October 9th, 2020

Real-Time Speech Workload Estimation for Intelligent Human-Machine Systems

Julian Fortune
Oregon State University
Corvallis, OR

Dr. Jamison Heard
Rochester Institute of Technology
Rochester, NY

Dr. Julie A. Adams
Oregon State University
Corvallis, OR

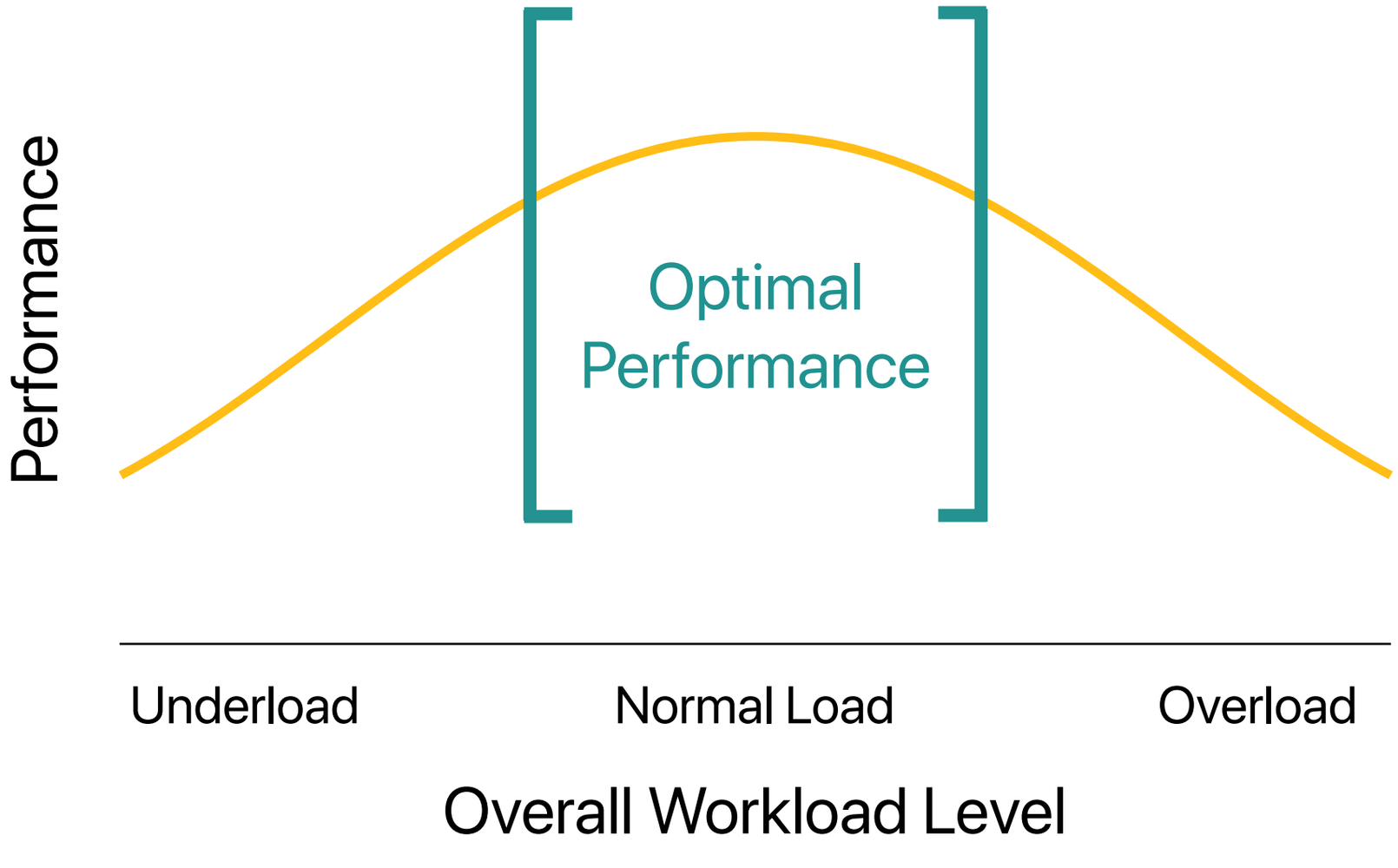
Problem statement

Find a means of estimating speech workload objectively in real-time.

- We want optimal performance from *human-machine teams*.
- The human-machine interface can maximize the human's performance by adapting interactions.

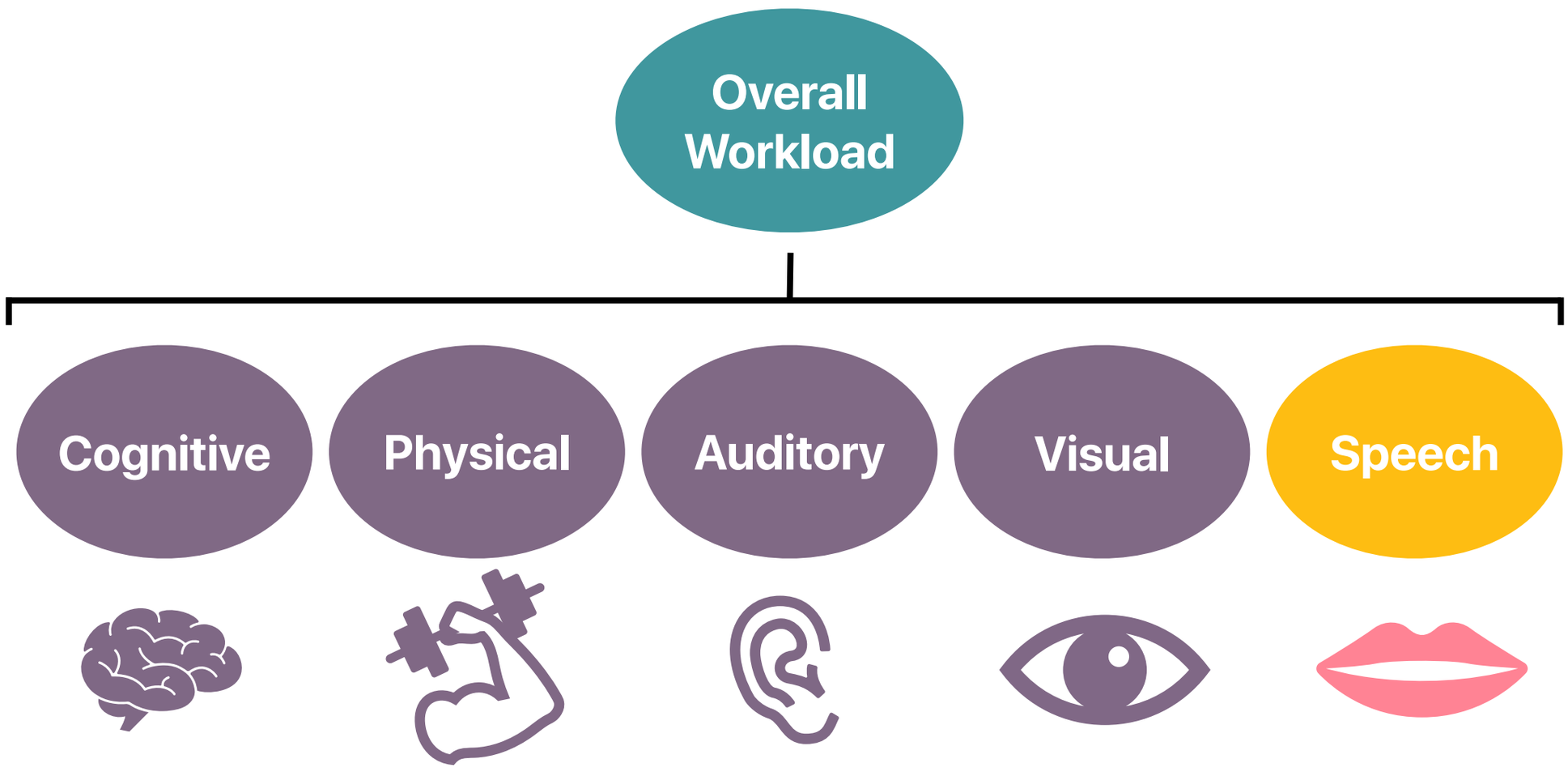
Performance and Workload

$$\text{Workload} = \frac{\text{Resources demanded}}{\text{Resources available}}$$



(Wickens, Lee, Liu, and Gordon-Becker, 2003)

Workload



(McCracken and Aldrich, 1984), (Mitchell, 2000)

Existing Speech Workload Algorithms

Speech Workload Estimation for Air Traffic Control
(Luig and Sontacchi, 2010)

PHYSIOPRINT
(Popovic, Stikic, Rosenthal, Klyde, and Schnell, 2015)

❌ Involved discrete classifications.

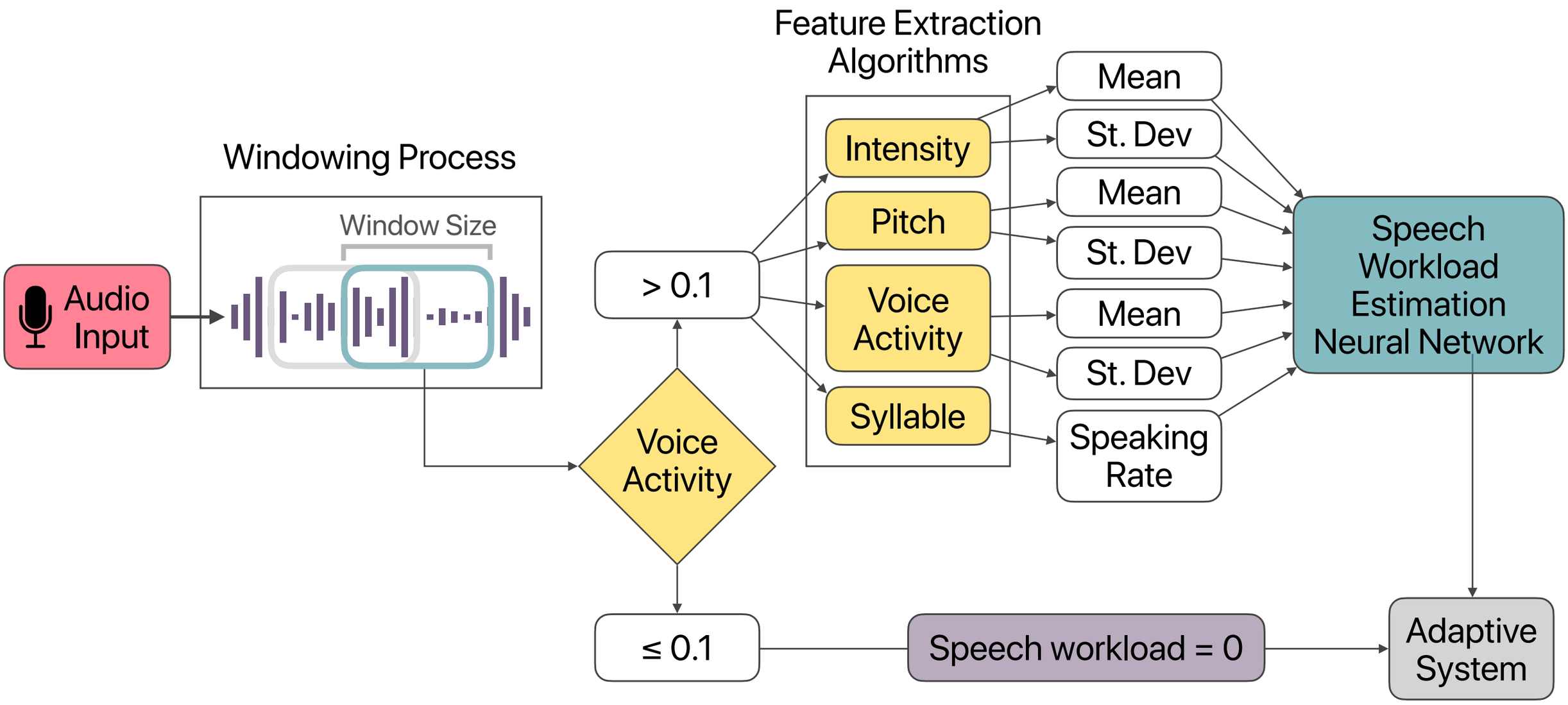
❌ Unable to demonstrate generalizability between individuals.

❌ Did not function in real-time.

Objective Speech Features

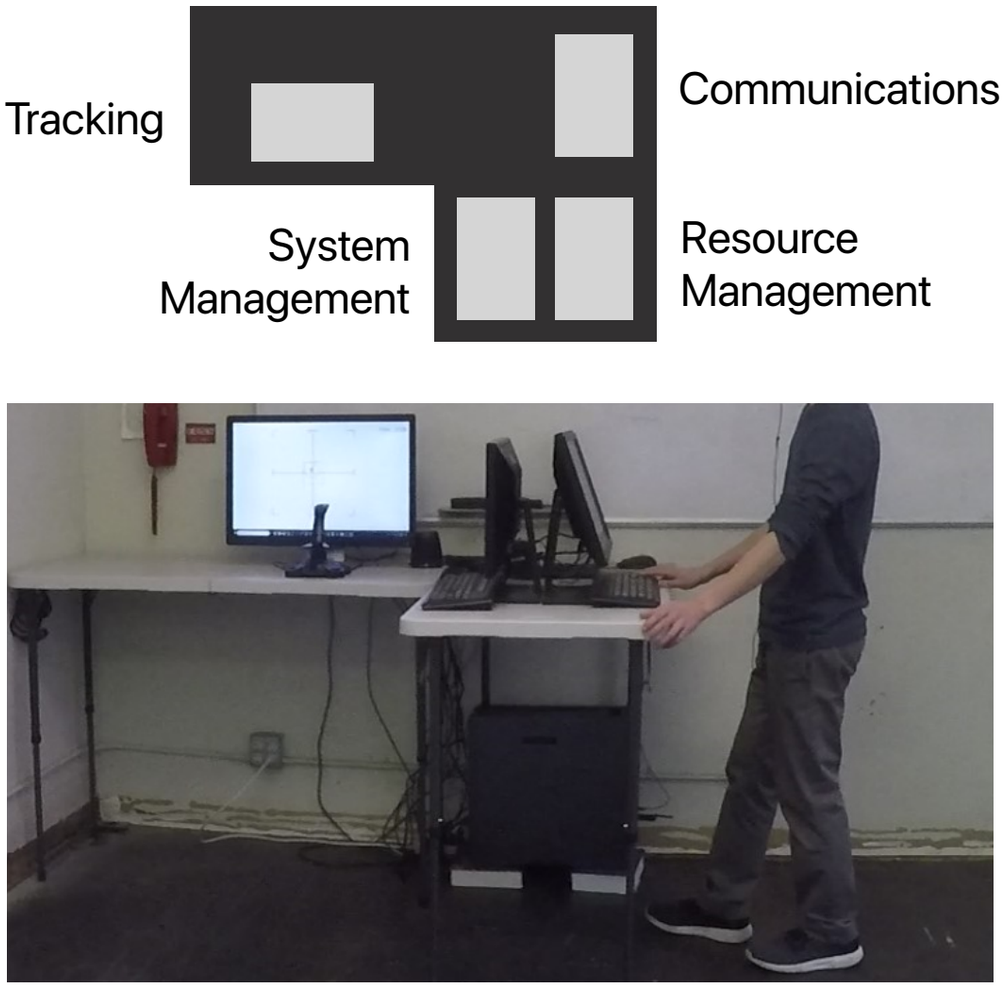
Feature	Correlation
Intensity	Increases
Pitch	Increases
Voice activity	Increases
Speaking rate (syllables/second)	Increases

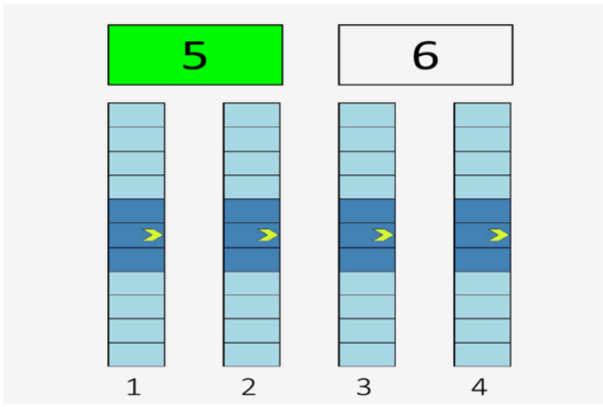
Speech Workload Estimation Algorithm



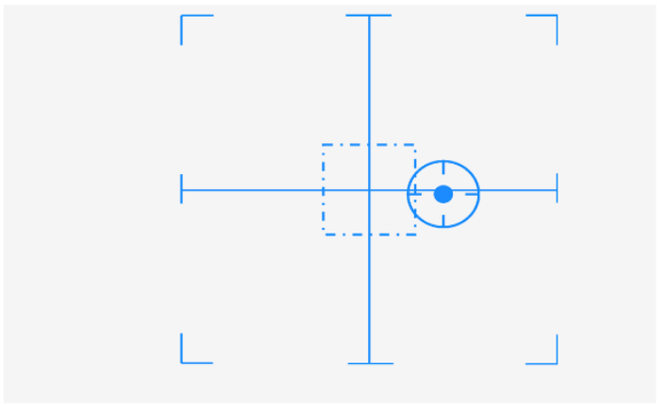
Real-Time Evaluation

- Physically separated NASA MATB-II.
- One 52.5-min trial: seven consecutive 7.5-min workload conditions.

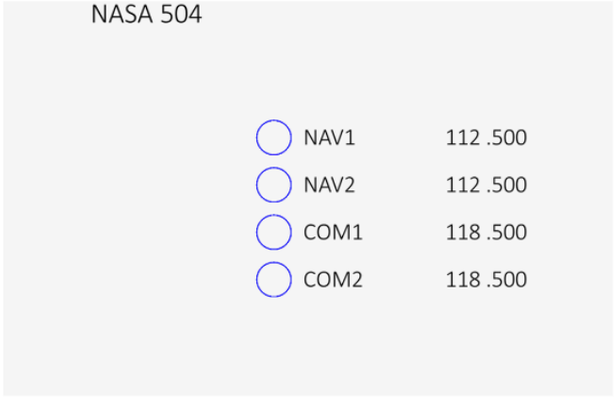




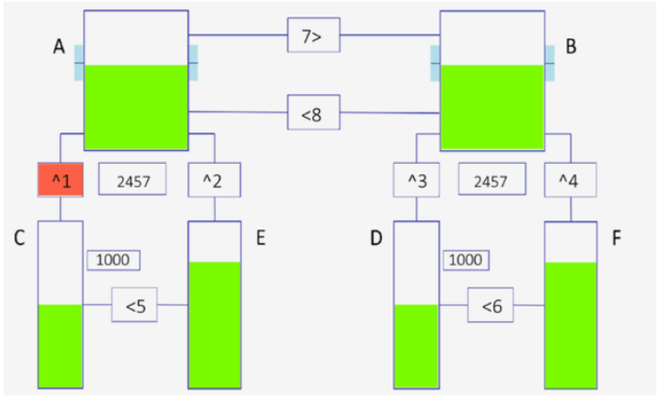
(a) System Management (SYSMAN).



(b) Tracking.



(c) Communications (COMM).



(d) Resource Management (RMAN).

(Heard, 2019)

Hypotheses

H_1: The **correlation** between the algorithm's estimates and the IMPRINT Pro speech workload predictions will **increase** as the window size increases.

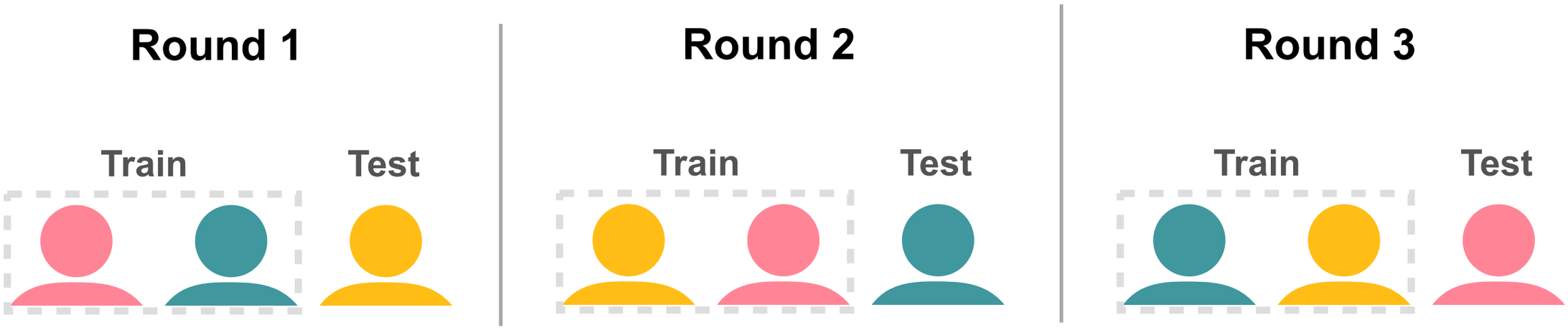
H_2: The **RMSE** of the algorithm's estimates, when compared to the IMPRINT Pro speech workload predictions, will **decrease** as the window size increases.

H_3: The **time** required to calculate the features will **increase** as the window size increases, but will **remain less than 1s**.

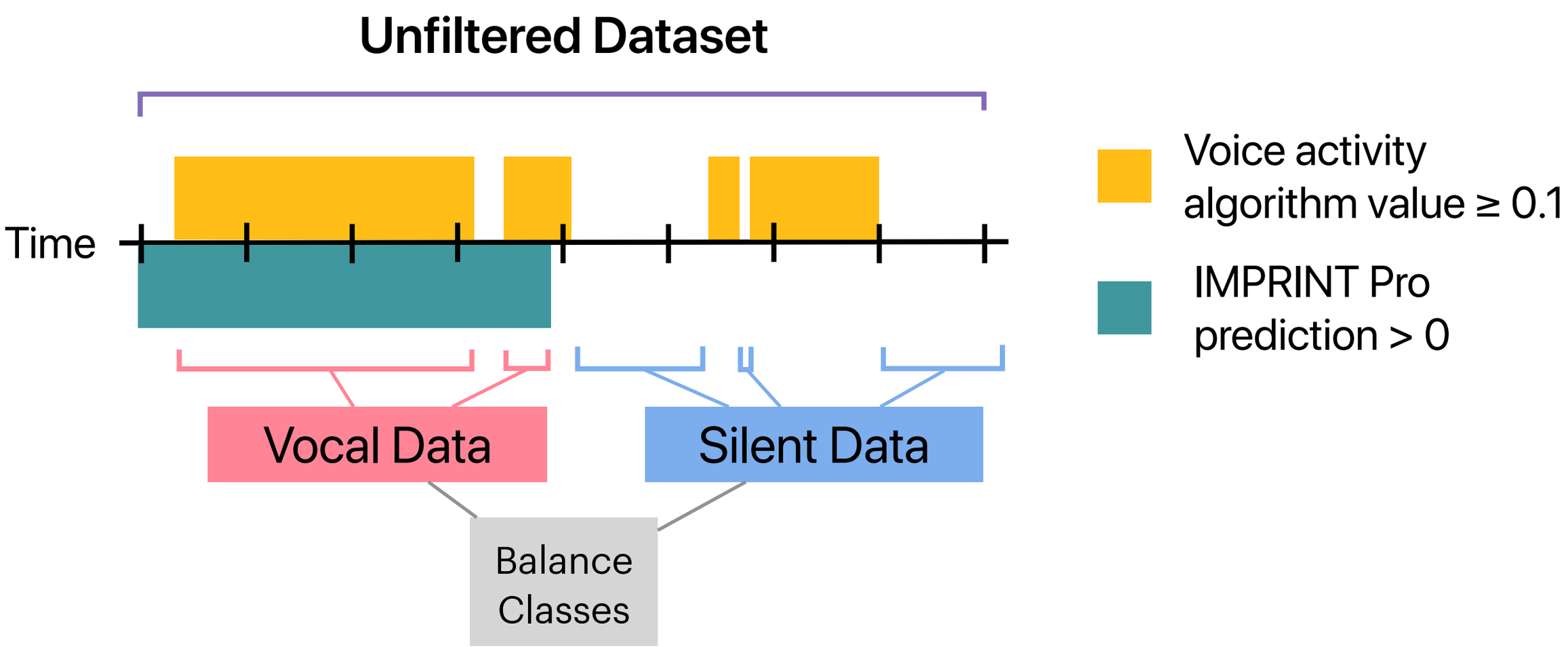
Methodology

- The investigated window sizes were 1s, 5s, 10s, 15s, 30s, and 60s.
- For each window size:
  - Run-time was recorded for all four features.
  - Estimation accuracy was assessed by *leave-one-participant-out cross-validation*.

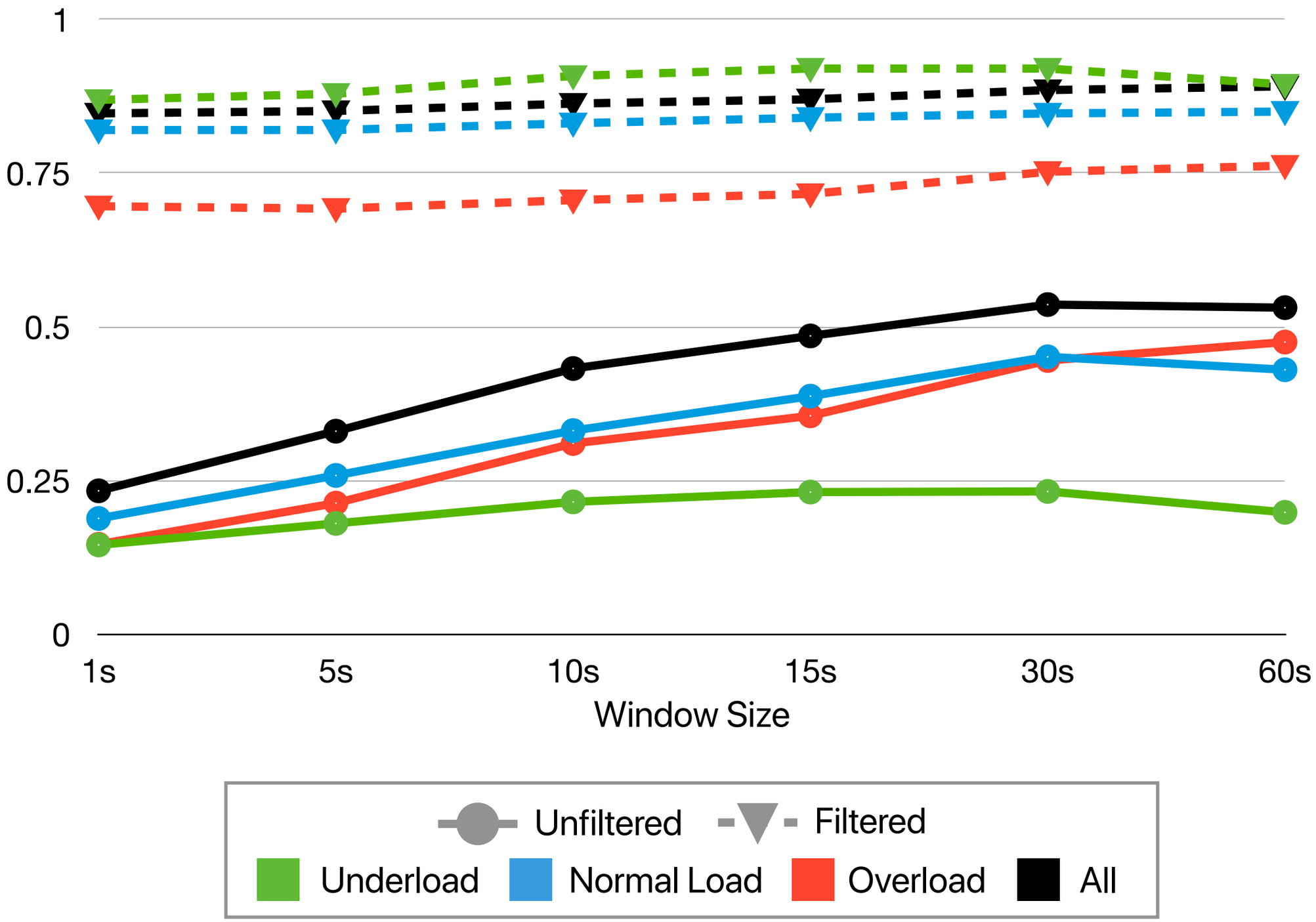
Leave-one-participant-out cross-validation example:



Filtered and Unfiltered Datasets



The Correlation Between the Algorithm's Estimates and the IMPRINT Pro Model Predictions by Window Size



Run-Time of Feature Extraction by Window Size

Feature	Window Size					
	1s	**5s**	**10s**	**15s**	**30s**	**60s**
Intensity	.001 (.00)	.007 (.00)	.013 (.00)	.019 (.00)	.038 (.00)	.069 (.01)
Pitch	.051 (.02)	.246 (.08)	.490 (.15)	.734 (.22)	1.46 (.44)	2.84 (.88)
Voice Activity	.004 (.00)	.024 (.00)	.049 (.00)	.074 (.00)	.149 (.00)	.286 (.02)
Speech Rate	.004 (.00)	.024 (.00)	.047 (.00)	.070 (.00)	.139 (.00)	.258 (.03)
All Features	.061 (.02)	.301 (.08)	.599 (.15)	.897 (.22)	1.78 (.44)	3.45 (.88)

Mean (St. Dev.) run-time in seconds. **Note**: Cells shaded in green *support* H_3 and unshaded cells *do not support* H_3.

Conclusion

H_1: The **correlation** between the algorithm's estimates and the IMPRINT Pro speech workload predictions will **increase** as the window size increases.

⚠️ **Partially supported**

H_3: The **time** required to calculate the features will **increase** as the window size increases, but will **remain less than 1s**.

🚫 **Not supported**

Lessons Learned

- Overall, a window size of **15s** is the most feasible size for **real-time** applications.
- A window size of **30s** is the most reliable for **offline** speech workload estimation.
- Speech workload estimation in real-time can be incorporated in human-machine systems.

Applications

Thank You

Live Q & A

Acknowledgments

This work was partially supported by NASA Cooperative Agreement No. NNX16AB24A and DoD Contract Number W81XWH-17-C-0252 from the CDMRP Defense Medical Research and Development Program.