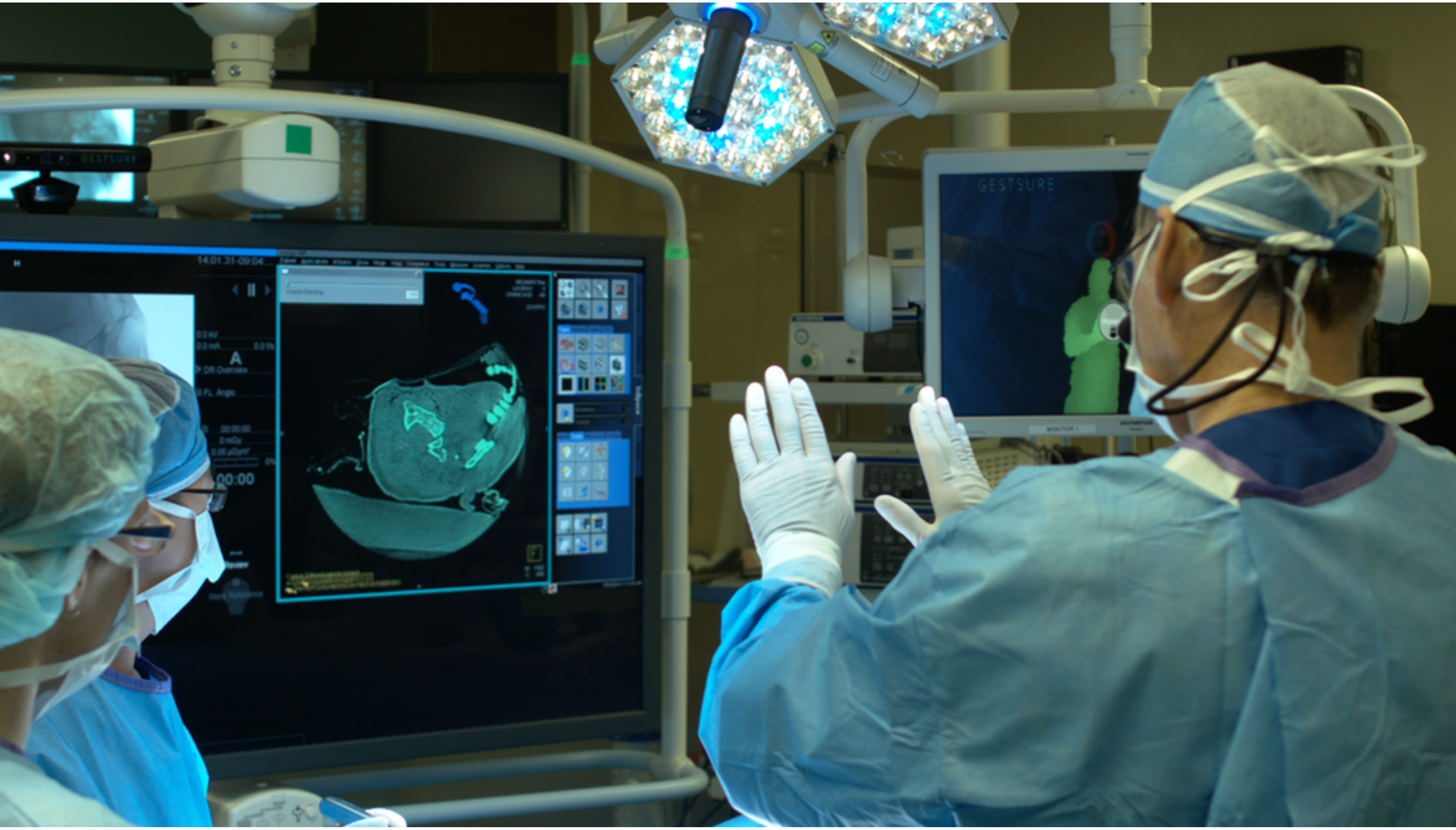
Combining Body Pose, Gaze and Gesture to Determine Intention to Interaction in Vision-Based Interfaces

Julia Schwarz, Charles Marais, Tommer Leyvand,
Scott E. Hudson, Jennifer Mankoff

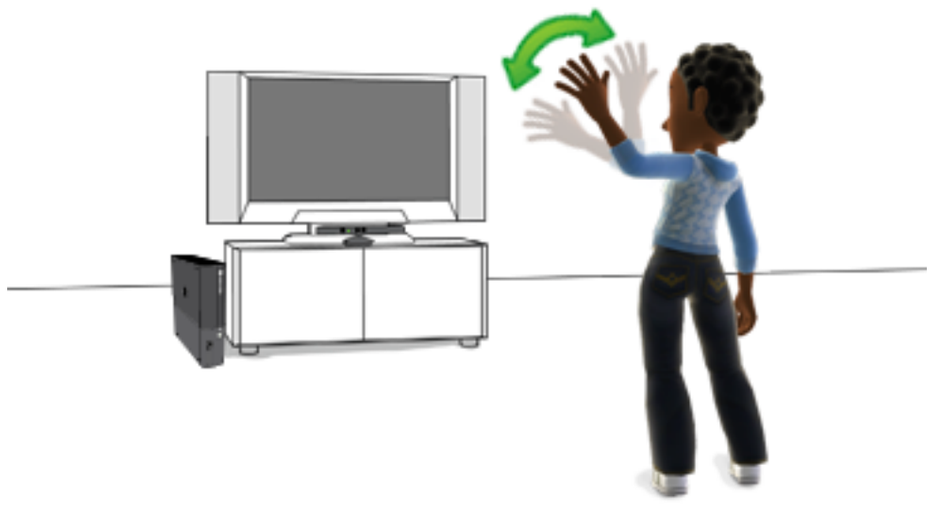
http://gamerinvestments.com/video-game-stocks/wp-content/uploads/2010/08/harry-potter-deathly-hallows-kinect-screens.jpg



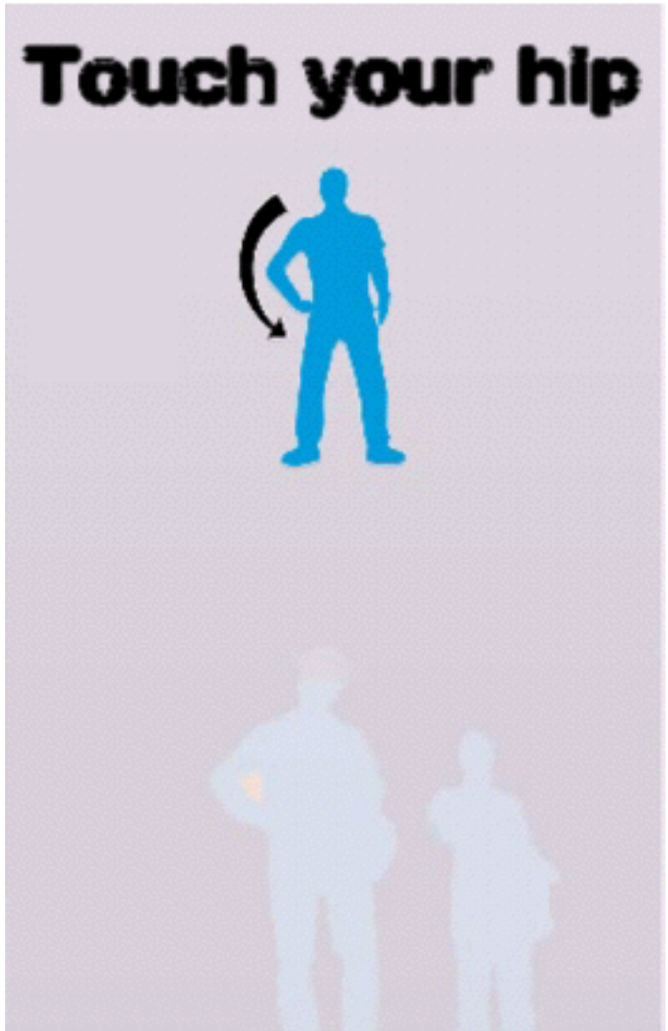
http://www.gestsure.com

http://blogs.msdn.com/cfs-filesystemfile.ashx/__key/communityserver-components-imagefileviewer/communityserver-blogs-components-weblogfiles-00-00-01-49-02/4150.BasicInteractions_5F00_Retail_5F00_520.jpg_2D00_520x0.jpg


0:54



Xbox 360 ‘wave to engage’



Walter et al. CHI ‘13

Contribution

Develop metric of user's intention to interact

Develop engagement algorithms using this metric

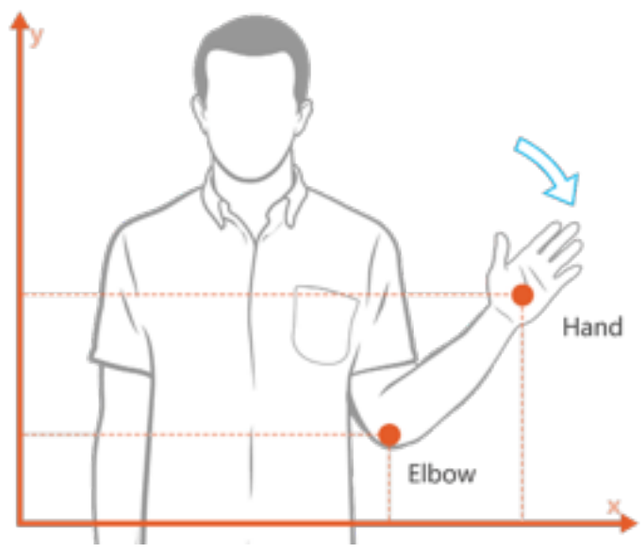
Study comparing different engagement algorithms

Study Findings

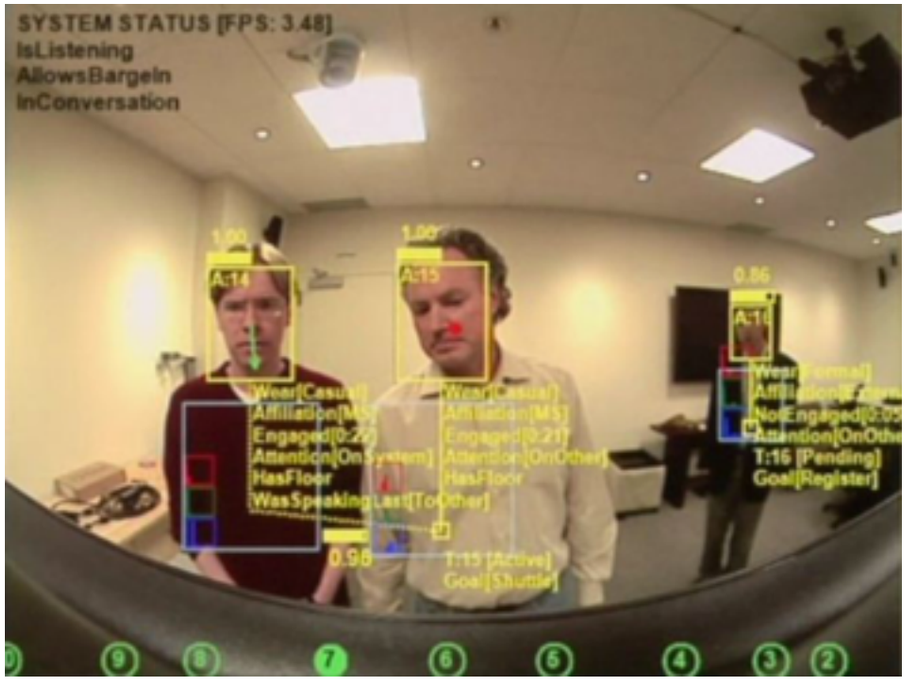
Combining intention to interact with
‘hand up and open’ gesture yielded best results.

Study Findings

12% accuracy improvement
20% reduction in time
compared to “wave to engage” on Xbox 360



http://pterneas.com/2014/01/27/implementing-kinect-gestures/



Bohus et al. ‘09

Intention to Interact Score

Classifiers: $\{C_1,...,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,...,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Intention to Interact Score

Classifiers: $\{C_1,...,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,...,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Intention to Interact Score

Classifiers: $\{C_1,\ldots,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,\ldots,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Intention to Interact Score

Classifiers: $\{C_1,...,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,...,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Intention to Interact Score

Classifiers: $\{C_1,...,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,...,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Intention to Interact Score

Classifiers: $\{C_1,...,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,...,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Engagement signals

Looking at camera

Hand above waist

Hand above head

Open posture

Body facing sensor

Waving

Engagement signals

Looking at camera

Hand above waist

Hand above head

Open posture

Body facing sensor

Waving

} gesture

6 Classifiers

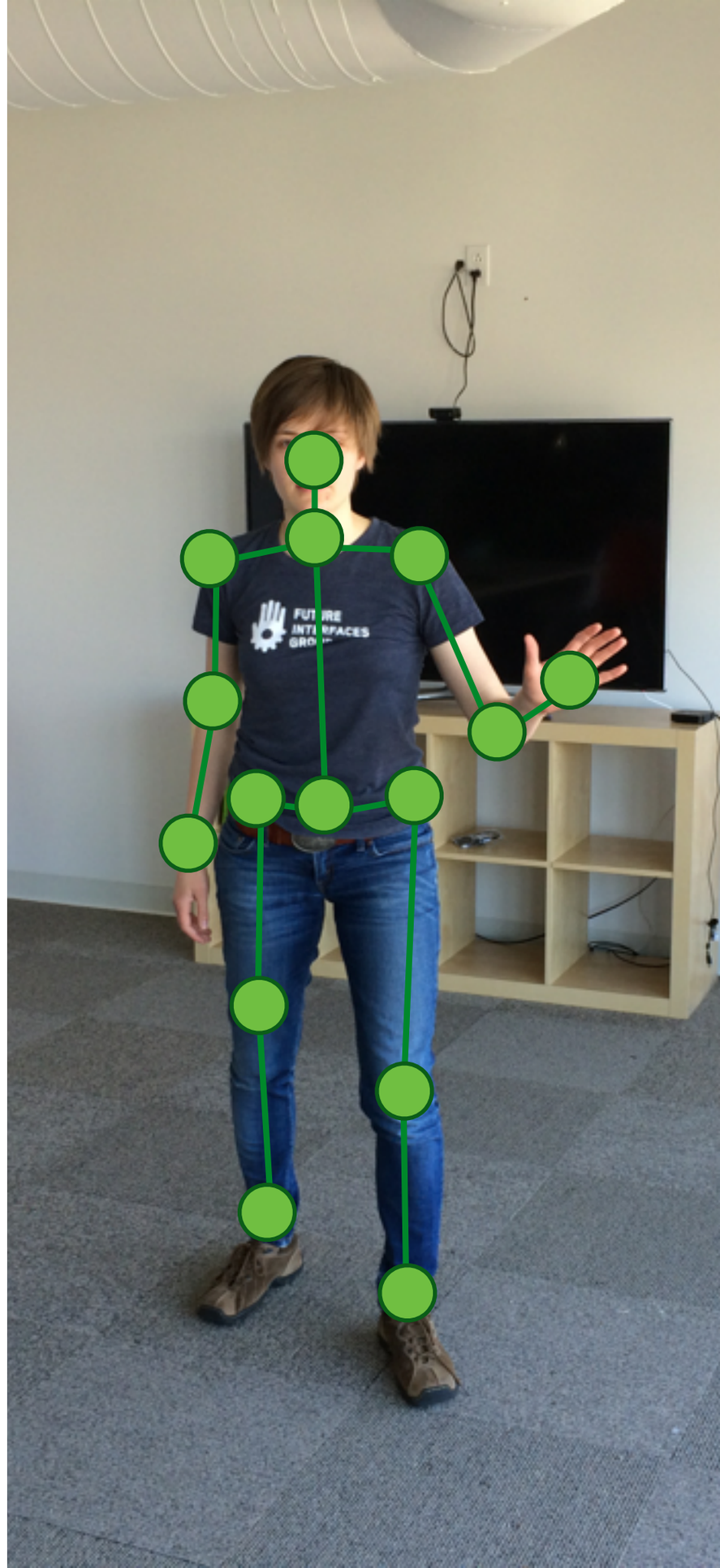
Features:

Relative joint positions

Relative joint movements

Ground truth: hand labeled frames

ADABoost, 1000 weak classifiers



Intention to Interact Score

Classifiers: $\{C_1,...,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,...,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Intention to Interact Score

Classifiers: $\{C_1,\ldots,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,\ldots,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

engage score = 0

engage score = 1

Model Weights

	model weight
Open stance	0.67*
Hand lifted above waist	0.15*
Looking at screen	0.12*
Waving	0.11*
Hand raised above head	0.08*
Body facing screen	-0.05*

Model Weights

	model weight
Open stance	0.67*
Hand lifted above waist	0.15*
Looking at screen	0.12*
Waving	0.11*
Hand raised above head	0.08*
Body facing screen	-0.05*

Intention to Interact Score

Classifiers: $\{C_1,\ldots,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,\ldots,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Intention to Interact Score

Classifiers: $\{C_1,...,C_n\}$ body signals/poses

Classifier Weights: $\{w_1,...,w_n\}$

Classification Result: $cr_i \in \{0,1\}$ 1 if body signal detected, else 0

Intention to Interact: $\sum_{i=1}^{n} w_i * cr_i$

Example

Look at screen = 1 * 0.12

Hand above waist = 1 * 0.15

Hand above head = 0 * 0.08

Open posture = 1 * 0.67

Body facing sensor = 1 * -0.05

Waving = 0 + 0.11

engage score = 0.89

Engagement Algorithms

1. engage_score > threshold
2. engage_score > threshold && hand_up_open
3. engage_score > threshold && 360_wave_detected

User Study

Compare engagement algorithms.

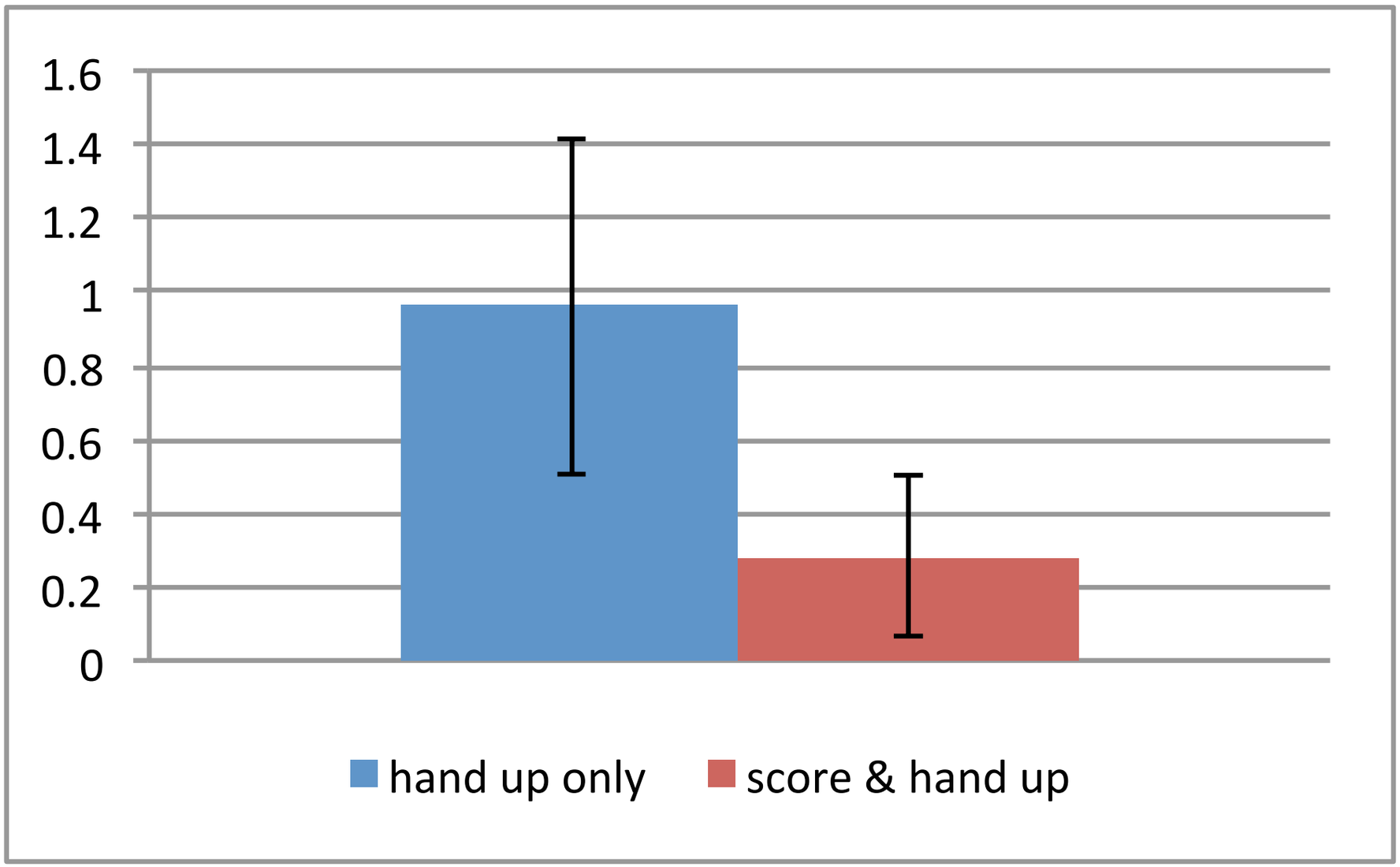
Evaluate engagement detection for single player.

Measure time to engage and accuracy.

Conditions

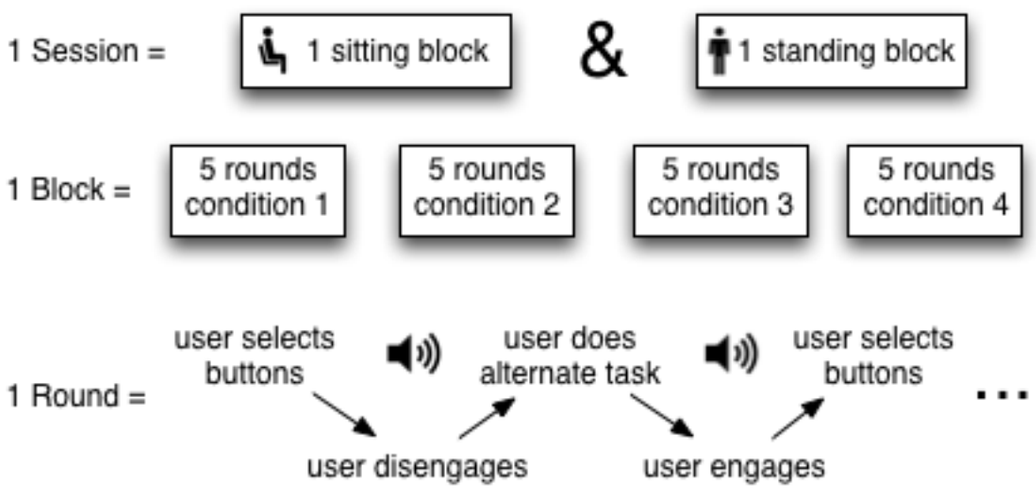
1. engage_score > threshold
2. engage_score > threshold && hand_up_open
3. engage_score > threshold && 360_wave_detected
4. (baseline) 360_wave_detected

Accidental engagements using hand up & open, with and without engage score



Experiment Design

30 person lab-study



Measures

accuracy: % of video frames correctly identified.

speed: time to engage and disengage.

Results

Conditions

Xbox 360 ‘wave’

wave only

engage_score > threshold

score only

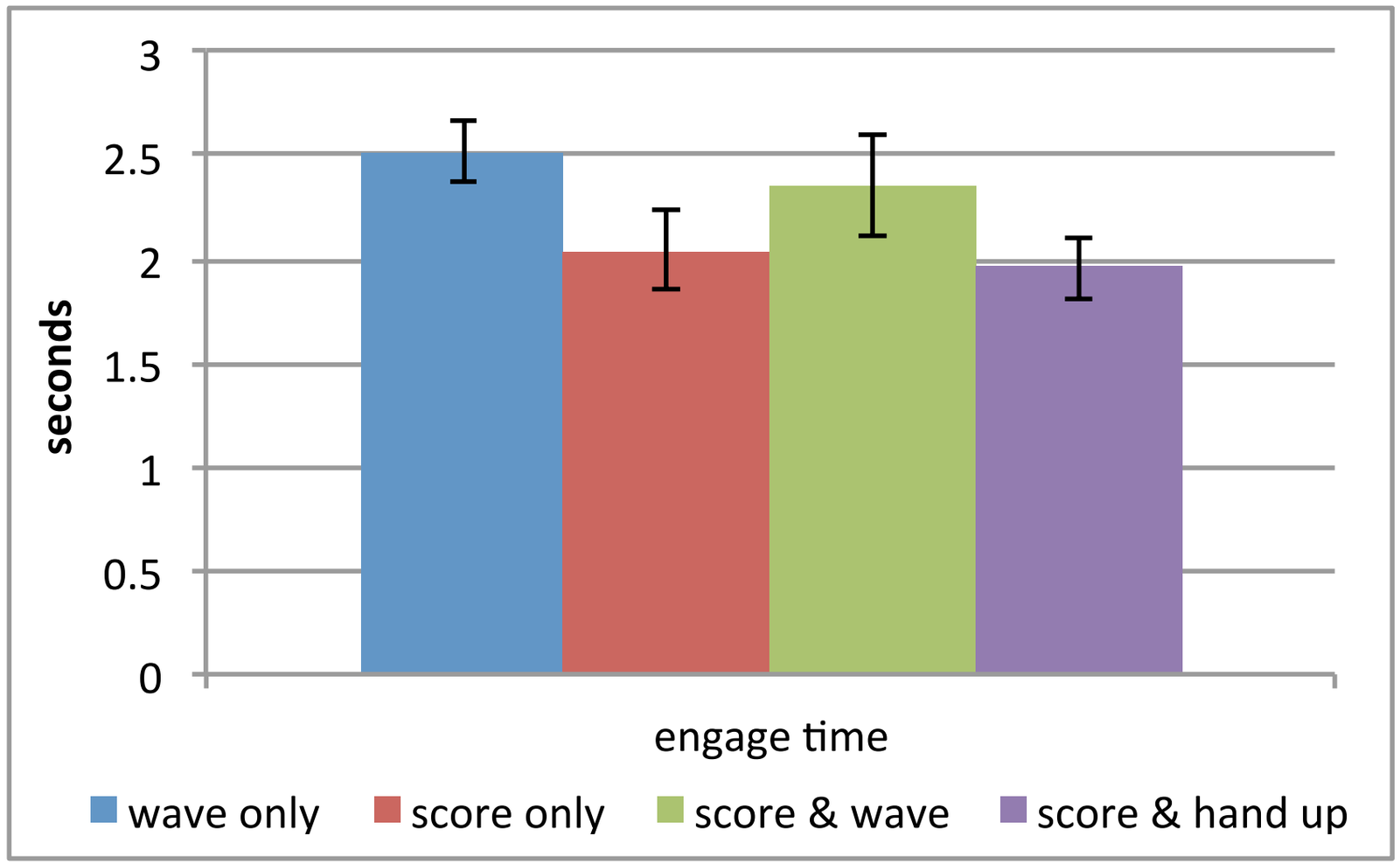
engage_score > threshold && wave

score & wave

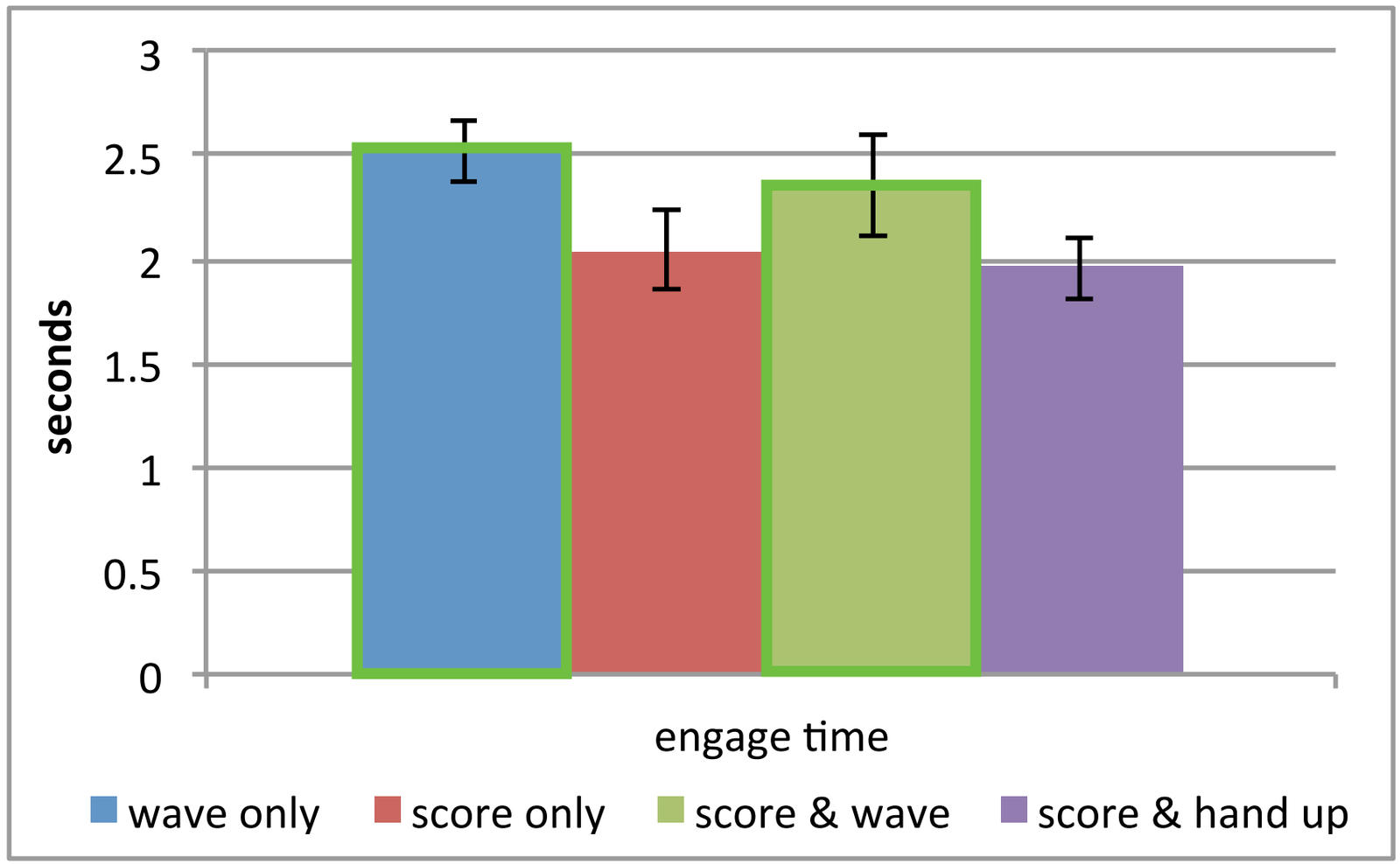
engage_score > threshold &&
hand up, open

score & hand up

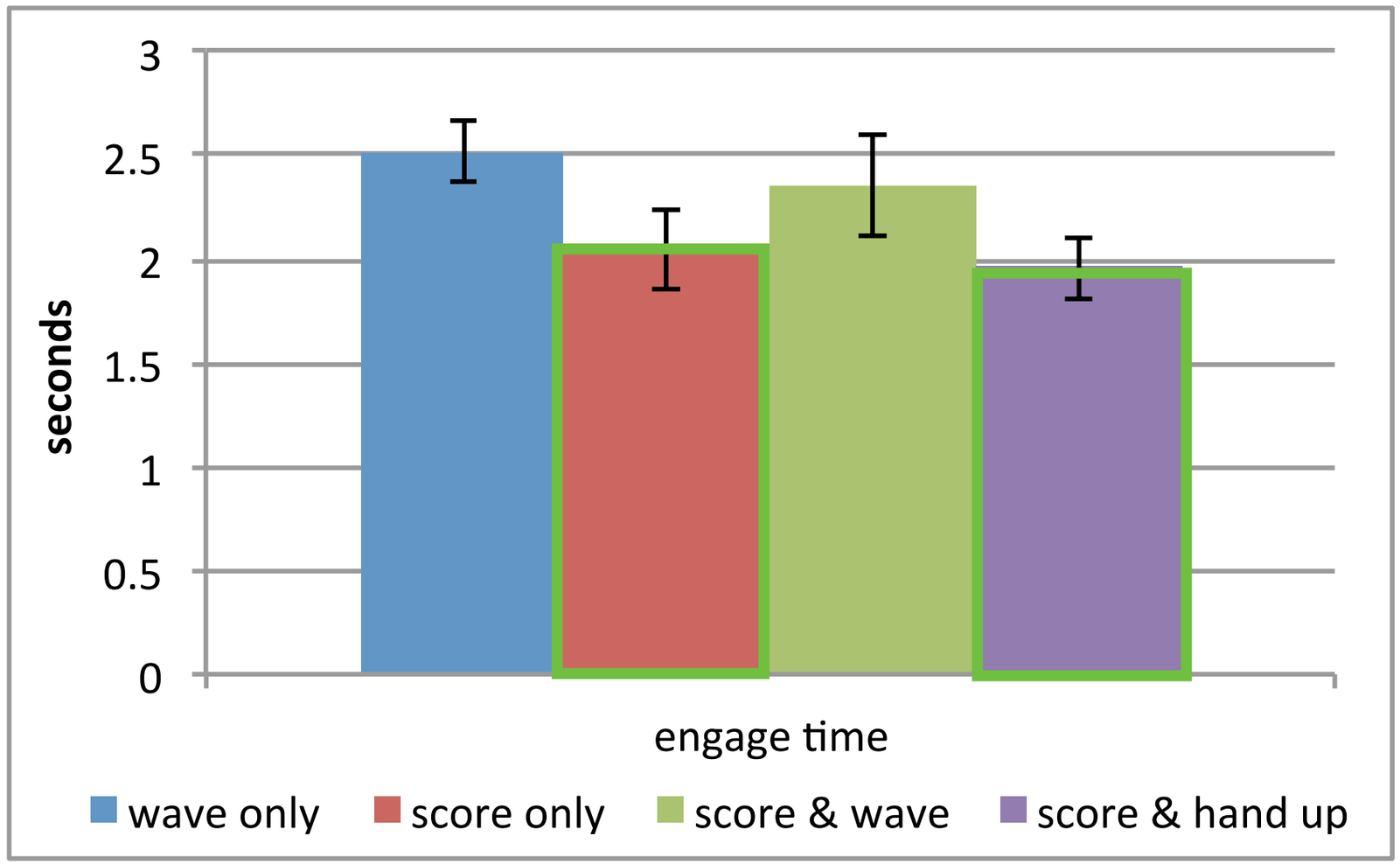
Mean time to engage



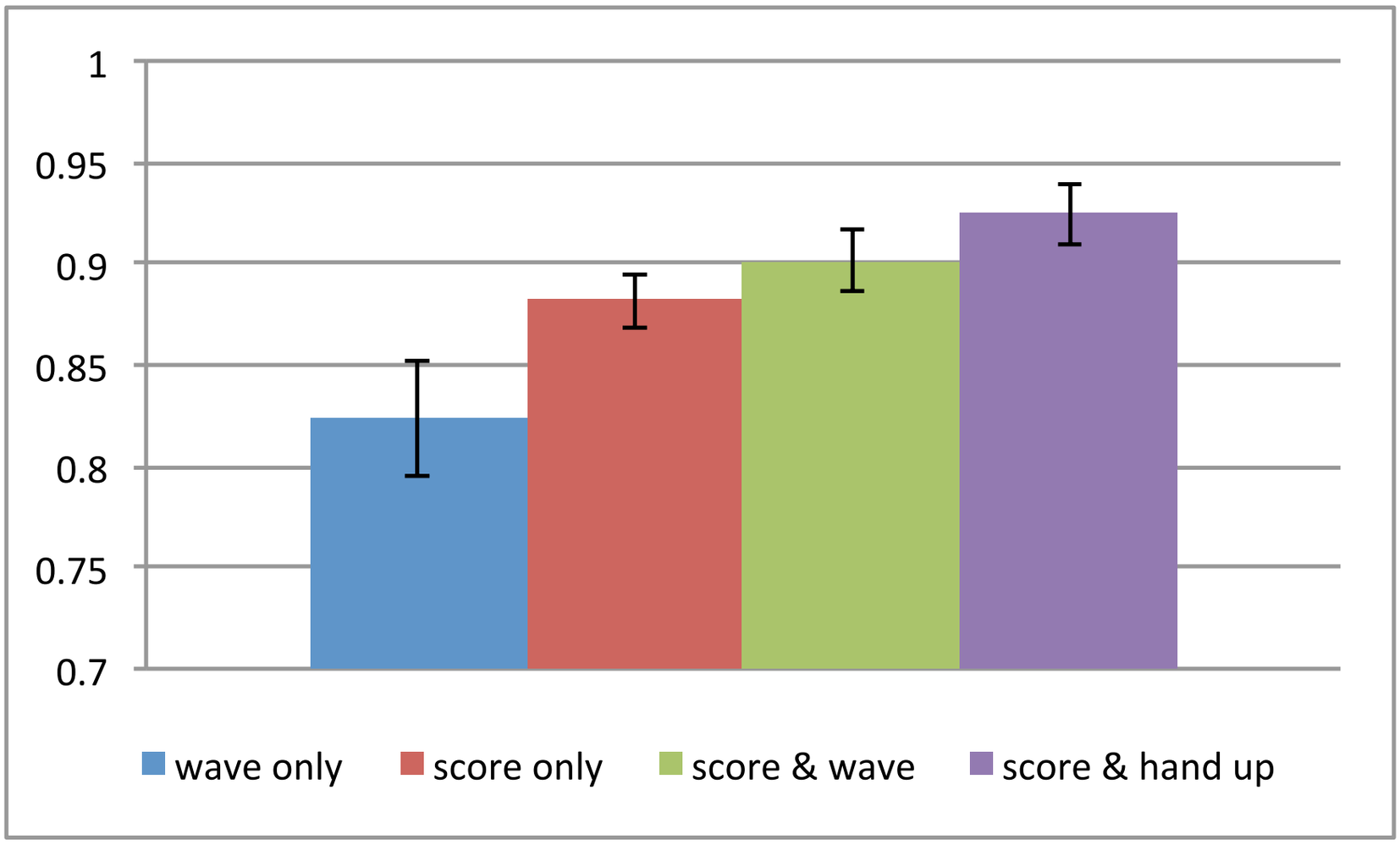
Mean time to engage



Mean time to engage

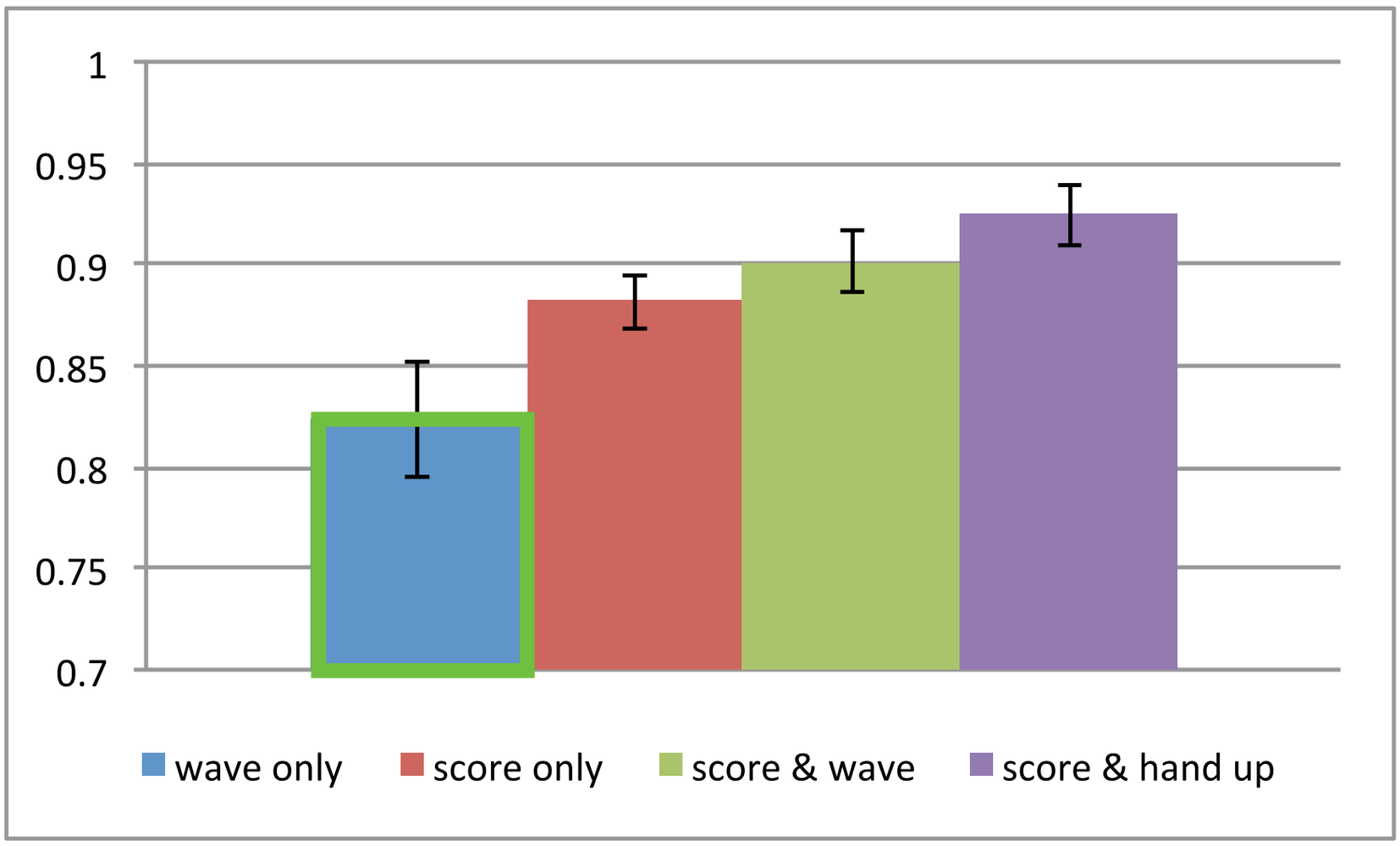


% of video frames correctly identified as ‘engaged’



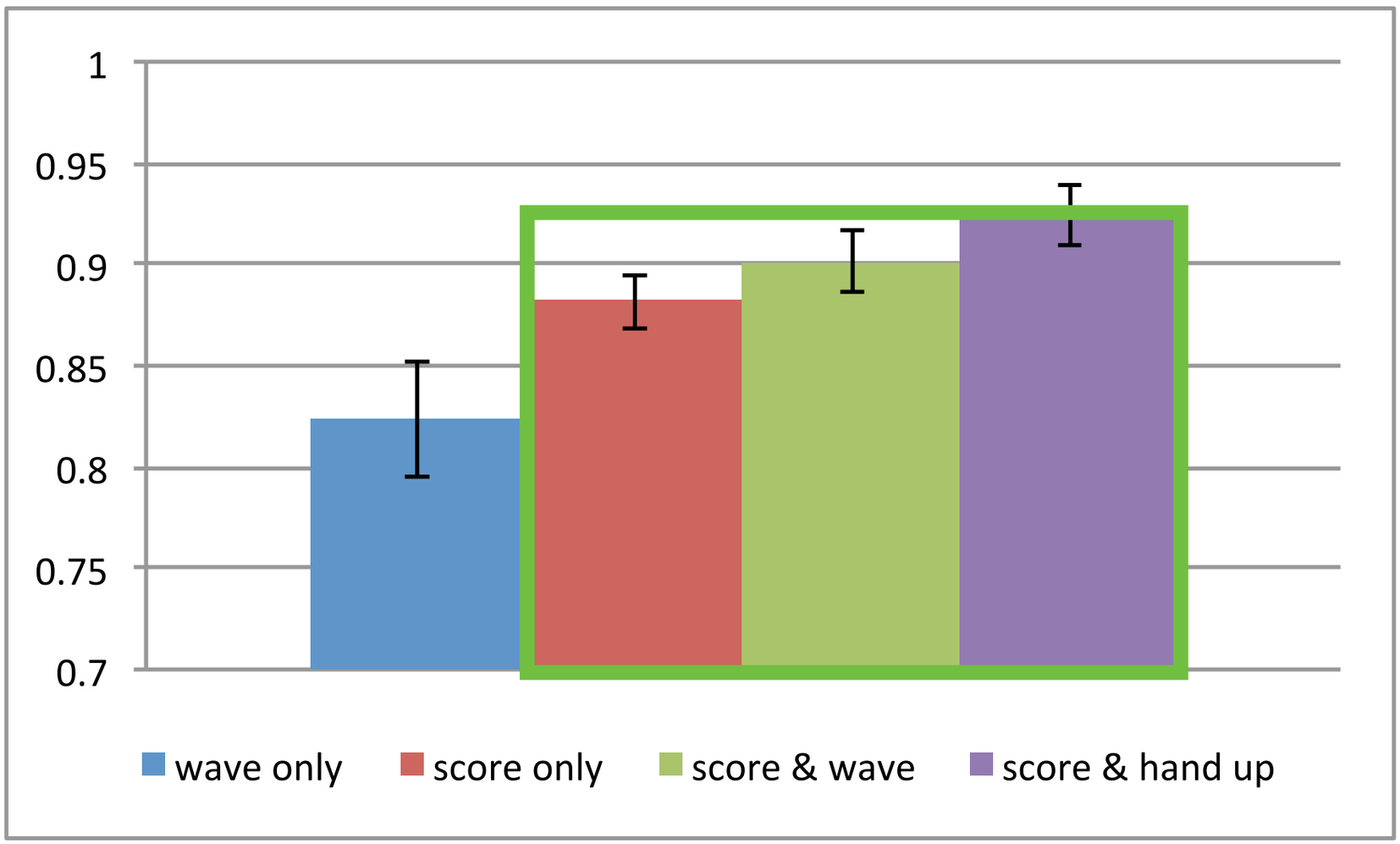
*bars represent 95% confidence interval

% of video frames correctly identified as ‘engaged’



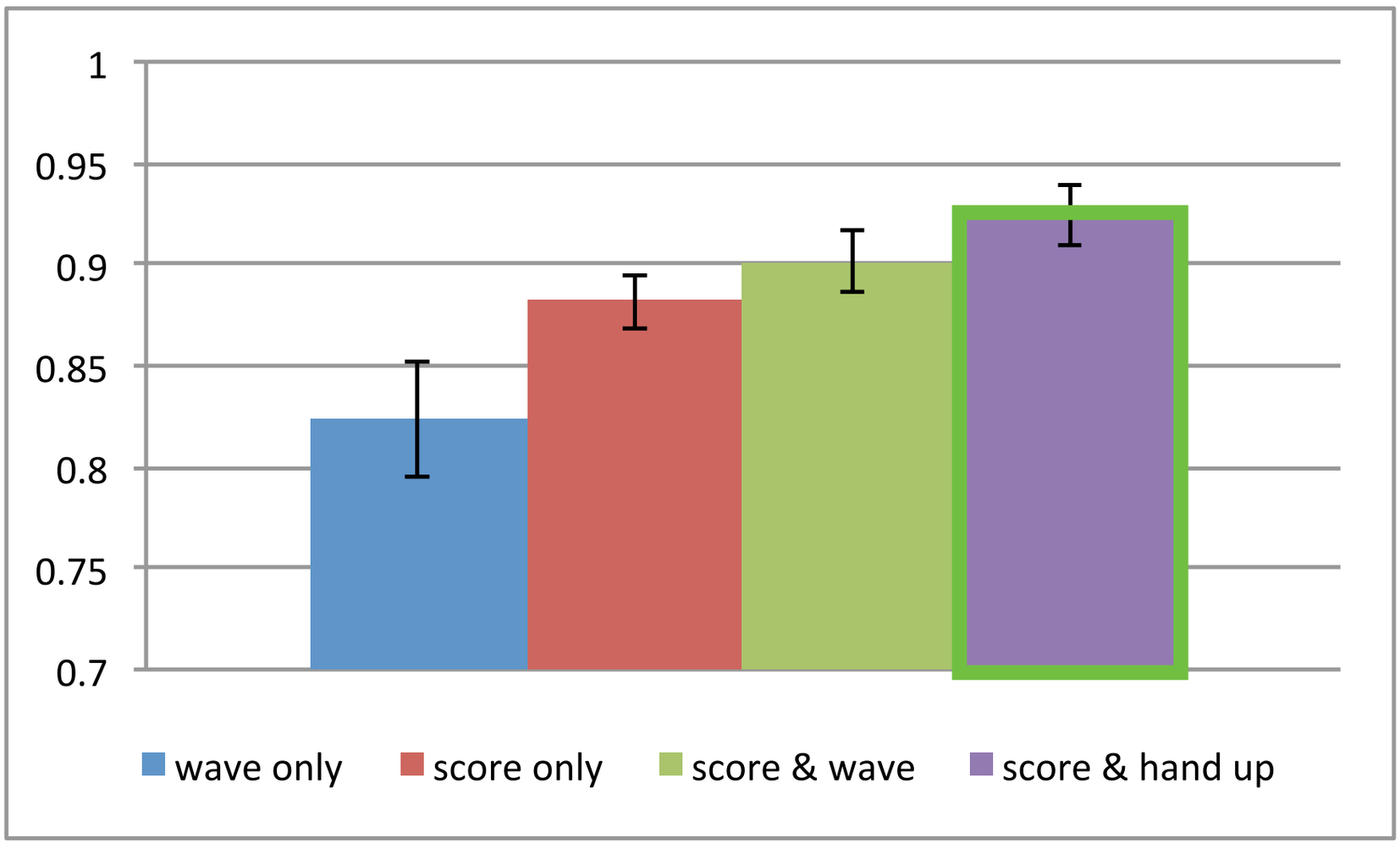
*bars represent 95% confidence interval

% of video frames correctly identified as ‘engaged’



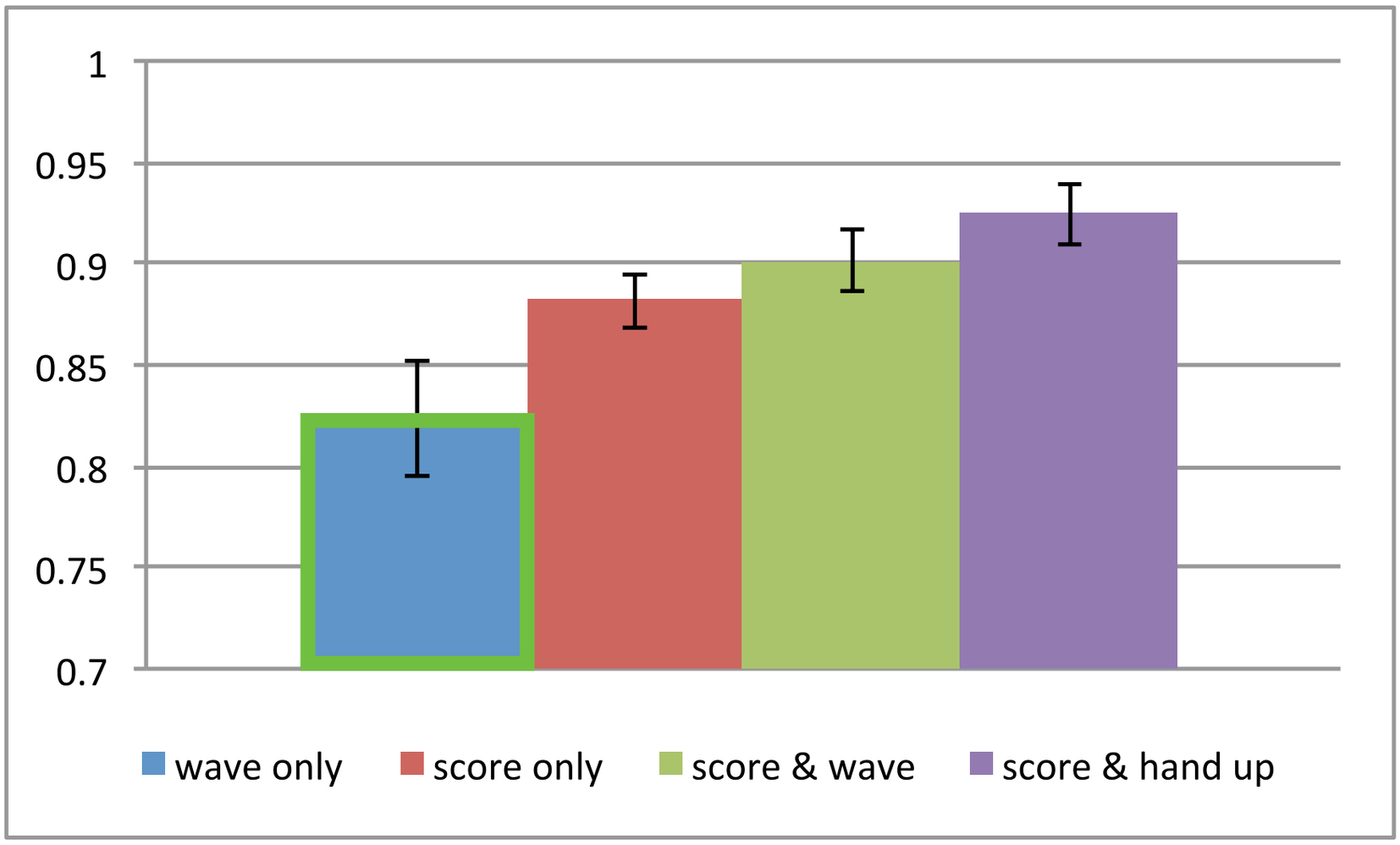
*bars represent 95% confidence interval

% of video frames correctly identified as ‘engaged’



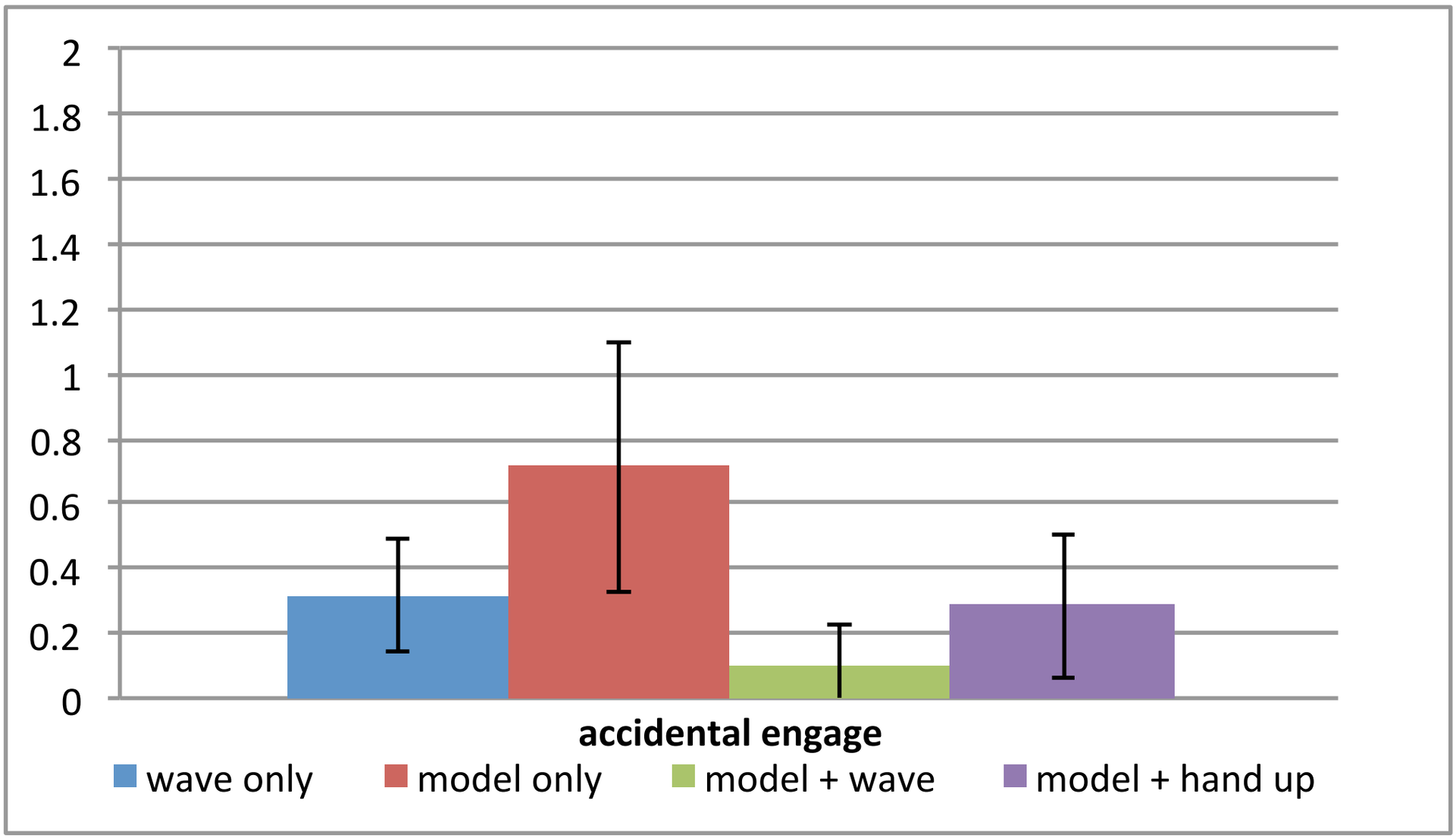
*bars represent 95% confidence interval

% of video frames correctly identified as ‘engaged’

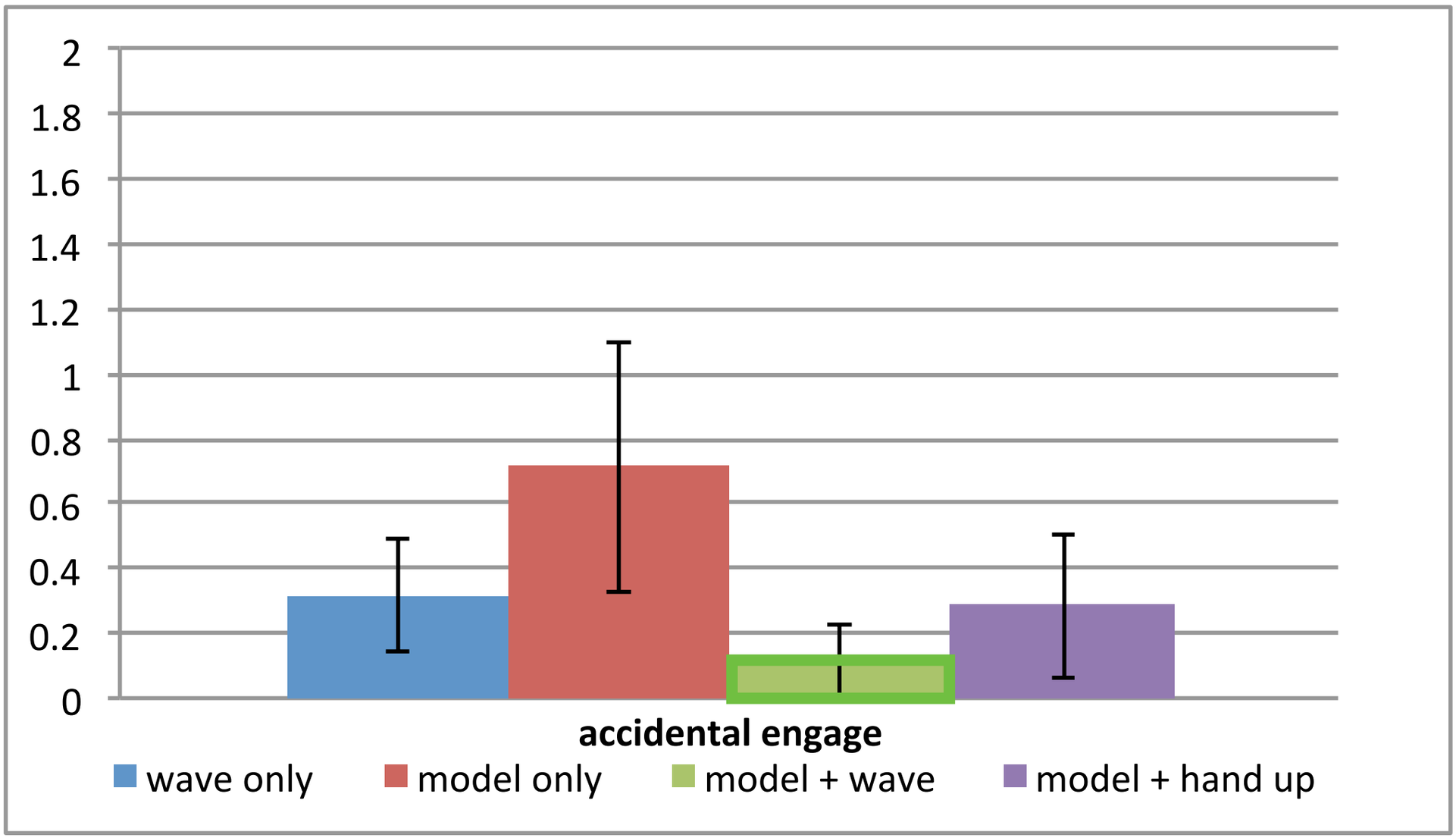


*bars represent 95% confidence interval

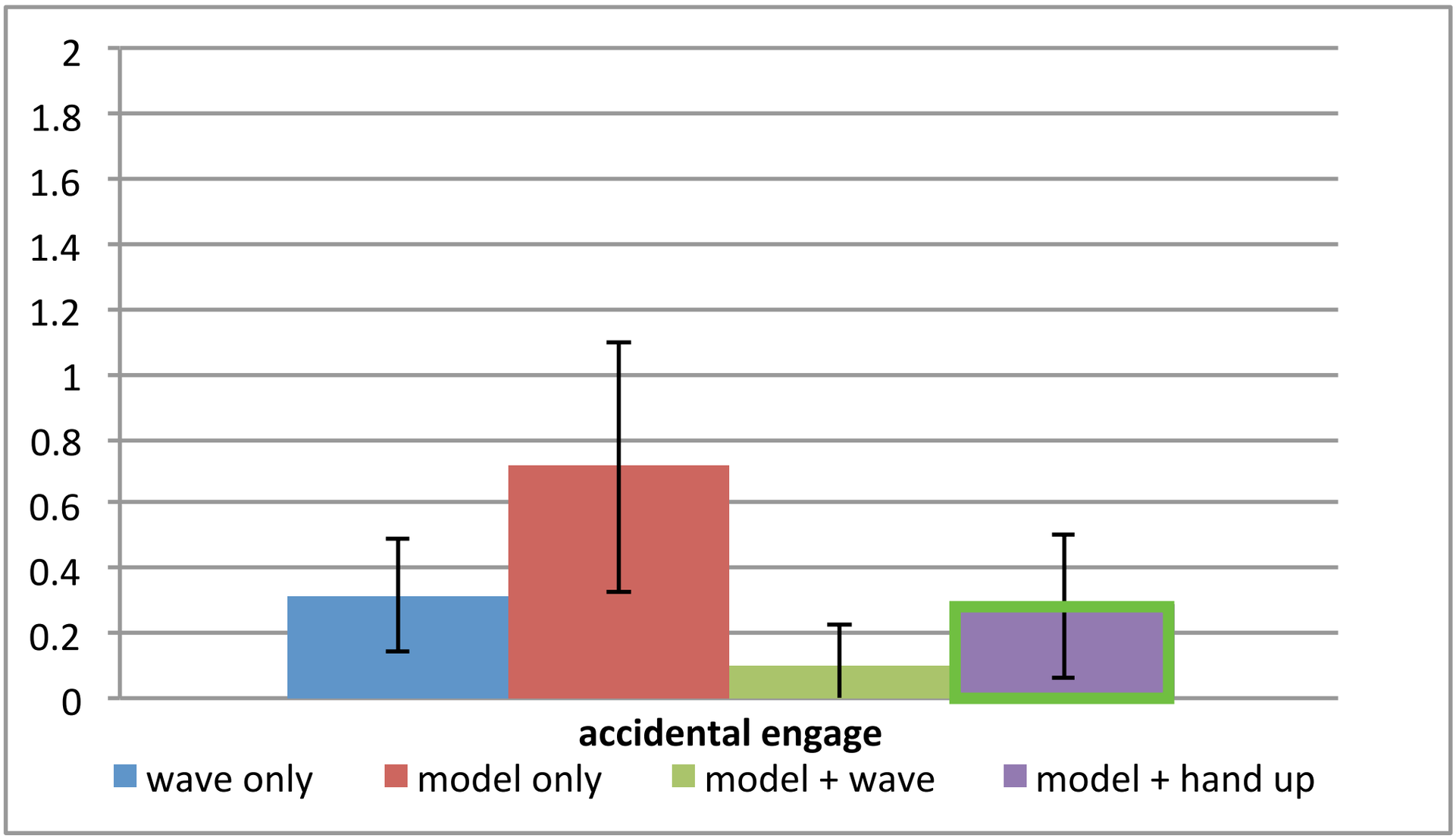
average # of accidental engagements, per user



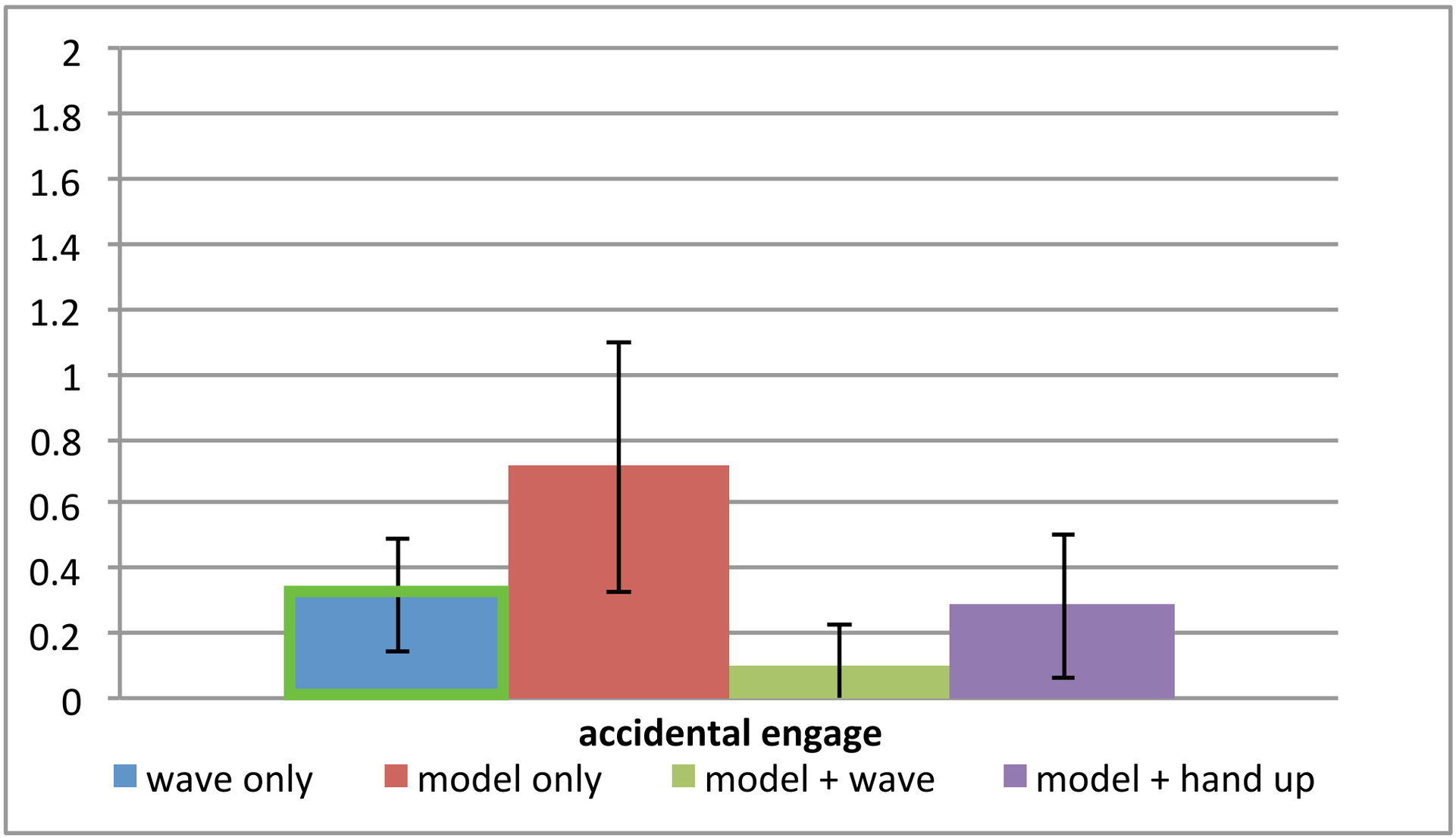
average # of accidental engagements, per user



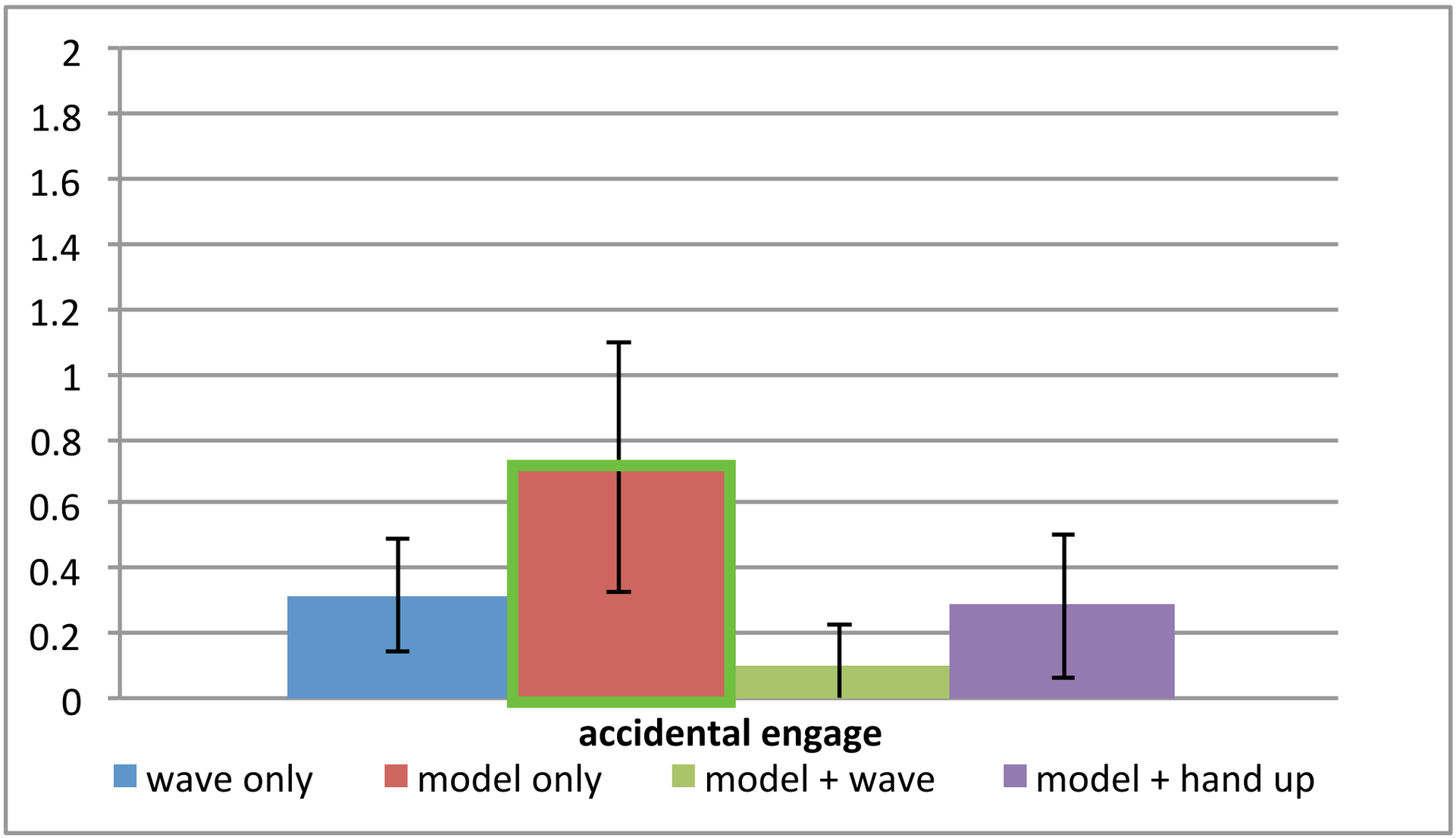
average # of accidental engagements, per user



average # of accidental engagements, per user

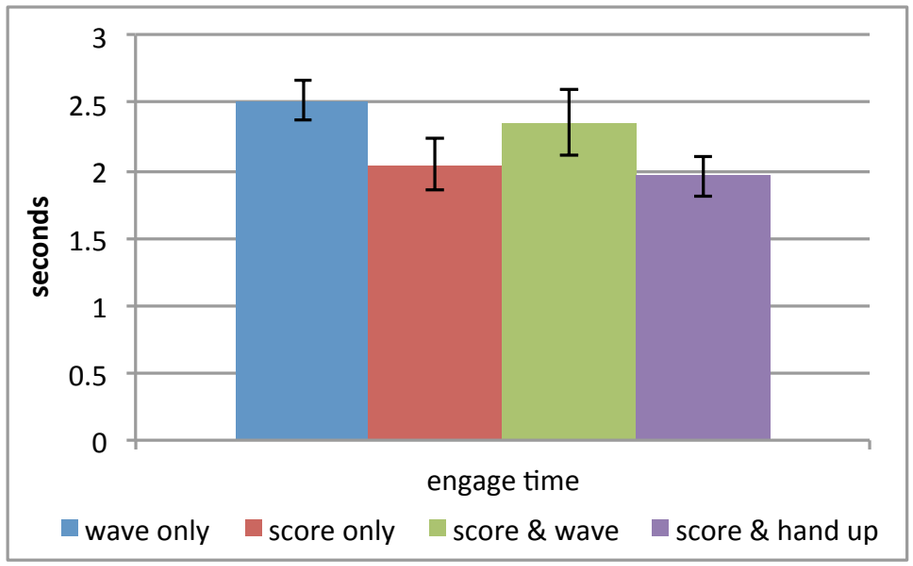


average # of accidental engagements, per user

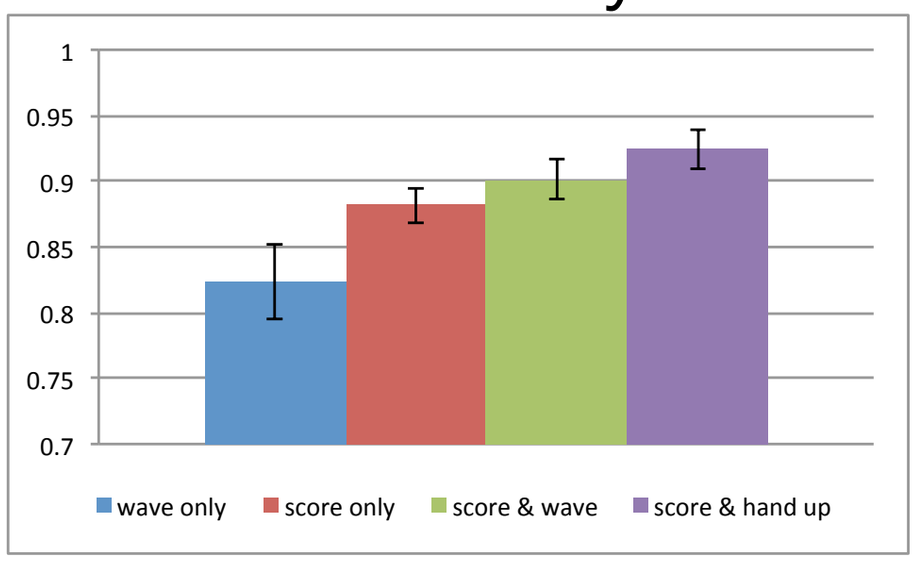


Results Summary

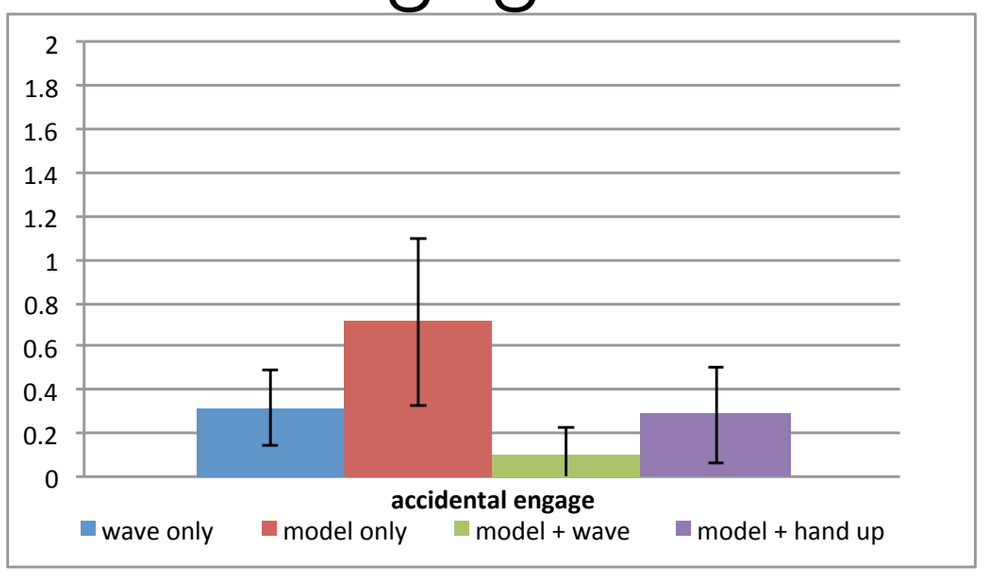
time to engage



accuracy

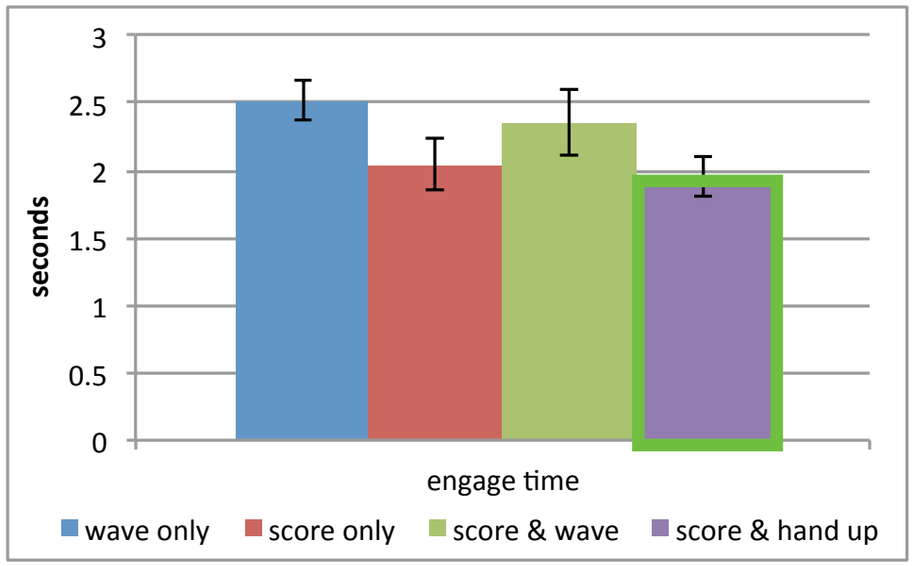


false engagements

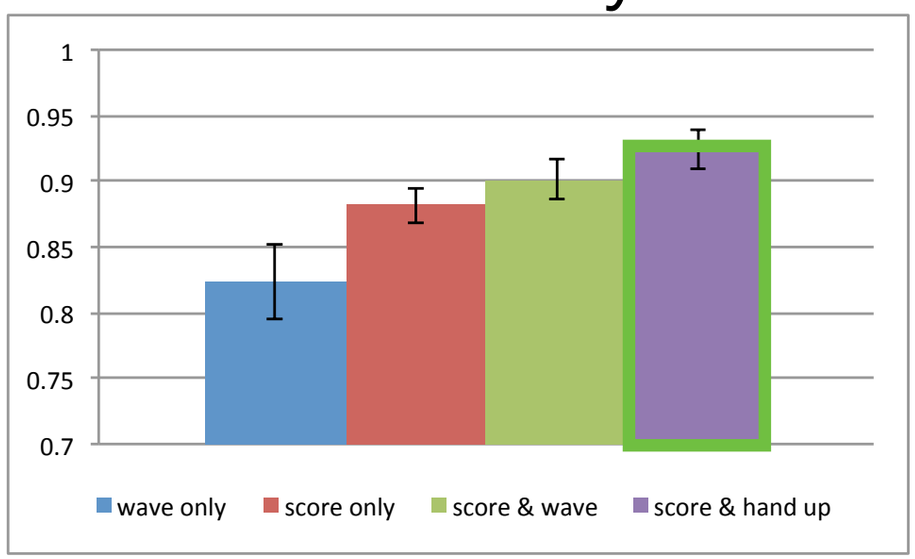


Results Summary

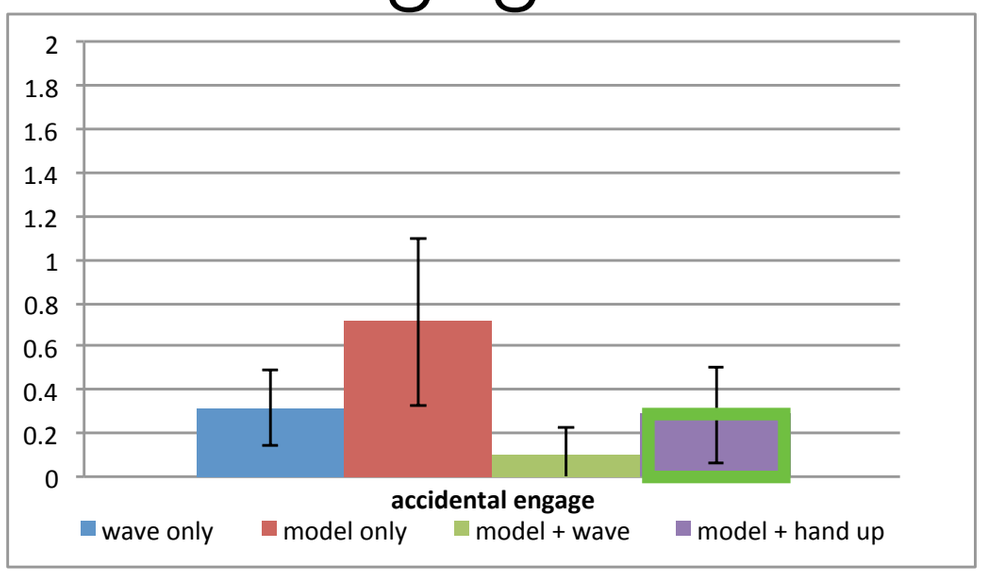
time to engage



accuracy

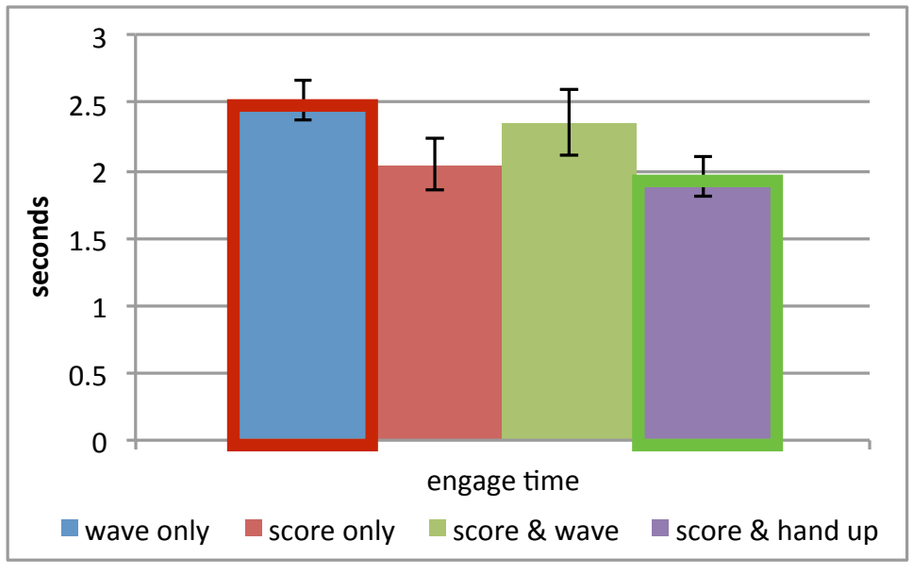


false engagements

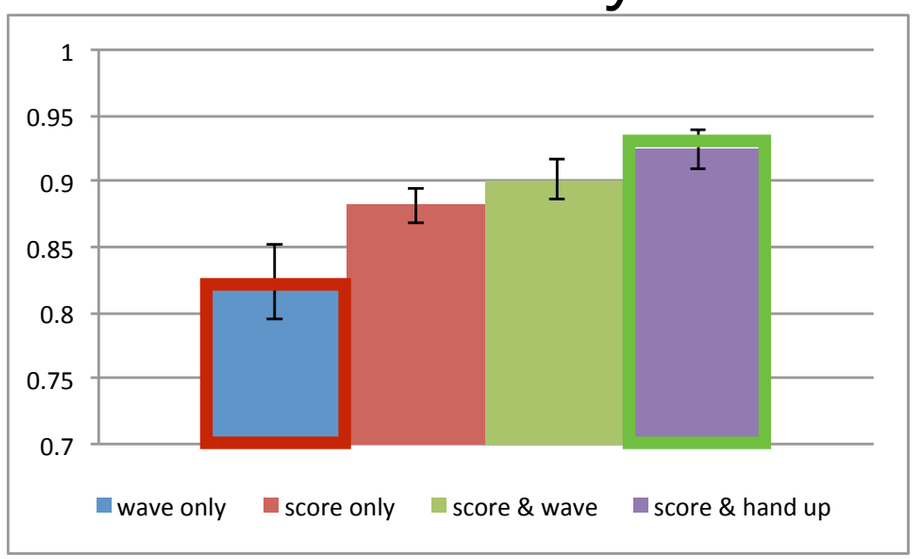


Results Summary

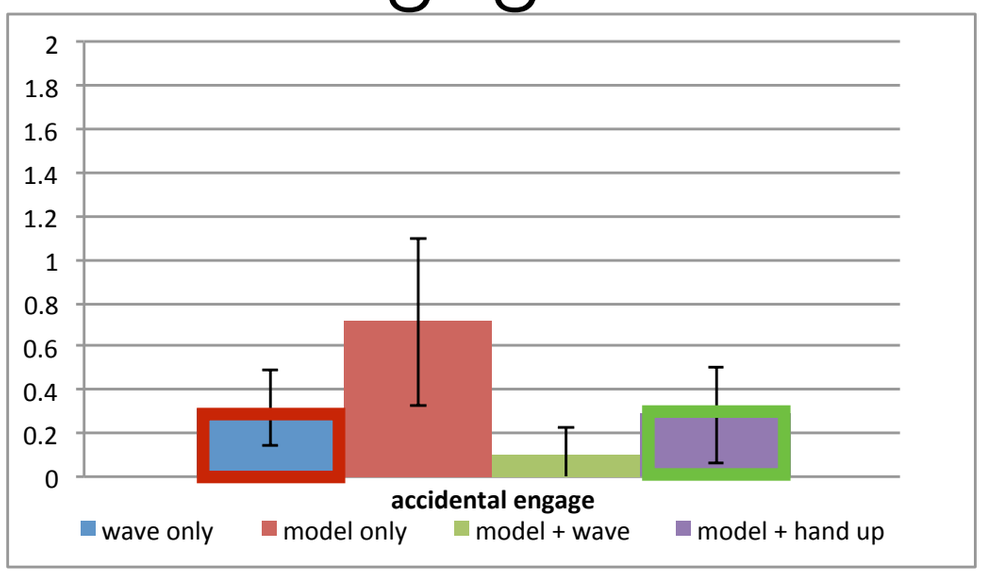
time to engage



accuracy



false engagements



Takeaways

Intention to interact can be used to improve accuracy of determining *user engagement.*

Best used when combined with a simple, explicit hand up and open gesture.

Thank you!

julia.schwarz@cs.cmu.edu