Audiovisual Semantic Relatedness of Real World Objects

Kira Wegner-Clemens[1], George L. Malcolm[2], Sarah Shomstein[1]

Department of Psychological & Brain Sciences, George Washington University; School of Psychology, University of East Anglia

How does semantics shape AV attention?

Semantics is crucial to daily life and influence various cognitive processes, including attention and memory, however it is difficult to quantify, especially across senses. Existing quantification methods include:

- Shared-category (e.g., fruits, kitchen items, animate) [1,2]
- Shared-source (e.g., a meow & a cat, same voice) [3,4]
- Distributional semantic models based on text corpora[5]

Humans have continuous understandings of semantic relatedness (e.g., rather than being related or not, items can have varying degrees of relateness) that continuous methods do not adaquately capture.

We created a database of continous audivisual relatedness values to investigate how continuous shapes audiovisual attention.

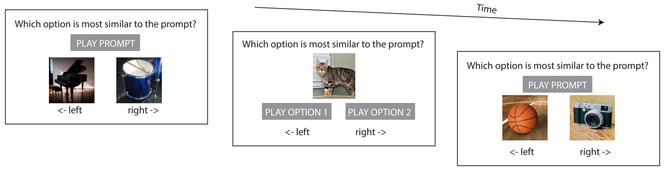
Shared-source sounds have been shown to speed search for visual targets (e.g., a meow speeds search for a cat), but only with this study has the role of more distant semantic relationships been investigated (e.g., will a bark speed search for a cat because cats and dogs are household animals?)

Methods

Participants: George Washington University undergraduates & mTurk workers

Stimuli: 30 items selected for recognizability in both sound & image through pilot testing. Images taken from THINGS database[6]. Sounds from various sources, trimmed to 1 sec, & normalized for loudness

Semantic judgement task: Participants (n=140) made AV similarity judgements. Each sound & image appeared as prompt & option. 20 trials for each stimulus trio.



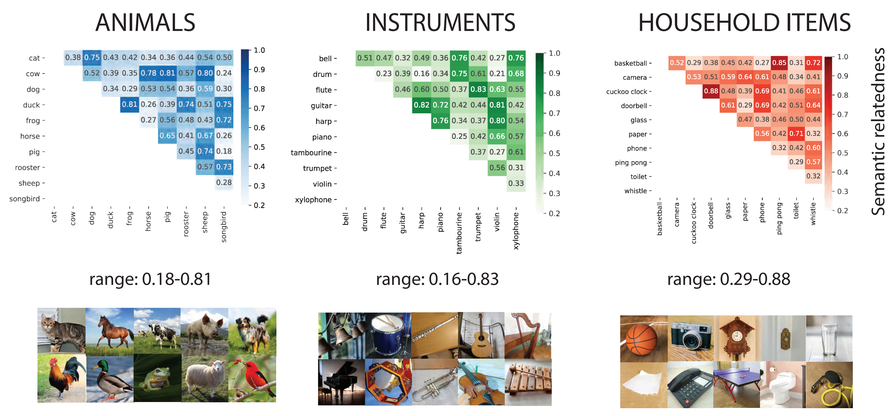
Calculating semantic relatedness values

1. Collect judgements for trios of stimuli (eg., harp, piano, drum)
2. For a pair, calculate likelihood of selecting piano if harp is the prompt, independent of second option
3. Average over trial, participant, prompt modality, prompt direction

x 20 repeats of each trio

Audiovisual search task: Participants (n=123) searched for a cued visual target while a sound in the same category played. Sounds ranged in semantic relatedness. Each visual image was from a different category.

Sight-Sound Semantics Database

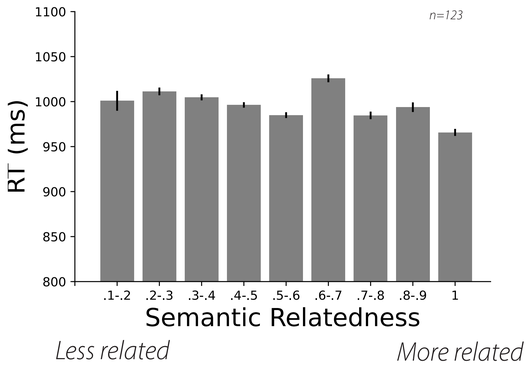


range: 0.18-0.81 range: 0.16-0.83 range: 0.29-0.88

Within every category, we observed a wide range of values, which allows for effective study of differences and suggests that participants have a shared representation of semantic similarity (e.g., a shared agreement of which items are more closely related).

A related sound helps you find a visual target

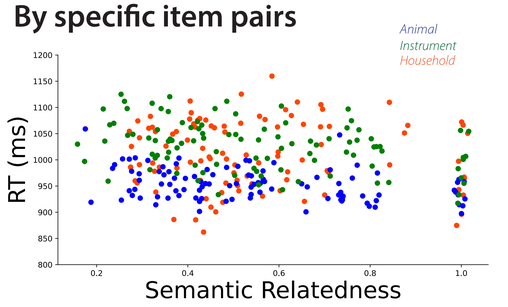
By binned semantic relatedness value



As semantic relatedness between the sound and visual target increases, search speeds decrease

1 way within subject rmANOVA: $p<0.0001$

By specific item pairs



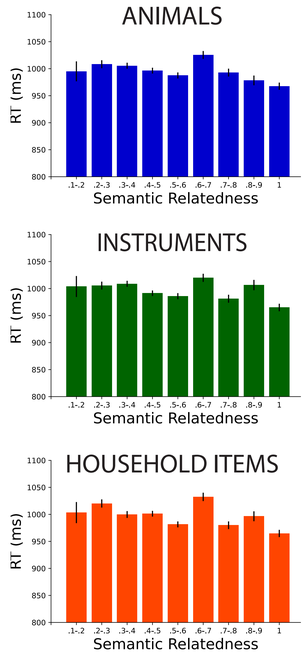
To investigate the effect of item, category, and participant, we conducted a linear mixed effect model

- Fixed: Semantic relatedness
- Random: Participant, category

RT ~ relatedness + 1|participant + 1|stimulus

Relatedness significantly predicted RT ($p<0.0001$)

Is the search benefit category specific?



In a trial, sounds and target images were always from the same category so the effect could have largely been driven by a subset of trials

However, the same pattern is observed in each category, with the fastest visual search times for the more closely related sounds (1 = exact match)

1 way within subject rmANOVA
- Animals: $p<0.0001$
- Instruments: $p<0.0001$
- Household: $p<0.0001$

In a second LME with category as a fixed factor and subject as a random factor, category did not predict RT.

Conclusion

Our database allows for a continuous range of semantic similarity values between images and sounds, based on judgements from human observers. This is the first audiovisual semantics database made available for research. By making this database publicly available on OSF, we hope it will be broadly useful to researchers studying semantics in audiovisual contexts.

With this database, we were able to investigate contiuous differences in how semantic relationships between sounds and visual targets moduates search efficiency.

Future work will investigate various remaining questions:

- Do the images need to be task relevant?
- Is there a "near match" repulsion effect?
- What are the underlying neural mechanism?

Through this work, we will produce a more robust understanding of the role of semantics in audiovisual attention.

References

(1) Moores, et al 2003 (2) Malcolm, et al 2016 (3) Iordanescu, et al 2008 (4) Kvasova, et al 2019 (5) Bhatia, et al 2019 (6) Hebart, et al 2019 (7) Mikolov, et al 2017

How does semantics shape AV atten

How does semantics shape AV atten