A Novel Generalization between Verbal Judgments and Perceptual Discrimination of 3D Space

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Prachi Mahableshwarkar, Lindsay Houck, John Philbeck, and Dwight Kravitz
Department of Psychological and Brain Sciences, The George Washington University

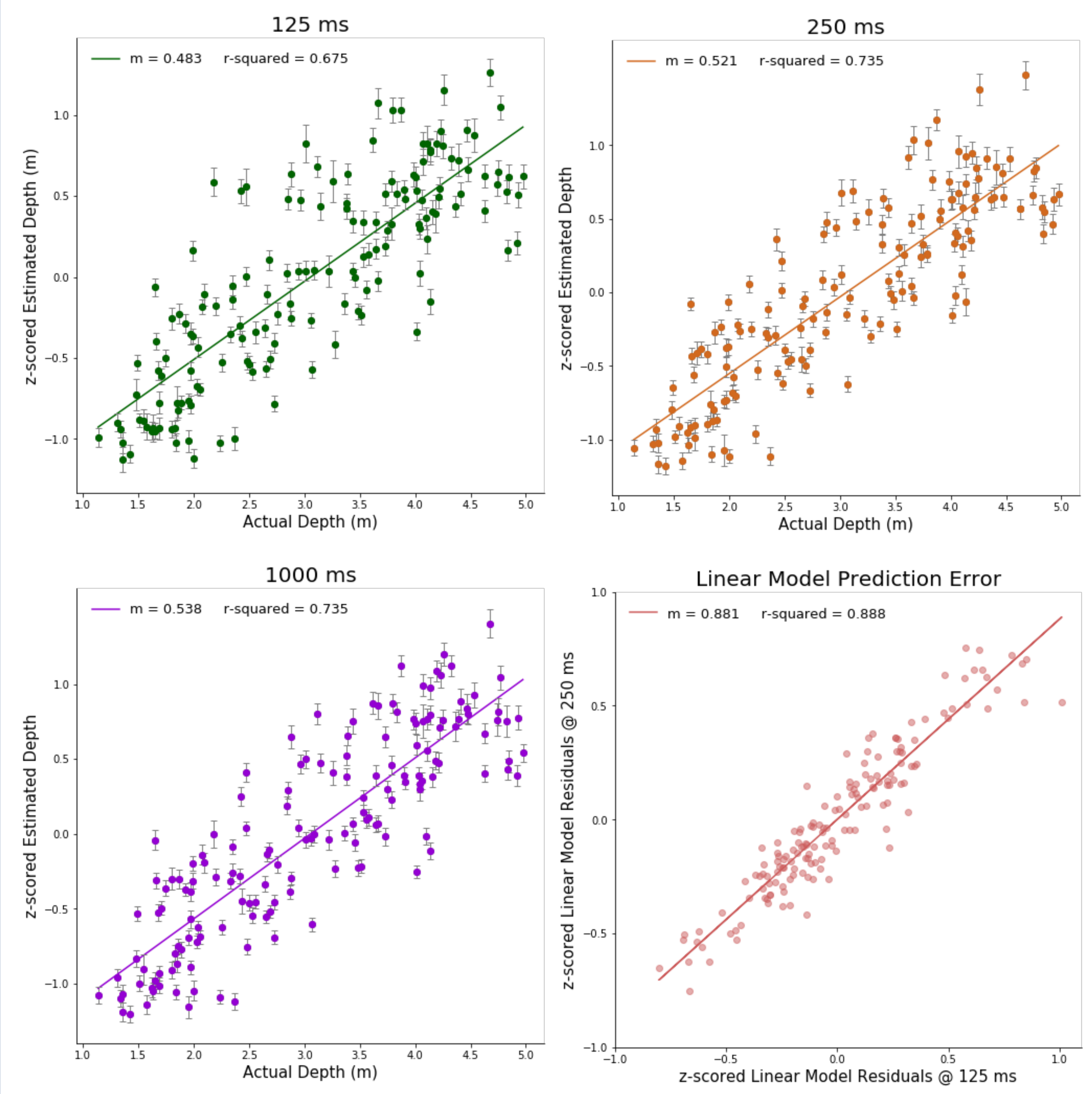
Introduction

Despite the lack of oculomotor and binocular disparity cues, humans are capable of visual distance perception in pictured scenes. Here we leverage crowd-sourced data collection to model the influence of spatial cues with temporal processing differences.

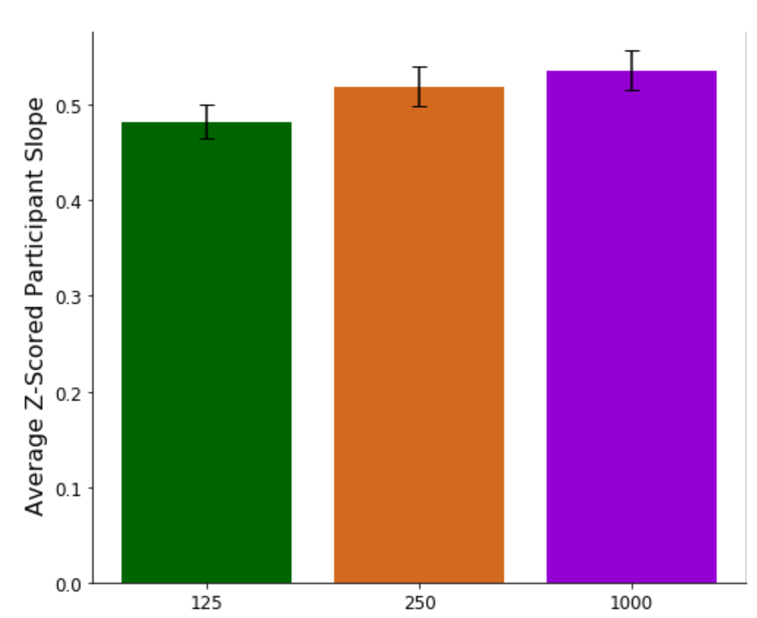
Stimuli

SUN-RGBD Sample Image [2]

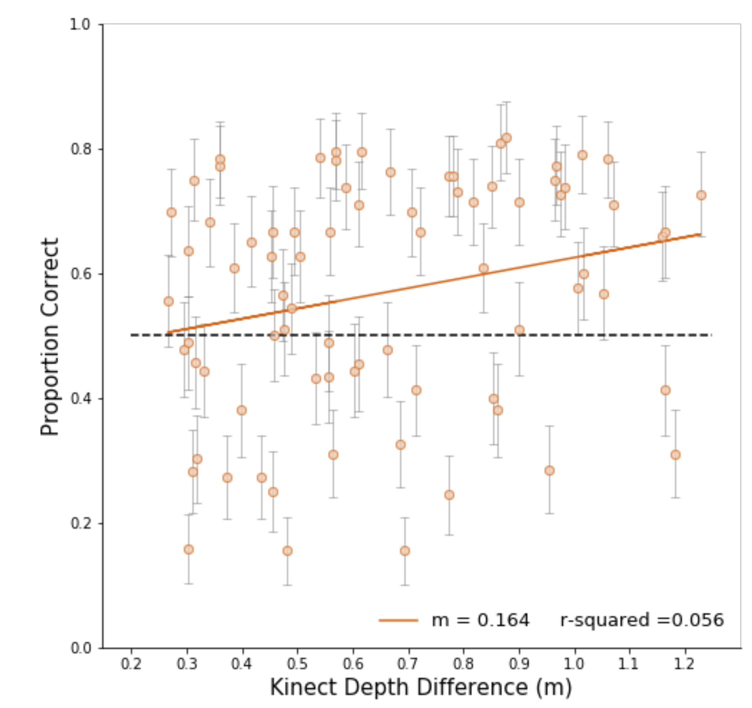
Verbal Estimate Results



The residuals for the linear model (Z-scored estimated depth x actual depth) at 125 ms and 250 ms for example are highly related, suggesting stable and **systematic biases** for individual stimuli across viewing duration even at brief glimpses.

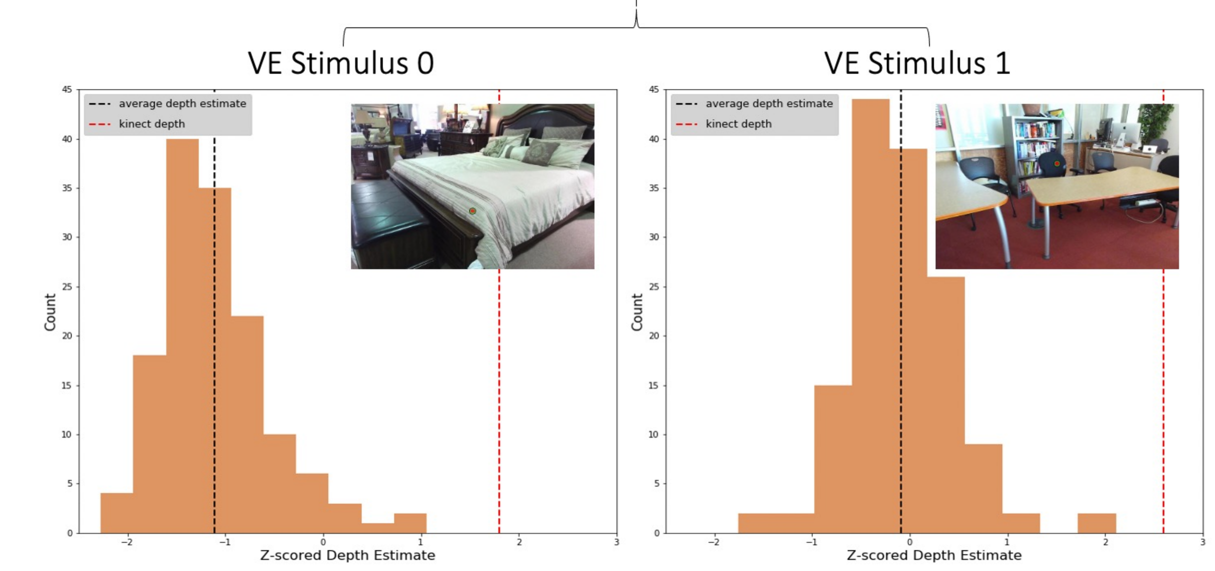


Discrimination Results

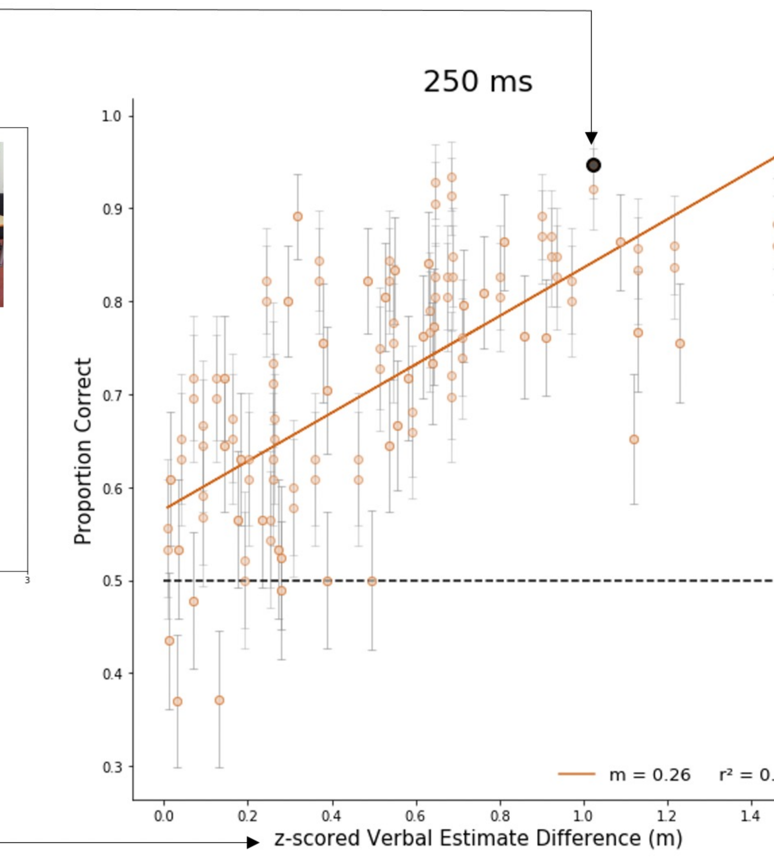


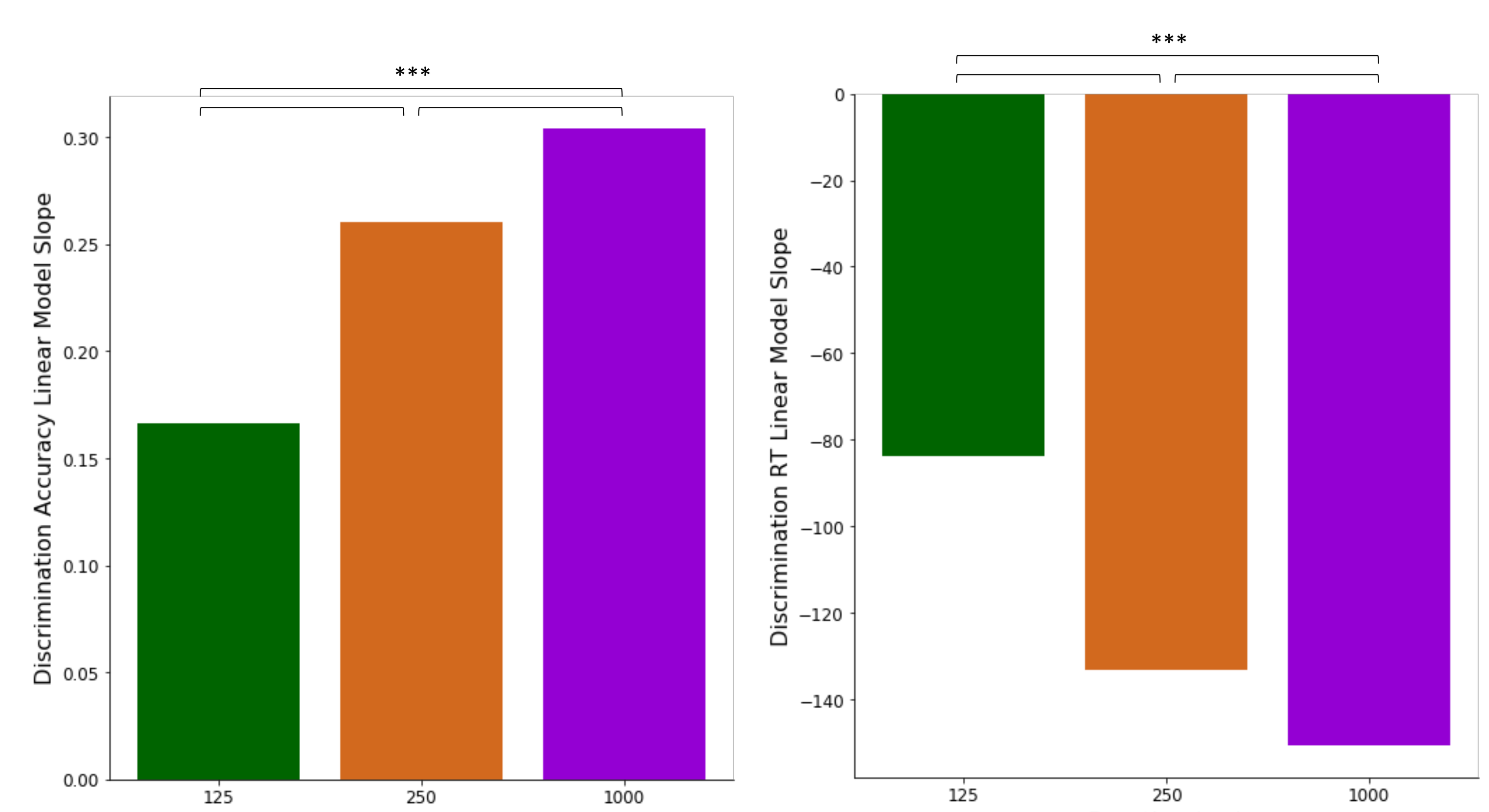
Discrimination performance (accuracy and RT) was more strongly predicted by **signal detection analyses** applied to the verbal estimates for the corresponding stimuli in each discrimination (R2 = .57, p < 10^-14) than Kinect tagged depths.

How predictable is discrimination performance from independent verbal judgments (Exp 1)?



$$\text{Z-scored Verbal Estimate Difference} = \left|\frac{1}{n}\sum_{i=1}^{n} s0_i - \frac{1}{n}\sum_{i=1}^{n} s1_i\right|$$

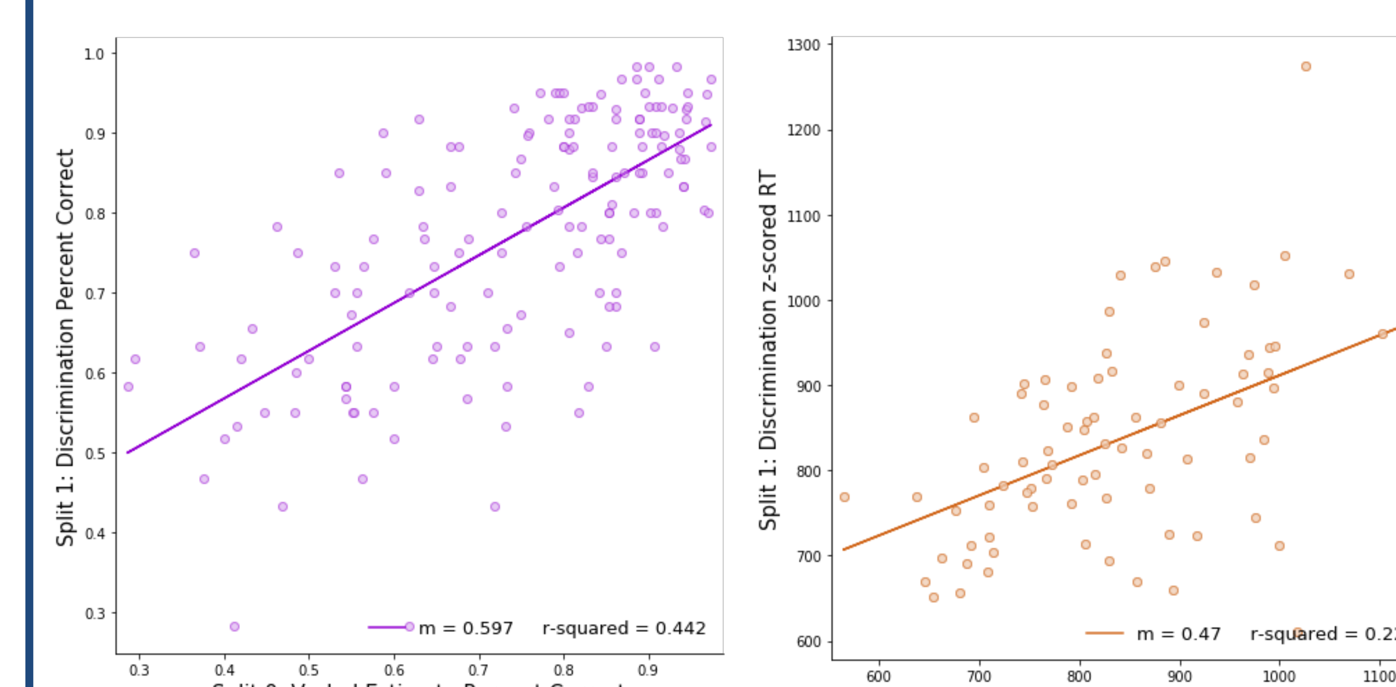


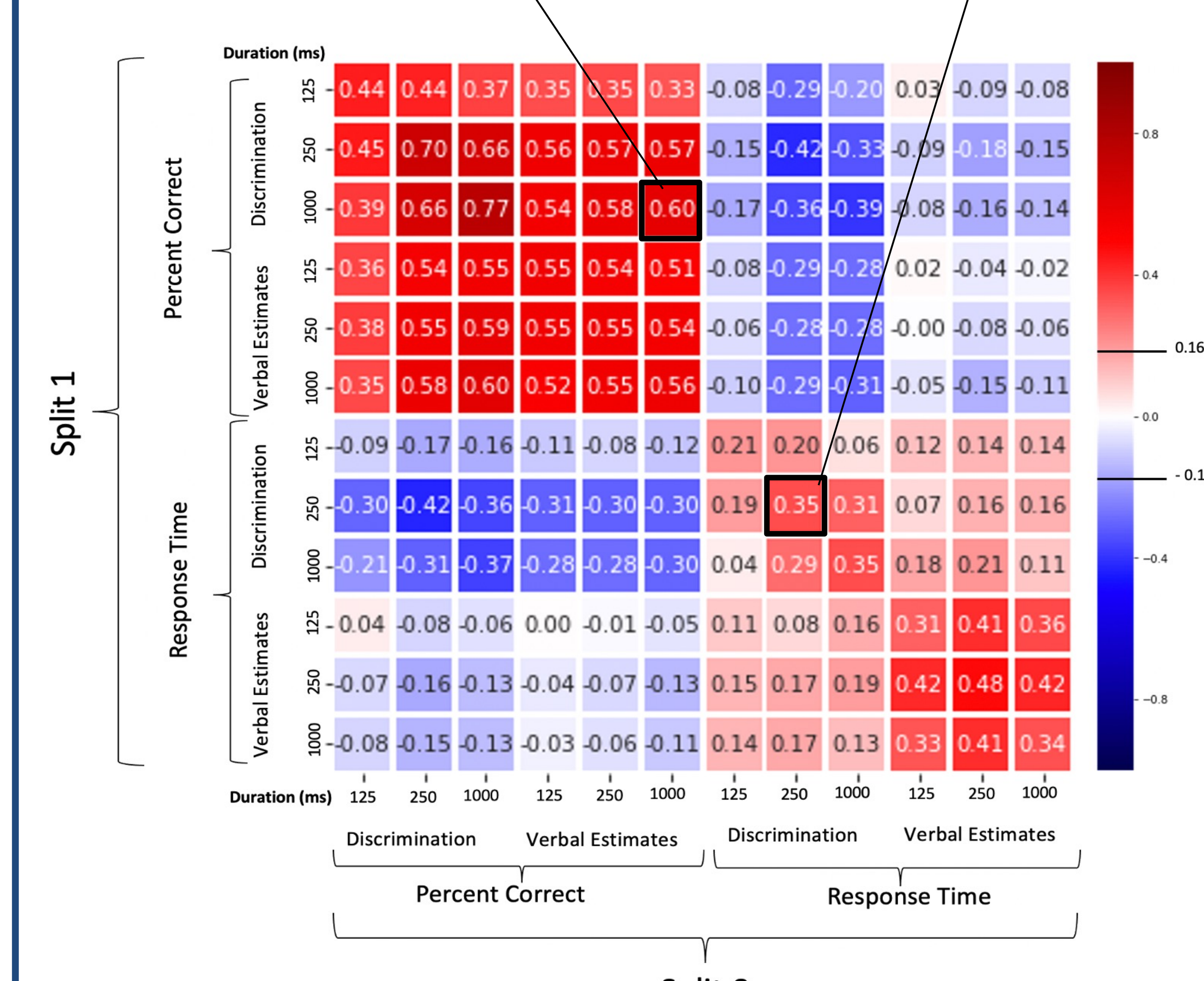


Discrimination accuracy, predicted by independent verbal judgments (VE), showed increasing sensitivity to distance with longer viewing durations (F = X, p < X). Discrimination response times show a significant negative linear relationship with VE depth differences, wherein larger differences result in quicker response times. This relationship strengthens as viewing duration increases.

Stimulus Driven Task Generalization

We found high reliability in the relationships between tasks, participants, and viewing duration in the complete dataset (top) and after repeatedly splitting the data into independent subsets and evaluating the average correlation matrix (bottom).



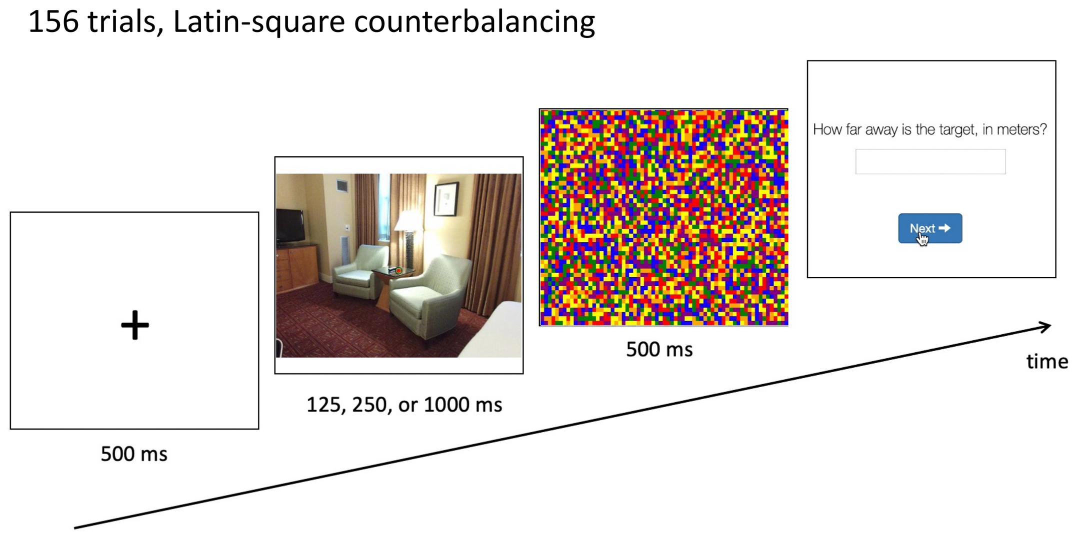


Experimental Paradigms

OSF PREREGISTRATION

Verbal Judgment (VE) of Distance

Amazon Mechanical Turk, 156 subjects
156 trials, Latin-square counterbalancing

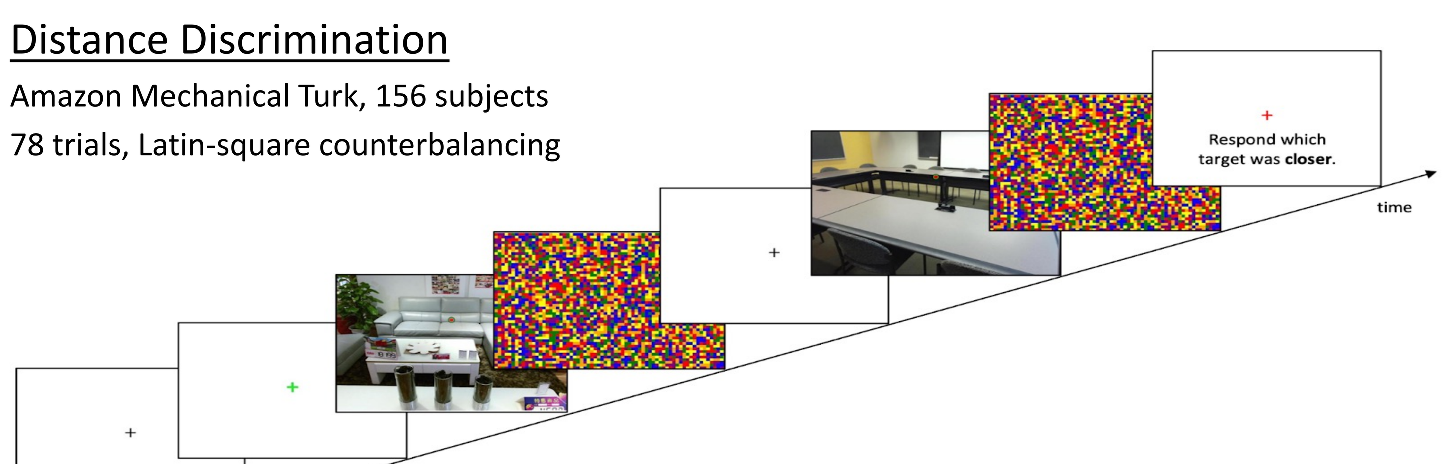


Independent groups of participants

1. VE @ 125 ms
2. VE @ 250 ms
3. VE @ 1000 ms
4. Discrimination @ 125 ms
5. Discrimination @ 250 ms
6. Discrimination @ 1000 ms

Distance Discrimination

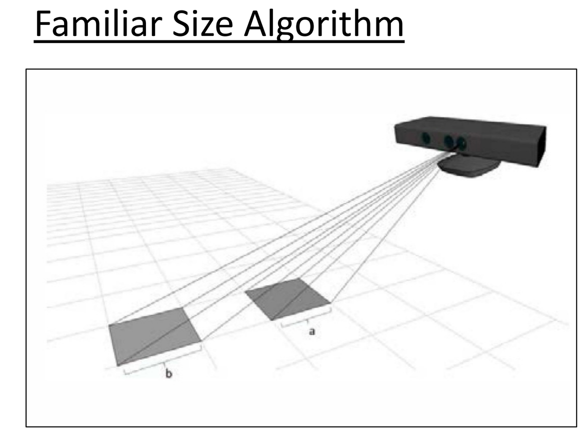
Amazon Mechanical Turk, 156 subjects
78 trials, Latin-square counterbalancing



Conclusion

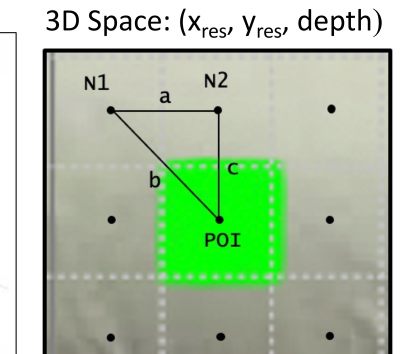
The generalization across tasks and participants suggests a stable and strong relationship between verbal report and discrimination, allowing for the creation of an explicit model that relates stimulus-based visual cues and 3D picture perception. Future work will quantify and model the influence of various visual cues, such as familiar size, on behavioral performance.

Familiar Size Algorithm



Calculating real world object dimensions from Kinect RGB-D image using dynamic resolution. [1]

3D Space: (x_{res}, y_{res}, depth)



$s = (a + b + c)/2$

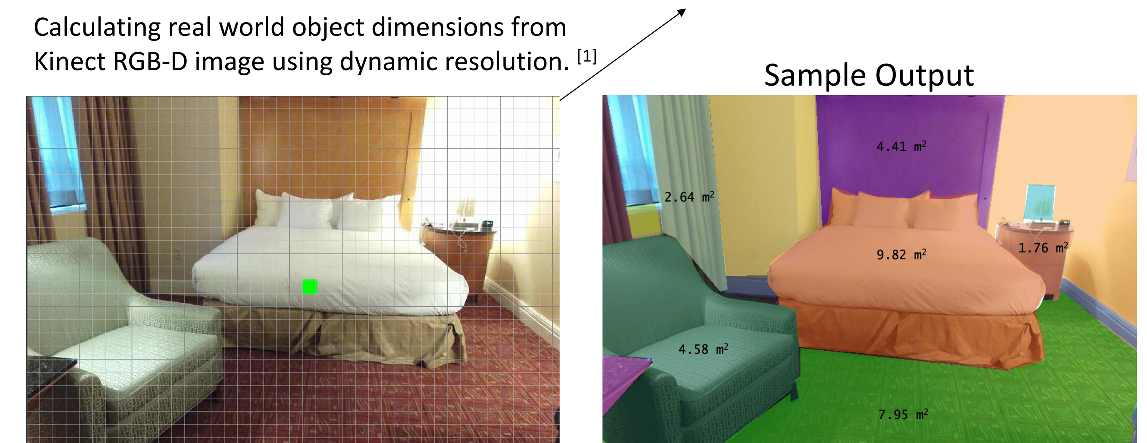
Pixel Surface Area

$SA = \sqrt{s(s-a)(s-b)(s-c)}$

Object Surface Area

$\sum_{i}^{n} SA_i = SA_1 + \ldots + SA_n$

Sample Output



References

[1] Anwer, A., Baig, A., & Nawaz, R. (2015). Calculating Real World Object Dimensions from Kinect RGB-D image using dynamic resolution. 2015 12th International Bhurban Conference on Applied Sciences and Technology (IBCAST).

[2] Song, S., Lichtenberg, S. P., & Xiao, J. (2015). Sun RGB-D: A RGB-D scene understanding benchmark suite. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Contact: pmahable@gwmail.gwu.edu