

Behavioral and Electrophysiological Investigation of Crossmodal Interaction in Object Recognition

Shlomit Yuval-Greenberg, Yuval Avivi, and Leon Y. Deouell
Human Cognitive Neuroscience Laboratory, Department of Psychology,
The Hebrew University of Jerusalem

Introduction

Normally we view the world as a coherent whole. However, information about real events comes in fact from multiple senses. Multisensory information converges to facilitate perception, which might be important in ambiguous situations. When conditions are not optimal (e.g. noise or multiple inputs) cross-modal conflicts might also occur. Crossmodal conflicts were studied so far mainly regarding object location and timing, but very rarely regarding object identity. Here we investigated the time course of visual-auditory interaction in recognition using Event Related Potentials (ERPs). In distinction from a previous study (Molholm et al. 2004), the study utilized object recognition rather than target detection.

Experiment 1

Subjects:

14 subjects (ages 20-29) were included.

Procedure:

A Stroop-like paradigm was used. Subjects were presented concurrently with a picture of an animal and a vocalization. Following the combined stimulus presentation, the name of an animal appeared with a question mark. In control blocks only visual or only auditory stimuli appeared.

Stimuli:

Visual - 24 pictures of 6 animals (cow, sheep, cat, dog, rooster and bird) and a neutral picture.

Auditory - 24 voices of the same 6 animals and a neutral voice.

Task:

In different blocks, to recognize the picture or recognize the voice of an animal by answering a yes/no question, and to ignore the irrelevant modality.

EEG Recording:

EEG was measured using an array of 50 scalp electrodes referenced to the nose, sampled at 250 Hz and amplified x20000.

Types of Blocks (Counterbalanced):

Auditory recognition block:

Congruent trials (60)
Incongruent trials (60)
Neutral trials (60)

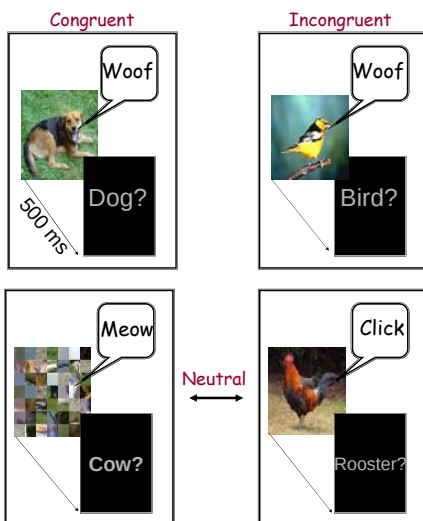
Auditory Only block (60 trials)

Visual recognition block:

Congruent trials (60)
Incongruent trials (60)
Neutral trials (60)

Visual Only block (60 trials)

Types of Audio-visual trials (balloons are for illustrative purposes only):

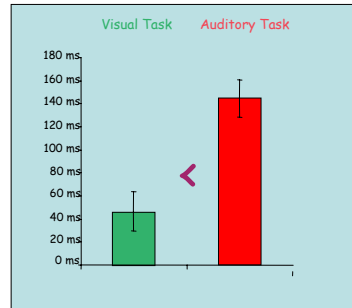


In half of the cases for each condition the correct answer was 'yes' and for the other half 'no', with quasi-random order.

Behavioral Results

The "incongruity effect" was significantly larger when the target was the sound was the target for recognition although both effects were significant.

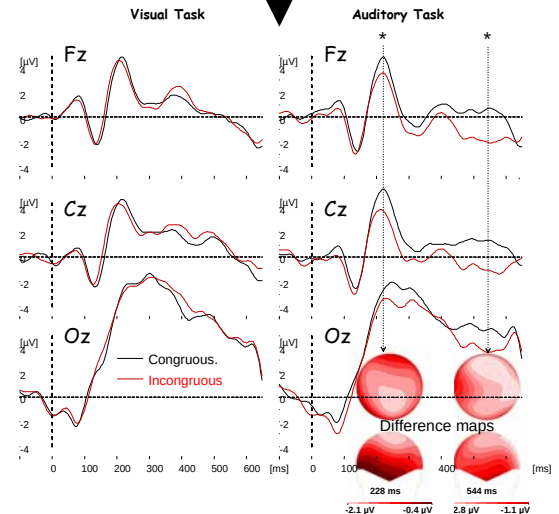
Incongruity Effect (Incongruent - Congruent)



Speed - accuracy tradeoff was ruled out. Comparison to the neutral showed both cost and benefits.

ERP Results

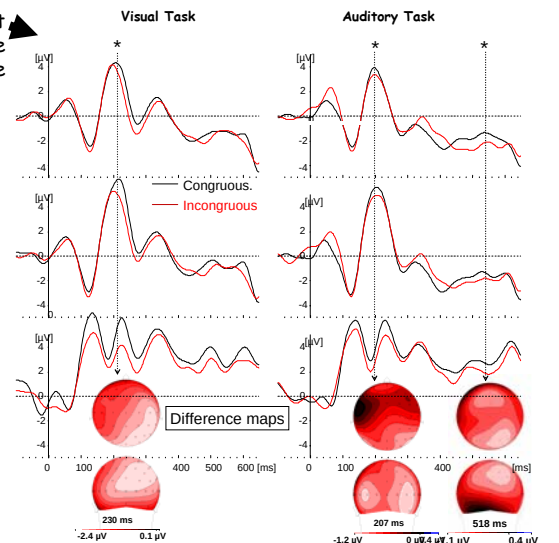
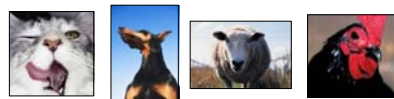
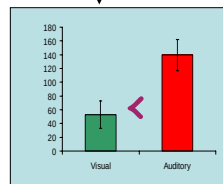
In the auditory task only, the response to incongruent presentations was more negative than to congruent conditions, with peak significant difference around the latency of P2 (~200ms), and later around 550ms (vertical lines). The maps show the distribution of incongruent - congruent difference



Experiment 2: To challenge visual recognition, we repeated the same design with a new set of pictures, of the same animals but from more unusual viewpoints. Consequently, whereas in Experiment 1 the responses to sounds alone were slower than for the pictures alone by 67 ms on the average, there was only 1 ms difference in the second experiment. Results below are from 8 consecutive subjects; 72 electrodes were used, in a new setting.

The early ERP effect was now significant for both tasks (vertical lines), but the late response was seen again only for the auditory task.

The speed of response in the auditory task was still more affected by the irrelevant modality than the response in the visual task.



Conclusions

- Sound and sight interact rather early (~200ms) even when subjects cannot prepare for a specific animal (as one could in a target detection task previously used). The latency is consistent with initial formation of object representation. The behavioral effect seems more related to the late effect.
- The cross-modal effect was sensitive to the task - in experiment 1, it was seen only in the auditory task, i.e., when the subject had to recognize sounds and ignore pictures, but not in the reverse task.
- Whereas Experiment 1 might indicate a dominance of visual over auditory information in case of conflict, Experiment 2 suggests that at least some of the effect depends on the perceptual challenge of the tasks. Early cross-modal effects do occur in both directions when the information is ambiguous. This is reminiscent of recent reports of reversed ventriloquism when visual information is degraded.

Yuvalsh@mcc.huji.ac.il

Leon.Deouell@huji.ac.il