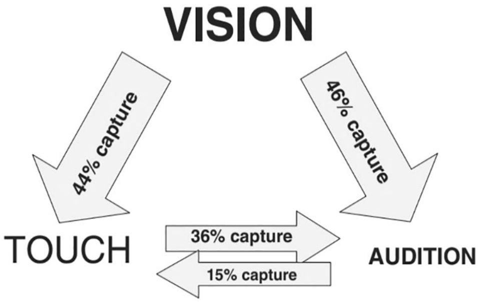

Figure 2. Schematic representation of the results of a study by Soto-Faraco and coworkers (Soto-Faraco, 2003; Soto-Faraco *et al.*, 2004), in which every possible pairing of target and distractor modality (across audition, vision and touch) was evaluated by asking participants to judge the direction of a moving stimulus in the target sensory modality while attempting to ignore a moving stimulus in the distractor modality. The number within each arrow represents the magnitude of the congruency effect.

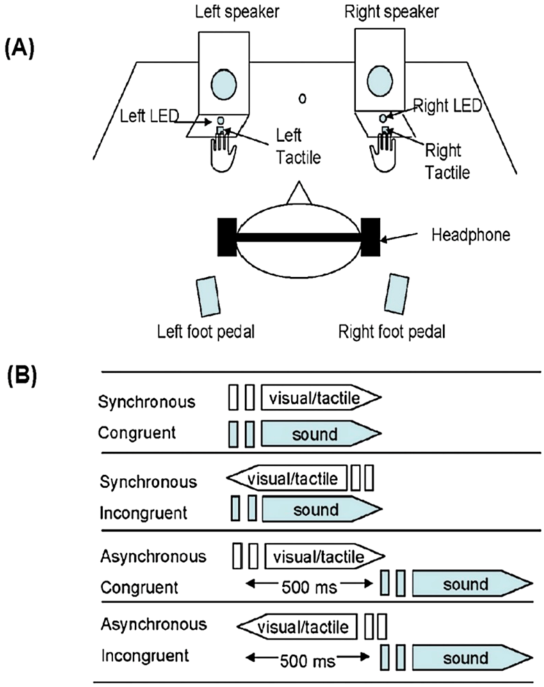

Figure 3. Experimental setup and temporal correspondence of crossmodal motion streams. (A) The participant placed two middle fingers on the tactile actuators, which were embedded into foams placed just in front of the two speakers, Two LEDs were collocated with the two actuators, respectively. (B) Spatial and temporal correspondences between auditory input and visual/tactile target stimuli. The auditory beeps could occur either congruently or incongruently with the target motion stream, simultaneously or 500 ms later with respect to the visual/tactile targets.

Figure 4. Presenters at the IMRF2016 Symposium 1: 40 years of the McGurk–MacDonald Effect. (Left) From left to right: John MacDonald, Julia Irwin, Michael Beauchamp, Salvador Soto-Faraco. Jean Vroomen presented a recorded video talk, due to unexpected visa issues and not being able to do an on-site presentation. (Right). John MacDonald giving the lecture.

understanding and validation of the principles governing multisensory integration. The McGurk effect is a pivotal phenomenon in multisensory research, demonstrating how the brain integrates auditory and visual information to perceive speech. When conflicting cues are presented (e.g., hearing ‘ba’ while seeing ‘ga’), the brain fuses them into a new perception, like hearing ‘da.’ This effect highlights the critical role of crossmodal integration in speech perception, proving that senses work together rather than independently. To celebrate the 40th anniversary of the seminal publication of McGurk effect, we specially invited Dr John McDonald to give a reviewing speech on the McGurk effect for IMRF 2016 (Figure 4). Dr. McDonald reviewed 40 years research of ‘McGurk’ in multiple facets, spanning from the factors affecting/not affecting the illusion, the progress of research methods, to developments of theoretical accounts.

Our research has now spanned across behavioral, neuroimaging, and computational approaches, significantly contributing to the multifaceted landscape of multisensory research, as initially outlined by McGurk and MacDonald in 1976.

3. Wave–Particle Duality: from Temporal Window to Temporal Averaging/Oscillation

In the conventional wisdom of multisensory research, the temporal window — the temporal proximity between crossmodal events — is deemed a critical factor for successful multisensory integration (MI). The boundaries of this window can be influenced by extensive perceptual learning, as suggested by Powers *et al.* (2009), or be subject to atypical development, as evidenced in studies by Chen *et al.* (2016), Feng *et al.* (2023a) and Meilleur *et al.* (2020). The temporal proximity between crossmodal (audiovisual) events or the intervals between paired crossmodal (audiovisual or audiotactile) events can be referred to as 'time particles'. When these particles fall within an optimal range, the integration effects, such as spatial and temporal ventriloquism, become perceptible, as demonstrated in works by Chen *et al.* (2010, 2011), Chen and Vroomen (2013), and Shi *et al.* (2010).

Perception and cognition are cyclical processes, reflecting the underlying oscillations known as perceptual cycles, as described by VanRullen (2016). In a review, we encapsulated how perception, timing, and action can be interpreted through the lens of gamma band oscillations, which represent the local activities of brain circuits, coupled to a specific phase of long-range low-frequency oscillations. This coupling proposes a temporal window of integration that optimizes action, as outlined by Gupta and Chen (2016).

Building on this, a recent review by Bauer *et al.* (2020) has synthesized the comprehensive role of neural oscillations in multisensory perception. They consider two mechanisms that facilitate crossmodal influences on sensory processing: crossmodal phase resetting and neural entrainment(Bauer *et al.*, 2020). In a timely paper, Senkowski and Engel (2024) introduced a novel concept of MI that underscores the pivotal role of neural dynamics across multiple timescales within and across brain networks. This concept enables the simultaneous integration, segregation, hierarchical structuring, and selection of information within various time windows.

The oscillation perspective, akin to the 'wave' metaphor, accommodates numerous empirical studies in MI, including neural modulation by Zuo and Wang (2024) and Keil and Senkowski (2018), as well as hierarchical causal inference by Rohe *et al.* (2019). This view is particularly promising and ecological in accounting for sensory integration across the temporal domain. We posit that in the more complex and dynamic scenarios of MI, the modulation of brain oscillations, including entrainment, as ‘waves’, could substantially account for the temporal averaging observed within the integration process.

4. Individual Differences: Synaesthesia, Aphantasia, Dyslexia and Autism

Multisensory integration is not an exception but a rule in the landscape of perceptual and cognitive sciences. It is subject to modulation by cognitive factors such as anxiety levels and empathy, as highlighted in research by Yiltiz and Chen (2015). This phenomenon also exhibits specificity in atypical developmental groups, including those with synaesthesia (Newell and Mitchell, 2016), aphantasia (Dawes *et al.*, 2020), dyslexia (Chen *et al.*, 2016), and autism spectrum disorder (ASD) (Feng *et al.*, 2021, 2023a, b).

The variability in MI behaviors can be linked to the Bayesian causal inference framework, which involves modified priors or likelihoods — two critical components that define an individual’s propensity to integrate sensory signals. One significant reason for the reduced integration capability in atypical development groups is the diminished sensory precision and reliability, as argued by Pellicano and Burr (2012). They suggest that weaker priors might underlie many perceptual symptoms of autism, a viewpoint further discussed by Friston *et al.* (2013) and Van de Cruys *et al.*(2014).

Neurodiagnostic studies indicate that individuals with ASD may exhibit a larger temporal binding window. For instance, children with ASD may not benefit as much from the additional visual information provided by a speaker’s face during a speech-in-noise task, as compared to typically developing (TD) children (Foxe *et al.*, 2018). These multisensory performance differences are associated with a reduced capacity to bind sensory information into a cohesive percept.

Foxe and colleagues’ findings imply that the integration disparity between individuals with and without ASD is most pronounced under low signal-to-noise ratio conditions, where MI is typically most advantageous (Baum *et al.*, 2015). Within the Bayesian predictive coding framework, which has gained popularity, poor predictive coding can be equated with a weak (flat) prior probability distribution, indicating an individual’s uncertainty about the likelihood of an event occurring in the real world (Lawson *et al.*, 2014).

We anticipate that in most atypical developmental conditions, such as synaesthesia, aphantasia, dyslexia, and autism, these individuals may possess atypical priors or likelihoods when compared to their typically developing peers. This perspective offers a nuanced understanding of how cognitive factors and atypical development can shape the intricate dynamics of MI.

5. Validation of MI Principles and Theoretical Contributions

As illustrated above, using traditional visual apparent motion paradigms (Chen *et al.*, 2018; Shen *et al.*, 2019) and neural entrainment behavioral paradigms, extensive behavioral experiments on auditory temporal averaging processing (Zheng and Chen, 2020) and visual Ternus apparent motion were conducted. A 'partial Bayesian integration' model that dynamically changes with trials was constructed, revealing the crossmodal perceptual organization rule of 'geometric averaging' in dealing with the relatively complex audiovisual scenario (Chen *et al.*, 2018). We have validated the multisensory dominance, as well as the inversed effectiveness, with spatial and temporal ventriloquism in a number of psychophysics tasks across different sensory modalities (audiovisual, audiotactile and auditory–olfactory) (Jiang and Chen, 2013; Liang *et al.*, 2022), and we have tested those principles with atypical participants, including individual with synaesthesia, aphantasia, dyslexia, and autism (Feng *et al.*, 2023a, b).

In a recently published magnetoencephalography (MEG) study on 'perception–action' coupling, we employed a 'perception–action' coupling research paradigm. Utilizing auditory temporal interval averaging processing (Zheng and Chen, 2020) and a task involving the reproduction of target durations (Guo *et al.*, 2024), we investigated the capacity for averaging encoding (averaging multiple temporal intervals within sound sequences) and the ability to immediately reproduce target durations within the sensory–motor loop through MEG experiments. These target durations could be half, equal to, or double the average temporal intervals of perceived auditory stimuli. The findings revealed that the central scalp's negative magnetic variation (CMV) and beta oscillations could predict the coupling relationship between perception and action in temporal averaging, uncovering the critical information flow pathway from the supplementary motor area to the superior temporal gyrus for the 'average interval' processing of auditory stimulus sequences and subsequent immediate perceptual decision-making. Preliminary work validates the previously proposed concept of different frequency-coupled brain oscillation patterns, indicating the intricate brain oscillations in MI (Gupta and Chen, 2016).

6. Interoceptive Processing: A New Start and Perspective on Multisensory Studies

Multisensory integration has historically focused on the 'outer' senses, yet interoception — the perception of one's internal physiological state — opens a new frontier. It taps into the realm of 'self-awareness', emphasizing a first-person perspective in sensory research (Gong *et al.*, 2022; Petzschner *et al.*, 2021; Schmitt and Schoen, 2022). This emerging awareness highlights the significance of studying the integration of interoceptive and exteroceptive signals to deepen our comprehension of cognitive processes(Candia-Rivera *et al.*, 2024).

A burgeoning area of research suggests that visual representations of interoceptive signals can heighten body self-consciousness and mitigate pain in chronic pain sufferers. However, the impact of the interplay between exteroceptive and interoceptive inputs on pain perception in individuals without pain conditions remains largely unexplored. To address this gap, we devised a variant of the rubber-hand illusion experiment, where an LED light on the rubber hand synchronized or desynchronized with the participants' heartbeats. This innovative setup enabled us to assess pain thresholds and brain responses (Gong *et al.*, 2022). Our findings revealed that visual interoceptive feedback diminished the P2 component, which is linked to pain processing, and that the rubber-hand illusion reduced alpha-band brain activity prior to stimulus presentation. Notably, body self-consciousness did not significantly modulate pain processing. These insights indicate that visual feedback of interoceptive signals can attenuate pain processing, an effect that appears to be largely independent of an individual's self-consciousness in healthy subjects.

The exploration of interoception has expanded the horizons of multisensory research, typically by incorporating aspects of consciousness. We posit that the traditional spatial and temporal rules, as well as the principle of inverse effectiveness, could be effectively applied to the interoceptive domain. Furthermore, the interplay between interoceptive and exteroceptive priors/likelihoods in a multisensory context could be reconceptualized in an innovative framework. This approach promises to enrich our understanding of how the brain weaves together the internal and external worlds to construct a cohesive perception of self and environment.

7. The Distinction Between Confused Concepts and Beyond

The PhD candidates who are new to my lab often ask me a fundamental question: what distinguishes crossmodal (multisensory) integration from crossmodal (multisensory) interaction? This inquiry delves into the core processes of multisensory research. The distinction between these concepts was a focal point of discussion during the IMRF 2011 in Fukuoka, Japan, particularly in Prof. Charles Spence's keynote address on the nuances of MI terminology (if I recall well). The themes of multisensory segregation and integration have been at the forefront of the field since the IMRF 2016, marking the first time the conference was hosted in China.

It is noteworthy that advanced AI models, such as ChatGPT, are now capable of addressing such complex questions, as suggested by Motoki *et al.*

(2024). With the inception of The China Brain Project (2021–2030) (Poo *et al.*, 2016), and within the research domain of neural mechanisms for MI, our focus has been on intricate aspects of multisensory processing. This includes the dynamics of segregation *versus* integration, binding *versus* oscillation, interaction and adaptation, as well as prediction and learning.

Dr Meredith elucidated that multisensory convergence is the inaugural, essential, and defining phase in the sequence of multisensory processing (as depicted in Figure I.1 of Meredith, 2012). Multisensory convergence allows the activity of one sensory input to influence another on the same neuronal membrane, thus generating multisensory responses. These responses, which manifest as enhancement, depression, or subthreshold modulation, represent the spectrum of activities that occur at the neuronal level during multisensory processing (Stein and Stanford, 2008).

In essence, ‘multisensory interaction’ encompasses a broader range of influences or interactions between different sensory systems, whereas ‘MI’ is a specific instance of multisensory interaction. It is a process where the brain synthesizes sensory information to forge a cohesive perception. Multisensory binding and integration, while related, address distinct facets of sensory processing. Multisensory binding pertains to the perceptual linkage of sensory inputs from various sources, suggesting that they belong together. In contrast, MI pertains to the actual merging of these inputs at multiple neural levels, which can refine or transform perception.

8. Concluding Remarks

‘The senses are a fundamental part of what makes us human’, a sentiment eloquently expressed by Velasco and Obrist (2020). But how do the principles of MI contribute to perceptual processes in the complex tapestry of real-world environments?

At the Sino-German multisensory symposium held in Beijing in September 2023 (Fig. 5), Soto-Faraco presented a thought-provoking review lecture titled ‘Multisensory Research in the Real World and Applications’. Dr Soto-Faraco addressed some studies taking well-known multisensory benefits, crossmodal effects on visual search and audiovisual speech. His research group also tested them under more complex conditions akin to what happens in real-world contexts. He also addressed a hot topic of the role of information conflict during perception, with two special cases of real-world stimuli: advertisement images and cinema. This lecture hence marked a significant shift, moving beyond the traditional boundaries of psychology and psychophysics laboratories. The aim was to explore the constraints on multisensory perception under conditions that, while not fully ecologically valid, are more naturalistic — what has been

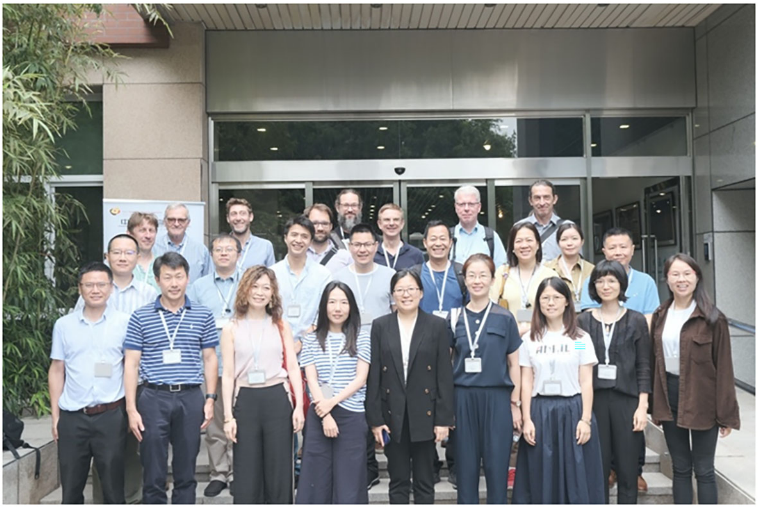

Figure 5. Group photo of presenters who gave talks at the Sino-German symposium on 'Multisensory Processing, Neural Mechanisms and Applications', Beijing, 17–23 September 2023.

termed 'naturalistic laboratory research' (Matusz *et al.*, 2019; Soto-Faraco, 2020; Soto-Faraco *et al.*, 2004; Spence, 2019).

This approach recognizes the need to understand how MI operates in environments that mimic the dynamic and unpredictable nature of the world outside the lab. By doing so, researchers can gain insights into how our senses work in concert to create a seamless and coherent perception of our surroundings, which is crucial for navigation, social interaction, and survival.

The symposium's focus on real-world applications also underscores the practical implications of multisensory research. From enhancing virtual reality experiences to improving assistive technologies for individuals with sensory impairments, understanding MI in more naturalistic contexts can lead to innovations that benefit society.

In essence, the symposium highlighted the importance of studying MI not just as an academic pursuit but as a key to unlocking the full potential of human perception in the rich and varied contexts of everyday life.

Indeed, the realm of multisensory research has expanded to encompass natural and dynamic scenarios, such as the 'cocktail party problem' (Ahmed *et al.*, 2023), where individuals must discern a target voice from a cacophony of competing speech streams. This challenge demands a significant capacity for selective attention and engages sophisticated cognitive processes to filter and focus on relevant auditory information amidst distractions.

The concept of 'spread of attention' in multisensory processing is another facet of attention mediation in environments rich with sensory stimuli, as

Figure 6. A public speech on 'Natural *vs* Artificial Intelligence: Perception, Decision and Action', delivered by Prof. Marc Ernst at Peking University, on 26 September 2018.

observers enrich their multisensory experiences (Mathias and von Kriegstein, 2023). The phenomenon of multimodal enrichment, where learning is enhanced by the presence of complementary information across various sensory or motor modalities, has been shown to significantly improve educational outcomes.

With the rise and widespread adoption of large language models (LLM), such as those discussed by Luo *et al*. (2024) the principles of MI are poised for further exploration and validation. In a public lecture in 2018 with the title "Natural and Artificial Intelligence: Perception, Decision-Making and Action", Prof. Marc Ernst provided a detailed introduction to the integration of multisensory information in human perception and motor behavior (Fig. 6). This included how to utilize prior knowledge for reasoning and understanding the coupling of perception and action within the Bayesian modeling framework. Importantly, examining how artificial intelligence agents integrate multisensory information — such as how they might respond to phenomena like the ventriloquist effect and the McGurk effect — could provide valuable insights into the fundamental principles of MI. These models (including LLMs typically under AI approach) provide a novel platform to simulate and analyze the complex interplay of sensory inputs and the cognitive processes underlying MI, offering new avenues for understanding how attention is directed and modulated across different sensory domains.

Acknowledgements

The research was supported by grants from STI2030-Major Project 2021ZD0202600 and grants from the Natural Science Foundation of China (T2192932, 62061136001, 31861133012).

References

Ahmed, F., Nidiffer, A. R., O'Sullivan, A. E., Zuk, N. J. and Lalor, E. C. (2023). The integration of continuous audio and visual speech in a cocktail-party environment depends on attention, *Neuroimage* **274**, 120143. DOI:10.1016/j.neuroimage.2023.120143.

Bauer, A.-K. R., Debener, S. and Nobre, A. C. (2020). Synchronisation of neural oscillations and cross-modal influences, *Trends Cogn. Sci.* **24**, 481–495. DOI:10.1016/j.tics.2020.03.003.

Baum, S. H., Stevenson, R. A. and Wallace, M. T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder, *Prog. Neurobiol.* **134**, 140–160. DOI:10.1016/j.pneurobio.2015.09.007.

Candia-Rivera, D., Engelen, T., Babo-Rebelo, M. and Salamone, P. C. (2024). Interoception, network physiology and the emergence of bodily self-awareness, *Neurosci. Biobehav. Rev.* **165**, 105864. DOI:10.1016/j.neubiorev.2024.105864.

Cespón, J., Hommel, B., Korsch, M. and Galashan, D. (2020). The neurocognitive underpinnings of the Simon effect: an integrative review of current research, *Cogn. Affect. Behav. Neurosci.* **20**, 1133–1172. DOI:10.3758/s13415-020-00836-y.

Chen, L. and Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review, *Atten. Percept. Psychophys.* **75**, 790–811. DOI:10.3758/s13414-013-0475-4.

Chen, L. and Zhou, X. L. (2011). Capture of intermodal visual/tactile apparent motion by moving and static sounds, *See. Perceiv.* **24**, 369–389. DOI:10.1163/187847511x584434.

Chen, L., Shi, Z. and Müller, H. J. (2010). Influences of intra- and crossmodal grouping on visual and tactile Ternus apparent motion, *Brain Res.* **1354**, 152–162. DOI:10.1016/j.brainres.2010.07.064.

Chen, L., Shi, Z. and Müller, H. J. (2011). Interaction of perceptual grouping and crossmodal temporal capture in tactile apparent-motion, *PLoS ONE* **6**, e17130. DOI:10.1371/journal.pone.0017130.

Chen, L., Wang, Q. and Bao, M. (2014). Spatial references and audio-tactile interaction in crossmodal dynamic capture, *Multisens. Res.* **27**, 55–70. DOI:10.1163/22134808-00002441.

Chen, L., Zhang, M., Ai, F., Xie, W. and Meng, X. (2016). Crossmodal synesthetic congruency improves visual timing in dyslexic children, *Res. Dev. Disabil.* **55**, 14–26. DOI:10.1016/j.ridd.2016.03.010.

Chen, L., Zhou, X. L., Müller, H. J. and Shi, Z. (2018). What you see depends on what you hear: temporal averaging and crossmodal integration, *J. Exp. Psychol. Gen.* **147**, 1851–1864. DOI:10.1037/xge0000487.

Dawes, A. J., Keogh, R., Andrillon, T. and Pearson, J. (2020). A cognitive profile of multisensory imagery, memory and dreaming in aphantasia, *Sci. Rep.* **10**, 10022. DOI:10022/10.1038/s41598-020-65705-7.

Feng, S., Lu, H., Wang, Q., Li, T., Fang, J., Chen, L. and Yi, L. (2021). Face-viewing patterns predict audiovisual speech integration in autistic children, *Autism Res.* **14**, 2592–2602. DOI:10.1002/aur.2598.

Feng, S., Lu, H., Fang, J., Li, X., Yi, L. and Chen, L. (2023b). Audiovisual speech perception and its relation with temporal processing in children with and without autism, *Read. Writ.* **36**, 1419–1440. DOI:10.1007/s11145-021-10200-2.

Feng, S. Y., Wang, Q. D., Hu, Y. X., Lu, H. Y., Li, T. B., Song, C., Fang, J., Chen, L. H. and Yi, L. (2023a). Increasing audiovisual speech integration in autism through enhanced attention to mouth, *Dev. Sci.* **26**, e13348. DOI:10.1111/desc.13348.

Foxe, J. J., Molholm, S., Baudouin, S. J. and Wallace, M. T. (2018). Explorations and perspectives on the neurobiological bases of autism spectrum disorder, *Eur. J. Neurosci.* **47**, 488–496. DOI:10.1111/ejn.13902.

Freeman, E. and Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound, *Curr. Biol.* **18**, 1262–1266. DOI:10.1016/j.cub.2008.07.066.

Friston, K. J., Lawson, R. and Frith, C. D. (2013). On hyperpriors and hypopriors: comments on Pellicano and Burr, *Trends Cogn. Sci.* **17**, 1. DOI:10.1016/j.tics.2012.11.003.

Gong, W., Gu, L., Wang, W. and Chen, L. (2022). Interoception visualization relieves acute pain, *Biol. Psychol.* **169**, 108276. DOI:10.1016/j.biopsycho.2022.108276.

Guo, L., Bao, M., Chen, Z. and Chen, L. (2024). Contingent magnetic variation and beta-band oscillations in sensorimotor temporal decision-making, *Brain Res Bull* **215**, 111021. DOI:10.1016/j.brainresbull.2024.111021.

Gupta, D. S. and Chen, L. (2016). Brain oscillations in perception, timing and action, *Curr. Opin. Behav. Sci.* **8z**, 161–166. DOI:10.1016/j.cobeha.2016.02.021.

Hirst, R. J., McGovern, D. P., Setti, A., Shams, L. and Newell, F. N. (2020). What you see is what you hear: tTwenty years of research using the Sound-Induced Flash Illusion, *Neurosci. Biobehav. Rev.* **118**, 759–774. DOI:10.1016/j.neubiorev.2020.09.006.

Jiang, Y. and Chen, L. (2013). Mutual influences of intermodal visual/tactile apparent motion and auditory motion with uncrossed and crossed arms, *Multisens. Res.* **26**, 19–51. DOI:10.1163/22134808-00002409.

Keil, J. and Senkowski, D. (2018). Neural Oscillations orchestrate multisensory processing, *Neuroscientist* **24**, 609–626. DOI:10.1177/1073858418755352.

Lawson, R. P., Rees, G. and Friston, K. J. (2014). An aberrant precision account of autism, *Front. Hum. Neurosci.* **8**, 302. DOI:10.3389/fnhum.2014.00302.

Liang, K., Wang, W., Lei, X., Zeng, H., Gong, W., Lou, C. and Chen, L. (2022). Odor-induced sound. localization bias under unilateral intranasal trigeminal stimulation, *Chem. Senses* **47**, bjac029. DOI:10.1093/chemse/bjac029.

Luo, X., Rechardt, A., Sun, G., Nejad, K. K., Yáñez, F., Yilmaz, B., Lee, K., Cohen, A. O., Borghesani, V., Pashkov, A., Marinazzo, D., Nicholas, J., Salatiello, A., Sucholutsky, I., Minervini, P., Razavi, S., Rocca, R., Yusifov, E., Okalova, T., Gu, N., Ferianc, M., Khona, M., Patil, K. R., Lee, P.-S., Mata, R., Myers, N. E., Bizley, J. K., Musslick, S., Bilgin, I. P., Niso, G., Ales, J. M., Gaebler, M., Ratan Murty, N. A., Loued-Khenissi, L., Behler, A., Hall, C. M., Dafflon, J., Bao, S. D. and Love, B. C. (2024). Large language models surpass human experts in predicting neuroscience results, *Nat. Hum. Behav.* **9**, 305–315. DOI:10.1038/s41562-024-02046-9.

Mathias, B. and von Kriegstein, K. (2023). Enriched learning: behavior, brain, and computation, *Trends Cogn. Sci.* **27**, 81–97. DOI:10.1016/j.tics.2022.10.007.

Matusz, P. J., Dikker, S., Huth, A. G. and Perrodin, C. (2019). Are we ready for real-world neuroscience? *J Cogn Neurosci* **31**, 327–338. DOI:10.1162/jocn_e_01276.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748. DOI:10.1038/264746a0.

Meilleur, A., Foster, N. E. V., Coll, S.-M., Brambati, S. M. and Hyde, K. L. (2020). Unisensory and multisensory temporal processing in autism and dyslexia: Aa systematic review and meta-analysis, *Neurosci. Biobehav. Rev.* **116**, 44–63. DOI:10.1016/j.neubiorev.2020.06.013.

Meredith, M. A. (2012). Commentary: Multisensory convergence: Where it all begins, in: *The New Handbook of Multisensory Processing*, B. E. Stein (Ed.), pp. 3–11. MIT Press, Cambridge, MA, USA. DOI:10.7551/mitpress/8466.003.0004.

Motoki, K., Spence, C. and Velasco, C. (2024). Colour/shape-taste correspondences across three languages in ChatGPT, *Cognition* **253**, 105936. DOI:10.1016/j.cognition.2024.105936.

Newell, F. N. and Mitchell, K. J. (2016). Multisensory integration and cross-modal learning in synaesthesia: a unifying model, *Neuropsychologia* **88**, 140–150. DOI:10.1016/j.neuropsychologia.2015.07.026.

Pellicano, E. and Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception, *Trends Cogn. Sci.* **16**, 504–510. DOI:10.1016/j.tics.2012.08.009.

Petzschner, F. H., Garfinkel, S. N., Paulus, M. P., Koch, C. and Khalsa, S. S. (2021). Computational Models of Interoception and Body Regulation, *Trends Neurosci* **44**, 63–76. DOI:10.1016/j.tins.2020.09.012.

Poo, M., Du, J., Ip, N. Y., Xiong, Z.-Q., Xu, B. and Tan, T. (2016). China Brain Project: basic neuroscience, brain diseases, and brain-inspired computing, *Neuron* **92**, 591–596. DOI:10.1016/j.neuron.2016.10.050.

Powers III, A. R., Hillock, A. R. and Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding, *J. Neurosci.* **29**, 12265–12274. DOI:10.1523/Jneurosci.3501-09.2009.

Rohe, T., Ehlis, A.-C. and Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception, *Nat. Commun.* **10**, 1907. DOI:10.1038/s41467-019-09664-2.

Schmitt, C. M. and Schoen, S. (2022). Interoception: a multi-sensory foundation of participation in daily life, *Front Neurosci* **16**, 875200. DOI:10.3389/fnins.2022.875200.

Senkowski, D. and Engel, A. K. (2024). Multi-timescale neural dynamics for multisensory integration, *Nat. Rev. Neurosci.* **25**, 625–642. DOI:10.1038/s41583-024-00845-7.

Shams, L., Kamitani, Y. and Shimojo, S. (2000). Illusions – What you see is what you hear, *Nature* **408**, 788. DOI:10.1038/35048669.

Shen, L., Han, B., Chen, L. and Chen, Q. (2019). Perceptual inference employs intrinsic alpha frequency to resolve perceptual ambiguity, *PLoS Biol.* **17**, e3000025. DOI:10.1371/journal.pbio.3000025.

Shi, Z., Chen, L. and Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: the influence of time interval, *Exp. Brain Res.* **203**, 723–735. DOI:10.1007/s00221-010-2286-3.

Soto-Faraco, S. (2020). Reply to C. Spence: Multisensory interactions in the real world, *Multisens. Res.* **33**, 693–699. DOI:10.1163/22134808-bja10005.

Soto-Faraco, S., Spence, C., Lloyd, D. and Kingstone, A. (2004). Moving multisensory research along: Motion perception across sensory modalities, *Curr. Dir. Psychol. Sci.* **13**, 29–32. DOI:10.1111/j.0963-7214.2004.01301008.x.

Spence, C., Soto-Faraco, S., Kvasova, D., Biau, E., Ikumi, N., Ruzzoli, M., Morís-Fernández, L. and Torralba, M. (2019). Multisensory Interactions in the Real World, *Perception* **49**, 240–242. DOI:10.1177/0301006619896976.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press, Cambridge, MA, USA.

Stein, B. E. and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron, *Nat. Rev. Neurosci.* **9**, 255–266. DOI:10.1038/nrn2331.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L. and Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism, *Psychol. Rev.* **121**, 649–675. DOI:10.1037/a0037665.

VanRullen, R. (2016). Perceptual cycles, *Trends Cogn. Sci.* **20**(10), 723–735. DOI:10.1016/j.tics.2016.07.006.

Velasco, C. and Obrist, M. (2020). *Multisensory Experiences: Where the Senses Meet Technology*. Oxford University Press, Oxford, United Kingdom. DOI:10.1093/oso/9780198849629.001.0001.

Yiltiz, H. and Chen, L. H. (2015). Tactile input and empathy modulate the perception of ambiguous biological motion, *Front. Psychol.* **6**, 161. DOI:10.3389/fpsyg.2015.00161.

Zheng, W. and Chen, L. (2018). The roles of attentional shifts and attentional reengagement in resolving the spatial compatibility effect in tactile Simon-like tasks, *Sci. Rep.* **8**, 8760. DOI:10.1038/s41598-018-27114-9.

Zheng, W. and Chen, L. (2020). Illusory perception of auditory filled duration is task- and context-dependent, *Br. J. Psychol.* **111**, 103–125. DOI:10.1111/bjop.12379.

Zuo, Y. and Wang, Z. (2024). Neural oscillations and multisensory processing, in: *Advances of Multisensory Integration in the Brain, Advances in Experimental Medicine and Biology, vol. 1437*, Y. Gu and A. Zaidel (Eds), pp. 121–137. Springer, Singapore. DOI:10.1007/978-981-99-7611-9_8.