

An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI

Ryan A. Stevenson · Sunah Kim · Thomas W. James

Received: 30 September 2008 / Accepted: 20 March 2009 / Published online: 8 April 2009
© Springer-Verlag 2009

Abstract It can be shown empirically and theoretically that inferences based on established metrics used to assess multisensory integration with BOLD fMRI data, such as superadditivity, are dependent on the particular experimental situation. For example, the law of inverse effectiveness shows that the likelihood of finding superadditivity in a known multisensory region increases with decreasing stimulus discriminability. In this paper, we suggest that Sternberg's additive-factors design allows for an unbiased assessment of multisensory integration. Through the manipulation of signal-to-noise ratio as an additive factor, we have identified networks of cortical regions that show properties of audio-visual or visuo-haptic neuronal convergence. These networks contained previously identified multisensory regions and also many new regions, for example, the caudate nucleus for audio-visual integration, and the fusiform gyrus for visuo-haptic integration. A comparison of integrative networks across audio-visual and visuo-haptic conditions showed very little overlap, suggesting that neural mechanisms of integration are unique to particular sensory pairings. Our results provide evidence for the utility of the additive-factors approach by demonstrating its effectiveness across modality (vision, audition,

and haptics), stimulus type (speech and non-speech), experimental design (blocked and event-related), method of analysis (SPM and ROI), and experimenter-chosen baseline. The additive-factors approach provides a method for investigating multisensory interactions that goes beyond what can be achieved with more established metric-based, subtraction-type methods.

Keywords Integration · Vision · Superadditivity · Perception · Neuroimaging · Auditory · Audio-visual · Speech recognition · Object recognition · Superior temporal sulcus

Introduction

The field of multisensory processing has grown from one largely based around behavioral measurements in humans and studies using single-unit recording in animals to one that is also informed directly about human neurophysiology by non-invasive measures like blood oxygenation-level dependent (BOLD) fMRI. Early studies of multisensory integration using fMRI noted that, because fMRI measures are derived from populations of neurons, the criteria for inferring convergence of sensory signals must be different for the two techniques (Calvert et al. 2000). Neuronal convergence, or the convergence of multiple sensory streams onto the same neuron (Meredith et al. 1992), is easily defined for single-unit recordings. If the response of a neuron to one sensory input is modulated by a second sensory input, that is evidence of neuronal convergence. If populations of neurons were homogenous in function, the assessment of neuronal convergence would be the same for populations; however, it cannot be assumed that the populations of neurons measured using fMRI are homogenous.

R. A. Stevenson (✉) · T. W. James
Department of Psychological and Brain Sciences,
Indiana University, 1101 East Tenth Street, Room 293,
Bloomington, IN 47405, USA
e-mail: stevenra@indiana.edu

R. A. Stevenson · S. Kim · T. W. James
Program in Neuroscience, Indiana University,
Bloomington, USA

S. Kim · T. W. James
Cognitive Science Program,
Indiana University, Bloomington, USA

As such, we must rule out the possibility that the BOLD response is not merely showing areal convergence, where sensory streams converge onto a brain region or voxel without interacting with each other. Distinguishing between areal and neuronal convergence with BOLD fMRI is a fundamental issue in functional neuroimaging. If multiple sensory streams converge on an area, but do not synapse onto the same neurons, the area should not be considered a site of integration.

In a majority of single-unit recording studies, multisensory integration in a single neuron is defined using the maximum rule. This rule has a clear analytic basis: an increase in spike count with a multisensory stimulus over and above the maximum count produced by a unisensory stimulus necessarily indicates that the cell was influenced by more than one modality of sensory input ($S_1 S_2 > S_1 \cap S_2$). The maximum rule, however, does not apply well to measurements like BOLD fMRI, which are pooled across multiple units. Animal models suggest that multisensory brain areas contain a heterogeneous mix of unisensory and multisensory neurons. It has been shown mathematically that the presence of two types of unisensory neurons in a population without any multisensory neurons is enough to elicit BOLD activation that exceeds the maximum rule (Calvert et al. 2000, 2001). Thus, although multisensory activation exceeding the maximum rule indicates that sensory streams converge on an area (areal convergence), it cannot verify the presence or absence of neuronal convergence (Stevenson et al. 2007).

Calvert et al. (2000) were the first to suggest the use of a superadditivity criterion or sum rule ($S_1 S_2 > S_1 + S_2$) to assess multisensory integration with functional neuroimaging measures. Because BOLD responses can be modeled as a time-variant linear system (Boynton et al. 1996; Dale and Buckner 1997; Glover 1999; Heeger and Ress 2002), the null hypothesis when using the superadditivity criterion is that a multisensory stimulus will produce activation equal to the linear sum of the activations with the component unisensory stimuli. The presence of multisensory neurons can be inferred if the activation with the multisensory stimulus exceeds the criterion. Although a few early studies (Calvert et al. 2000, 2001) made good use of superadditivity, later studies suggested that the criterion was too strict and should be replaced by more liberal criteria such as the maximum and mean rule. These suggestions were driven from empirical evidence that the false-negative rate with superadditivity was too high for voxels in known multisensory areas (Beauchamp 2005). Thus, although the superadditivity criterion has been the clear choice for researchers based on theoretical grounds, it proves difficult to use in practice. In the remainder of this article, we suggest theoretical grounds for why superadditivity (and

in fact all of the rules described above) are inappropriate for use with BOLD fMRI. We also provide an alternative criterion for assessing neuronal convergence, and present new findings based on that criterion.

Neuronal spike counts are measured on a ratio scale, a scale that has an absolute zero. BOLD responses, however, are not. Instead, BOLD responses measure only the relative change from a control condition or baseline (the BOLD level from which the relative change is measured will henceforth be referred to as ‘baseline’). For BOLD measurements, ‘zero’ is not absolute, but is entirely dependent on what each particular experimenter chooses to use as their experimental baseline (Binder et al. 1999; Stark and Squire 2001). Thus, BOLD signals are measured on an interval scale at best (Stevens 1946). The use of an interval scale affects the interpretation of the superadditivity metric due to the fact that measuring the superadditivity criterion is reliant upon summing two unisensory responses and comparing to a single multisensory response. Because the responses are measured relative to an arbitrary baseline, the baseline has a different effect on the summed unisensory responses than on the single multisensory response. Superadditivity for audio-visual stimuli is described according to the following equation:

$$A + V < AV \quad (1)$$

but, it is more accurately described by:

$$(A - \text{baseline}) + (V - \text{baseline}) < (AV - \text{baseline}), \quad (2)$$

This can be rewritten as:

$$A + V - 2 * \text{baseline} < AV - 1 * \text{baseline}. \quad (3)$$

Equation 3 clearly shows that the baseline activation level has twice the influence on the left side of the than on the right, making the sensitivity of superadditivity reliant on the experimenter-chosen baseline. As the activation level of the chosen baseline approaches the activation level of the stimulus conditions, superadditivity becomes more liberal (see Fig. 4). There are many parameters in neuroimaging studies that can affect the difference in activation between the stimulus and baseline conditions. Thus, this property of the superadditivity criterion may explain why similar experiments from different laboratories produce different findings when that criterion is used.

Another approach to assessing neuronal convergence with BOLD fMRI is to use the method of additive factors (see Appendix 1 for an in-depth description) (Sternberg 1969b). Take, for example, an experiment in which accuracy is measured for detecting the presence of a cue, and that cue can be auditory, visual, or a combined audio-visual cue. Dependence or independence of the two sensory processes could be inferred from a comparison of the

unisensory and multisensory conditions, but that inference would be based on several assumptions about accuracy measurements. Adding an additional orthogonal factor to the experimental design, such as varying the detectability of the cues, allows for assessment of the dependence of the two processes with fewer assumptions. If the added factor alters the relationship between the two modalities, then the two processes are dependent. If there is no interaction, then they are independent. The additive-factors method can be applied to any dependent measurement, including BOLD activation (Sartori and Umiltà 2000), and provide a significant improvement over subtraction methods (Donders 1868) for assessing interactions between two processes (Townsend 1984).

The additive-factors method is based on examining *relative differences* in activation across the levels of the added factor; therefore, the results are not influenced by the numerical value of the baseline. The relative differences across two levels of an added factor for an audio-visual multisensory integration experiment would look like this when an interaction between processes was present:

$$\Delta_A + \Delta_V \neq \Delta_{AV}. \quad (4)$$

This expression can be rewritten as follows:

$$(A_1 - A_2) + (V_1 - V_2) \neq (AV_1 - AV_2), \quad (5)$$

where A_1 , A_2 , V_1 , V_2 , AV_1 , and AV_2 represent the modality-specific stimuli across two levels of the added factor. As with superadditivity, each of these BOLD measures is actually a change from a baseline, which when written explicitly, make the equation:

$$\begin{aligned} & ((A_1 - \text{baseline}) - (A_2 - \text{baseline})) \\ & + ((V_1 - \text{baseline}) - (V_2 - \text{baseline})) \neq \\ & ((AV_1 - \text{baseline}) - (AV_2 - \text{baseline})). \end{aligned} \quad (6)$$

The important point is that the baseline variables cancel out for the additive-factors design. Thus, in contrast to superadditivity, the numerical value of the difference between stimulus and baseline conditions has no effect when the method of additive factors is used.

Here, we use the additive-factors approach to assess neuronal convergence with data from four experiments. In Experiments 1 and 2, using a blocked additive-factors design, we found evidence of integration in new brain regions, including the caudate nucleus and fusiform gyrus. Findings from these experiments in conjunction with Experiments 3 and 4 demonstrate the reliability of the additive-factors design across blocked and event-related experimental designs, different stimulus types, and combinations of sensory modalities (i.e., audio-visual and visuo-haptic).

Methods and materials

Experiment 1: audio-visual blocked design

Participants

Participants included 11 right-handed subjects (6 female, mean age = 26.5). One participant was excluded from analysis due to excessive motion during scanning. All participants signed informed consent, and the study was approved by the Indiana University Institutional Review Board.

Stimuli

Two-second AV recordings of manual-tool stimuli were presented in audio-only, video-only, or AV conditions. The additive factor in this experiment was signal-to-noise ratio (SNR). Three levels of SNR were used, associated with behavioral accuracy levels of 72, 87, and 100% recognition on a two-alternative forced-choice (2AFC) task. Levels of stimulus saliency were created by varying the root mean squared contrast of the visual stimulus and root mean squared amplitude of the auditory stimulus while holding external noise constant (Stevenson et al. 2007; Stevenson and James 2009). Visual noise was Gaussian, clipped at two standard deviations. Auditory noise was the ambient acoustic noise produced by the MRI, which also approximates a clipped Gaussian distribution. Stimuli were the same 2-s AV recordings of manual tools previously used in Stevenson and James (2009).

Scanning procedures

Each imaging session included two phases: functional localizer runs and experimental runs. Functional localizers consisted of high-SNR stimuli presented in a blocked stimulus design while participants completed an identification task. Each of two localizer runs began with the presentation of a fixation cross for 12 s followed by six blocks of A, V, or AV stimuli. Each run included two blocks of each stimulus type, with blocks consisting of eight, 2-s stimulus presentations, separated by 0.1 s inter-stimuli intervals (ISI). New blocks began every 28 s separated by fixation. Runs ended with 12 s of fixation. Experimental runs were identical in design to the localizer runs, but varied in SNR levels. Each run included A, V, and AV blocks at one SNR level. Four runs were presented at high SNR, four at medium, and four at low, for a total of 12 experimental runs. Block orders were counterbalanced across runs and run orders were counterbalanced across participants.

Imaging parameters and analysis

Imaging was carried out using a Siemens Magnetom Trio 3-T whole body scanner, and collected on an eight-channel phased-array head coil. The field of view was $22 \times 22 \times 9.9$ cm, with an in plane resolution of 64×64 pixels and 33 axial slices per volume (whole brain), creating a voxel size of $3.44 \times 3.44 \times 3$ mm, re-sampled at $3 \times 3 \times 3$ mm. Images were collected using a gradient echo EPI sequence (TE = 30 ms, TR = 2,000 ms, flip angle = 70°) for BOLD imaging. High-resolution T1-weighted anatomical volumes were acquired using a Turbo-flash 3-D sequence (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 224×256 mm (voxel size = $1 \times 1 \times 1$ mm).

Imaging data were pre-processed using Brain VoyagerTM 3-D analysis tools. Anatomical volumes were transformed into a common stereotactic space (Talarach and Tournoux 1988). Functional data were aligned to the first volume of the run closest in time to the anatomical data collection. Each functional run was then 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) of the group data were calculated using Brain VoyagerTM general linear model (GLM) procedure.

Experiments 2, 3, and 4: visuo-haptic blocked design and audio-visual event-related designs

Methods for Experiment 2 were reported previously in Kim and James (2009, under review), where the resulting data from individual ROIs ($n = 7$) were submitted to standard analyses. Stimuli were three-dimensional novel objects consisting of four geometric features, sized such that they could be held in one hand and manipulated. For visual presentation, subjects viewed images of the objects projected onto a rear-projection screen. The additive factor in this experiment was also SNR, or stimulus quality. Two levels of stimulus quality were used, associated with behavioral accuracy levels of 71 and 89% recognition on a two-alternative forced-choice (2AFC) categorization task. Visual stimuli were degraded in the same way as Experiment 1, that is, external noise was added and stimulus contrast was lowered. Performance thresholds of 71 and 89% were found for individual subjects by manipulating stimulus contrast. For haptic presentation, the experimenter placed each object on an angled “table” on the subject’s midsection. The subject

then haptically explored the object. Haptic stimuli were degraded by having subjects wear gloves and by covering the objects with layers of thick felt fabric. Performance thresholds of 71 and 89% were achieved by manipulating the number of layers of felt. Although conceptually the details of the haptic degradation are different than the visual degradation, practically, both methods were successful at lowering performance by degrading stimulus quality. For visuo-haptic presentation, visual and haptic presentations were performed simultaneously with matched performance levels across sensory modality. Other methodological details were the same as detailed under Experiment 1.

Methods for Experiments 3 and 4 were reported previously in Stevenson and James (2009), where the resulting data were submitted to a standard analysis using the superadditivity metric. In short, 2-s AV recordings of manual tools (Experiment 3, $n = 11$) or speech (Experiment 4, $n = 11$) stimuli were presented in audio-only, video-only, or AV conditions. Tool stimuli were a hammer and a paper cutter, with the visual component including the actor’s hand performing a functionally appropriate action with the tool. Speech stimuli were single-word utterances, with the visual component being video of the whole face of the speaker. The additive factor in these two experiments was also SNR, defined in the same way as Experiment 1. Five levels of SNR were used, associated with behavioral accuracy levels of 55, 65, 75, 85, and 95% recognition on a two-alternative forced-choice (2AFC) task. SNR levels associated with these behavioral accuracies were found for each individual participant in a pre-imaging session utilizing an interleaved psychophysical staircase described previously in Stevenson and James (2009). Also, a noise condition was included in which noise in the absence of a stimulus was presented in both audio and visual modalities. In Experiment 4, the fixation condition was also calculated as a baseline in order to assess the effects that differential baselines may have on the superadditivity criterion. Other methodological details were the same for all four experiments and are described in detail under Experiment 1.

Results

Whole-brain analysis in Experiments 1 and 2

In order to identify regions of interaction across modality type and saliency level in Experiment 1, a conjunction of the following three contrasts was used (with H, M, and L, referring to high, medium, and low saliency, respectively):

$$(A_H - A_M) + (V_H - V_M) > (AV_H - AV_M)$$

$$(A_H - A_L) + (V_H - V_L) > (AV_H - AV_L)$$

$$(A_M - A_L) + (V_M - V_L) > (AV_M - AV_L)$$

A number of regions were identified in which the summed unisensory difference and multisensory differences were not equal. Regions with both positive ($\Delta_A + \Delta_V > \Delta_{AV}$, inverse effectiveness) or negative ($\Delta_A + \Delta_V < \Delta_{AV}$) interactions were identified (see Fig. 1; Table 1) with a minimum voxel-wise t value of 4.0 ($P < 0.003$), but with the additional statistical constraint of a cluster threshold correction of 10 voxels (270 mm³). The cluster threshold correction technique used here controls false positives, with a relative sparing of statistical power (Forman et al. 1995), which was important for studying the small effect sizes seen between our experimental conditions. Support for statistical criteria similar to these has been documented elsewhere (Thirion et al. 2007).

The null hypothesis of sensory independence predicts that the summed unisensory difference would be equal to the multisensory difference across the added factor of stimulus saliency ($\Delta_A + \Delta_V = \Delta_{AV}$), while significant differences ($\Delta_A + \Delta_V \neq \Delta_{AV}$) indicate an interaction between sensory streams. In our data, two distinct patterns of interaction were found. The first type of interaction,

$\Delta_A + \Delta_V > \Delta_{AV}$, indicates that as a stimulus becomes more degraded, the BOLD response amplitude with unimodal stimuli is reduced to a greater extent than with multisensory stimuli (given the direction of subtraction is high quality – low quality). In all regions exhibiting this interaction (see Tables 1, 2), the BOLD response (or effectiveness) with both unisensory and multisensory stimuli was directly proportional to the stimulus quality, that is, as stimulus quality increased, the BOLD response also increased. A BOLD response combining these two patterns—direct proportionality between stimulus quality and BOLD activation and greater multisensory gain with low-quality stimuli—reflects a phenomenon seen in single-unit recordings known as inverse effectiveness.

The second type of interaction, $\Delta_A + \Delta_V < \Delta_{AV}$, indicates that as a stimulus becomes more degraded, the BOLD response amplitude with unimodal stimuli is reduced to a lesser extent than with multisensory stimuli (again given the direction of subtraction is high quality – low quality). In all regions exhibiting this interaction (see Table 1), the BOLD response (or effectiveness) with both unisensory and multisensory stimuli was *indirectly* proportional to the stimulus quality, that is, as stimulus quality increased, the BOLD response *decreased*. A BOLD response combining these two patterns—*indirect* proportionality between

Fig. 1 Audio-visual regions of neuronal convergence. Results from Experiment 1 identified regions showing a positive interaction between modality and stimulus saliency ($\Delta_A + \Delta_V > \Delta_{AV}$, in orange) or a negative interaction ($\Delta_A + \Delta_V < \Delta_{AV}$, in blue) as identified by a whole-brain SPM analysis

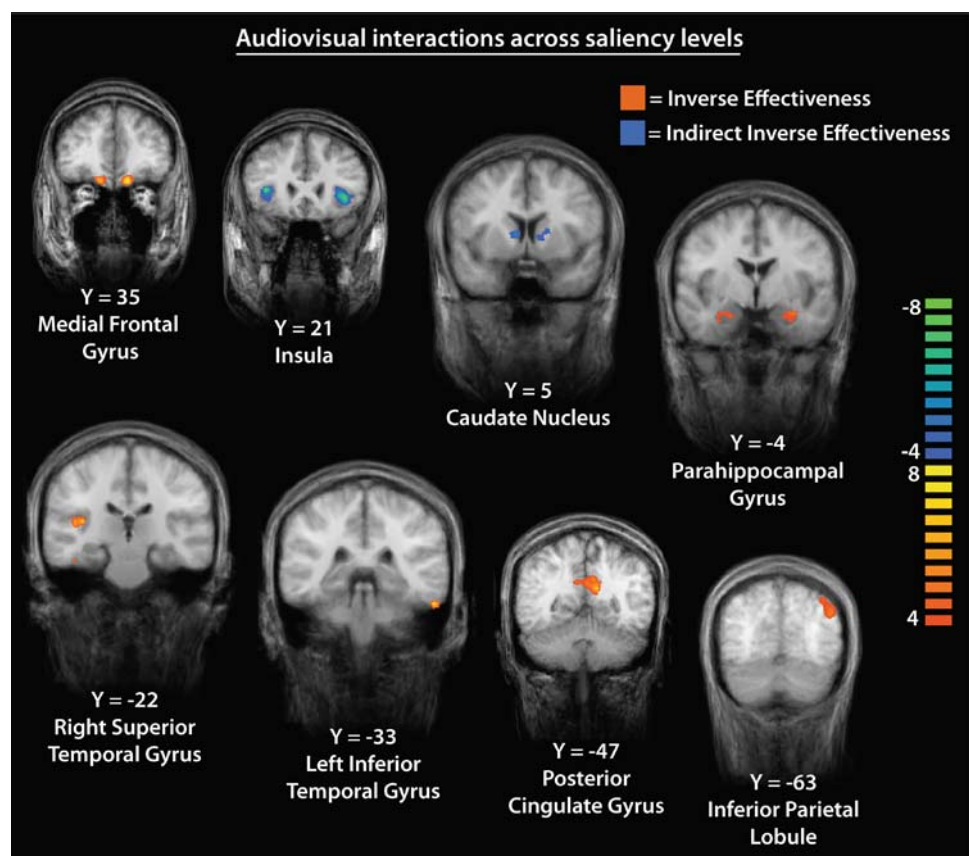


Table 1 Regions of audio-visual neuronal convergence in Experiment 1

| Region | Talaraich coordinate | <i>t</i> value | <i>P</i> value |
|--|----------------------|----------------|----------------|
| $\Delta_A + \Delta_V > \Delta_{AV}$ inverse effectiveness | | | |
| R medial frontal gyrus | 13, 36, -12 | 4.4 | <0.002 |
| L medial frontal gyrus | -11, 35, -12 | 4.7 | <0.001 |
| R parahippocampal gyrus | 34, -11, -19 | 5.1 | <0.0009 |
| L parahippocampal gyrus | -25, -7, -19 | 4.9 | <0.001 |
| R superior temporal gyrus | 41, -20, 10 | 4.7 | <0.001 |
| L inferior temporal gyrus | -55, -32, -21 | 4.3 | <0.002 |
| R posterior cingulate gyrus | 7, -48, 28 | 4.5 | <0.002 |
| L posterior cingulate gyrus | -10, -51, 25 | 6.5 | <0.0001 |
| L inferior parietal lobule | -45, -64, 33 | 4.8 | <0.001 |
| $\Delta_A + \Delta_V < \Delta_{AV}$ indirect inverse effectiveness | | | |
| R insula | 35, 19, 10 | 6.0 | <0.0003 |
| L insula | -34, 22, 4 | 6.3 | <0.0002 |
| R caudate nucleus | 10, 6, 9 | 4.0 | <0.003 |
| L caudate nucleus | -9, 5, 7 | 4.0 | <0.003 |
| R anterior cingulate cortex | 6, 13, 39 | 3.4 | <0.009 |
| L anterior cingulate cortex | -4, 13, 45 | 3.4 | <0.009 |

stimulus quality and BOLD activation and *smaller* multisensory gain with low-quality stimuli—has not been previously described in either single-unit or BOLD fMRI studies. Because BOLD activation, or effectiveness, and multisensory gain are inversely related, this interaction is similar to the previously described phenomenon of inverse effectiveness. However, this interaction differs from inverse effectiveness because stimulus quality and BOLD response are *indirectly* related. Therefore, we will call this effect *indirect* inverse effectiveness.

In Experiment 2, regions of interaction across modality type and saliency level were identified using the contrast $(V_H - V_L) + (H_H - H_L) > (VH_H - VH_L)$. Regions were identified with a minimum voxel-wise threshold of $t = 4.0$ ($P < 0.00006$) and a minimum cluster-size filter of 10 voxels (270 mm³) to control for multiple comparisons (as with Experiment 1). All of these regions showed inverse effectiveness ($\Delta_A + \Delta_V > \Delta_{AV}$, see Fig. 2; Table 2).

Experiments 3 and 4

In Experiments 3 and 4, Stevenson and James (2009) identified three regions of interest, an audio-visual region (STS) defined as a conjunction of regions that showed activation with unisensory audio and unisensory visual stimuli, a visual-only region (lateral occipital complex) defined according to greater activation with visual-only than with audio-only stimuli, and an audio-only region (secondary auditory cortex) defined according to greater

Table 2 Regions of visuo-haptic neuronal convergence in Experiment 2

| Region | Talaraich coordinate | <i>t</i> value | <i>P</i> value |
|---|----------------------|----------------|----------------|
| $\Delta_H + \Delta_V > \Delta_{HV}$ inverse effectiveness | | | |
| L medial frontal gyrus | -13, 56, 0 | 4.3 | <0.00001 |
| R fusiform gyrus | 33, -40, -18 | 4.8 | <0.00002 |
| R anterior cerebellum | | 4.2 | <0.00002 |
| L anterior cerebellum | | 4.6 | <0.00004 |

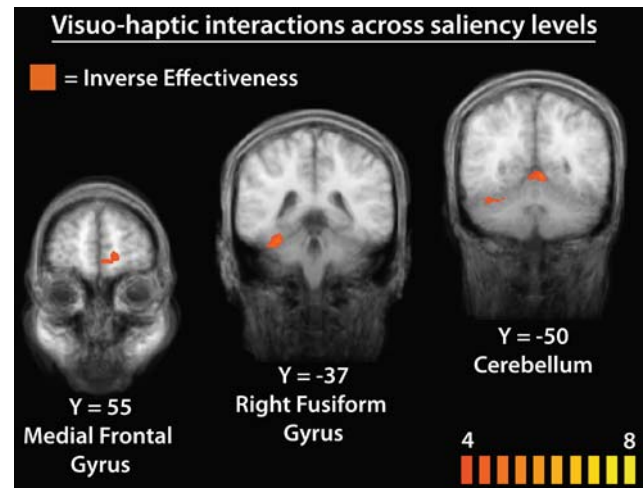


Fig. 2 Visuo-haptic regions of neuronal convergence. Results from Experiment 2 identified regions showing a positive interaction between modality and stimulus saliency ($\Delta_H + \Delta_V > \Delta_{HV}$, in orange) as identified by a whole-brain SPM analysis

activation with audio-only than with visual-only stimuli. Data from these ROIs were re-analyzed according to our new additive-factors analysis. BOLD response amplitudes were calculated for each condition in each ROI. A linear regression of SNR on BOLD response amplitudes was performed and showed a highly significant linear trend for all three stimulus types, [for Experiment 3: A ($R^2 = 0.78$), V ($R^2 = 0.92$), and AV ($R^2 = 0.95$); for Experiment 4: A ($R^2 = 0.91$), V ($R^2 = 0.92$), and AV ($R^2 = 0.92$)]. Within the AV ROIs, SNR had a strong effect on BOLD activation in the A , V and AV conditions, with high-SNR trials producing the greatest BOLD activation. Pairwise differences in BOLD activation between all neighboring rank-ordered SNR levels were calculated, and mean differences were calculated for each subject from those pairwise differences. (note, mean pairwise differences were used only because of the remarkable linear trend across SNR levels. If the trend had not been linear, then separate metrics would have been calculated for each pairwise difference, instead of collapsing). Summed mean unisensory differences ($\Delta_A + \Delta_V$)

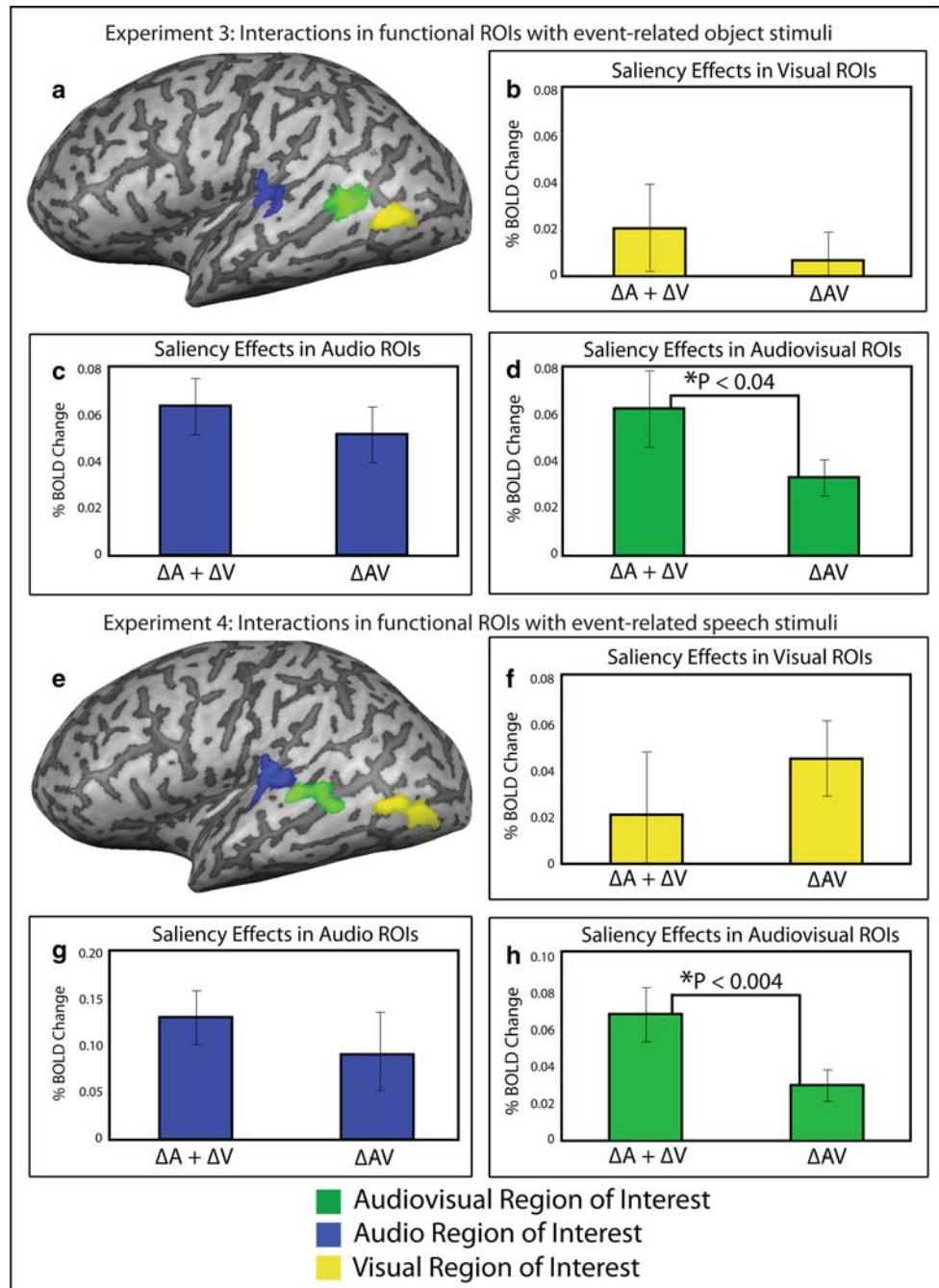
were compared with the mean multisensory difference (Δ_{AV}) in each ROI, in both experiments.

Application of this analysis to the AV STS ROIs in both Experiments 3 and 4 revealed that the multisensory difference was significantly less than the summed unisensory difference ($\Delta_A + \Delta_V > \Delta_{AV}$) (for Experiment 3: $P < 0.04$, Fig. 3d; for Experiment 4: $P < 0.004$, Fig. 3h), implying inverse effectiveness and as such, neuronal convergence.

($\Delta_A + \Delta_V = \Delta_{AV}$) (see Fig. 3c, b, f, g, for Experiments 3 and 4, respectively). This null result implies a lack of convergence in those brain regions.

In order to assess the effect of baseline on the metric of superadditivity, summed mean unisensory differences and mean multisensory differences in Experiment 4 were calculated in reference to a second baseline, a fixation condition (in addition to the analysis with the AV noise-only condition used as baseline as described above) (see Fig. 4). When fixation was used as baseline, the 65%- and

Fig. 3 Analysis of function ROIs according to an additive-factors analysis. Audio (blue), visual (yellow), and audio-visual (green) ROIs from Stevenson and James (2009) were identified in Experiment 3 with manual tools (a) and in Experiment 4 with speech (b). Responses revealed no interaction between modalities in audio (c, g) or visual (b, f) regions, implying a lack of neuronal convergence. Audio-visual regions in STS (d, h) showed a positive interaction ($\Delta_A + \Delta_V < \Delta_{AV}$) or inverse effectiveness, implying neuronal convergence



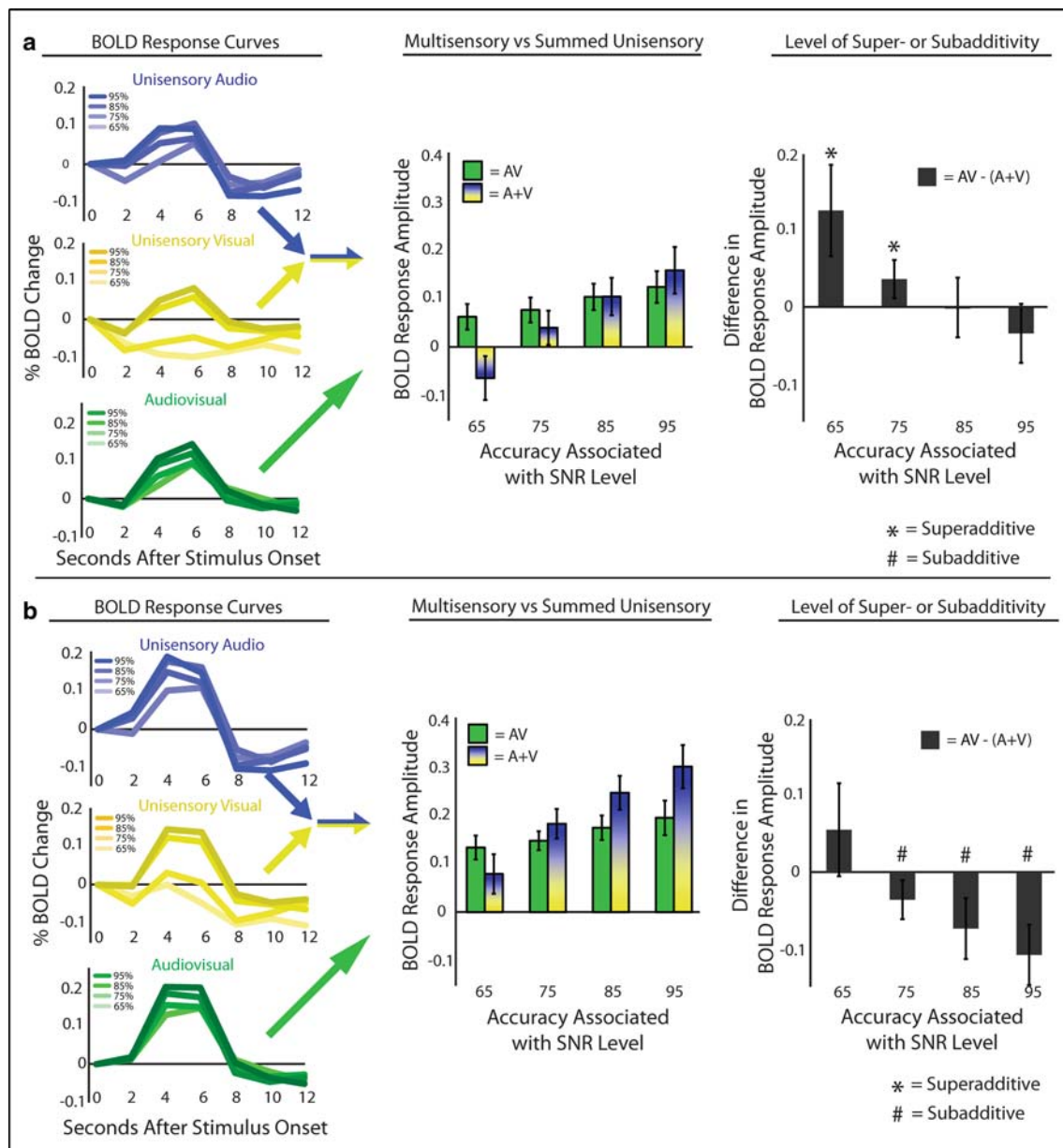


Fig. 4 The effect of experimenter-chosen baseline on findings of superadditivity. Results from Experiment 4 assessed using the superadditivity criterion with two distinct baselines, fixation (**a**) and noise (**b**). The baseline measure disproportionately affects the

summed unisensory responses, resulting in changes in the level of superadditivity depending upon which baseline is used. The higher the baseline relative to the signal being measured, the more liberal the criterion of superadditivity becomes

75%-accuracy conditions showed significant superadditivity ($P < 0.05$) and the 85%- and 95%-accuracy conditions did not significantly differ from additivity. When the AV-noise condition was used as the experimenter-chosen baseline, none of the conditions showed superadditivity. In fact, the 75–95% conditions exhibited significant *subadditivity* ($P < 0.05$). Changing the baseline had no effect on the interaction seen with our additive-factors approach, and as such results and significance values are identical to those reported above.

Discussion

The use of an additive-factors design in which unisensory and multisensory stimuli were systematically varied across levels of stimulus salience produced four key findings. First, multisensory integration of audio-visual object stimuli occurred throughout a network of brain regions, not just the established multisensory superior temporal sulcus (STS). Second, integration of visuo-haptic object stimuli occurred throughout a network of brain regions that was

distinct from (i.e., non-overlapping with) the audio-visual network. Third, as predicted, the superadditivity metric was influenced by the experimental situation, whereas the additive-factors design was invariant. Finally, generalization of the additive-factors approach was demonstrated across different stimulus types, different baseline conditions, and different experimental design protocols, without influencing its reliability.

The additive-factors design was developed by Sternberg (1969a, b) as an alternative to Donders' subtraction method (Donders 1868). By adding an orthogonal factor to an existing experimental design, the researcher can better assess the dependence (or interaction) of two processes (in our case, sensory modalities). If the added factor alters the relationship between the two processes, then the two processes are dependent. Here, we found evidence for two distinct patterns of BOLD activation across sensory modalities by our added factor, SNR. The first effect has been previously described in single-unit recordings as inverse effectiveness. As SNR decreases, BOLD activation with the three stimulus conditions decreases, but activation with the multisensory stimulus decreases much less than predicted based on the decrease in activation seen with the unisensory component stimuli ($\Delta_A + \Delta_V > \Delta_{AV}$; $\Delta_H + \Delta_V > \Delta_{HV}$). This effect was seen with both audio-visual and visuo-haptic stimulus combinations, although the brain networks that showed the effect did not overlap spatially at the statistical thresholds used.

The multisensory audio-visual brain network is shown in Fig. 1. In recent years, the superior colliculus and the STS have been investigated extensively for multisensory attributes with both single-unit recordings and fMRI, often, it seems, to the exclusion of other brain regions. However, our results, in addition to recent reviews of the literature by other groups (Doehrmann and Naumer 2008; Driver and Noesselt 2008), emphasize that integration of audio-visual stimuli involves a wide-spread network of cortical and subcortical regions. The regions found in our studies include the medial frontal gyrus (MFG) (Giard and Peronnet 1999; Calvert et al. 2001; Molholm et al. 2002; Senkowski et al. 2007), superior temporal gyrus (STG) (Kreifelts et al. 2007), inferior temporal lobe (IT) (Senkowski et al. 2007), left inferior temporal gyrus (ITG) (Dolan et al. 2001; Macaluso et al. 2004; Kreifelts et al. 2007), and inferior parietal lobule (IPL) (Lewis et al. 2000; Calvert et al. 2001; Macaluso et al. 2004; Sestieri et al. 2006; Senkowski et al. 2007).

The multisensory visuo-haptic brain network is shown in Fig. 2. With the possible exception of the MFG, this network was essentially non-overlapping with the audio-visual network (Fig. 1). And, although evidence for multisensory activation was found in MFG in both experiments, the actual coordinates of those activations were

relatively distant from each other. In addition to the MFG, evidence of visuo-haptic integration was found the fusiform gyrus (FG). The FG has been previously shown to be involved in visual face recognition (Puce et al. 1995; Kanwisher and Yovel 2006), and also haptic face recognition (Kilgour et al. 2005; James et al. 2006); however, this is the first study to suggest that inputs from the visual and haptic modalities may converge and be integrated in the FG.

The second additive-factors effect that was detected was indirect inverse effectiveness. As SNR decreased, BOLD activation with the three stimulus conditions increased. This indirect relationship between stimulus quality and BOLD response resulted in an inverted relationship between stimulus quality and multisensory gain ($\Delta_A + \Delta_V < \Delta_{AV}$). This effect was only seen with audio-visual stimulus combinations. The brain regions that showed this effect (Fig. 1) are the insula, which has long been considered multisensory, and the caudate nucleus (CN). While the CN has not been considered multisensory in humans, the rat CN has been shown to respond to stimuli presented in somatosensory, auditory, and visual sensory modalities (Chudler et al. 1995; Nagy et al. 2005, 2006), and the non-human primate CN receives direct input from ITG, middle temporal gyrus (MTG), STG, and IPL (Yeterian and Van Hoesen 1978), regions that themselves are integrative. Also exhibiting this effect at a lower statistical threshold ($P < 0.01$) was the anterior cingulate cortex (ACC). This network of brain regions has been previously shown to respond more when stimuli contain less information or conflicting information, according to the error-likelihood hypothesis (Brown and Braver 2005; Brown and Braver 2007). As predicted by this hypothesis, these regions show an *increase* in BOLD activation with a *decrease* in stimulus quality. This response pattern is the opposite of areas showing inverse effectiveness, which show an increase in BOLD response with an increase in stimulus quality. It is this response pattern, where degraded stimuli result in higher BOLD responses, that drives our finding of indirect inverse effectiveness.

Our discussion up to this point has focused on the two interactions seen in these data. It should be noted that a third type of interaction has been previously reported in the BOLD signal for visuo-haptic integration (Kim and James 2009, under review) and in single-unit recordings for audio-visual integration (Allman et al. 2008). Enhanced effectiveness is seen when stimulus quality is directly proportional to the BOLD activation (as seen with our regions exhibiting inverse effectiveness), but shows an increase in multisensory gain as stimulus quality increases (as seen in indirect inverse effectiveness). This interaction, however, was not seen in our analysis. This discrepancy arises from differences in analysis techniques in these two

reports. The previous findings of enhanced effectiveness with visuo-haptic integration were seen within the functional ROIs of individuals, defined as object-selective regions by an object–texture contrast, while our analysis was on the group level. The use of ROIs allows for a less stringent correction for multiple comparisons, resulting in increased statistical power. This discrepancy, where significant interactions are seen in ROIs of individuals but not group-averaged SPMs, is generally known to occur (Saxe et al. 2006), and has more specifically been reported in other multisensory studies (Stevenson et al. 2007; Stevenson and James 2009).

The additive-factors approach provides a much more reliable and rigorous differential assessment of areal and neuronal convergence (or integration) than the use of metrics such as superadditivity and the maximum rule. A re-analysis of the data in Experiment 4 provides an illustration of this point (Fig. 4). Experiment 4 was conducted with two different baseline conditions, fixation combined with MR acoustic noise, and visual Gaussian noise combined with MR acoustic noise. To illustrate the effect that changes in baseline activation can have on the superadditivity metric, we re-analyzed the data to explicitly compare superadditive results with these two different baselines. When the AV noise condition was used as the baseline, the 65% condition met the superadditive criterion, whereas the 75, 85, and 95% conditions did not (see Fig. 4b). When fixation was used as the baseline, however, the results changed: both the 65 and 75% conditions met the superadditive criterion (see Fig. 4a). Thus, using the superadditivity criterion, the experiment with the fixation baseline would have produced different findings than the experiment with the noise baseline. Specifically, the use of a baseline condition that produced greater activation (fixation in this case), and thus reduced the difference in activation between stimulus and baseline conditions, makes the superadditivity criterion more liberal. Because the additive-factors approach uses relative differences to assess independence, it is invariant to incidental or experimenter-chosen differences in the baseline condition.

The additive-factors approach, however, does pose limitations on experimental design. Specifically, the number of cells in the experimental design is multiplied by the number of levels of the additive factor. In other words, for an imaging session of standard length, the number of trials per condition is divided by the number of levels of the additive factor. Also, to ascertain the functional relationship between BOLD response and the additive factor, a dynamic range of factor levels should be used. In other words, if only a few levels of the additive factor are used, and those levels are confined to a narrow interval of the possible levels, the relationship between BOLD response and the effect of the additive factor may be misrepresented. In the current

studies, the relationship between SNR and BOLD response was linear, which was established in Experiments 3 and 4 by measuring BOLD responses at 5 levels, resulting in 15 total conditions. Once this linear relationship was established, the number of levels was dropped to 3 in Experiment 1 and two in Experiment 2, which increased the number of trials per stimulus condition, and hence increased statistical power to the point where, when combined with a blocked experimental design, there was enough to prepare meaningful whole-brain SPMs while still controlling for the multiple comparison problem.

The additive factor used in our experiments was a manipulation of SNR, but other manipulations should produce similar results. For instance, an early multisensory integration study by Calvert et al. (2000) reported a region in STS that responded to congruent AV speech stimuli in a manner that met the superadditive criterion. The same region also produced sub-additive activation when the multisensory combination stimulus was created from incongruent unisensory component stimuli. It is unclear why this particular study found activation in STS that met the superadditivity criterion, while other studies (Beauchamp et al. 2004; Stevenson and James 2009) have not. What is clear, however, is that if the pattern of relative difference between multisensory and unisensory conditions was different across the congruent and incongruent presentations, that by itself would imply an interaction of the sensory streams in that area, and this assessment could be made whether or not the congruent or incongruent conditions showed either superadditivity or sub-additivity. In Calvert's experiment, congruency could have acted as the additive factor. We believe that less reliance on metrics such as superadditivity and the maximum rule and more reliance on the experimental manipulation of factors such as semantic congruence, SNR, attention, spatial congruence, temporal synchrony, and perceptual learning will benefit the field of multisensory research. The additive-factors approach provides a methodological framework within which to apply and vary those factors.

In summary, we have applied an additive-factors approach to the study of multisensory integration with BOLD fMRI, and through the manipulation of our additive factor, SNR, we have identified networks of both audio-visual and visuo-haptic integrative regions that show properties of neuronal convergence. These regions not only contain previously identified multisensory regions, but also new regions, including the caudate nucleus and the fusiform gyrus. We provide evidence for its utility across sensory modality (vision, audition, and haptics), stimulus type (speech and non-speech), experimental design (blocked and event-related), method of analysis (SPM and ROI), and experimenter-chosen baseline condition. The additive-factor design provides a method for investigating

multisensory interactions that goes beyond what can be achieved using more established metric-based, subtraction-type methods.

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. Thanks to Laurel Stevenson, June Young Lee, and Karin Harman James for their support, to David Pisoni, Luiz Hernandez, and Nicholas Altieri for the speech stimuli; Andrew Butler, Hope Cantrell, and Luiz Pessoa for assistance with Experiment 1; and James Townsend and the Indiana University Neuroimaging Group for their insights on this work.

Appendix 1

The additive-factors method

Sternberg's (1969a, b, 1975) additive-factors method was originally proposed as a reaction-time paradigm that improved upon Donders' subtraction method (1868). The subtraction method was originally used to measure the time taken to perform individual 'stages' of a given task (suppose the task consists of stages a , b , and c). It supposed that the difference in time between performing a task with and without a given stage (b) was equal to the processing time of that stage:

$$RT(abc) - RT(ac) = RT(b). \quad (7)$$

The subtraction method assumes that the insertion of processing stage b does not in any way effect processing stages a or c . The additive-factors method avoids this fallacy of 'pure insertion' (among others) assumed by the Dondersian subtraction method (1868) by instead relying upon the assumption of selective influence, which supposes that if there is an experimental factor that selectively influences process a without influencing process b , then the two processes are independent. Sternberg's idea was to manipulate such an experimental factor that changed the processing time of a process, let us say process a (where a' includes a manipulated experimental factor). If two processes (a and b) are not interactive, the manipulation of the experimental factor will have an additive effect: that is, the manipulation of the experimental factor will have the same effect on a condition with processes a and b as it would on process a alone:

$$RT(ab) - RT(a) = RT(a'b) - RT(a'). \quad (8)$$

This would suggest that the processes a and b were indeed separate processes, with the experimental factor selectively influencing process a . However, if the processes are not selectively influenced, there will be an interaction between the processes:

$$RT(ab) - RT(a) \neq RT(a'b) - RT(a'). \quad (9)$$

Such interactive findings indicate a lack of selective influence of the experimental factor and thus suggest that processes a and b may not be independent.

Since its inception, the additive-factors method has been employed by a wide range of disciplines, including proposed usage in fMRI (Sartori and Umiltà 2000), and has been extended and more rigorously generalized (Taylor 1976; Schweickert 1978; Townsend and Ashby 1980; Ashby 1982; Pieters 1983; Townsend 1984; Ashby and Townsend 1986; Townsend and Thomas 1994; Wenger and Townsend 2000; Sternberg 2001). For a more in-depth overview of the additive-factor method, see "Discovering mental processing stages: the method of additive factors" (Sternberg 1998).

References

- Allman BL, Keniston LP, Meredith MA (2008) Subthreshold auditory inputs to extrastriate visual neurons are responsive to parametric changes in stimulus quality: sensory-specific versus non-specific coding. *Brain Res* 1242:95–101
- Ashby FG (1982) Testing the assumptions of exponential, additive reaction time models. *Mem Cogn* 10:125–134
- Ashby FG, Townsend JT (1986) Varieties of perceptual independence. *Psychol Rev* 93:154–179
- Beauchamp MS (2005) Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93–113
- Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Rao SM, Cox RW (1999) Conceptual processing during the conscious resting state. A functional MRI study. *J Cogn Neurosci* 11:80–95
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4221
- Brown JW, Braver TS (2005) Learned predictions of error likelihood in the anterior cingulate cortex. *Science* 307:1118–1121
- Brown JW, Braver TS (2007) Risk prediction and aversion by anterior cingulate cortex. *Cogn Affect Behav Neurosci* 7:266–277
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14:427–438
- Chudler EH, Sugiyama K, Dong WK (1995) Multisensory convergence and integration in the neostriatum and globus pallidus of the rat. *Brain Res* 674:33–45
- Dale AM, Buckner RL (1997) Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* 5:329–340
- Doehrmann O, Naumer MJ (2008) Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res* 1242:136–150
- Dolan RJ, Morris JS, de Gelder B (2001) Crossmodal binding of fear in voice and face. *Proc Natl Acad Sci USA* 98:10006–10010

- Donders FC (1868) Over de Snelheid van Psychische Processen. In: *Onderzoekingen Gedachten in het Psychologisch Laboratorium der Utrechtsche Hoogeschool*, pp 92–120
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 57:11–23
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 33:636–647
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490
- Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9:416–429
- Heeger DJ, Ress D (2002) What does fMRI tell us about neuronal activity? *Nat Rev Neurosci* 3:142–151
- James TW, Servos P, Kilgour AR, Huh E, Lederman S (2006) The influence of familiarity on brain activation during haptic exploration of 3-D facemasks. *Neurosci Lett* 397:269–273
- Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc Lond B Biol Sci* 361:2109–2128
- Kilgour AR, Kitada R, Servos P, James TW, Lederman SJ (2005) Haptic face identification activates ventral occipital and temporal areas: an fMRI study. *Brain Cogn* 59:246–257
- Kim S, James TW (2009) Enhanced Effectiveness in visuo-haptic object-selective brain regions with increasing stimulus saliency (under review)
- Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D (2007) Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage* 37:1445–1456
- Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb Cortex* 10:873–888
- 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
- Meredith MA, Wallace MT, Stein BE (1992) Visual, auditory and somatosensory convergence in output neurons of the cat superior colliculus: multisensory properties of the tecto-reticulo-spinal projection. *Exp Brain Res* 88:181–186
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14:115–128
- Nagy A, Paroczky Z, Norita M, Benedek G (2005) Multisensory responses and receptive field properties of neurons in the substantia nigra and in the caudate nucleus. *Eur J Neurosci* 22:419–424
- Nagy A, Eordeghe G, Paroczky Z, Markus Z, Benedek G (2006) Multisensory integration in the basal ganglia. *Eur J Neurosci* 24:917–924
- Pieters JPM (1983) Sternberg’s additive factor method and underlying psychological processes: some theoretical considerations. *Psychol Bull* 93:411–426
- Puce A, Allison T, Gore JC, McCarthy G (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J Neurophysiol* 74:1192–1199
- Sartori G, Umiltà C (2000) The additive factor method in brain imaging. *Brain Cogn* 42:68–71
- Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: a defense of functional localizers. *Neuroimage* 30:1088–1096 discussion 1097–1099
- Schweickert R (1978) A critical path generalization of the additive factor method: analysis of a Stroop task. *J Math Psychol* 18:105–139
- Senkowski D, Saint-Amour D, Kelly SP, Foxe JJ (2007) Multisensory processing of naturalistic objects in motion: a high-density electrical mapping and source estimation study. *Neuroimage* 36:877–888
- Sestieri C, Di Matteo R, Ferretti A, Del Gratta C, Caulo M, Tartaro A, Olivetti Belardinelli M, Romani GL (2006) “What” versus “where” in the audiovisual domain: an fMRI study. *Neuroimage* 33:672–680
- Stark CE, Squire LR (2001) When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc Natl Acad Sci USA* 98:12760–12766
- Sternberg S (1969a) The discovery of processing stages: extensions of Donders’ method. *Acta Psychol* 30:315–376
- Sternberg S (1969b) Memory-scanning: mental processes revealed by reaction-time experiments. *Am Sci* 57:421–457
- Sternberg S (1975) Memory scanning: New findings and current controversies. *Exp Psychol* 27:1–32
- Sternberg S (1998) Discovering mental processing stages: the method of additive factors. In: Scarborough D, Sternberg S (eds) *An invitation to cognitive science: vol 4, methods, models, and conceptual issues, vol 4*. MIT Press, Cambridge, pp 739–811
- Sternberg S (2001) Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychol* 106:147–246
- Stevens SS (1946) On the theory of scales of measurement. *Science* 103:677–680
- 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
- Stevenson RA, Geoghegan ML, James TW (2007) Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Exp Brain Res* 179:85–95
- Talaraich J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. Thieme Medical Publishers, New York
- Taylor DA (1976) Stage analysis of reaction time. *Psychol Bull* 83:161–191
- Thirion B, Pinel P, Meriaux S, Roche A, Dehaene S, Poline JB (2007) Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage* 35:105–120
- Townsend JT (1984) Uncovering mental processes with factorial experiments. *J Math Psychol* 28:363–400
- Townsend JT, Ashby FG (1980) Decomposing the reaction time distribution: pure insertion and selective influence revisited. *J Math Psychol* 21:93–123
- Townsend JT, Thomas RD (1994) Stochastic dependencies in parallel and serial models: effects on systems factorial interactions. *J Math Psychol* 38:1–34
- Wenger MJ, Townsend JT (2000) Basic response time tools for studying general processing capacity in attention, perception, and cognition. *J Gen Psychol* 127:67–99
- Yeterian EH, Van Hoesen GW (1978) Cortico-striate projections in the rhesus monkey: the organization of certain cortico-caudate connections. *Brain Res* 139:43–63