Do sounds affect how you see?

Oscar Solis[1], with supervision from Cameron Kyle-Davidson[2] & Dr Karla K. Evans[2]
[1]School of Natural Sciences, University of York, [2]Department of Psychology, University of York

RATIONALE

Semantically congruent sounds speed up visual search for related object targets in arrays [1] and videoclips of naturalistic scenes [2].

Interactions between visual and auditory inputs could occur via:

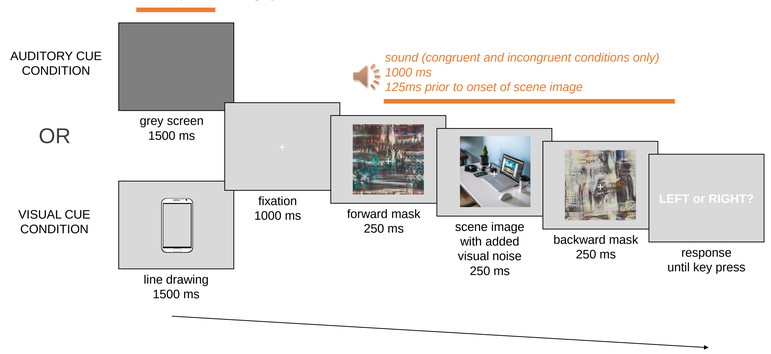
- direct route – sounds directly prime a visual representation [3] or
- indirect route – through an intermediate semantic representation [4].

QUESTIONS

1. Does a semantically congruent sound to a visual target object guide visual search in real-world static scenes?
2. Is an audio-visual cross-modal semantic association the result of a direct link between the modalities or via intermediate a semantic representation?
3. What can a neural network model tell us about cross-modal interactions at the semantic level?

METHOD – BEHAVIOURAL STUDY

- 30 participants (14 males, 16 females, aged 19 to 38).
- 240 real-world scenes, 16 target objects and 21 real-world sounds.
- 2 cue conditions (auditory, visual) for two experimental groups.
- 3 experimental conditions (congruent, incongruent, absent) across 240 trials.



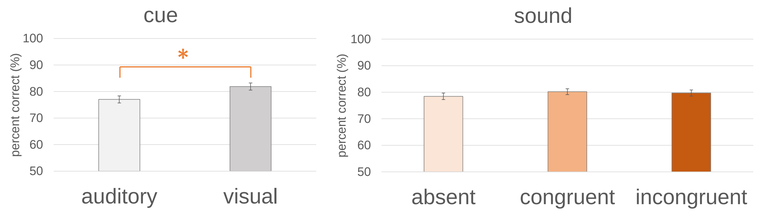
CONCLUSIONS

- Semantically congruent sounds enhance the salience of a visual object, effectively guiding attention during visual search.
- Sounds may activate a visual representation indirectly through an intermediate semantic representation.
- Semantic congruency may only facilitate visual search when audio-visual stimulus is task-relevant [1, 2].
- Semantic audio-visual interactions are the result of learned associations from repeated experiences of semantically congruent audio-visual stimuli.

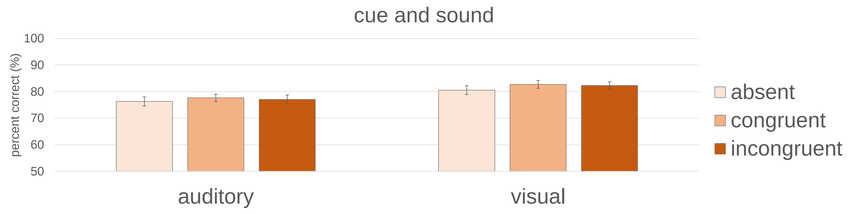
RESULTS – BEHAVIOURAL STUDY

Accuracy:

- Significant main effect of cue condition, F(1, 28) = 7.150, p = .012, η_p^2 = .203.
- No significant effect of sound condition, F(2, 56) = 1.472, p = .238, η_p^2 = .050.

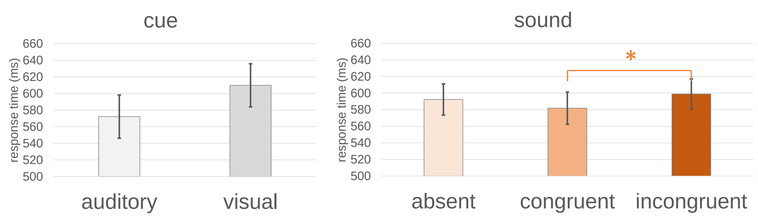


- No significant interaction effect between cue and sound conditions, F(2, 56) = .130, p = .878, η_p^2 = .005.

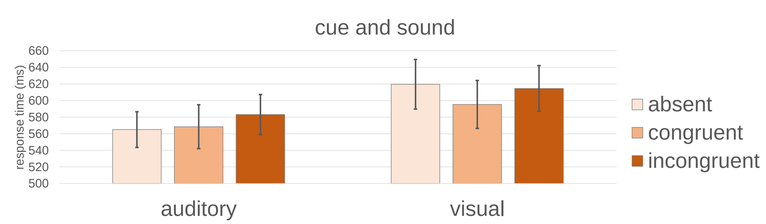


Response Times (RTs):

- No significant main effect of cue, F(1, 28) = 1.046, p = .315, η_p^2 = .036.
- Significant effect of sound, F(2, 56) = 3.346, p = .042, η_p^2 = .107.
- Planned contrasts revealed that RTs for congruent sounds were not significantly faster than for absent sound, F(1, 28) = 2.965, p = .096, η_p^2 = 0.96, but congruent sounds were significantly faster than for incongruent sounds, F(1, 28) = 7.069, p = .013, η_p^2 = .202.

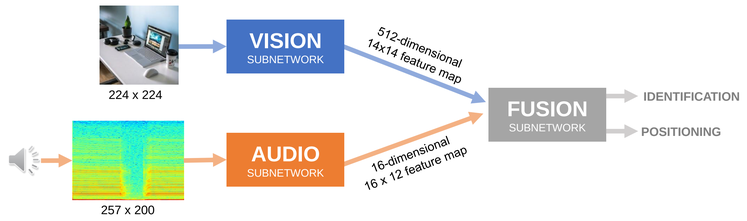


- Interaction effect between cue and sound on RTs for correct responses was close to statistical significance, F(2, 56) = 2.537, p = .088, η_p^2 = .083.
- However, pattern of results indicate a different effect of audio-visual interaction on search depending on the target cue modality.



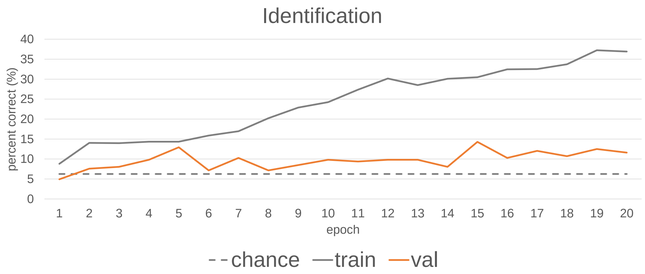
METHOD – COMPUTATIONAL MODEL

- Stimuli from behavioural study (256 images, 21 sounds) were used.
- Neural network inspired by AVOL-Net [5].

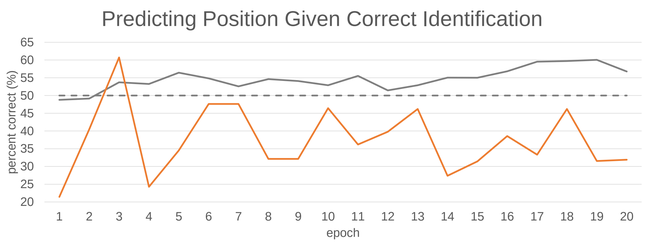


RESULTS – COMPUTATIONAL MODEL

- Model is unable to generalise to the validation set.
- Accuracy for identification during validation does not increase in the same way as during training indicating the network memorises but does not learn.



- Accuracy for predicting position of correctly identified visual targets is at chance.



REFERENCES

1. Iordanescu L, Guzman-Martinez E, Grabowecky M, Suzuki S. Characteristic sounds facilitate visual search. Psychon Bull Rev. 2008;15(3).
2. Kvasova D, Garcia-Vernet L, Soto-Faraco S. Characteristic Sounds Facilitate Object Search in Real-Life Scenes. Front Psychol. 2019;10.
3. Vallet GT, Riou B, Versace R, Simard M. The Sensory-dependent nature of audio-visual interactions for semantic knowledge. In: Proceedings of the 33rd Annual Conference of the Cognitive Science Society. 2011.
4. Brandman T, Avancini C, Leticevscaia O, Peelen M v. Auditory and Semantic Cues Facilitate Decoding of Visual Object Category in MEG. Cerebral Cortex. 2020;30(2).
5. Arandjelović R, Zisserman A. Objects that Sound. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2018.

Many thanks to the Laidlaw foundations for funding this project and supporting undergraduate research.
Address correspondence to: Oscar Solis, email address: jots500@york.ac.uk