Characterizing Time Spent in Video Object Tracking Annotation Tasks: A Study of Task Complexity in Vehicle Tracking

Amy Rechkemmmer[1], Alex C. Williams[2], Matthew Lease[2,3], and Li Erran Li[2]

Purdue University[1] AWS AI, Amazon[2] The University of Texas at Austin[3]

HCOMP 2023 Delft, Netherlands

PURDUE UNIVERSITY | Department of Computer Science

amazon | science

The Data Annotation Industry has Widespread Implications

In-house, trained workforce performing complex annotation tasks

Ex: Amazon SageMaker Ground Truth, Scale AI, Sama

Meta

TOYOTA

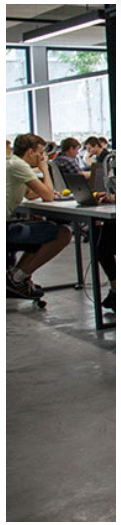
AstraZeneca

T-Mobile

U.S. ARMY

NFL

The Data Annotation Industry has Widespread Implications



MARKETSANDMARKETS™

YOTA

Data Annotation and Labeling Market worth $3.6 billion by 2027, growing at a CAGR of 33.2%: Report by MarketsandMarkets™

In-

NFL

U.S. ARMY

Ex: Amazon SageMaker Ground Truth, Scale AI

The Data Annotation Industry has Widespread Implications

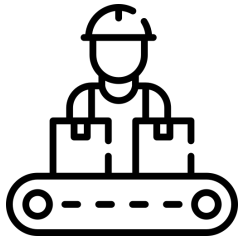
… but the work is time-consuming and tedious

This data is often proprietary, making it unsuitable for crowd workers.

1. Only label apples that are a certain distance from the ground
2. Only label apples that are of a sufficient size

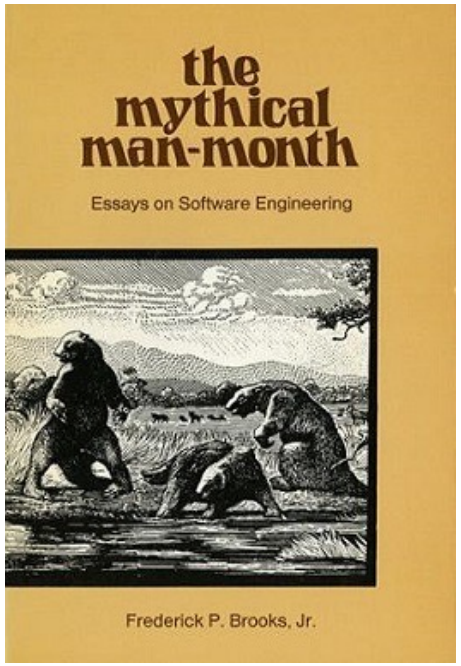
Understanding Productivity is Challenging

Output/Time



Number of boxes moved per hour

Number of lines of code written per hour?

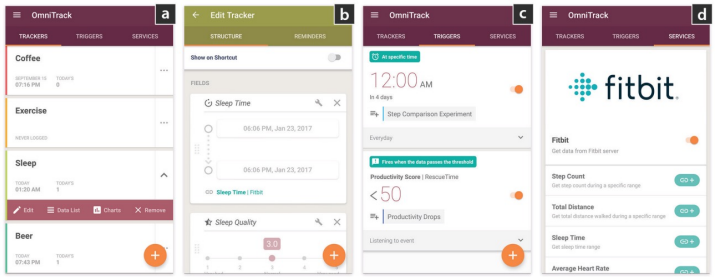


For knowledge-based tasks, the threshold for acceptable output and the path to get there are often unclear

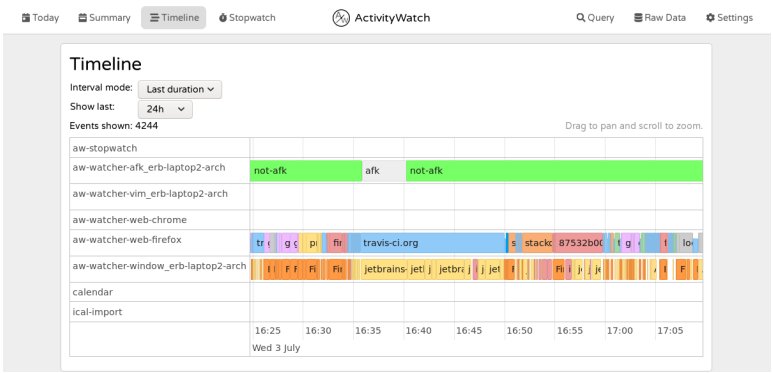
Determining quality is difficult/subjective --> Understanding time spent is a common proxy

Lessons on Understanding Time in Knowledge Work

Understanding How Time is Spent



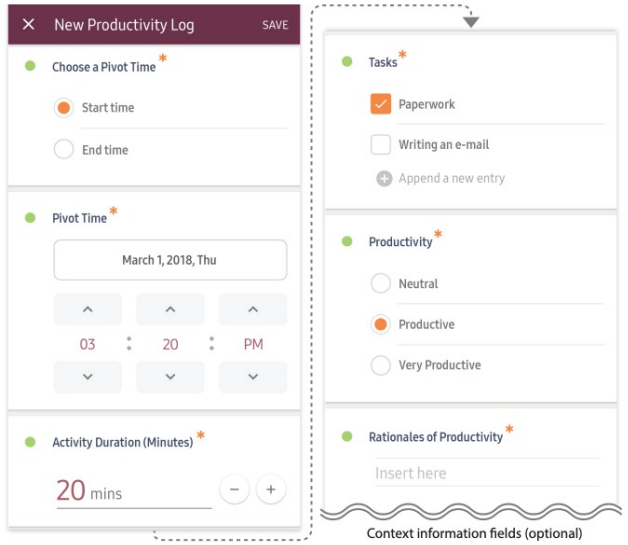
Kim et al. "OmniTrack: a flexible self-tracking approach leveraging semi-automated tracking." UIST (2017).



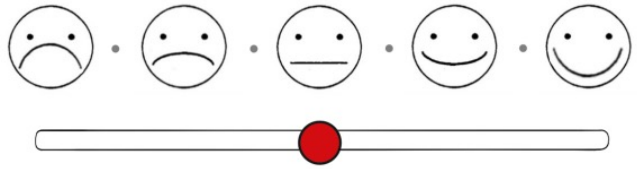
https://activitywatch.net/

Apple's ScreenTime

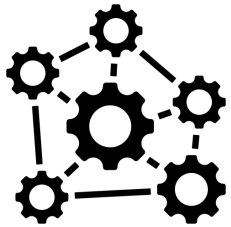
Understanding How Workers Perceive their Time Spent



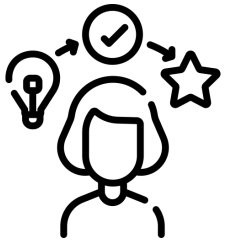
Kim et al. "Understanding personal productivity: How knowledge workers define, evaluate, and reflect on their productivity." CHI (2019).



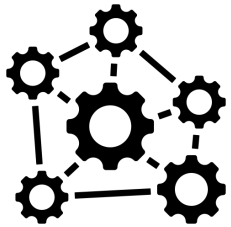
Guillou et al. "Is your time well spent? reflecting on knowledge work more holistically." CHI (2020).

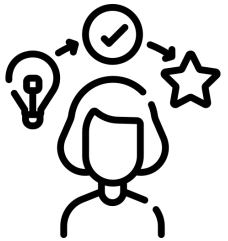


task complexity



worker experience

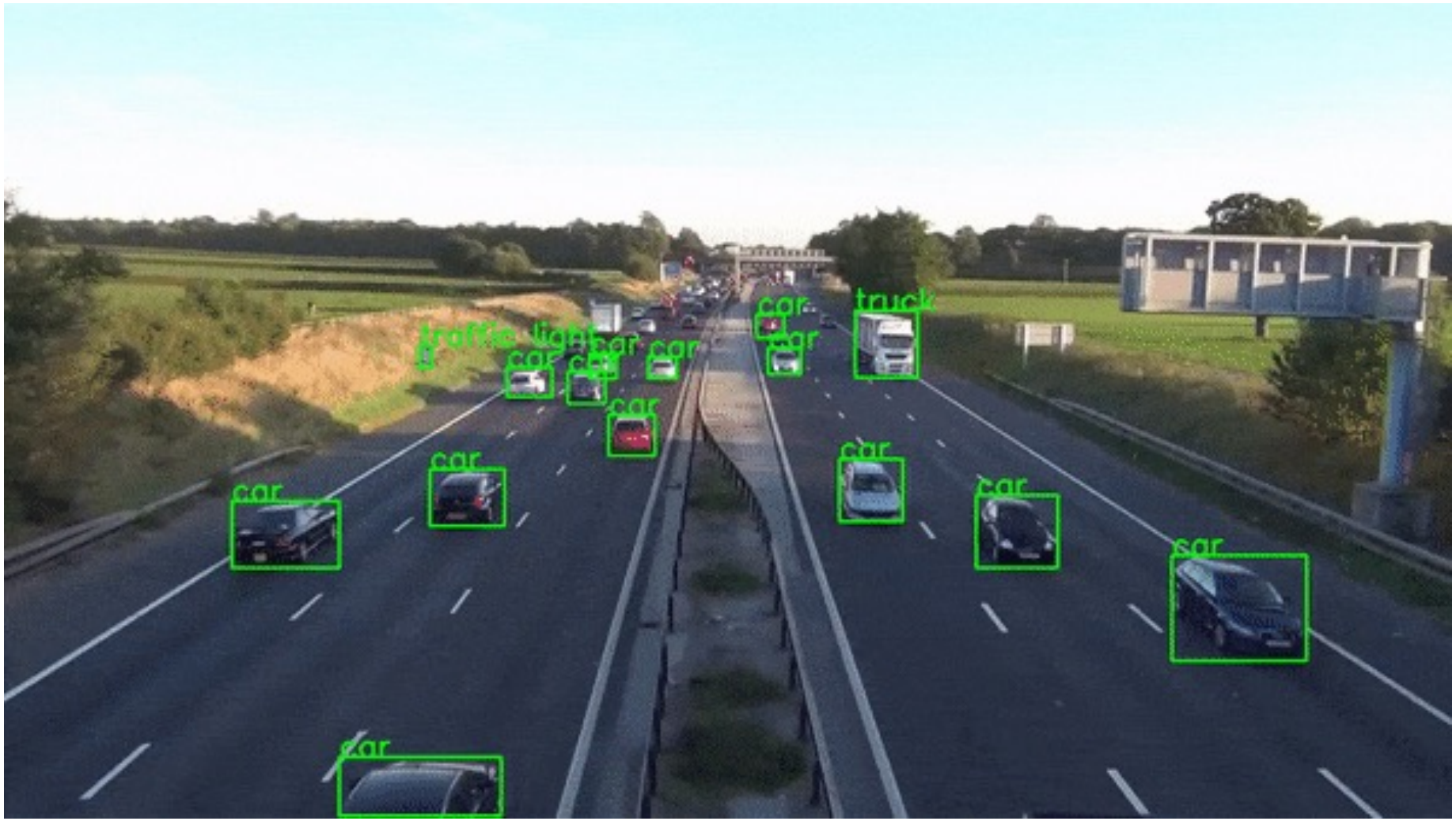




How do **task complexity** and **worker experience** impact:

1. Time spent performing data annotation

2. Worker perceptions of time spent performing data annotation and other assistive activities

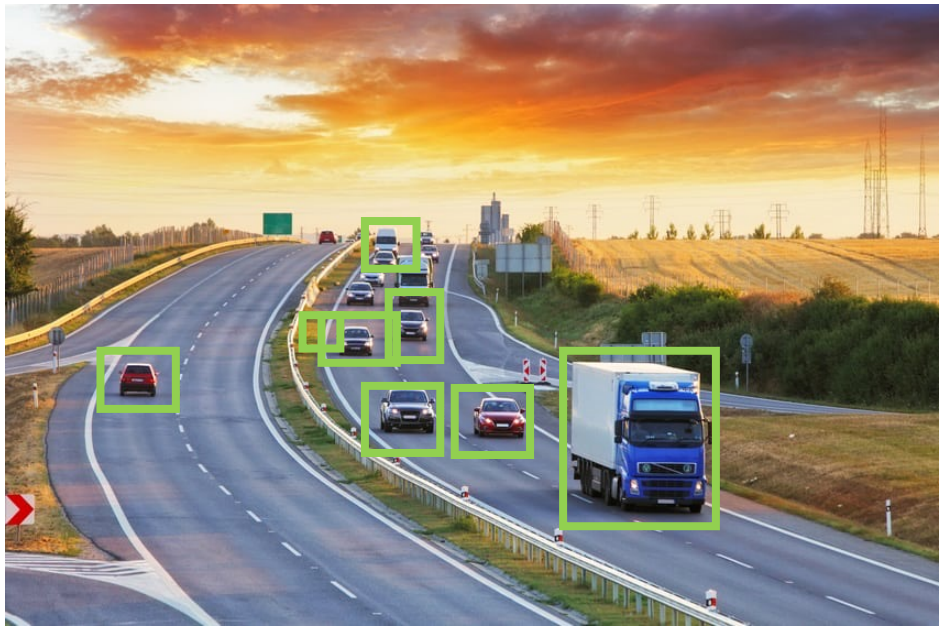

traffic light
car
car
car
car
car
car
car
car
car
car
car
car
truck

40 full-time annotators at a large-scale technology corporation

Working on a real-world proprietary task involving vehicle tracking on city streets and highways

Interface Functionality

- Manual Creation
- Edit
- Delete
- Predict Next
- Copy Frame
- Zoom In
- Zoom Pan
- Play Video
- Frame Navigation (Forward and Backward)



Low Label Constraint

Annotate All Vehicles

Constraints based on requirements of real proprietary task used

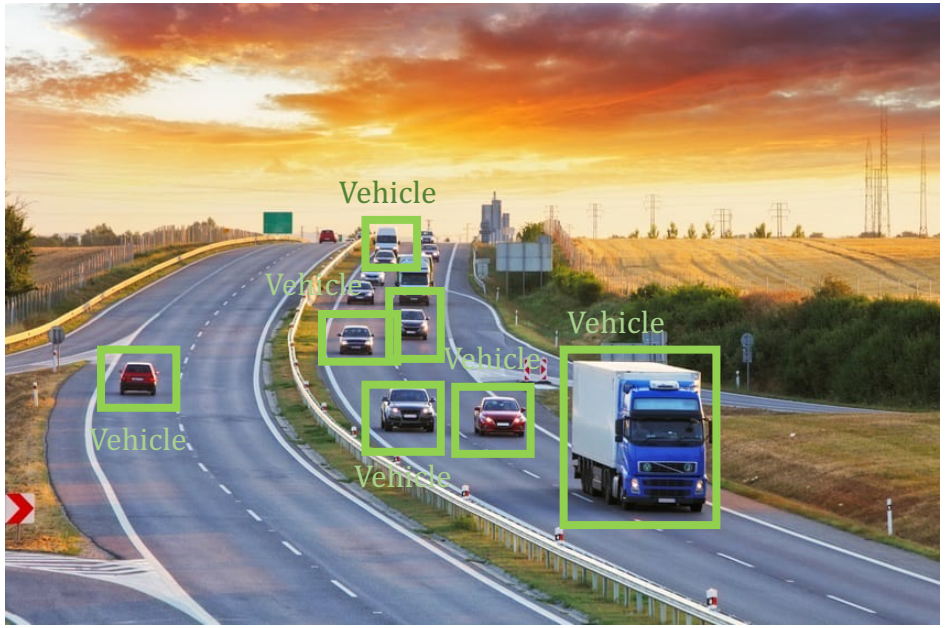
High Label Constraint

Local

1. Are all of this vehicle's brake lights currently visible in this frame?
2. Is this vehicle in the same lane as the POV vehicle or at most on lane over?
3. Is this vehicle moving in the same direction as the POV vehicle?

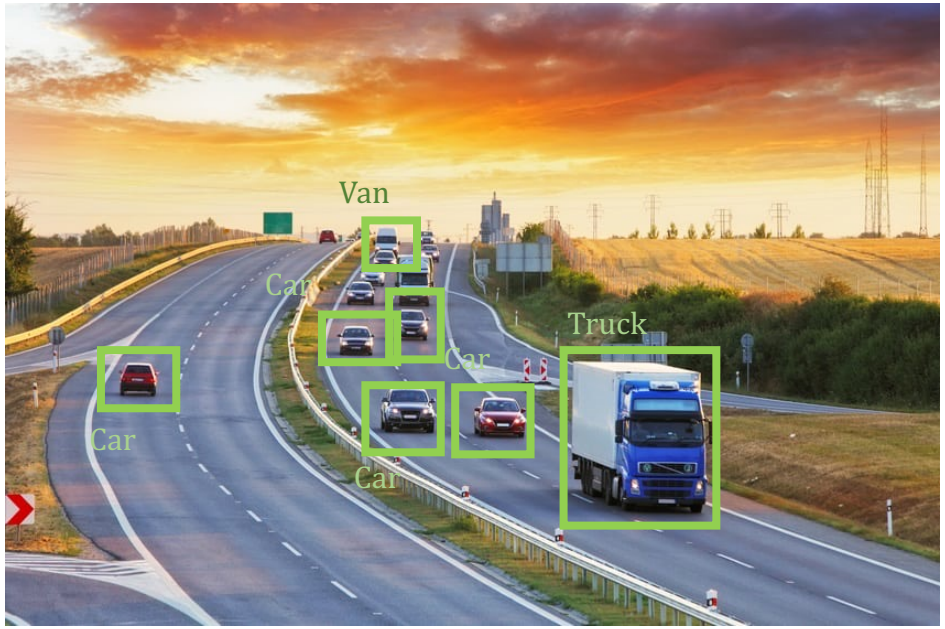
Global

4. Does this vehicle have its brake lights on at least once at any point in this video?



Low Label Granularity

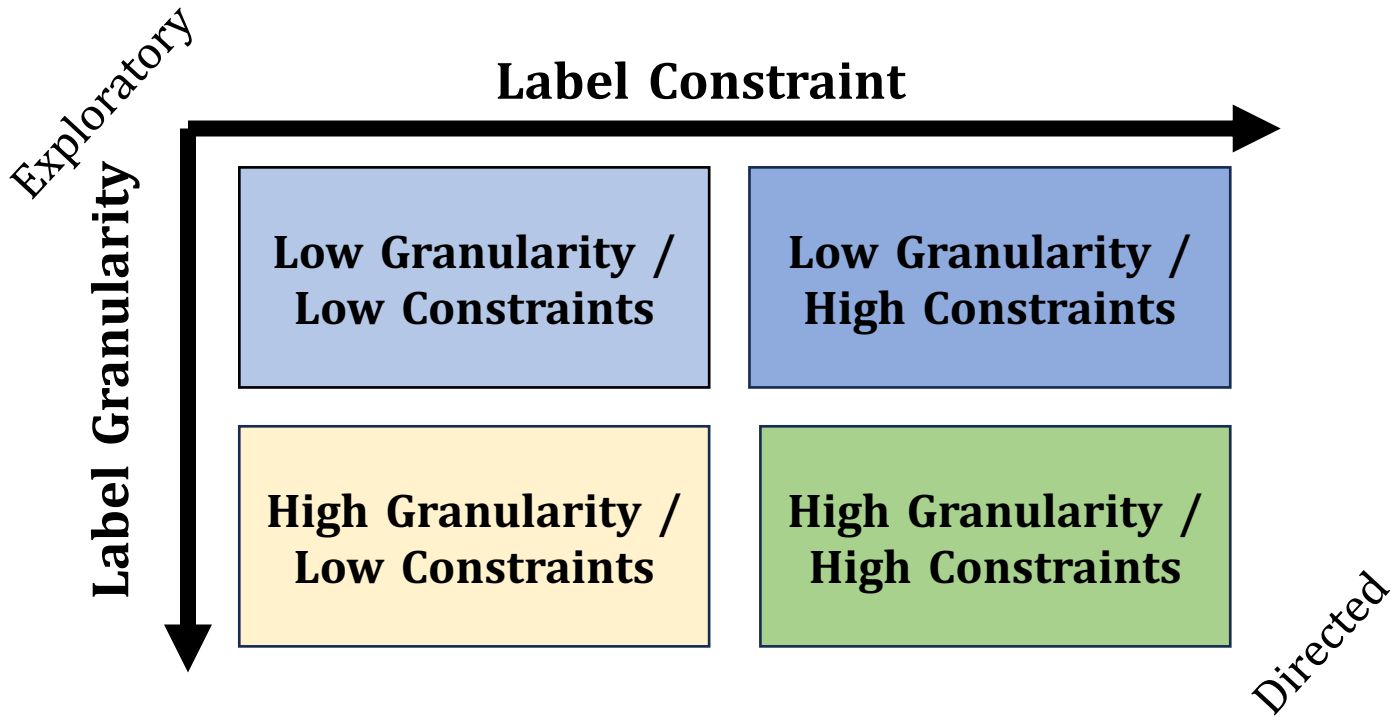
Vehicle



High Label Granularity

Car, Truck, Van, Motorbike, Bus

2 x 2 Task Complexity
Experimental Design
(10 workers per condition)

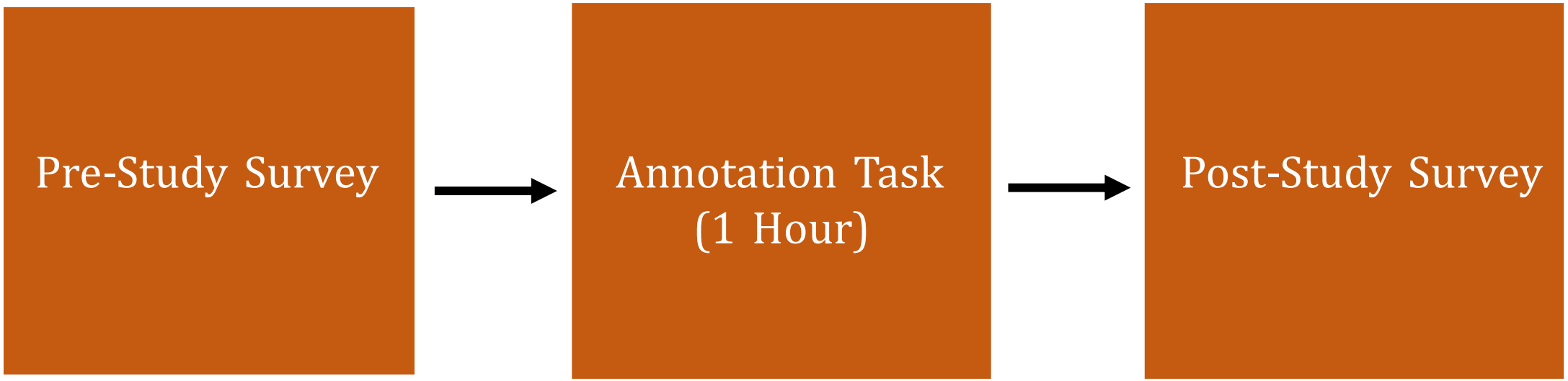


Worker Experience

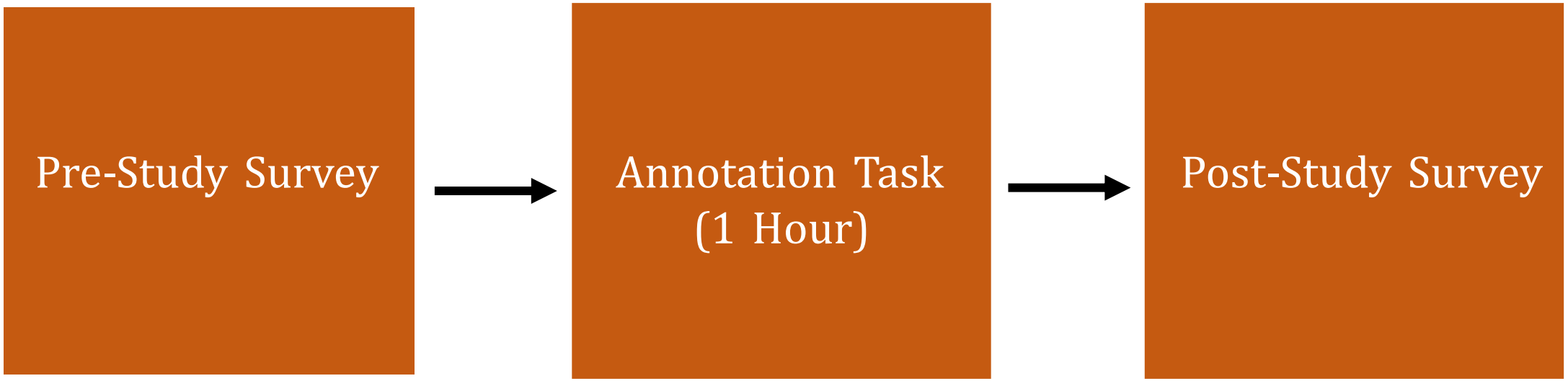
High Experience: >= 6 months VOT Experience
(14 workers)

Low Experience: < 6 months VOT Experience
(26 workers)

Procedure



Procedure

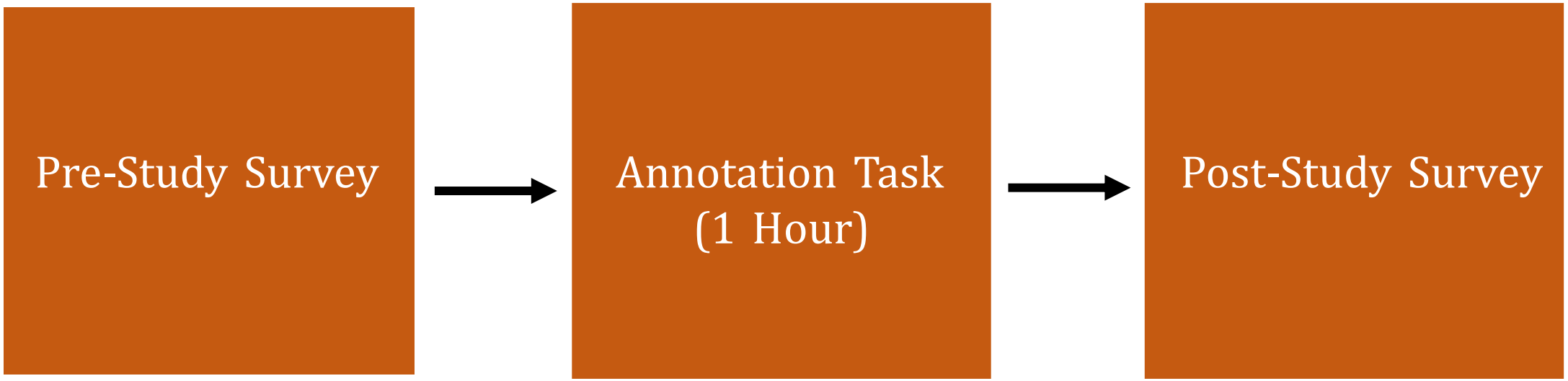


- Demographics
- Experience with VOT
- Estimates of task activity time (Likert)

Procedure

Varied by Task Complexity

Label Constraint x Label Granularity



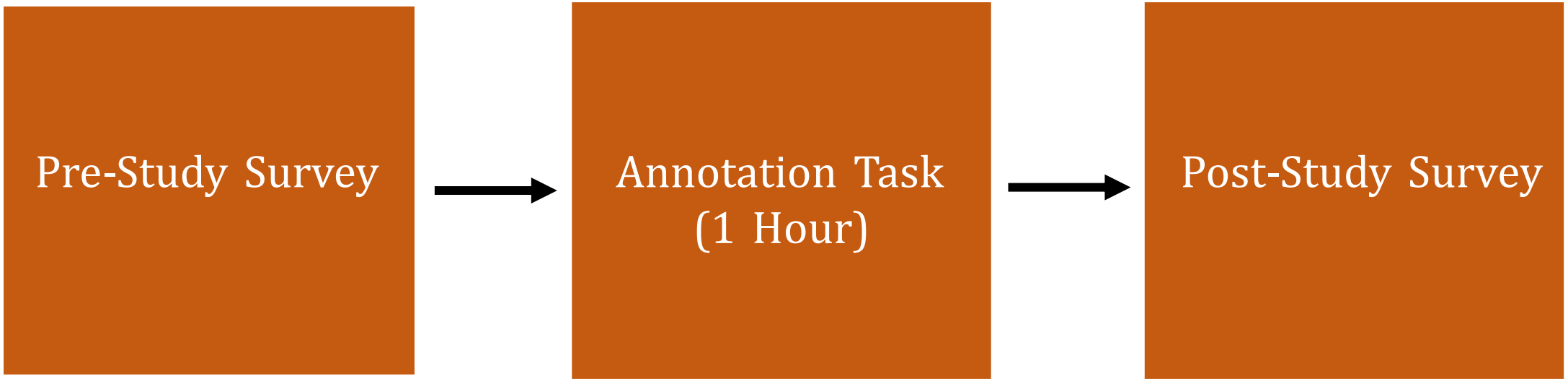
- Demographics
- Experience with VOT
- Estimates of task activity time (Likert)

- Counts and durations of telemetry events collected through task interface

Procedure

Varied by Task Complexity

Label Constraint x Label Granularity



- Demographics
- Experience with VOT
- Estimates of task activity time (Likert)

- Counts and durations of telemetry events collected through task interface (more details on next slide)

- Perceptions of time spent preparing for task (Likert)
- Perceptions of time spent during task (Likert)

Telemetry Counts

- Manual Creation
- Edit
- Delete
- Predict Next
- Copy Frame
- Zoom In
- Zoom Pan
- Frame Navigation (Forward and Backward)

LABEL

ZOOM

NAVIGATE

Telemetry Durations

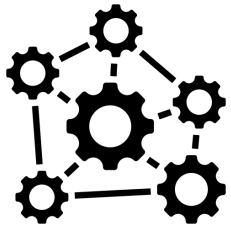
- Manual Creation
- Edit
- Play Video

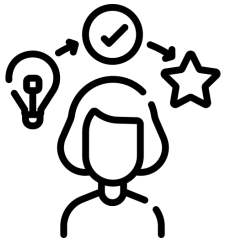
LABEL

NAVIGATE

We ran GLMs to analyze the impact of task complexity and worker experience (ind. vars) on our recorded count and duration variables (dep. vars).

We also developed state diagrams to visualize the most common paths between annotation activities by different sets of workers.



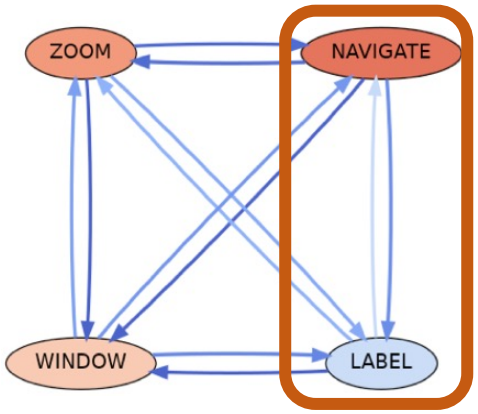


How do **task complexity** and **worker experience** impact:

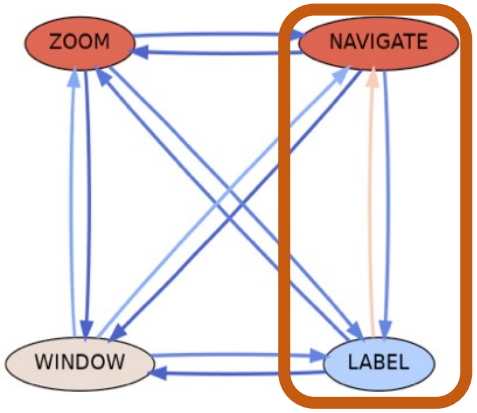
1. Time spent performing data annotation

2. Worker perceptions of time spent performing data annotation and other assistive activities

The Effect of Label Constraint on Time Spent



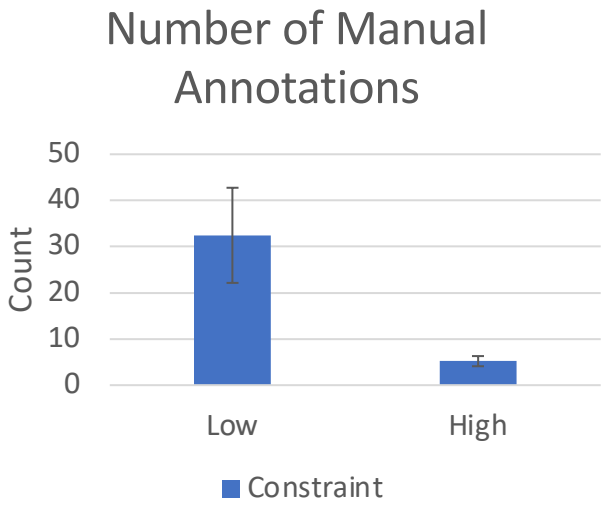
Low Constraint

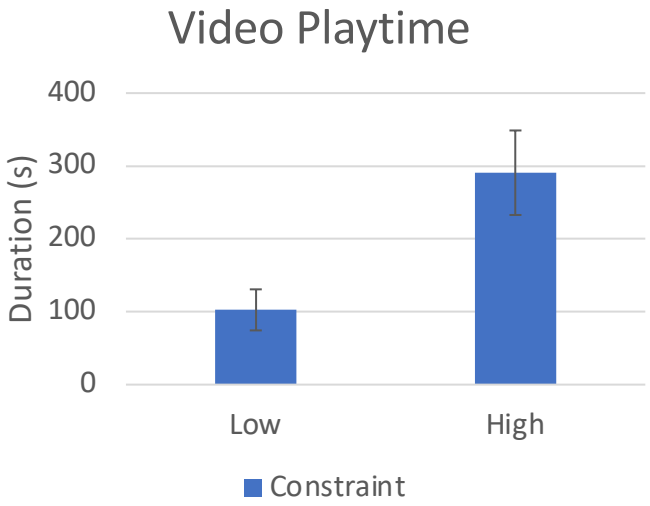


High Constraint

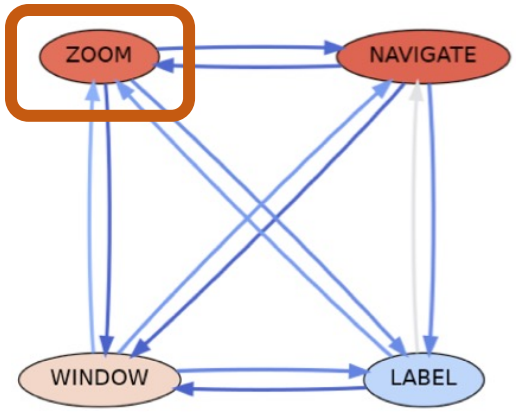
Workers assigned to a high constraint task created **less manual annotations** and spent **more time navigating** through the frames.

High constraint tasks also led to workers transitioning directly from labeling to navigation more frequently.

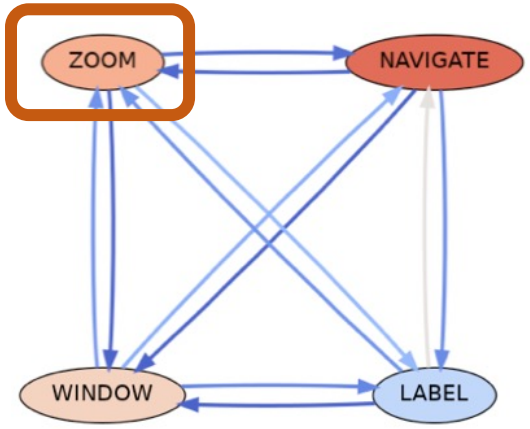




The Effect of Label Granularity on Time Spent

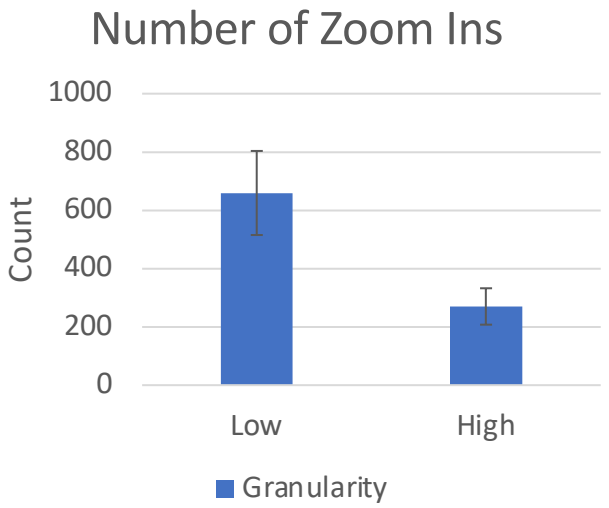


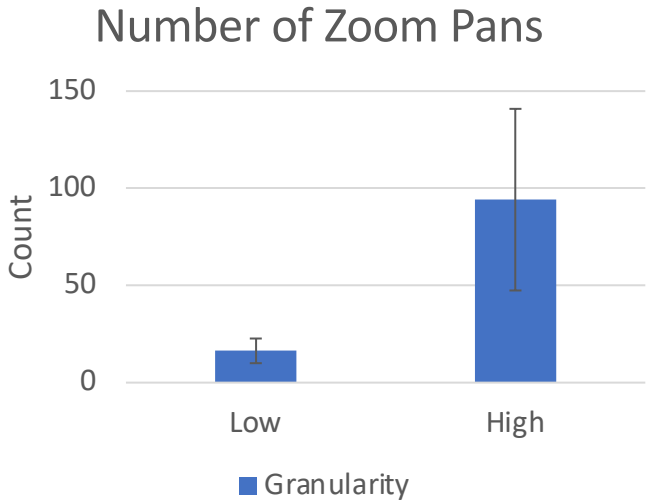
Low Granularity



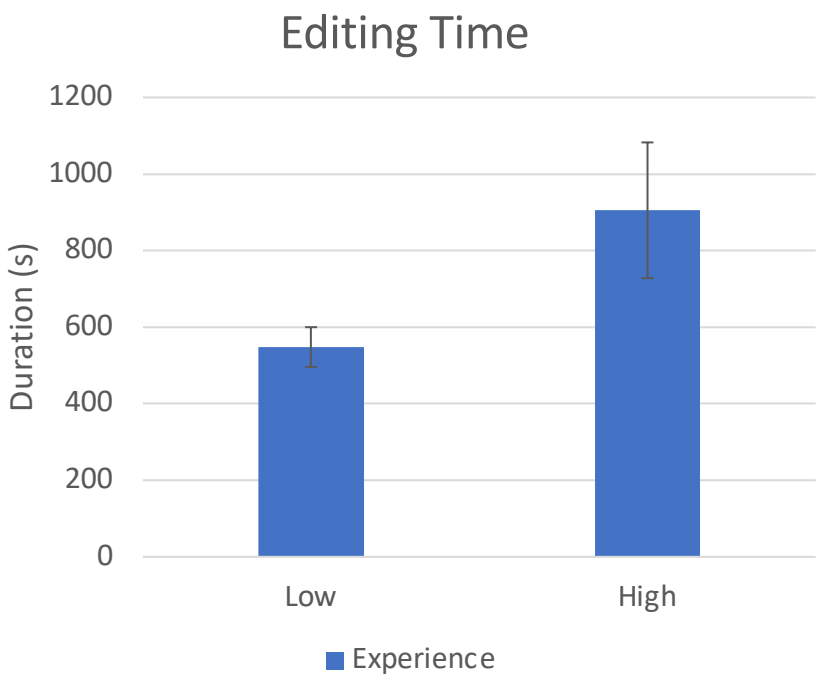
High Granularity

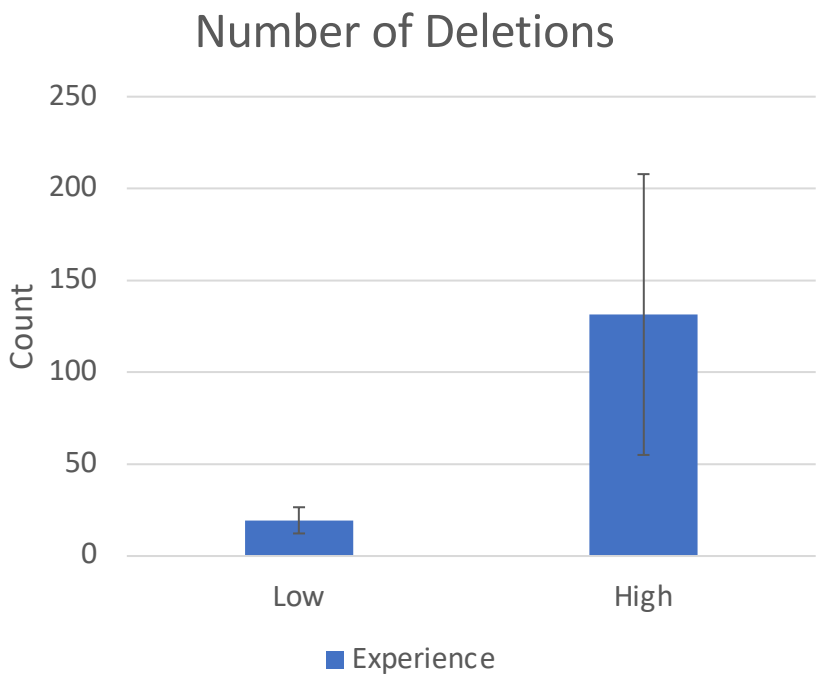
Granularity impacted worker zoom behaviors.



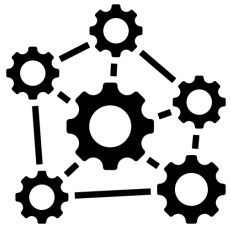


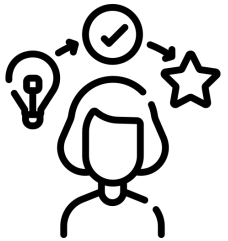
The Effect of Worker Experience on Time Spent





Workers with higher experience with VOT tasks spent **more time editing** annotations and **deleted more annotations**





How do **task complexity** and **worker experience** impact:

1. Time spent performing data annotation

2. Worker perceptions of time spent performing data annotation and other assistive activities

Worker Perceptions of Time and Assistive Activities

- The top 3 activities that workers perceived to spend their time on were:
  - (1) zooming/panning in the interface,
  - (1) playing through the video frames, and
  - (3) performing assistive activities *prior* to working on the task

- Even though the average participant spent 3 times the amount of time editing annotations than playing through the video, editing annotations was overall ranked 5th in terms of time consumption

- 90% of participants engaged in assistive activities prior to the task, and 65% engaged in these activities during the task

- We found no impact of task complexity or worker experience on worker perceptions of time during the task or during assistive activities

Immediate Implications

- Providing helpful nudges to workers who are starting out or having trouble with what to do next
- Better incorporation of resources directly into task interface for easier reference
- Greater understanding needed of the relationship between time spent and label quality, especially for experienced annotators

Discussion

- Rethinking how we capture time spent
- Tracking time vs. surveillance
- Assisting worker productivity while considering worker differences

Thank you!

contact: arechke@purdue.edu

Read the full paper here!

Alex C. Williams
AWS AI, Amazon

Matthew Lease
The University of Texas at Austin and Amazon

Li Erran Li
AWS AI, Amazon

I will be looking for postdoctoral positions within the next 1-1.5 years. Please come chat with me if you are interested in designing systems for crowd worker well-being, collective action, or social computing that make use of crowd intelligence!