Investigating the effect of bimodal stimuli in relation to reaction times and accuracy in audio-visual trials

Introduction

Many times a day, the human brain uses multisensory integration to analyse and process countless information from our surroundings. Almost all of this multisensory integration is done completely effortlessly and subconsciously by cooperation between the nervous system and the brain, and assists in providing information to the conscious brain to make better decisions.

This effect manifests in two ways. Firstly, through the intimate dialogue among individual neurons interacting on multiple levels in a shared semantic system [1,5], enhancing or degrading their responses [1]. Secondly, it manifests by producing alterations in performance equivalent and proportionate to the neurological response [1].

The most common ways this effect is measured and tested is through the individual use of audio, visual and tactile stimuli, as well as any combination of the stimuli. The methods and conclusions of other studies into audio-visual stimuli can be used to provide some insight into what sort of results are expected.

In the experiment of Noesselt *et al.* 2008, the presence of an audio cue served to increase the prominence and visibility of visual events [3] by helping to filter out the “background noise” of other less relevant stimuli from the target event. In the trials, participants were to indicate which of two visual stimuli were momentarily blinked off, which when compared with control experiments indicated that the auditory cue combined with the visual event “significantly enhanced subjects’ detection ability” [3]. Alternatively, it was also observed that providing an additional visual cue had the effect of hindering the participants ability when compared to a no-cue condition.

A similar effect can be seen within Chen, Y. C., & Spence, C. (2010) in where a visual event was briefly shown and then rapidly covered by an incoherent mask. Patients were then asked to identify the pictures based on what they had seen (see, figure 1). Occasionally the pictures were accompanied by one of three sounds: semantically congruent (one which made sense in relation to the image), semantically incongruent (one that made no sense in relation to the picture) and a neutral white noise. As with the previous example, when a semantically congruent sound was played simultaneously with the onset of the picture, it improved the visual identification performance of the participants, whereas an incongruent sound proved to impair the results [5].

Not only do we see this effect of increased visual recognition and reduced background noise, but we also see an increase in reaction time with participants responding “significantly more rapidly, but no less accuracy” [4] when presented with an additional auditory or otherwise stimulus. This effect isn’t limited to audio-visual stimulus as is seen in Ngo *et al.* (2012). Participants where provided with not only a visual cue, but a “simultaneously presented auditory, vibrotactile, or audiotactile cue” [4].

This is further evidenced in both Stein & Stanford (2006) [1] and Diederich, A., & Colonius, H. (2004) [2]. In the former, neural response to audio-visual stimuli was measured to be significantly larger than the sum of neural response in relation to visual and auditory stimuli individually [1]. In the latter “responses to trimodal stimulus combinations were faster than those to bimodal combinations, which in turn were faster than reactions to unimodal stimuli” [2].

Throughout studies the consistent conclusion is that multisensory integration enhances not only visual recognition capabilities [5], but also the neurological response time [1,2] and that “the enhancement did not depend on an exact temporal alignment” [3] of the stimuli, with the multisensory integration effect being observed with up to 300ms between stimuli [5]. With all this information in mind, I expect the results obtained to be in line with other findings, with the audio-visual trials having a significantly faster response time when compared to a unimodal stimuli.

Method

The experiment itself was coded and ran within PsychoPy3, comprising of 4 blocks of trials. The investigation is of true experimental research design, relying on statistical analysis to prove or disprove the hypothesis as well as having control trials with no alterations to the stimulus given. Furthermore, it is of a within-subject study design as the same people are exposed to both unimodal and bimodal trials. A trial is considered bimodal if both a visual stimulus and auditory stimulus are provided, otherwise it is considered unimodal.

Both the block order and trials within the block were randomised. The independent variables of the experiment include the sound frequency (Hz) of the auditory task, the frequency of alternating white and black lines within the visual task, and whether the trial is unimodal or bimodal. The dependant variables are the reaction time as well as the accuracy of the participants response.

As previously stated, the experiment consists of 4 blocks of trials comprised of alternating auditory trial blocks and visual trial blocks, each containing 48 trials for a total of 192 trials over the course of the experiment. For both types of trial, a fixation cross is shown in the centre of the screen for 500ms, proceeded by a random interval between 0.9, 1.1, 1.3 and 1.5 seconds. After the interval, the corresponding stimulus was then shown on screen or played for 200ms. When the stimulus is given, there is a 50% chance for there to be an accompanying stimulus, turning it into a bimodal trial (see, figure 2).

The visual task consists of showing one of two visual stimuli: a high frequency grating, with 6 alternating black and white lines, or a low frequency grating with 2 alternations. Additionally, there is a 50% chance for a sound of 840 Hz to be played simultaneously. The auditory task similarly has two stimuli, a high frequency sound of 1440Hz and a low frequency sound of 220 Hz, again having a 50% chance to play an accompanying visual stimulus with 3 alternations of black and white lines.

Upon execution of the program, the participant is asked for their name or participant number, their gender and age. They are then instructed of the tasks and the reminder to have their volume up and to a comfortable level, with headphones or earphones being recommended. Following the continuation of the program due to user input, they will begin 16 practice trials, being informed of the high and low sound/line frequency and to press “H” to indicate high frequency and “L” to indicate low frequency. After the practice trials have been completed, the program will begin to run the actual trials in the format specified above.

References

[1] Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. Nature reviews neuroscience, 9(4), 1, 255-266.

[2] Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Perception & psychophysics, 66(8), 1, 1388-1404.

[3] Noesselt, T., Bergmann, D., Hake, M., Heinze, H. J., & Fendrich, R. (2008). Sound increases the saliency of visual events. Brain research, 1220, 157-163.

[4] Ngo, M. K., Pierce, R. S., & Spence, C. (2012). Using multisensory cues to facilitate air traffic management. Human factors, 54(6), 1, 1093-1103.

[5] Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. Cognition, 114(3), 389-404.

[6] C. Cinel, (2022, January 18). Lecture notes. CE171, Neural Engineering Research Methods. University of Essex

Appendix

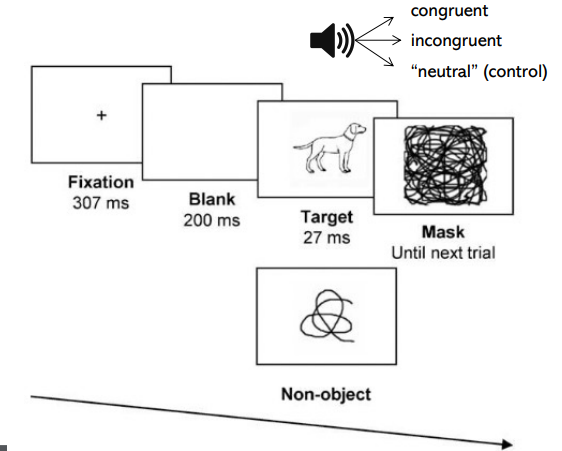


Figure 1: Chen, Y. C. & Spence, C. (2010) When hearing the bark helps identify the dog [5]

Diagram

Description automatically generated

Figure 2: C. Cinel. (2022, January 18). Lecture notes. CE171, Neural Engineering Research Methods. University of Essex [6]