



Multisensory perception and action: development, decision-making, and neural mechanisms

Zhuanghua Shi^{1*} and Hermann J. Müller^{1,2}

¹ Department of Psychology, Experimental Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

² Department of Psychological Science, Birkbeck College, University of London, London, UK

*Correspondence: shi@psy.lmu.de

Edited by:

Sidney A. Simon, Duke University, USA

Keywords: multisensory perception, multisensory timing, multisensory development, multisensory learning, multisensory neural mechanisms

Surrounded by multiple objects and events, receiving multisensory stimulation, our brain must sort through relevant and irrelevant multimodal signals to correctly decode and represent the information from the same and different objects and, respectively, events in the physical world. Over the last two decades, scientific interest has increased dramatically in how we integrate multisensory information and how we interact with a multisensory world, as evidenced by exponential growth of the relevant studies using behavioral and/or neuro-scientific approaches.

The Special Issue topic of “Multisensory perception and action: psychophysics, neural mechanisms, and applications” emerged from a scientific meeting dedicated to these issues: the *Munich Multisensory Perception Symposium* held in Holzhausen am Ammersee, Germany (June 24–26, 2011). This volume, which collects research articles contributed by attendees of the symposium as well as the wider community, is organized into three interrelated sections:

- (I) Development, learning, and decision making in multisensory perception
- (II) Multisensory timing and sensorimotor temporal integration
- (III) Electrophysiological and neuro-imaging analyses of multisensory perception

DEVELOPMENT, LEARNING, AND DECISION-MAKING IN MULTISENSORY PERCEPTION

Many multisensory studies, ranging from spatial (e.g., Ernst and Banks, 2002; Alais and Burr, 2004) to temporal integration (e.g., Burr et al., 2009; Chen et al., 2010; Shi et al., 2013b), reveal that our brain combines multisensory signals if they are closely relevant to the task, in order to boost overall performance. Senses, however, are not the only source for decision-making. Prior, contextual, and symbolic cues can also contribute as an extra source of information to improve performance (Jazayeri and Shadlen, 2010; Petzschner and Glasauer, 2011; for a review, see Shi et al., 2013a). Accordingly, Petzschner et al. (2012) set out to examine how auxiliary contextual cues, such as symbolic “short” and “long” cues, are used optimally in a distance production-reproduction task. Their findings indicate that humans are capable of using symbolic cues for final estimates, even though the mapping of the symbolic cue onto the stimulus dimension has to be learned during the experiment.

With respect to learning, one prominent question in multisensory integration concerns when and how we acquire the capacity to optimally integrate multisensory cues. Some recent studies suggest that this capacity is not present at birth, but rather develops after about 8 years of age (e.g., Gori et al., 2008). Gori et al. (2012) expanded this line of research by examining audio-visual temporal and spatial bisection tasks in young children, finding that young children exhibit strong unisensory dominance over multisensory integration of audiovisual signals, with audition dominating audiovisual time perception and vision dominating space perception. Both dominance effects reflect a process of cross-sensory calibration of developing systems, where the more accurate sense calibrates or teaches the other, rather than fusing with it. In another study, Wismeijer et al. (2012) showed that our brain also exhibits remarkable ability to learn cue-associations, such as an arbitrary association of visual gloss and touch softness, and use the learned associate-cues for judgments—with learning being more efficient from touch-to-vision than from vision-to-touch, which is in line with earlier evidence of touch teaching vision for size discrimination in young children (Gori et al., 2008).

Multisensory signals, compared to separate unisensory signals, not only enhance overall performance, but also facilitate the speed of responses. Based on their previously developed framework of the time-window-of-integration (TWIN), Colonius and Diederich (2012) provided further qualitative and quantitative predictions of the TWIN model regarding how the probability of multisensory integration would affect response facilitation differently in the crossmodal-signals and the focused-attention paradigm. In the reverse direction Hong et al. (2012) examined response impairments arising from conflicting crossmodal stimuli or configurations that engender multisensory illusions, in particular, the hand-reversal illusion.

MULTISENSORY TIMING AND SENSORIMOTOR TEMPORAL INTEGRATION

Time perception is susceptible to a wide range of factors (Shi et al., 2013a), in particular with multisensory inputs. A number of authors examining this set of issues have attempted to pin down key factors in multisensory timing. With regard to the perception of multisensory durations, Shi et al. (2012) showed that high-arousal affective pictures have differential impacts on subsequent tactile duration judgments, with pictures that evoke threat meanings expanding subjective duration, whereas pictures that evoke disgust meanings exhibiting no effects on tactile temporal

judgments—indicative of the importance of crossmodal connections in the processing of multisensory timing. Ganzenmüller et al. (2012) further demonstrated that delayed onset of auditory signals generated by participants' manual button press immediately lengthens the reproduced duration, whereas offset delays did not—showing that multisensory timing relies differentially on sensory and motor signals in duration reproduction. Using apparent motion as an implicit measure of perceived duration, Zhang et al. (2012) reported another differential adaptation effect in multisensory timing: adaptation to a short auditory or visual interval resulting in a consistent negative aftereffect for Ternus apparent motion, whereas adaptation to a long interval yielded an aftereffect only for the auditory, and not the visual, condition.

Similar to multisensory duration, multisensory temporal-order processing is also influenced by many factors. For example, to identify key physical changes associated with the articulation of consonants and vowels that may influence the temporal integration window for audiovisual speech, Vatakis et al. (2012) examined the perception of audiovisual synchrony using video clips uttered by different speakers with differential audiovisual signal saliencies (with auditory saliency measured by a combination of three acoustic features: instantaneous energy of the most active filter, instantaneous amplitude, and frequency of the dominant filter's output; and visual saliency computed by intensity, color, and motion). They found that the (degree of) saliency of visual-speech signals can modulate the lead of visual over auditory signals that is necessary for them to be perceived as simultaneous, the lead typically found in audiovisual speech perception. These findings thus support the “information reliability hypothesis,” on which the perception of a multisensory feature is dominated by the modality that provides the most reliable information (Welch and Warren, 1980; Ernst et al., 2004). Similarly, Hendrich et al. (2012) found that not only stimuli features, but also task requirements, such as dual tasks, could affect audiovisual temporal-order judgments, arguing that the influence of dual tasks on crossmodal temporal processing is mainly on the perceptual, rather than the response-selection, stage.

ELECTROPHYSIOLOGICAL AND NEURO-IMAGING ANALYSES OF MULTISENSORY PERCEPTION

The neural mechanisms underlying integrative and interactive functions are central to understanding multisensory perception. Quite a number of studies concerned with these functions have been designed to elucidate how information that comes from different sensory modalities are processed and integrated in the brain.

Several studies provide found evidence that multisensory signals are integrated at a very early stage. Naci et al. (2012), for example, found that higher-order regions in anterior temporal (AT) and inferior prefrontal cortex (IPC) performed audiovisual integration 100 ms earlier than a sensory-driven region in the posterior occipital (pO) cortex, suggesting the brain represents familiar and complex multisensory objects through early interactivity between higher-order, and sensory-driven regions. Stekelenburg and Vroomen (2012) also showed that spatial congruity between auditory and visual signals modulates audiovisual interactions reflected in early ERP components, namely,

the N1 and P2. Early integration may boost the saliency of the multisensory signals, even when the multisensory signals are irrelevant distractors, causing an attentional shift toward the multisensory distractor, as measured by steady-state visual evoked potentials (SSVEP) in an audiovisual speech task (Krause et al., 2012). Instead of using multisensory signals, Töllner et al. (2012) presented separate auditory and visual signals in a dual-task paradigm requiring both auditory and visual discriminations, to investigate influences of task order predictability (TOP) and inter-task onset asynchrony (SOA) on perceptual, and motor processing stages, two stages indexed, respectively, by two EEG components: the Posterior-Contralateral-Negativity (PCN) and the Lateralized-Readiness-Potential (LRP). Töllner et al. found TOP to interact with inter-task SOA in determining the speed of perceptual processing, providing electrophysiological evidence of central capacity limitations in the processing of auditory and visual dual tasks.

Using functional MRI imaging techniques, two other studies examined brain regions involved in multisensory perception. Noesselt et al. (2012) investigated the neural basis of the perception of synchrony/asynchrony for audiovisual speech stimuli, and found a distinct pattern of modulations within the multisensory superior temporal sulcus complex (mSTS-c): “auditory leading (AL)” and “visual leading (VL) areas” lie closer to “synchrony areas” than to each other, suggesting the presence of distinct sub-regions within the human STS-c for the maintenance of temporal relations for audiovisual speech stimuli, with differential functional connectivity with prefrontal regions. Beer et al. (2013), on the other hand, found bimodal presentation of audiovisual speech and audiovisual movement stimuli, compared to unimodal stimulation, engaged a temporal-occipital brain network including the multisensory superior temporal sulcus (msSTS), the lateral superior temporal gyrus (ISTG), and the extrastriate body area (EBA). Moreover, brain areas involved in multisensory processing showed little direct connectivity with primary sensory cortices; rather these brain areas were connected to early sensory cortices via intermediate nodes of the STS and the inferior occipital cortex (IOC).

Taken together, this collection provides a broad-spectrum but overall coherent addition to the rapidly growing field of multisensory perception and action. Of course, more work needs to be carried out and many open questions and issues (some of which are identified in the present collection) remain to be addressed in order to achieve a full understanding the functions and neural mechanisms of multisensory perception and action. We would like to thank all the authors, the expert reviewers, and the Frontiers staff for helping to make this Special Issue possible. We hope this collection can act as a catalyst for some of the future work, and we look forward to further explorations of multisensory perception and action.

REFERENCES

- Alais, D., and Burr, D. C. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. doi: 10.1016/j.cub.2004.01.029
- Beer, A. L., Plank, T., Meyer, G., and Greenlee, M. W. (2013). Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Front. Integr. Neurosci.* 7:5. doi: 10.3389/fnint.2013.00005

- Burr, D. C., Banks, M. S., and Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Exp. Brain Res.* 198, 49–57. doi: 10.1007/s00221-009-1933-z
- 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
- Colonius, H., and Diederich, A. (2012). Focused attention vs. crossmodal signals paradigm: deriving predictions from the time-window-of-integration model. *Front. Integr. Neurosci.* 6:62. doi: 10.3389/fnint.2012.00062
- Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a
- Ernst, M. O., Bühlhoff, H. H., and Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169. doi: 10.1016/j.tics.2004.02.002
- Ganzenmüller, S., Shi, Z., and Müller, H. J. (2012). Duration reproduction with sensory feedback delay: differential involvement of perception and action time. *Front. Integr. Neurosci.* 6:95. doi: 10.3389/fnint.2012.00095
- Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr. Biol.* 18, 694–698. doi: 10.1016/j.cub.2008.04.036
- Gori, M., Sandini, G., and Burr, D. (2012). Development of visuo-auditory integration in space and time. *Front. Integr. Neurosci.* 6:77. doi: 10.3389/fnint.2012.00077
- Hendrich, E., Strobach, T., Buss, M., Müller, H. J., and Schubert, T. (2012). Temporal-order judgment of visual and auditory stimuli: modulations in situations with and without stimulus discrimination. *Front. Integr. Neurosci.* 6:63. doi: 10.3389/fnint.2012.00063
- Hong, S. W., Xu, L., Kang, M.-S., and Tong, F. (2012). The hand-reversal illusion revisited. *Front. Integr. Neurosci.* 6:83. doi: 10.3389/fnint.2012.00083
- Jazayeri, M., and Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nat. Neurosci.* 13, 1020–1026. doi: 10.1038/nn.2590
- Krause, H., Schneider, T. R., Engel, A. K., and Senkowski, D. (2012). Capture of visual attention interferes with multisensory speech processing. *Front. Integr. Neurosci.* 6:67. doi: 10.3389/fnint.2012.00067
- Naci, L., Taylor, K. I., Cusack, R., and Tyler, L. K. (2012). Are the senses enough for sense. Early high-level feedback shapes our comprehension of multisensory objects. *Front. Integr. Neurosci.* 6:82. doi: 10.3389/fnint.2012.00082
- Noesselt, T., Bergmann, D., Heinze, H.-J., Münte, T., and Spence, C. (2012). Coding of multisensory temporal patterns in human superior temporal sulcus. *Front. Integr. Neurosci.* 6:64. doi: 10.3389/fnint.2012.00064
- Petzschner, F. H., and Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *J. Neurosci.* 31, 17220–17229. doi: 10.1523/JNEUROSCI.2028-11.2011
- Petzschner, F. H., Maier, P., and Glasauer, S. (2012). Combining symbolic cues with sensory input and prior experience in an iterative bayesian framework. *Front. Integr. Neurosci.* 6:58. doi: 10.3389/fnint.2012.00058
- Shi, Z., Church, R. M., and Meck, W. H. (2013a). Bayesian optimization of time perception. *Trends Cogn. Sci.* 17, 556–564. doi: 10.1016/j.tics.2013.09.009
- Shi, Z., Ganzenmüller, S., and Müller, H. J. (2013b). Reducing bias in auditory duration reproduction by integrating the reproduced signal. *PLoS ONE* 8:e62065. doi: 10.1371/journal.pone.0062065
- Shi, Z., Jia, L., and Müller, H. J. (2012). Modulation of tactile duration judgments by emotional pictures. *Front. Integr. Neurosci.* 6:24. doi: 10.3389/fnint.2012.00024
- Stekelenburg, J. J., and Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Front. Integr. Neurosci.* 6:26. doi: 10.3389/fnint.2012.00026
- Töllner, T., Strobach, T., Schubert, T., and Müller, H. J. (2012). The effect of task order predictability in audio-visual dual task performance: just a central capacity limitation. *Front. Integr. Neurosci.* 6:75. doi: 10.3389/fnint.2012.00075
- Vatakis, A., Maragos, P., Rodomagoulakis, I., and Spence, C. (2012). Assessing the effect of physical differences in the articulation of consonants and vowels on audiovisual temporal perception. *Front. Integr. Neurosci.* 6:71. doi: 10.3389/fnint.2012.00071
- Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 88, 638–667. doi: 10.1037/0033-2909.88.3.638
- Wismeijer, D. A., Gegenfurtner, K. R., and Drewing, K. (2012). Learning from vision-to-touch is different than learning from touch-to-vision. *Front. Integr. Neurosci.* 6:105. doi: 10.3389/fnint.2012.00105
- Zhang, H., Chen, L., and Zhou, X. (2012). Adaptation to visual or auditory time intervals modulates the perception of visual apparent motion. *Front. Integr. Neurosci.* 6:100. doi: 10.3389/fnint.2012.00100

Received: 28 October 2013; accepted: 04 November 2013; published online: 21 November 2013.

Citation: Shi Z and Müller HJ (2013) Multisensory perception and action: development, decision-making, and neural mechanisms. *Front. Integr. Neurosci.* 7:81. doi: 10.3389/fnint.2013.00081

This article was submitted to the journal *Frontiers in Integrative Neuroscience*.

Copyright © 2013 Shi and Müller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.