ASSESSING MULTISENSORY INTEGRATION WITH ADDITIVE FACTORS AND FUNCTIONAL MRI

Thomas W. James, Ryan A. Stevenson and Sunah Kim
Department of Psychological and Brain Sciences, Indiana University, 1101 E Tenth St, Bloomington, IN 47401, USA
thwjames@indiana.edu; stevenra@indiana.edu; suakim@indiana.edu

Abstract

The topic of this presentation is the use of additive-factor designs in combination with functional MRI to assess multisensory integration. Unisensory and multisensory stimuli were presented across two different pairings of sensory systems, audio-visual (AV) and visuo-haptic (VH). In addition to stimulus modality, signal-to-noise ratio (SNR) was included as an additive factor. Previous research investigating the effect of SNR on sensory integration has documented an effect called inverse effectiveness, where the multisensory gain increases with decreasing SNR. Potential sites of multisensory convergence were mapped for each sensory pairing and were found to be non-overlapping, suggesting that the neural mechanisms of integration are specialized for each unique pairing of sensory systems. Evidence of inverse effectiveness was found at all convergence sites, regardless of whether they were AV or VH. This result suggests that inverse effectiveness is a general characteristic of multisensory integration, regardless of the sensory pairing. The results also showed that a single-factor additive model of multisensory integration produced different outcomes at different levels of SNR. Based on this last result, we conclude that an additive-factors approach to assessing multisensory integration will provide more reliable inferences than single-factor designs.

Advances in neuroimaging in the last two decades have provided methods for studying the neural mechanisms of cognitive function in humans non-invasively. Functional MRI and other neuroimaging techniques have provided unique insights about the relations between cognitive function and brain circuits. The relations between fMRI measurements, such as blood oxygenation-level dependent (BOLD) response, and neural activity measurements, such as action potentials or local field potentials, however, are not fully understood. Thus, caution must be exercised when making inferences based on neuroimaging measurements.

The study of multisensory perception has a long history in science (Molyneux, 1688), but recently there has been increased interest in multisensory phenomena, and especially their neural mechanisms (Amedi, Von Kriegstein, Van Atteveldt, Beauchamp, & Naumer, 2005; Stein & Stanford, 2008). Our research program focuses specifically on multisensory object perception, including the perception of object identity through vision, touch, and hearing. The focus of this presentation is twofold. First, to examine whether or not there are unique neural substrates involved in integrating information from unique pairings of sensory systems or, more specifically, are the sites of integration for visuo-haptic pairings different from the sites for audio-visual pairings.

The second focus is to evaluate the utility of using additive-factors designs to assess multisensory integration with BOLD fMRI measurements. The additive factor used was stimulus signal-to-noise ratio (SNR). This factor was chosen because it is commonly used in multisensory research to study a phenomenon called ‘inverse effectiveness’ (Meredith & Stein, 1986). In both single-unit recordings and behavioral measures (accuracy and RT), inverse effectiveness is observed when multisensory gain increases as stimulus SNR decreases (Holmes, 2007). In other words, as the stimuli become more difficult to discriminate, there is a greater benefit to integrating across sensory sources.

Method

Audio-visual Experiment

There were 11 right-handed, native-English speaking subjects (6 female, mean age = 25.9). Speech stimuli were selected from a previously published stimulus set, The Hoosier Audiovisual Multi-Talker Database. Ten speech tokens from a single female talker (F1) were used. Words were monosyllabic, had the highest levels of accuracy on both visual-only and audio-only discrimination and resided in low-density neighborhoods. The 10 words were selected such that they fell into two easily distinguishable semantic categories, and had mean word length approximately equal across categories. The two categories were body part words (face, leg, mouth, neck, and teeth) and environmental feature words (beach, dirt, rain, rock, and sand).

Prior to imaging, subjects' individual psychophysical thresholds were found. Subjects performed two-alternative forced-choice (2AFC) task, deciding whether the word was a body part or an environmental feature. Separate staircases were conducted for audio-only and visual-only (lip reading) performance. Noise variance (RMS) was held constant while contrast (RMS) varied in a two-up one-down staircase. The data gathered during the staircases were fitted with psychometric Weibull functions.

Stimulus levels used during the fMRI testing procedure were derived from the fitted functions for each participant for audio (A) and visual (V) stimuli at four SNR levels corresponding to 65, 75, 85, and 95% accuracy. SNR of AV presentations, which combined A and V stimulus components into a multisensory combination stimulus, were not determined from AV staircases. Instead, AV presentations were derived by combining A and V components of the same signal level.

Visuo-haptic Experiment

There were 14 right-handed subjects (7 female; mean age 27.2). Object stimuli were tangible plastic cubes with different amounts of curvature on the top surface. That is, the top surface varied in curvature from very curved to only slightly curved. Subjects performed a 2AFC task to discriminate which of two objects was more curved. On a single trial, subjects were presented with a single object, on which they made the 2AFC decision. Within blocks of trials the trial stimulus was chosen randomly from a pair of stimuli of different curvature. The relative difference in curvature between the pair determined the difficulty of the block of trials. Based on pilot experiments, two pairs of stimuli were chosen that produced large differences in difficulty across all subjects. Although this difficulty variable is operationally different from SNR, we will nonetheless refer to this variable as SNR below, for consistency with the AV experiment

During fMRI testing, subjects were presented with objects in three different sensory modality conditions. For each block of trials, subjects were signaled with a short audio cue as to the stimulus modality condition. For visual (V), subjects opened their eyes to view the object and also moved their index finger over the object, but did not touch it. For haptic (H), subjects left their eyes closed and explored the object with their index finger. For VH, subjects opened their eyes and explored the curvature with their index finger. All exploration was done using the right hand and responses (more or less curved) were made with left-hand button presses.

General Procedures

The study was approved by the Indiana University Institutional Review Board. Written informed consent was obtained prior to the experiments. Each imaging session included two phases: functional localizer runs and experimental runs. Localizer runs used high-SNR stimuli presented in a blocked stimulus design and subjects performed a simple task to maintain attention. During experimental runs, stimuli were presented in a rapid event-related design and subjects performed the same 2AFC task used to find thresholds. For both the AV and VH experiments, approximately 40 trials were collected for each combination of stimulus modality and SNR level (3x4 for AV and 3x2 for VH experiment).

Imaging was carried out using a Siemens Magnetron Trio 3-T whole body scanner, with eight-channel phased-array head coil. The field of view was 22 x 22 x 9.9 cm, with an in plane resolution of 64 x 64 pixels and 33 axial slices per volume (whole brain), creating a voxel size of 3.44 x 3.44 x 3 mm. Images were collected using a gradient echo EPI (TE = 30 ms, TR = 2000 ms, flip angle = 70°) for BOLD imaging. High-resolution T1-weighted anatomical volumes were acquired using Turbo-flash 3-D (TI = 1,100 ms, TE = 3.93 ms, TR = 14.375 ms, Flip Angle = 12°) with 160 sagittal slices with a thickness of 1 mm and field of view of 224 x 256 (voxel size = 1 x 1 x 1 mm). Imaging data were pre-processed using Brain Voyager ™ 3-D analysis tools. Anatomical volumes were transformed into a common stereotactic space (Talaraich and Tournoux, 1988). Functional data were aligned to the transformed anatomical volumes, transforming the functional data to a common stereotactic space across participants. Functional data underwent a linear trend removal, 3-D spatial Gaussian filtering (FWHM 6 mm), slice scan time correction, and 3-D motion correction. Whole-brain, random-effects statistical parametric maps (SPM) were calculated using Brain Voyager™ general linear model (GLM) procedure. Event-related averages (ERA), consisting of aligning and averaging all trials from each condition to stimulus onset, were created based on stimulus type for both the localizer and the experimental study. BOLD response amplitudes were defined as the arithmetic mean of the time course within a time window 6–16 s after block onset for the localizer runs, and a window of 4-6 s after trial onset for the rapid event-related experimental runs.

Results and Discussion

In the first analysis phase, potential sites of sensory convergence were identified using functional localizer data. Figure 1 shows an overlay of the group-average statistical parametric maps (SPM) for both experiments on an inflated representation of the cortical sheet. Sites of AV convergence and the sites of VH convergence were non-overlapping. Based on previous research (Amedi, Malach, Hendler, Peled, & Zohary, 2001; Calvert, Campbell, & Brammer, 2000), we chose to perform further analyses on regions of interest (ROI) in the posterior superior temporal cortex (STC) and the lateral occipical tactile visual area (LOtv). These results suggest the existence of integration mechanisms that are specific to certain sensory pairings. Integration mechanisms may be specialized to make optimal use of redundant or complimentary information across multiple sensory streams. The nature of those redundancies and how they can be exploited are likely specific to the particular unique pairing of sensory systems.

In the second analysis phase, BOLD measurements collected during the experimental runs and extracted from the specific ROIs described above were examined for evidence of inverse effectiveness. In Figure 1, graphs on the left and right are for the AV and VH pairings, respectively. Graphs on the top row show raw BOLD percent signal change as a function of the three stimulus modalities and the different levels of SNR. Across all six stimulus modalities, increasing SNR or difficulty decreases the BOLD response.

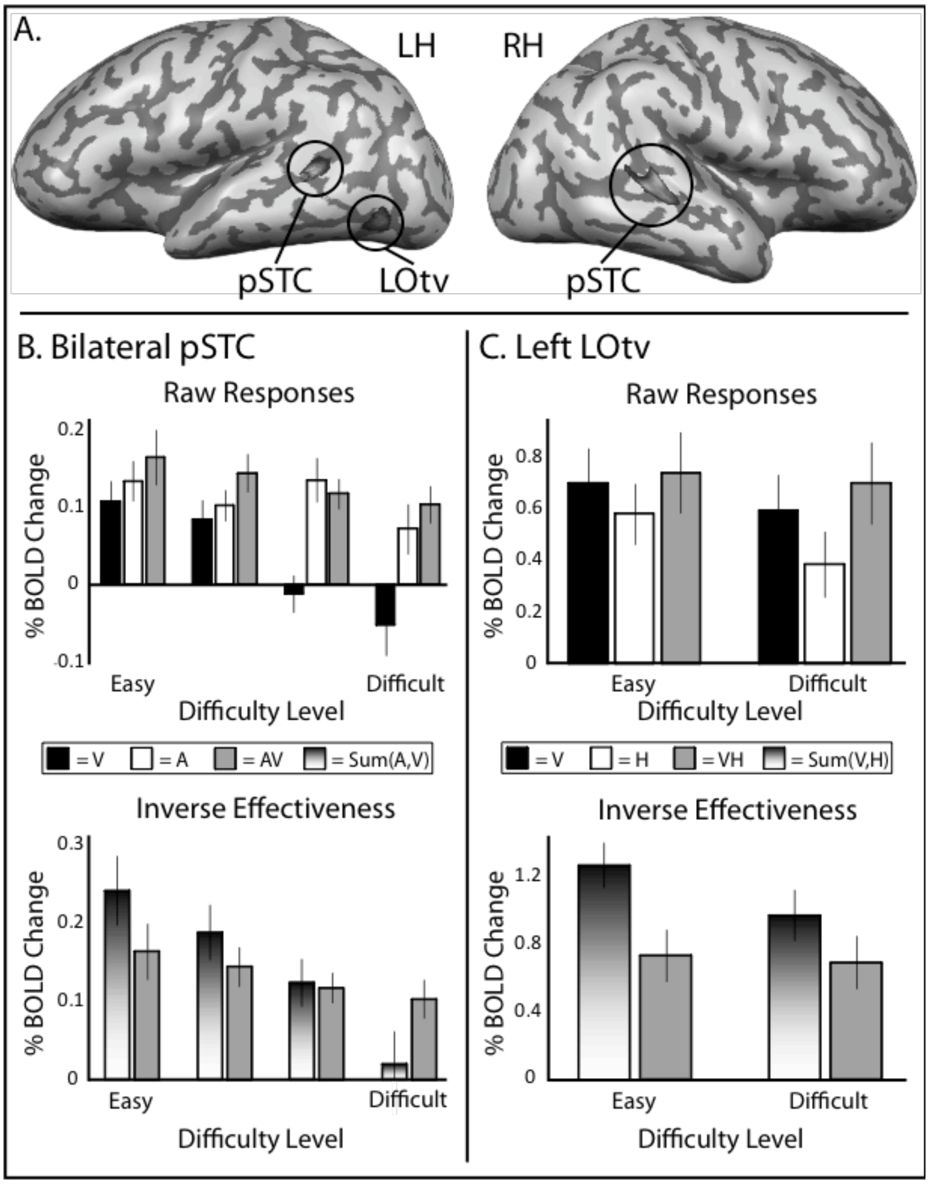

Fig. 1. Assessment of inverse effectiveness at sites of multisensory convergence.

The established model for assessing multisensory integration with BOLD measurements is an additive model (Calvert, et al., 2000). The use of the additive model is based on two supportable premises. First, BOLD activation can be modeled as a time-invariant linear system, that is, activation produced by two stimuli presented together can be modeled by summing the activity produced by those same two stimuli presented alone (Ashby & Waldschmidt, 2008). Second, the null hypothesis to be rejected is that the neuronal population does *not* contain multisensory neurons (Calvert, et al., 2000). BOLD responses are measured from the vasculature that supplies blood to a heterogeneous *population* of neurons. When modeling the underlying activity that produces BOLD responses, it is tempting to consider that all of the neurons in that population have similar response properties. However, neuronal populations within multisensory brain regions contain a mixture of unisensory neurons from different sensory modalities and different types of multisensory neurons (Allman, Keniston, & Meredith, 2009; Stein & Stanford, 2008). Using the additive criterion, the presence of multisensory neurons can be inferred (and the null hypothesis is rejected) if the activation with the multisensory stimulus exceeds the additive criterion (i.e., superadditivity). A time-invariant linear system that spatially sums across a population of neurons that contains only unisensory neurons will produce a BOLD response with a multisensory stimulus that is the sum of the BOLD responses with its unisensory components.

Graphs on the bottom row of Figure 1 show this model prediction with the graded black/white bars; the height of the bars represents the sum of the two unisensory conditions, either A+V or V+H. These model predictions are compared to the observed BOLD data for the multisensory condition, either AV or VH. The observed BOLD response is sub-additive at high levels of SNR and is superadditive at the lowest levels of SNR. This change from a sub-additive to a superadditive pattern across changes in SNR is consistent with the pattern of inverse effectiveness. The multisensory gain increased with decreasing SNR. The pattern of inverse effectiveness was the same for AV and VH stimulus pairings, even though the sites of multisensory convergence for AV and VH pairings were non-overlapping, suggesting that inverse effectiveness maybe a general characteristic of multisensory integration.

In the third analysis phase, a statistical parametric mapping analysis was conducted on the experimental runs to assess the whole brain for evidence of inverse effectiveness. With both AV and VH pairings, a network of brain regions was found that showed a pattern of inverse effectiveness. However, consistent with our other analyses, the networks were non-overlapping (Stevenson, Kim, & James, in press).

The patterns of inverse effectiveness shown in Figure 1 can be interpreted in the larger context of additive-factors designs (Sartori & Umilta, 2000). To assess inverse effectiveness, more than one level of SNR is needed. To predict a BOLD response with a multisensory stimulus using the additive model, however, requires observed data from only one level of SNR. Most fMRI experiments employ the latter design rather than the former. The results clearly show that the outcome of designs using only a single level of SNR is dependent on the difficulty. This may explain the inconsistency of results interpreted with the additive model (Beauchamp, 2005). By employing a factorial design, where the factor of interest is crossed with an additional factor, more reliable and rigorous inferences can be made from the results. Here, SNR was used as an additive factor, which revealed sub-networks of brain regions involved in multisensory integration that also showed patterns of inverse effectiveness. The use of other additive factors, such as temporal synchrony or spatial congruence, would potentially reveal other sub-networks with different specializations.

We have focused here on the interaction pattern called inverse effectiveness. In previous work, however, other types of interaction effects have been found between stimulus modality and SNR. In a previous VH experiment, LOtv showed ‘enhanced effectiveness’ when visual and haptic sources were spatially incongruent (Kim & James, in press). BOLD response with unisensory V and H stimuli decreased slightly with decreasing SNR, but BOLD response with multisensory VH stimuli decreased dramatically. Thus, multisensory gain increased with *increasing* SNR, rather than decreasing SNR. A third interaction effect has also been found. Brain regions such as the anterior cingulate cortex (ACC), which are involved in executive control and performance monitoring, usually show increased BOLD response with decreasing SNR, which is the reverse of sensory brain regions such as STC and LOtv. In these regions, decreasing SNR produces a stronger effect on unisensory BOLD than multisensory BOLD (Stevenson, et al., in press), an effect that is analogous to inverse effectiveness, but with changes of the opposite sign.

The existence in the literature of three interaction effects that are related to, but distinct from, inverse effectiveness, suggests that a nomenclature should be established to distinguish them. We suggest a naming convention that describes the direction of the relationship between SNR and BOLD response and the influence that increasing SNR has on multisensory gain. Inverse effectiveness would be called direct multisensory gain suppression (or direct suppression). Enhanced effectiveness would be called direct multisensory gain enhancement (or direct enhancement). The third effect described in anterior cingulate cortex would be called indirect suppression. A fourth type of interaction, which to our knowledge has yet to be shown, would be called indirect enhancement.

Acknowledgments

This research was supported in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc., the IUB Faculty Research Support Program, and the Indiana University GPSO Research Grant.

References

Allman, B. L., Keniston, L. P., & Meredith, M. A. (2009). Not just for bimodal neurons anymore: the contribution of unimodal neurons to cortical multisensory processing. *Brain Topogr, 21*(3-4), 157-167.

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience, 4*(3), 324-330.

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal indentification and object recognition. *Experimental Brain Research, 166*, 559-571.

Ashby, F. G., & Waldschmidt, J. G. (2008). Fitting computational models to fMRI. *Behavior Research Methods, 40*(3), 713-721.

Beauchamp, M. S. (2005). Statistical criteria in FMRI studies of multisensory integration. *Neuroinformatics, 3*(2), 93-113.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol, 10*(11), 649-657.

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability. *Neuropsychologia, 45*(14), 3340-3345.

Kim, S., & James, T. W. (in press). Enhanced effectiveness in visuo-haptic object-selective brain regions with increasing stimulus salience. *Human Brain Mapping*.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol, 56*(3), 640-662.

Molyneux, W. (1688). Letter to John Locke. In E. S. de Beer (Ed.), *The correspondence of John Locke* (Vol. 3). Oxford: Clarendon Press.

Sartori, G., & Umilta, C. (2000). The additive factor method in brain imaging. *Brain Cogn, 42*(1), 68-71.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci, 9*(4), 255-266.

Stevenson, R. A., Kim, S., & James, T. W. (in press). An additive-factors design to disambiguate neuronal and areal convergence: Measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Experimental Brain Research*.