Multisensory Information Processing for Enhanced Human-Machine Symbiosis

Frederick D. Gregory(✉) and Liyi Dai

U.S. Army Research Office, Durham, NC, USA
{frederick.d.gregory5.civ,liyi.dai.civ}@mail.mil

Abstract. Multisensory information processing is a basic feature of neural systems and has been exploited to facilitate development of Army systems that augment Soldier performance through multisensory displays. However, the full potential of these systems has yet to be determined and will require understanding fundamental features of the underlying neurophysiology of multisensory processing, the neuroergonomics of multisensory machine interface and analytical methods for neural signal analysis, dimensionality reduction and pattern recognition. Here, findings from basic and applied research efforts will be presented that have focused on various aspects of human (brain)-computer interfaces to uncover understanding in these areas and mediate recent technological developments in multisensory display technology, passive mental state detection, attention/orientation detection, and human activities recognition from video in general. Based on the knowledge of multisensory processes acquired from these efforts there are emerging opportunities for creating new human gesture-controlled recognition systems based upon multimodal data analysis which will allow for unprecedented human-machine symbiosis.

Keywords: Human-machine interfaces · Brain-computer interface · Data analysis · Human activity monitoring · Multisensory cueing

1 Introduction

The modern Army is quickly transforming into a highly networked force with integrated platforms that will enable vast amounts of on-demand multimodal data. Individual soldiers will be responsible for unprecedented information management duties while ensuring personal and team situational awareness, decision-making and overall mission effectiveness. Strategies that mitigate the impact of information overload on the soldier are vital and must inform future system designs. Head mounted displays for the dismounted soldier [1], unmanned autonomous aerial and ground sensors [2–5] and communication platforms [6] could all simultaneously push information to the soldier through smaller and lighter displays. Therefore, strategies for ideal presentation of information to a user must continue to be an area of active research. Symbiosis of the soldier with machines is envisioned as a mutually-interdependent, tightly-coupled relationship that maximally exploits human and machine strengths in a seamless interface. Research communities have shown growing interest in this symbiosis due in part to recent progress in modern computing capabilities combined with the availability

© Springer International Publishing Switzerland 2015
S. Yamamoto (Ed.): HIMI 2015, Part I, LNCS 9172, pp. 354–365, 2015.
DOI: 10.1007/978-3-319-20612-7_34

of ubiquitous sensing modalities for capturing information about the human user in non-laboratory conditions [7, 8]. In this paper, we highlight a few technologies that have potential to be important components of human-machine interfaces and present new scientific opportunities.

2 Multisensory Information Processing

Our brains generate a unified percept of the world through partially redundant sensory information about an object or event. We watch movies and derive enjoyment even though we are aware that the sounds from people and objects on-screen originate from television or movie theater speakers. We readily perceive that voices in the movie are coming from the actor's lips. This sensory illusion, the ventriloquist effect, is a result of our innate ability to integrate auditory and visual information which results in the perceptual alteration of speech sound location [9]. The McGurk effect [10], another audio-visual illusion, occurs when lip movements alter the phoneme that is perceived. Sensory illusions are important tools for elucidating the neural processes underlying multisensory integration. Behavioral studies have suggested that when two sensory cues are separated by even 200 ms, the advantage of multisensory integration and perceptual consequences of ventriloquism are greatly reduced [11]. However, multisensory cells such as those recorded in the superior colliculus [12] and cortex [13] still show integrative responses to sensory cue separation of 600 ms and longer. The relationship between the temporal dynamics of single unit responses in the brain to behavior must be linked with multisensory neural network activity to inform multisensory information presentation and display technology.

2.1 Multisensory Displays

Dynamic and highly adverse operational environments often present scenarios where sensory information is degraded or obstructed. Multisensory cueing has been demonstrated as an effective strategy for orienting attention under non-ideal conditions [14, 15]. Multisensory cueing has also shown to be an effective strategy for offsetting performance decrements due to stress [16]. Delivery of temporally congruent information is being actively explored for multisensory displays with combined audio-visual and other multisensory interactions for augmenting human performance [17–19]. While some studies have reported less effective impacts of combined sensory cues for specific tasks [20], the emerging and unified view is that cueing underused sensory streams provides an overall performance advantage [20, 21]. Human multisensory integration is suggested to rely upon correlations between converging sensory signals that result in statistically optimal input to the nervous system and behavioral outputs [22]. However, the manner in which congruent multisensory information impacts a user's nervous system in real world situations has yet to be fully exploited. Emerging applications for navigation, covert communication and robotic control will benefit by further understanding how the underlying neurophysiological mechanisms of multisensory processing relate to the statistics of behavior.

2.2 Multisensory Information Processing in the Brain

The integration of information from multiple senses was originally thought to occur in high level processing areas in the frontal, temporal or parietal lobes [23–25]. More recent anatomical, neurophysiological and neuroimaging studies in non-human primates and functional brain studies in humans lead to the emerging view that multisensory processing involves a diversity of cortical and sub-cortical neural networks [26–28]. Based on behavioral studies with multisensory cueing, the neural coding strategy within multisensory integrative neural networks must be biased by the extent of spatial and temporal congruency of incoming sensory information [29]. Preliminary findings suggest that converging synaptic signaling by pre-cortical sensory integrating neurons of the thalamus show augmented output to the primary auditory cortex [30]. Both excitatory and inhibitory signals are strengthened by these congruent sensory inputs and highlight the diversity of computational modifications occurring within multisensory integrating networks [31]. Decoding multisensory neural network activities could potentially serve as feedback commands for closing the human-machine interaction loop.

The underlying neural codes of multisensory processes must be considered within the context of mathematical and theoretical models in order to best define pathways for improving multisensory interfaces. Feed-forward convergence of information from simultaneous senses (sensory organ to cortex) is accompanied by feed-back input from unisensory processing cortical areas onto lower-level multisensory integrating sites [32]. This view of multisensory processing builds upon the modality appropriateness hypothesis which offers the proposal that the greater acuity sensory modality for a particular discrimination task, ultimately dominates perception in a winner-take-all competition [33]. A similar, and complementary, view is that multisensory integration obeys Bayesian probability statistics [34, 35] and most closely resembles the properties of a maximum likelihood integrator [22, 36, 37]. An alternative view is that multisensory enhancement of information processing is a result of temporal or spectral multiplexing, where, for example, spike timing information from single neurons and activity from network oscillations interact in time and lead to an enhanced multiplexed code [38]. The complexity of multisensory integration-induced modifications of the neural code require improved signal processing approaches for decoding multiscale neural activity combined with appropriate theoretic frameworks and mathematical modeling to fully realize the potential of multisensory information processing for informing advanced display technologies.

3 Complementary Approaches

Three areas of active research are utilizing methods and creating technologies that can support multisensory information display technologies.

- Brain State Awareness
- Human Activity Monitoring
- Direct Brain-Computer Interfaces

Together, these areas lay foundations for next-generation systems that exploit principles of human cognition to mediate ergonomically enhanced human-system interfaces that maximally augment performance. Here we review example technologies by superficially highlighting potential opportunities.

3.1 Brain State Awareness

Performing tasks under complex, dynamic, and time-pressured conditions is troublesome for maintenance of operational tempo. Mental workload is a topic of increasing importance to human factors and significant effort has been devoted to developing innovative approaches to objectively assessing cognitive load in real-time. Stress is another topic of significant importance for the deleterious impact on performance of the user of any display technology [14]. Strategies are sought to offer fatigue offsetting interventions like selecting the best information content and format for presentation of information to the human operator. Data relevant for mental state detection include facial features, involuntary gestures, tactile signals, brain neural signals, and physiological signals (e.g., speech, heart rate, respiration rate, skin temperature, and perspiration). Mental states such as anxiety or fatigue often lead to temporal changes in biophysiological signals that might be classified by machine learning algorithms. For example, anxiety may result in increased rate of heartbeat and increased blood pressure relative to physiological signals of an individual's "normal" mental state. A major challenge is the lack of precise quantitative metrics that define mental states and the difficulty for cross-subject validation.

Stress, Anxiety, Uncertainty and Fatigue (SAUF). Recent attempts have been made to detect stress, anxiety, uncertainty and fatigue from visual and infrared images of a human face [39, 40]. An infrared image, either long wave or mid wave IR, captures the thermal signatures of the skin. Mental states, like stress or anxiety, generate subtle changes in local blood flow beneath the skin, reflected as changes in skin temperature. Thermal imagery is rather sensitive to such physiological changes although the changes may be invisible to the naked eye in certain groups of individuals [39, 41]. Non-invasive detection methods are highly desirable and offer a simple and affordable computer interface solution. State detection from imaging modalities allow for a passive means of detection without interfering with the operator's normal activities or requiring operator cooperation, which could be amenable to real-world applications.

For visual/thermal video based SAUF detection, the first step is to determine facial landmarks such as mouth corners, eye inner and outer corners, nasal tip, eyebrow start and end points. These landmark points are algorithmically tracked so that spatial and temporal information, called features, can be extracted from both the visual and thermal videos and subsequently used in pattern classification. Features include eye and/or mouth movement and physiological features such as the temperatures of these facial points. The data size is typically huge: Frame rates for visual and thermal videos can be 30 fps or higher. Recording from hours of thermal and visible videos are needed for algorithm training. In [40], the authors described the development of a computer system for SAUF detection using both visual and thermal videos in real-time. The

system achieved detection errors in the range of 3.84 %–8.45 % for anxiety detection. In addition to algorithmic accuracy, errors may also be due to view changes (resulting in face deformation), full or partial occlusion, or individual variation. While this approach may not serve as a single source solution, non-invasive imaging provides an alternative and complementary approach for brain state detection that can accompany brain signal-based detection of mental states [42].

3.2 Human Activity Analysis and Prediction

The objective of human activity analysis and prediction is to understand the physical behavior of a human operator. Near term activity analysis focuses on understanding what the operator is doing and predicting the operator's intention for imminent action. Long term activity analysis aims at recognizing an operator's habits and personality such as right-handed person or left-handed person, or patterns of keyboard strokes for identity confirmation. Intention recognition and high level activity recognition are active research areas in artificial intelligence [43–49]. The methods for visual data analysis are general and applicable to a wide range of applications including human-machine interfaces as well as surveillance across a wide-span geographical region.

Visual data contains rich information for activity analysis and understanding. An adult can recognize activities from an image or a video segment with little effort. However, visual activity analysis and understanding by a computer has proven extremely difficult. The key challenges are that spatiotemporal features in imagery or video are typically high dimensional, noisy, ambiguous, and lie on (unknown) non-linear manifolds. There is a lack of robust methods for detecting the underlying patterns. Human activities occur in a wide variety of contexts and at wide range of scales. In many cases, contextual information is essential for understanding human activities but often unavailable. Conceptually, vision based human activity analysis and understanding consists of several components: action representation, action recognition, activity recognition and prediction although the boundary between action and activity may not be analytically definable.

Action Representation. Activity analysis and understanding is typically carried out in a general hierarchical framework. The low-level, atomic components are "actions" or "actionlets", i.e. primitive motion patterns typically lasting for a short duration of time, such as turning of the head or lifting the left arm. An activity is a temporal, typically complex composition of multiple actions. For example, "making a phone call" can be decomposed into four actions. At the low signal level, actions are characterized by spatiotemporal features and potentially distinguishable through pattern classification of the features. A component based hierarchical model was proposed to account for articulation and deformation of the human body due to factors such as view change or partial occlusion [48, 50, 51].

Action Recognition. Motion is a critical attribute for action recognition and spatio-temporal features can be extracted from multiple sequential frames in a video. Examples of spatial features include Scale-Invariant Feature Transform points,

Histograms of Oriented Gradients and Histograms of Optical Flow. Algorithms are used for frame registration and landmark or object tracking in video to extract temporal motion information. The spatial features and temporal information are combined to feed into pattern analysis and classification algorithms for action recognition.

Activity Recognition and Prediction. Activity recognition typically requires behavior modeling and high level reasoning, which is essential for activity or near real-time intention prediction. Parametric models like Hidden Markov Model or Petri Nets and non-parametric models such as Bayesian methods for inference require the incorporation of prior knowledge learned from past data or to be manually coded. Such frameworks are flexible to allow the incorporation of novel action dependencies for human activities. A general framework for human activity analysis and prediction has been developed [49, 52, 53] and supplemented by a hierarchical framework that can automatically detect contextual information and incorporate it in activity understanding [54].

3.3 Direct Brain-Computer Interface

Machines and humans, unfortunately, do not have an inherent common language for engaging in the human-computer interaction loop. In order for the human in the loop to derive maximal benefit from the interface the computational framework on the other end must be able to accurately determine user intent in real-world settings. This includes when the user is under duress and is placed into a dynamic physiological and/or neural state. Software specifications like those used in Controlled Natural Languages may provide a possible solution [55, 56]. However, these methods have mainly been tested for simple interfaces. Complex operational environments will require other complementary solutions.

Brain-Computer Interface Methods. Brain-computer interfaces permit direct communication of user intent to machine interfaces. The general framework for open-loop brain-computer interface system control originates from the detection of brain activity related to user intent. Electroencephalography (EEG), electrocorticography (ECoG) and intracortical (single unit) recording configurations are some of the technologies currently in use for brain-computer interfaces. Other sensing modalities include magnetoencephalography (MEG), Positron Emission Tomography (PET), functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS). These imaging modalities together are complementary in information attributes, spatial-temporal resolution and degree of invasiveness. For example, EEG provides high temporal but low spatial resolution while fMRI provides low temporal but high spatial resolution. ECoG is a semi-invasive technique and intracortical recordings are invasive. Following analog to digital conversion, advanced signal processing and machine learning algorithms can then be deployed to classify neural activity information and derive user intent or state.

Detection of Silent Speech. A recent effort attempted to develop a brain-based communication and orientation system using EEG and ECoG signals [57, 58]. The objective was to create signal processing methods that allow detection of imagined

speech for communication and determining directional attention for orientation from brain signals. One key challenge was a lack of understanding how imagined speech related to overt speech brain function. In order to be successful this study also had to overcome the limited understanding about the interaction among networked neurons in speech processing pathways, the difficulty of determining a baseline for imagined speech and the existence of noise in the neural recording. Based upon the existing real-time software system BCI2000, algorithms were generated that are capable of extracting electrophysiological features on a single-trial basis. Based on chance accuracy of 25 %, ECoG-based decoding showed overall ~40 % performance levels for detection of vowels and consonants during both overt and covert speech [57, 58]. The results indicate higher than chance likelihood of correctly decoding imagined consonants and vowels.

For detecting attention and orientation, the setup is similar to that for imagined speech detection. Each subject was presented with visual cues and stimuli on a computer screen with built-in eye tracker, which verified ocular fixation on the central cross during data acquisition. The system achieved average detection accuracy of 84.5 % for attention engagement and 48.0 % for attention locus [59, 60] from ECoG data. While this line of work has only been able to achieve recognition of phonemes, a multisensory information processing approach may be taken to improve algorithm performance. Communication inherently involves multisensory processes which may be exploited to elucidate a new regime of neural network activity that might drive classification schemes of future brain-computer interfaces. Exploration of this idea may offer an opportunity to advance research in fundamental mechanisms of the neural processing of speech and close the loop in brain-computer interface design to facilitate performance for applications like covert communication and device control.

4 Vision for Future Multisensory Information Displays

Advances in functional neuroimaging combined with signal processing capabilities have led to new opportunities to identify spatial and temporal features of neural processing during real world experimentation [7, 8]. Research on human-machine interfaces has also considered methods for combining physiological data (e.g., respiration rate, heart rate, blood pressure and temperature) and behavioral information (e.g., posture, eye movements, gesture, and visual/thermal facial expression). The larger neural real estate devoted to multisensory processes and the diversity of signaling mechanisms available open new opportunities for human machine interfaces. Signal processing and data analytic advances can be devoted to decoding information related to this complex signaling and modification as a result of presentation of sensory information through multisensory displays. Brain-computer interface research has largely focused on the presentation of information to one of a user's senses while decoding brain activity with open-loop pattern classification, i.e. using electroencephalography while watching a visual display. The research has demonstrated utility in direct brain-computer communications for simple choices like user control of a cursor on a screen but state-of-the-art pattern classification algorithms only show limited performance for complex tasks such as decoding intended speech. Recent

advances point to an emerging opportunity for a paradigm shift. To understand how simultaneous information presentation modifies behavioral response we need to determine where and how information from different senses is combined in the brain and what are the neural computational advantages rendered by these processes.

4.1 A Lesson from Sensory Deprivation

Sensory deprivation can lead to improvements in perceptual abilities in the intact senses for the blind or deaf. For example, individuals with early onset blindness show improved temporal and spectral frequency discrimination when compared to those with late-onset blindness or those who are sighted [61]. The early-blind have also been demonstrated to show enhanced sound localization ability relative to sighted individuals [62]. Surprisingly, in the study of Lessard et al., a group of blind subjects that had maintained some level of residual peripheral vision showed degraded sound localization ability relative to the completely blind. These observations together highlight the complicated mechanisms mediating multisensory processing when information is missing or corrupted in one sensory stream. This may be relevant to situations when only degraded sensory information is available in a high attentional load operational environment to a person with full sensory capabilities. A more recent study showed that by depriving normal sighted mice of light for as little as two days was enough to elicit potentiation of specific pre-cortical inputs from the thalamus into the auditory [30] or somatosensory cortices [63]. More work is needed in this area but the underlying neurophysiological mechanisms that mediate responses to sensory deprivation, not from disease or injury, may be relevant and provide inspiration for novel neuroplasticity-based approaches to advanced human-machine interfaces capabilities and augmented cognition.

5 Conclusion

The state-of-the-art view of multisensory displays has shown advantages of multisensory stimulation and has highlighted the need to understand the underlying neural bases mediating cueing-induced behavioral improvements. New approaches leading to higher resolution multimodal data as a result of developments in sensor technologies are an enabling tool but pose significant computational challenges. However, statistical modeling approaches and advancing computational analysis capabilities are providing new methodologies to facilitate the availability of neural information for direct human-computer interaction. There is a fundamental need to study human cognitive behavior under real-world conditions and multisensory information displays offer a unique capability to engage humans while they perform outside the laboratory.

State-of-the-art advances have not completely approached the vision of closed-loop human-machine symbiosis, but have paved the way for more sophisticated theories and technologies that will enable the attainment of this vision. Here we have described example technologies that provide emerging opportunities to exploit advances in understanding the underlying principles governing neural processing of information

from simultaneous sensory streams to create systems that interface with the human in intuitive and, potentially, seamless ways. Multisensory displays show great potential to support future soldier-machine technologies and future designs should be created based on principles grounded in data and theory from basic cognitive neuroscience and neurophysiology. The future military operational environment will be more complex and require more from the human operator as she interacts with soldier systems. In order to take full advantage of scientific opportunities presented by multisensory information processing, a deep understanding of how the human brain, body, and sensory systems work in concert to accomplish tasks is required in order to close the loop in human-systems interactions.

6 Disclaimer

The views and opinions contained in this paper are those of the authors and should not be construed as an official Department of the Army position, policy, or decision.

References

1. Rash, C.E., Russo, M.B., Letowski, T.R., Schmeisser, E.T.: Helmet-Mounted Displays: Sensation, Perception and Cognition Issues. Army Aeromedical Research Laboratory, Fort Rucker (2009)
2. Murphy, D.W., Gage, D.W., Bott, J.P., Marsh, W.C., Cycon, J.P.: Air-Mobile Ground Security and Surveillance System (AMGSSS) Project Summary Report. NRAD-TD-2914. Naval Command Control and Ocean Surveillance Center RDT&E Div, San Diego (1996)
3. Wargo, C.A., Church, G.C., Glaneueski, J., Strout, M.: Unmanned Aircraft Systems (UAS) research and future analysis. In: 2014 IEEE Aerospace Conference, pp. 1–16 (2014)
4. Mitchell, D.K., Brennan, G.: Infantry Squad Using the Common Controller to Control a Class 1 Unmanned Aerial Vehicle System (UAVS) Soldier Workload Analysis. ARL-TR-5012. U.S. Army Research Laboratory, Aberdeen Proving Ground (2009)
5. Mitchell, D.: Soldier Workload Analysis of the Mounted Combat System (MCS) Platoon's Use of Unmanned Assets. ARL-TR-3476. U.S. Army Research Laboratory, Aberdeen Proving Ground (2005)
6. Goldberg, D.H., Vogelstein, R.J., Socolinsky, D.A., Wolff, L.B.: Toward a wearable, neurally-enhanced augmented reality system. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) FAC 2011. LNCS, vol. 6780, pp. 493–499. Springer, Heidelberg (2011)
7. Liao, L.D., Lin, C.T., McDowell, K., Wickenden, A.E., Gramann, K., Jung, T.P., Chang, J.Y.: Biosensor technologies for augmented brain–computer interfaces in the next decades. Proc. IEEE **100**, 1553–1566 (2012)
8. McDowell, K., Lin, C.T., Oie, K.S., Jung, T.P., Gordon, S., Whitaker, K.W., Hairston, W.D.: Real-world neuroimaging technologies. IEEE Access **1**, 131–149 (2013)
9. Howard, I.P., Templeton, W.B.: Human Spatial Orientation. Wiley, Oxford (1966)
10. McGurk, H., MacDonald, J.: Hearing lips and seeing voices. Nature **264**, 746–748 (1976)
11. Jack, C.E., Thurlow, W.R.: Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. Percept. Mot. Skills **37**(3), 967–979 (1973)

12. Meredith, M.A., Nemitz, J.W., Stein, B.E.: Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J. Neurosci. **7**(10), 3215–3229 (1987)
13. Wallace, M.T., Meredith, M.A., Stein, B.E.: Integration of multiple sensory modalities in cat cortex. Exp. Brain Res. **91**(3), 484–488 (1992)
14. Hancock, P.A., Szalma, J.L. (eds.): Performance Under Stress. Ashgate Publishing, Burlington (2008)
15. Merlo, J.L., Duley, A.R., Hancock, P.A.: Cross-modal congruency benefits for combined tactile and visual signaling. Am. J. Psychol. **123**(4), 413–424 (2010)
16. Hancock, P.A., Warm, J.S.: A dynamic model of stress and sustained attention. Hum. Factors J. Hum. Factors Ergon. Soc. **31**(5), 519–537 (1989)
17. Oron-Gilad, T., Downs, J.L., Gilson, R.D., Hancock, P.A.: Vibrotactile guidance cues for target acquisition. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**(5), 993–1004 (2007)
18. Myles, K., Kalb, J.T.: Guidelines for Head Tactile Communication. ARL-TR-5116. U.S. Army Research Laboratory, Aberdeen Proving Ground (2010)
19. Hancock, P.A., Mercado, J.E., Merlo, J., Van Erp, J.B.: Improving target detection in visual search through the augmenting multi-sensory cues. Ergonomics **56**(5), 729–738 (2013)
20. Santangelo, V., Spence, C.: Assessing the automaticity of the exogenous orienting of tactile attention. Percept. London **36**(10), 1497–1506 (2007)
21. Prewett, M.S., Elliott, L.R., Walvoord, A.G., Coovert, M.D.: A meta-analysis of vibrotactile and visual information displays for improving task performance. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **42**(1), 123–132 (2012)
22. Parise, C.V., Spence, C., Ernst, M.O.: When correlation implies causation in multisensory integration. Curr. Biol. **22**(1), 46–49 (2012)
23. Jones, E.G., Powell, T.P.S.: An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain **93**(4), 793–820 (1970)
24. Schroeder, C.E., Foxe, J.: Multisensory contributions to low-level, 'unisensory' processing. Curr. Opin. Neurobiol. **15**(4), 454–458 (2005)
25. Schroeder, C.E., Foxe, J.J.: Multisensory Convergence in Early Cortical Processing. The Handbook of Multisensory Processes, pp. 295–309. MIT Press, Cambridge (2004)
26. Ghazanfar, A.A., Schroeder, C.E.: Is neocortex essentially multisensory? Trends Cogn. Sci. **10**(6), 278–285 (2006)
27. Driver, J., Noesselt, T.: Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron **57**(1), 11–23 (2008)
28. Cappe, C., Rouiller, E.M., Barone, P.: Multisensory anatomical pathways. Hear. Res. **258**(1), 28–36 (2009)
29. Calvert, G.A., Thesen, T.: Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol. Paris **98**(1), 191–205 (2004)
30. Petrus, E., Isaiah, A., Jones, A.P., Li, D., Wang, H., Lee, H.K., Kanold, P.O.: Crossmodal induction of thalamocortical potentiation leads to enhanced information processing in the auditory cortex. Neuron **81**(3), 664–673 (2014)
31. Stein, B.E., Stanford, T.R.: Multisensory integration: current issues from the perspective of the single neuron. Nat. Rev. Neurosci. **9**(4), 255–266 (2008)
32. Driver, J., Spence, C.: Multisensory perception: beyond modularity and convergence. Curr. Biol. **10**(20), R731–R735 (2000)
33. Welch, R.B., Warren, D.H.: Immediate perceptual response to intersensory discrepancy. Psychol. Bull. **88**(3), 638–667 (1980)
34. Battaglia, P.W., Jacobs, R.A., Aslin, R.N.: Bayesian integration of visual and auditory signals for spatial localization. JOSA A **20**(7), 1391–1397 (2003)

35. Sato, Y., Toyoizumi, T., Aihara, K.: Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. Neural Comput. **19**(12), 3335–3355 (2007)
36. Alais, D., Burr, D.: The ventriloquist effect results from near-optimal bimodal integration. Curr. Biol. **14**(3), 257–262 (2004)
37. Ernst, M.O., Banks, M.S.: Humans integrate visual and haptic information in a statistically optimal fashion. Nature **415**(6870), 429–433 (2002)
38. King, A.J., Walker, K.M.: Integrating information from different senses in the auditory cortex. Biol. Cybern. **106**(11–12), 617–625 (2012)
39. Puri, C., Olson, L., Pavlidis, I., Levine J., Starren, J.: Stress cam: non-contact measurement of users' emotional state through thermal imaging. In: Proceedings of the 2005 ACM Conference on Human Factors in Computing Systems (CHI), pp. 1725–1728 (2005)
40. Zhu, M., Wu, Y., Li, Q., Contrada, R., Ji, Q.: Non-intrusive Stress and Anxiety Detection by Thermal Video Analysis. U.S. Army Research Office Final Report (2014)
41. O'Kane, B.L., Sandick, P., Shaw, T., Cook, M.: Dynamics of human thermal signatures. In: Proceedings of the InfraMation Conference (2004)
42. Haynes, J.D., Rees, G.: Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. **7**, 523–534 (2006)
43. Huang, T., Koller, D., Malik, J., Ogasawara, G.H., Rao, B., Russell, S.J., Weber, J.: Automatic symbolic traffic scene analysis using belief networks. In: AAAI-94, pp. 966–972 (1994)
44. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: a survey. Comput. Vis. Image Underst. **108**(1), 116–134 (2007)
45. Ryoo, M.S., Aggarwal, J.K.: Recognition of composite human activities through context free grammar based representation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1709–1718 (2006)
46. Schmidt, C., Sridharan, N., Goodson, J.: The plan recognition problem: an intersection of psychology and artificial intelligence. Artif. Intell. **11**, 45–83 (1978)
47. Turaga, P., Chellappa, R., Subrahmanian, V.S., Udrea, O.: Machine recognition of human activities: a survey. IEEE Trans. Circuits Syst. Video Technol. **18**(11), 1473–1488 (2008)
48. Wang, C., Wang, Y., Yuille, L.: An approach to pose based action recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 915–922 (2013)
49. Wu, Y., Huang, T.S.: Vision-based gesture recognition: a review. In: Braffort, A., Gibet, S., Teil, D., Gherbi, R., Richardson, J. (eds.) GW 1999. LNCS (LNAI), vol. 1739, pp. 103–116. Springer, Heidelberg (2000)
50. Chen, X., Yuille, A.L.: Articulated pose estimation with image-dependent preference on pairwise relations. In: Advances in Neural Information Processing Systems 27 (NIPS 2014) (2014)
51. Fidler, S., Mottaghi, R., Yuille, A.L., Urtasun, R.: Bottom-up segmentation for top-down detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3294–3301 (2013)
52. Li, K., Fu, Y.: Prediction of human activity by discovering temporal sequence patterns. IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) **36**(8), 1644–1657 (2014)
53. Yao, Y., Zhang, F., Fu, Y.: Real-time hand gesture recognition using RGB-D sensor. In: Shao, L., Han, J., Kohli, P., Zhang, Z. (eds.) Computer Vision and Machine Learning with RGB-D Sensors, pp. 289–313. Springer, Cham (2014)
54. Ma, Z., Yang, Y., Li, X., Pang, C., Hauptmann, A.G., Wang, S.: Semi-supervised multiple feature analysis for action recognition. IEEE Trans. Multimedia **16**(2), 289–298 (2014)

55. Fuchs, N.E., Schwitter, R.: Specifying logic programs in controlled natural language. In: Proceedings on Computational Logic for Natural Language Processing, vol. 95, pp. 1–16 (1995)
56. Kuhn, T.: A survey and classification of controlled natural languages. Comput. Linguist. **40** (1), 121–170 (2014)
57. Pei, X., Barbour, D.L., Leuthardt, E.C., Schalk, G.: Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J. Neural Eng. **8** (4), 046028 (2011)
58. Pei, X., Hill, J., Schalk, G.: Silent communication: toward using brain signals. IEEE Pulse Mag. **3**(1), 43–46 (2012)
59. Gunduz, A., Brunner, P., Daitch, A., Leuthardt, E.C., Ritaccio, A.L., Pesaran, B., Schalk, G.: Neural correlates of visual–spatial attention in electrocorticographic signals in humans. Front. Hum. Neurosci. **5**, 89 (2011)
60. Gunduz, A., Brunner, P., Daitch, A., Leuthardt, E.C., Ritaccio, A.L., Pesaran, B., Schalk, G.: Decoding covert spatial attention using electrocorticographic (ECoG) signals in humans. Neuroimage **60**(4), 2285–2293 (2012)
61. Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R.J., Belin, P.: Neuropsychology: pitch discrimination in the early blind. Nature **430**(6997), 309 (2004)
62. Lessard, N., Pare, M., Lepore, F., Lassonde, M.: Early-blind human subjects localize sound sources better than sighted subjects. Nature **395**(6699), 278–280 (1998)
63. Jitsuki, S., Takemoto, K., Kawasaki, T., Tada, H., Takahashi, A., Becamel, C., Takahashi, T.: Serotonin mediates cross-modal reorganization of cortical circuits. Neuron **69**(4), 780–792 (2011)