



Cognitive Processing Special Issue on Multisensory Processes

LETTER TO THE EDITOR

Interest and validity of the additive model in electrophysiological studies of multisensory interactions

Julien BESLE, Alexandra FORT, Marie-Helene GIARD

Mental Process and Brain activation Lab, U280 INSERM,
151 Cours Albert Thomas, 69003 Lyon, France

Over the past decades, the Stein group has provided a fundamental neural model of multisensory integration at the single neuron level in animals. They have shown in cat and monkey that when inputs from different modalities are presented in close temporal and spatial proximity, multisensory neurons in the superior colliculus (SC) can increase their firing rate to a level exceeding that predicted by summing the responses to each unimodal cue (review in Stein & Meredith, 1993).

Although this supra-additive effect applies to the single neuron, it has inspired a wider model that has been used at the integrated level of cortical populations (brain sites) in various functional brain imaging (ERP, MEG, fMRI) studies of multisensory integration. The rationale is that, *under certain conditions* that will be described below, neural activities induced by a bimodal stimulus (e.g. audiovisual, AV) should be equal to the sum of the responses generated separately by the two unisensory stimuli (e.g. auditory, A and visual, V), if the two dimensions of the stimulus were to be independently processed. Hence, any neural activity departing from the mere summation of unimodal activities should be attributed to the bimodal nature of the stimulation, that is to interactions between the inputs from the two modalities. Using this model, it is therefore possible to estimate the crossmodal interactions in the differences between the brain responses to bimodal stimuli and the algebraic sum of the unimodal responses.

$AV \text{ Interactions} = \text{Response to } (AV) - [\text{Response to } (A) + \text{Response to } (V)]$

Note that these interactions may include modulations of unimodal responses as well as new activities in sensory or polysensory areas.

This procedure, first used by Berman (1961) in event-related corticograms of cat, was later more formally expounded by Barth et al. (1995) in a study in which they identified the brain regions that responded (evoked potentials) uniquely to bimodal AV stimuli in rat cortex: "The model assumes that if subpopulations of cells that respond separately to auditory and visual stimulation do not respond uniquely to multisensory stimuli, their contribution to the [AV-ERP] will be the linear sum of their contributions to the [A-ERP] and [V-ERP] respectively. This assumption is valid for extracellular volume conducted potentials in a purely resistive extracellular media, and is based on the law of superposition of electrical fields. The sum [A-ERP] + [V-ERP] was then subtracted from the actual [AV-ERP] to obtain a difference waveform complex [AV-(A+V)]. The [AV-(A+V)] complex was used to determine cortical regions that were uniquely activated by polysensory stimulation." (Barth et al., 1995 p.179)

Although this model theoretically can be applied to any measures of human brain activity, it has been mainly used in electrophysiological data (scalp ERP and magneto-encephalography, MEG: Miniussi et al., 1998; Giard & Peronnet, 1999; Foxe et al., 2000; Raij et al., 2000; Fort et al., 2002a, 2002b; Molholm et al., 2002; Klucharev et al., 2003; Möttönen et al., 2004). On the other hand, its use has been recurrently criticized (Teder-Sälejärvi et al., 2002; Calvert & Thesen, in press) because of the multiple biases it can generate in the estimation of the crossmodal interactions if several important conditions are not fulfilled. We discuss in this note what are these biases and how to avoid or minimize

them, with particular emphasis on electromagnetic (EEG/MEG) recordings. Finally, we explain why, in spite of its strict conditions of application, the superadditive model is particularly interesting in ERP/MEG studies of multisensory integration.

Potential biases and artifacts generated by the additive model

1. The additive model is valid only when the brain responses that are analyzed do not include activity common to all conditions. Indeed these activities will be added only once but subtracted twice in the $[AV-(A+V)]$ model, which would confound the derivation of the multisensory interaction.

“Common activity” may be of several types. One type is neural responses related to late semantic processes, target processing (e.g. N2b/P3 waves in ERP/MEG recordings), response selection or motor processes. ERP literature has shown that these activities usually arise after about 200 ms post-stimulus whereas earlier latencies are characterized by sensory-specific responses (review in Hillyard et al., 1998). One way to avoid this problem is to restrict the analysis period to the early time frame (<200 ms) of stimulus processing. While this procedure is very simple in ERP/MEG recordings since their time resolution is of the order of the millisecond, sorting the response components according to their latency is still virtually impossible in haemodynamic imaging techniques.

Second, in paradigms requiring speeded responses with rapidly presented stimuli, “anticipatory” slow responses may arise before each (unimodal and bimodal) stimulus and continue for a time after stimulus onset. These anticipatory responses appear, when present in ERP/MEG recordings, as slow ramp-like deflections in the prestimulus and early post-stimulus periods. These deflections can thus give rise to spurious residual effects in the $[AV-(A+V)]$ signals that may be confused with early cross-modal interactions (Teder-Sälejärvi et al., 2002). Note that such anticipatory processes are independent of the technique used and can also be included in fMRI/PET responses. At the level of the experimental design, a procedure that may be applied whatever the neuroimaging technique to avoid or strongly reduce anticipatory processes is to present the stimuli at random interstimulus time intervals during data acquisition. In ERP/MEG signal analysis, two methods have been further proposed to control for these effects: modify the latency of the prestimulus period that will be used as the reference baseline, and/or high-pass the data (e.g. 2-Hz cut-off frequency) to remove the slow wave effects.

2. Several functional imaging studies using block-designed paradigms have reported a decrease of activation in sensory-specific cortices (e.g. the auditory cortex) when subjects were presented with continuous stimulation in another (e.g. visual) modality (Haxby et al., 1994; Kawashima et al., 1995; Laurienti et al., 2002). There are two possibilities to explain this. First these effects may reflect cross-sensory driving and/or inhibition of lower-order sensory areas via direct projections from one sensory cortex to another (Falchier et al., 2002; Rockland & Ojima, 2003; review in Schroeder et al., this issue). There is however no experimental evidence that such “crossmodal effects” in unimodal conditions may be seen at the integrated level of scalp ERP/MEG or fMRI signals irrespective of the task or stimulus delivery context. In addition, even in this case, the additive model should still apply since any difference in these processes between a unimodal and a bimodal condition should appear - if strong enough - as low-level crossmodal interactions in the model, and further represents one possible neural mechanism for multisensory integration.

A second, more likely explanation is that when a particular sensory cortex is continuously and exclusively activated during a whole block while the other nonmatching cortices are not activated, the attentional resources are dedicated to the relevant modality (even in passive or little attention-demanding tasks) whereas the other modalities are more or less voluntarily ignored (deactivated) to optimize the processing in the relevant sensory cortex (see also Ghatan et al., 1998; Kawashima et al., 1999, for similar attentional effects). In studies of multisensory integration, the $[AV-(A+V)]$ model should therefore not be used in experiments based on block-designed paradigms, since these unimodal deactivations would be subtracted from the bimodal activations, resulting in artificial increases of the “crossmodal” effects. One way to eliminate or considerably reduce such attention-related deactivations in unisensory cortices is to consider paradigms in which the stimuli are randomly and equiprobably delivered across all modality conditions (e.g. Giard & Peronnet, 1999; Calvert et al., 2000; Foxe et al., 2000; Raji et al., 2000; Calvert et al., 2001; Fort et al., 2002a, 2002b; Molholm et al., 2002; Wright et al., 2003).

3. Random mixing of conditions may however not be sufficient for a correct control of attention. Although a classical design to avoid attentional biases in the additive model is to require the same task in the three modalities, in some paradigms, the task may be easier and require less effort in one unimodal condition than in the other. This problem can be overcome by equating the levels of difficulty across unimodal conditions (by equating the behavioral performance in both unimodal conditions, e.g. Giard & Peronnet, 1999). However, in some particular cases, this may not be possible and using the same task across all the conditions can lead to noticeable spurious effects in the computation of interactions.

Consider for example speech stimuli (lip movements associated with syllable sounds) randomly presented in the three A, V and AV conditions: if a discrimination task (e.g. respond to target syllables) is required in the three modality conditions, the processing of syllables in the lip reading condition alone will include an important visual attention effect that will not be eliminated in the AV-(A+V) derivation, since in speech perception (unlike what is likely to occur for bimodal non-speech objects), normal subjects will *naturally* engage much less visual attention to process AV than V stimuli. Alternatively, if the subjects are required to respond only whenever they hear (A and AV conditions) a target syllable, their (selective) auditory attention effect will be expressed rather similarly for A and AV stimuli and eliminated in the [AV-(A+V)] model; in the same way, a lesser (if any) effect of visual attention (rather similar for V and AV stimuli) should be mostly eliminated in the model.

A general principle in dealing with attentional problems is therefore, in addition to systematically mix conditions, to equate the attentional load between each unimodal condition and the bimodal condition (but not necessarily between the two unimodal conditions).

Advantages of the additive model in ERP/MEG studies of crossmodal interactions

All the examples above show that non-biased estimation of multisensory interactions in the human cortex using the additive model requires taking important precautions both in the experimental design and in data analysis. While the constraints relative to the control of attention may be easily respected whatever the neuroimaging technique used, caveats concerning the temporal selection of the response components to be analyzed can be overcome only in EEG/MEG approaches, because of the excellent time information provided by these techniques.

In addition, the additive model has a further fundamental interest in ERP/MEG analysis of crossmodal interactions. Indeed, unlike what is observed at the voxel level in fMRI or PET signals, a significant value at a particular electrode (sensor) in ERP/MEG recordings does not mean that the structure beneath the electrode/sensor is active. Rather what is recorded at the scalp surface results from the diffusion of electrical currents inside the brain originating from *distant* “generators”, and the interpretation of the surface signals needs to take into account these volume conduction factors (using topographic analysis, generator modeling...). Interestingly, the additive [AV-(A+V)] model in ERP/MEG has the fundamental property to free oneself from the problem of overlaps of volume conduction effects in the different subcomponents of the bimodal response, by removing the conduction effects of the corresponding unimodal responses. In this respect, the additive model is not a mere application of the single-cell model used by Stein's group and other authors: it applies not only at the *local* structure level (single cell, voxel), but also at the *distant* electrode/sensor level (volume conduction effects) because it is based on the superposition principle of electrical fields, in which the potentials from separate current sources in a conductive medium sum linearly. If its conditions of application are fulfilled, the additive model will therefore isolate the (volume conduction) effects specifically related to the interactions (that will have to be analyzed in turn in terms of topography and generators).

We therefore believe that the additive model is particularly well suited for ERP/MEG study of multisensory interactions in humans, and that its multiple advantages make it worth to deal with the several constraints it imposes. Provided that its conditions of application are respected, the model can reveal the existence of genuine crossmodal interactions without making a-priori assumptions on the congruent/incongruent character of the bimodal inputs, or introducing supra-additive/sub-additive criteria for integration (e.g. Calvert, 2001; Calvert et al., 2001). Rather the additive model allows one to access the dynamics of the multisensory interactions, the possibility of observing both supra-additive and sub-additive modulations of unimodal activities in sensory-specific cortices – that appear to form a highly flexible network of crossmodal operations – as well as new processes specifically activated by the bimodal nature of the stimulus.

REFERENCES

- Barth, D. S., Goldberg, N., Brett, B., & Di, S. (1995). The spatiotemporal organization of auditory, visual and auditory-visual evoked potentials in rat cortex. *Brain Research*, *678*, 177-190.
- Berman, A. L. (1961). Interaction of cortical responses to somatic and auditory stimuli in anterior ectosylvian gyrus of cat. *Journal of Neurophysiology*, *24*, 608-620.
- Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110-1123.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*, 649-657.
- Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the bold effect. *Neuroimage*, *14*, 427-438.
- Calvert, G. A., & Thesen, T. (in press). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of physiology (Paris)*.
- Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience*, *22* (13), 5749-5759.
- Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002a). Dynamics of cortico-subcortical crossmodal operations involved in audio-visual object detection in humans. *Cerebral Cortex*, *12* (10), 1031-1039.
- Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002b). Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*, *14*, 20-30.
- Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, *10*, 77-83.
- Ghatan, P. H., Hsieh, J. C., Petersson, K. M., Stone-Elander, S., & Ingvar, M. (1998). Coexistence of attention-based facilitation and inhibition in the human cortex. *Neuroimage*, *7*, 23-29.
- Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans : A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11* (5), 473-490.
- Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: A pet-rcbf study of selective attention to faces and locations. *The Journal of Neuroscience*, *14* (11 Pt 1), 6336-6353.
- Hillyard, S. A., Teder-Sälejärvi, W. A., & Munte, T. F. (1998). Temporal dynamics of early perceptual processing. *Current Opinion in Neurobiology*, *8*, 202-210.
- Kawashima, R., Imaizumi, S., Mori, K., Okada, K., Goto, R., Kiritani, S., et al. (1999). Selective visual and auditory attention toward utterances—a pet study. *Neuroimage*, *10*, 209-215.
- Kawashima, R., O'Sullivan, B. T., & Roland, P. E. (1995). Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: Closing the "mind's eye". *Proceedings of the National Academy of Sciences of the U S A*, *92*, 5969-5972.
- Klucharev, V., Mottonen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, *18* (1), 65-75.
- Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, *14* (3), 420-429.
- Miniussi, C., Girelli, M., & Marzi, C. A. (1998). Neural site of the redundant target effect: Electrophysiological evidence. *Journal of Cognitive Neuroscience*, *10*, 216-230.
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*, *14* (1), 115-128.
- Möttönen, R., Schurmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: A magnetoencephalographic study. *Neuroscience Letters*, *363* (2), 112-115.
- Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, *28* (2), 617-625.

- Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, *50* (1-2), 19-26.
- Schroeder, C., Molholm, S., Lakatos, P., Ritter, W., & Foxe, J. J. (this issue). Human-simian correspondence in the early cortical processing of multisensory cues. *Cognitive Processing*.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: The MIT Press.
- Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (erp) recordings. *Cognitive Brain Research*, *14* (1), 106-114.
- Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, *13* (10), 1034-1043.