

INTERDISCIPLINARY STUDIES OF MULTISENSORY INTEGRATION: FROM SINGLE UNITS TO PERCEPTION

Mark T. Wallace, Ph.D.

Vanderbilt Brain Institute

Departments of Hearing & Speech Sciences, Psychology and Psychiatry

Vanderbilt University

Abstract

The integration of information from multiple sensory sources is a ubiquitous and highly adaptive process that can dramatically alter behavior and shape perception. Neurophysiological studies in animal models and in human subjects have begun to reveal the combinatorial rules by which neurons and networks combine this multisensory information to create a coherent perceptual unity. In addition, recent work has highlighted that the development of multisensory circuits is strongly shaped by early sensory experience, and has the capacity for significant plasticity. Most recently, studies have begun to focus on the plastic potential of adult multisensory processes. This work has found that significant plasticity can be engendered in adult multisensory systems, but only by coupling changes in the statistical relations of the combined stimuli with reinforcement. Such findings have significant implications for our understanding of basic integrative processes and have shed important light in bridging between neuronal and perceptual levels of analysis.

An Introduction to Multisensory Processes

We live in a world rich in sensory information. This information is conveyed via various forms of stimulus energy (e.g., chemical, mechanical, electromagnetic, thermal, etc.), and our sensory systems have evolved specialized peripheral organs to transduce these energies into a common neural code (i.e., action potentials). Heroic efforts in neuroscience research over the past 50 years have focused on understanding the nature of these transduction processes, and on elucidating the “neural code” for each of the sensory systems. Intriguingly, this work has largely focused on understanding these processes within the individual sensory systems, and as a result we have a detailed understanding of the mechanics of the transduction and encoding events within the visual, auditory, somatosensory, vestibular, gustatory and olfactory systems. However, intimate knowledge of these events in each of these sensory systems is ultimately inadequate for understanding the nature of our perceptual gestalt, since this unity is built also from the synthesis of information across the different senses.

Numerous examples serve to highlight the powerful ability that the different sensory systems have to influence one another, and ultimately to shape our view of the world around us. One of the most entertaining of these examples is that of the ventriloquist, in which discordant visual cues (i.e., the movements of the dummy’s lips and head) can dramatically alter our judgments about the source of an auditory signal (i.e., the ventriloquist’s voice; see Thurlow and Jack 1973). In this particular example we refer to the ability of the visual cues to bias our localization of an auditory signal. In addition to these biasing effects under discordant conditions, visual cues can also dramatically improve the intelligibility of an auditory signal when presented in spatial concordance. For example, in a crowded and noisy room, the ability to see a speaker’s lip movements confers an effective 15 dB gain in our ability to comprehend their speech (Sumbly and Pollack 1954). Finally, and as first described

by William James (1890), visual information can alter the content of the spoken word. The grouping of the seemingly nonsensical French words *'pas de lieu Rhone que nous'* suddenly becomes recognized as 'paddle your own canoe' when accompanied by English subtitles. Although these illustrations are rich in contextual and semantic information, compelling influences between the senses can be seen even with very simple stimuli. One example of this is the so-called "flash-beep" illusion, in which the pairing of a single visual stimulus (i.e., a flash) with multiple auditory cues (i.e., the beeps) strongly biases the observer toward the perception of multiple flashes (Shams et al. 2000). Finally, although each of these examples highlights the interplay between vision and audition, multisensory interactions have been described for almost all sensory combinations (for review see Calvert et al. 2004).

Although these perceptual examples are compelling, they are not the reasons why the different sensory systems have evolved with the capacity to influence one another. These reasons are undoubtedly rooted more in changes in behavioral processes mediated by multiple sensory inputs. An example of this is in the speeding of simple reaction times seen under multisensory conditions. Thus, the average reaction time to a combined visual-auditory stimulus is significantly faster than the responses to either the visual or auditory stimuli when delivered alone (Hershenson 1962). In addition to faster responses, a number of studies have shown improvements in response accuracy under multisensory circumstances. For example, behavioral studies in cats have shown that the pairing of visual and auditory stimuli that, when presented alone, result in few correct responses, dramatically improves response accuracy (Stein et al. 1989).

Despite the ubiquity of such multisensory phenomena, and their utility in improving our ability to extract information from and respond appropriately to the external world, surprisingly little is known about the neural bases for these interactions when compared with the individual sensory systems. Only in the last 25 years has this become an active area of investigation, and these studies have begun to reveal the nature of the multisensory encoding process.

The Neural Bases of Multisensory Integration: Lessons from the Single Neuron

Anatomical studies have long shown that inputs from the different senses converge at numerous sites in the brain, making these regions good candidates for mediating the behavioral and perceptual changes seen under multisensory conditions. In an effort to better understand the encoding processes that take place in these convergent zones, neurophysiological studies targeted one of best characterized of these structures, the superior colliculus (SC). Along with serving as a watershed for the convergence of visual, auditory and somatosensory (i.e., tactile) inputs, this midbrain structure had long been known to play an important role in the control of gaze, making it an ideal model for revealing how sensory cues are transformed into appropriate motor commands (for review see Stein and Meredith 1993).

These neurophysiological studies found that many neurons in the SC are multisensory – responsive to or influenced by stimuli from two or more sensory modalities. However, of greatest interest was the fact that these neurons did far more than simply respond to these stimuli – they actively integrated these inputs and gave rise to transformed responses that differed substantially from the individual unisensory responses. For example, a multisensory neuron might respond to a given visual stimulus with an average of 5 action potentials per stimulus presentation, and it might respond to a given auditory stimulus with an average of 3 action potentials. One could predict that under combined visual-auditory stimulation the neuron might respond with either 5 (the best unisensory response) or 8 (the sum of the visual and auditory responses) action potentials. Instead, a typical response might be 15-20 action potentials, reflecting a strong nonlinear gain in the integrative process.

The interactive products of these multisensory combinations were subsequently found to abide by a very simple set of integrative principles grounded in the physical characteristics of the combined stimuli (see Stein and Meredith 1993). Thus, the relative effectiveness of the individual stimuli was found to be an important determinant of the response seen to their pairing. Whereas the pairing of two weak (i.e., eliciting few action potentials) unisensory stimuli typically results in a large response gain, the pairing of highly effective stimuli generally resulted in little gain. This principle, labeled “inverse effectiveness,” makes a good deal of intuitive sense, since strongly effective stimuli in one sense are highly salient, whereas the real benefits of multisensory cue combinations are seen when the individual stimuli are at or near threshold. In addition to effectiveness, the spatial relationship of multisensory stimuli is critically important for the resultant interaction, such that if the paired stimuli are in close spatial proximity, large response gains are the typical outcome. Conversely, stimuli that are spatially disparate typically result in no interaction, or an actual depression in the response. Finally, in parallel with this spatial principle, the temporal relationship of the stimuli was found to be important, such that stimuli occurring close together in time result in large enhancements, and as the temporal interval between them increases the interaction declines. These spatial and temporal principles also have strong ethological validity, in that stimuli that originate from the same location at the same time have a high probability of being derived from the same event.

Together, these integrative principles, first derived from recordings from single neurons in the midbrain, have been found to be applicable across a wide range of brain structures and animal species. Indeed, and as detailed below, the universality of these principles has been found to extend to include recordings from much larger neuronal ensembles (e.g., fMRI), as well as to hold in a variety of psychophysical and perceptual contexts.

The Principles of Multisensory Integration: From Neurons to Networks and Beyond

Contemporary neuroimaging approaches are beginning to be used to better elucidate the neural networks that are involved in multisensory processes. Several of these studies have employed a stimulus structure derived from single unit neurophysiological analyses – thus manipulating the relative effectiveness of the paired stimuli along with their spatial and temporal relationships. For example, a recent study by Stevenson and James (2009) sought to examine inverse effectiveness in a canonical cortical structure for audiovisual convergence and integration, the superior temporal sulcus (STS). Parametric manipulation of the effectiveness of the visual and auditory stimuli by changing the signal to noise ratio (SNR) resulted in changes in the BOLD signal consistent with inverse effectiveness, in that low SNR pairings resulted in superadditive interactions, whereas higher ratios did not. Similarly, other studies that have focused on manipulating the spatial and temporal structure of a multisensory stimulus complex have found strong concordance with the established principles in a network of cortical areas that include the STS, prefrontal, posterior parietal and insular cortices (Bushara et al. 2003; Macaluso et al. 2004).

Along with these examinations into the neural bases of multisensory processes, a number of studies have looked at these principles in the context of human behavior, assessing changes in performance and perception as they relate to the physical characteristics of the paired stimuli. For example, a study of human saccades in response to visual and auditory targets found that responses were speeded (i.e., of shorter latency) when targets were spatially- and temporally-coincident, and that this facilitation in response times gradually disappeared as the stimuli were separated in space and/or time (Frens et al. 1995). In a detection study, Bolognini and colleagues (2005) found similar facilitations in target detection

that also appeared to abide by inverse effectiveness, in that these gains were seen preferentially for stimuli at or below threshold detection levels. As alluded to earlier, these behavioral advantages conferred by multisensory stimulus presentations make a great deal of adaptive sense, in that the information about an object or event is amplified by the use of multiple sensory channels.

Plasticity in Multisensory Representations: Lessons from Development

In recent years a great deal of interest has arisen in the development of multisensory circuits and the behavioral and perceptual processes that they mediate. These studies have been motivated, at least in part, by the strong discussions and debates in the human developmental literature as to the multisensory state of the newborn brain. Whereas one extreme holds that the newborn is an obligate synesthete, merging all sensory information into a single perceptual representation, the other extreme holds that multisensory convergence and integration only develops after the appropriate associations between the senses are made during early life (see Lewkowicz [1994] for a review). Although evidence can be put forth in support of both of these views, recent neurophysiological studies have begun to examine the maturational timetable for multisensory neurons and their integrative properties in several of the well-established model systems (e.g., the SC). This work has shown that multisensory neurons are very immature early in development, and only gradually acquire their adult-like features during postnatal life (Wallace and Stein 1997).

The prolonged maturational timeline for these multisensory networks strongly suggested that a critical element in the developmental process was the acquisition of the appropriate sensory experiences during a “critical” or “sensitive” period. To test the importance of early sensory experience in these events, recent studies have eliminated or altered these experiences and examined their consequent impact on multisensory processes. Whereas eliminating experience in one sensory modality (i.e., vision) abolishes all visual-nonvisual integration in multisensory neurons (Wallace et al. 2004), altering the statistical relations of the pairing changes the integrative process so that it reflects these modified experiences (Wallace and Stein 2007). For example, animals that are reared in an environment in which visual and auditory stimuli are always presented at the same time but consistently from disparate locations (i.e., always separated by 30°) form multisensory populations that amplify the neural signal associated with this pairing. Together, these results have strongly reinforced the importance of early sensory experience in shaping developing multisensory representations.

In contrast to these development effects, similar manipulations in adult animals fail to result in significant reorganization, a result that is similar to what has been observed within the individual sensory systems (and which gave rise to the concept of a critical period, see Hubel and Wiesel 1998). However, recent explorations within unisensory systems has suggested that plasticity can indeed be engaged in adult sensory systems, but that simple manipulation of the sensory statistics is insufficient to drive these changes.

Adult Multisensory Plasticity: Bridging Single Neurons to Perceptual Processes

These adult plasticity studies (see Blake et al. 2006) have shown it necessary for there to be an association between changes in the sensory world and reinforcement processes for significant plasticity to be seen in the adult. In an effort to extend this framework to multisensory systems, we have recently embarked on parallel experiments in which adult animals are rewarded for responses to paired visual-auditory cues that are presented from the same location but that are separated by a fixed temporal interval (e.g., 100 ms). After several

weeks in such an altered environment, neurophysiological examinations of multisensory neuronal populations found that the temporal tuning functions of these neurons had shifted so as to reflect the new timing relationship between these stimuli. In contrast, control animals raised under identical sensory conditions but in the absence of reward failed to show changes. Hence, as is seen in the individual sensory systems, the cognitive association between stimulus and reward appears essential to engage multisensory plasticity in the adult brain.

In an effort to bridge between these neuronal findings and adult perceptual processes, we have recently conducted a study to examine whether similar rules govern multisensory perceptual plasticity. The design of this study is predicated on the concept of a temporal “window” of multisensory integration, within which the brain strives to bind together sensory events from different modalities. Intriguing, this temporal window, indexed in a number of human perceptual studies, seems to span several hundred milliseconds, a number very close to the width of the temporal tuning functions described above for individual neurons (for a review of this literature see Powers et al. 2009). The experiment employed a simultaneity judgment task in which participants had to make judgments as to the simultaneity of an audiovisual stimulus pairing (both two-alternative and two-interval designs were used). Plotting reports of simultaneity as a function of temporal offset between the stimuli results in a distribution that provides a proxy for the multisensory temporal window, which averaged approximately 300 ms prior to training. Participants were then divided into two groups, one that underwent feedback training while performing the task and the other that received the same stimuli but with no feedback. The results were dramatic, with a nearly 40% reduction in the size of the window in trained subjects and no reduction in the passive exposure group.

These results are very exciting in a number of domains. First, they represent an important bridge between neurophysiological and perceptual levels of analysis, and provide key clues as to the brain networks important in multisensory perceptual processes and plasticity. Second, they extend work in the individual sensory systems to illustrate the striking plastic potential of the brain. Finally, they offer great hope for use as remediation tools in circumstances in which multisensory function has gone awry, and evidence is accumulating for disruptions in multisensory temporal binding in conditions such as autism, dyslexia and schizophrenia.

References

- Blake DT, Heiser MA, Caywood M, Merzenich MM (2006) Experience-dependent adult cortical plasticity requires cognitive association between sensation and reward. *Neuron* 52:371-381.
- Bolognini N, Frassinetti F, Serino A, Ladavas E (2005) "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp Brain Res* 160:273-282.
- Bushara KO, Hanakawa T, Immisch I, Toma K, Kansaku K, Hallett M (2003) Neural correlates of cross-modal binding. *Nat Neurosci* 6:190-195.
- Calvert GA, Spence C, Stein BE, eds (2004) *The Handbook of Multisensory Processes*. Cambridge, MA: The MIT Press.
- Frens MA, Van Opstal AJ, Van der Willigen RF (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys* 57:802-816.
- Hershenson M (1962) Reaction time as a measure of intersensory facilitation. *J Exp Psychol* 63:289-293.
- Hubel DH, Wiesel TN (1998) Early exploration of the visual cortex. *Neuron* 20:401-412.
- James W (1890) *Principles of Psychology*. Cambridge, MA: Harvard University Press.

- Lewkowicz DJ (1994) Development of intersensory perception in human infants. In: The development of intersensory perception: comparative perspectives (Lewkowicz DJ, Lickliter R, eds), pp 165-203. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Macaluso E, George N, Dolan R, Spence C, Driver J (2004) Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage* 21:725-732.
- Powers AR, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. *J Neurosci* (In Press)
- Shams L, Kamitani Y, Shimojo S (2000) Illusions. What you see is what you hear. *Nature* 408:788.
- Stein BE, Meredith MA (1993) *The Merging of the Senses*. Cambridge, MA: MIT Press.
- Stein B, Meredith M, Huneycutt W, McDade L (1989) Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J Cogn Neurosci* 1:12-24.
- Stevenson RA, James TW (2009) Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44:1210-1223.
- Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212-215.
- Thurlow WR, Jack CE (1973) Certain determinants of the "ventriloquism effect". *Percept Mot Skills* 36:1171-1184.
- Wallace MT, Perrault TJ, Jr., Hairston WD, Stein BE (2004) Visual experience is necessary for the development of multisensory integration. *J Neurosci* 24:9580-9584.
- Wallace MT, Stein BE (1997) Development of multisensory neurons and multisensory integration in cat superior colliculus. *J Neurosci* 17:2429-2444.
- Wallace MT, Stein BE (2007) Early experience determines how the senses will interact. *J Neurophysiol* 97:921-926.