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

Abstract

This investigation used a combination of unimodal (audio or visual) and bimodal (audio and visual) stimuli in trials to test the effect of multisensory integration on accuracy and reaction time, hypothesising a positive effect on both on bimodal trials. 12 participants did 192 trials each, half being unimodal the other half being bimodal. It was found that while initial processed data supported the hypothesis, the inferential statistics did not disprove the null hypothesis. However, since this goes against the majority of scientific research done around the topic, it was concluded that the fault most likely lies within this investigation.

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) [5] 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].

This study is an investigation into the role of multisensory integration on accuracy and reaction time, specifically audio-visual stimuli. In line of the results of the other studies mentioned, my hypothesis is that there is a significant decrease in reaction time and accuracy when comparing unimodal and bimodal 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.

Results

After processing the data and classifying them into different categories, the data groups that were useful to the investigation was unimodal data, bimodal data, correct unimodal data and correct bimodal data. While looking at the data however, I decided to extend my hypothesis to include the comparison of the audio task bimodal data and the visual task bimodal data to see if there was any significant difference.

As is visible in table 1 (see, Table 1), unimodal data had the highest mean accuracy with 94.79% as well as the lowest standard deviation within the accuracy being 22.22%, however also had the highest mean reaction time of 0.523s and highest standard deviation within the reaction time being 1.529, almost a full second above all other categories. I believe this is due to one or two outliers being above 10 seconds whereas most other reaction times fall within the range of 0-5 seconds.

When observing the data in a histogram (see, figures 3,4) it was clear that the data had a positive skew, meaning the data was not normally distributed. Since the data was paired, a Wilcoxon signed rank test (see, table 2) was used to compare data, with the null hypothesis being that there is no significant difference between the two sets of data. The rest of the results section will be split into sections based on the three groups of comparisons chosen: unimodal and bimodal, correct unimodal and correct bimodal, as well as audio bimodal and visual bimodal (see figures 4-7).

Unimodal vs Bimodal

As previously stated, the unimodal data has a higher accuracy (94.79%) than the bimodal data (93.23%) and a smaller standard deviation, with a difference of 3%. Contrarily, the bimodal data has a faster mean reaction time and a much lower standard deviation. When compared using the Wilcoxon test, the p values for both the accuracy comparison and the reaction time comparison both failed to reach the level of significance necessary to suggest a level of statistical difference.

Correct Unimodal vs Correct Bimodal

A comparison for accuracy is unnecessary here as the group is made up of solely accurate responses and therefore the mean accuracy and the standard deviation is 100% and 0% respectively. In terms of reaction time, the bimodal data has a faster mean reaction time as well as a almost half the standard deviation, however once again when compared using the Wilcoxon test there was no significant difference between the two.

Audio Bimodal vs Visual Bimodal

Firstly, let me clarify what each category contains; as described in the method section, the tasks were split into two blocks – visual and audio. The bimodal trials within the audio block comprise the audio bimodal category and vice versa.

In terms of accuracy, audio trials had a 2% higher mean accuracy as well as a 5% lower standard deviation (see, table 1). When comparing reaction time, it’s the visual data that has the faster of the two and the smallest standard deviation in the entire dataset. When compared using the Wilcoxon test, neither the accuracy nor reaction time comparison returned as significantly different.

Discussion

To recap, the purpose of this experiment was to investigate the relationship between unimodal and bimodal stimulus and the effectiveness of multisensory integration. The hypothesis formed from research was that bimodal stimuli would produce higher accuracy and lower reaction times when compared to unimodal stimuli.

On the contrary, the inferential statistics show that there is no significant increase or decrease in reaction time or accuracy for any of the pairwise comparisons I did (see, figure 7). This is opposing to all the results of the studies researched prior to conducting this investigation, however there are more than a few reasons why this might be.

While the data set we collected was relatively large with over 2300 trials completed in total, the actual number of participants was quite small, being only 12. With this limited pool, it becomes much harder form a generalised conclusion based on the data received as well as leaving the data more susceptible to anomalies. This could easily be remedied by having a larger number of participants.

Not only this, but the experiments were conducted wherever the participant saw fit, with little to no control over the experiment environment or participant engagement. This causes even more problems as the participant may be more prone to distractions, confusion at the instructions, or even technical difficulties by the device they use. This is another easy fix of having a controlled environment and equipment provided to participants, as well as someone nearby to advise if necessary.

Finally, another possibility for the results is by human error during the processing and analysing of the data. Due to my unfamiliarity with the Wilcoxon test, as well as various issues with differing array sizes during the testing, it is possible that the results could either be tested wrong or analysed incorrectly. This problem can be solved by peer review and/or collaborating with others when doing the study in future.

Conclusion

Research into multisensory integration in many forms has almost always resulted in seeing a positive effect on things like reaction time, decision making, and observation. While the descriptive statistics initially support the hypothesis, when using more statistically powerful testing the null hypothesis can not be disproven. Since this contradicts all other research material used for this investigation [1-6], it is likely that the flaw lies within the investigation itself, due to one of the multiple reasons listed.

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

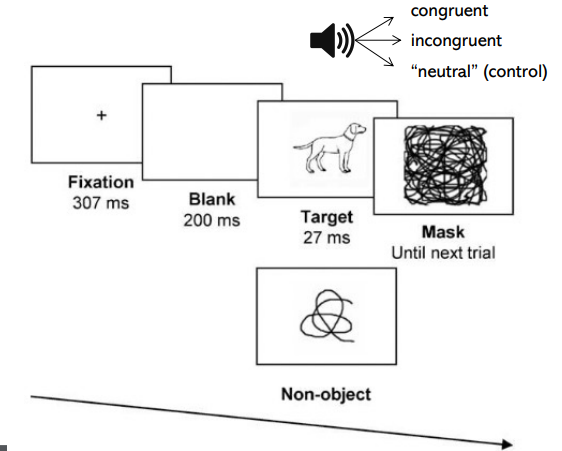


Figure 1: Chen, Y. C. & Spence, C. (2010) When hearing the bark helps identify the dog [5]

Diagram

Description automatically generated

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

Chart, histogram

Description automatically generated

Figure 3: Histogram of unimodal data

Chart, histogram

Description automatically generated

Figure 4: Histogram of bimodal data

Table

Description automatically generated

Table 1: Descriptive statistics for data used

Table

Description automatically generated

Table 2: Wilcoxon signed rank p values for pairwise comparisons

Chart, bar chart

Description automatically generated

Figure 5: Bar chart of accuracy statistics

Chart, bar chart

Description automatically generated

Figure 6: Bar chart of reaction time statistics

Chart, bar chart

Description automatically generated

Figure 7: Bar chart of Wilcoxon p values for the pairwise comparisons in relation to the 0.05 p value

A picture containing text

Description automatically generatedText

Description automatically generatedText

Description automatically generated

*Figure 8: Code used to process data*