COLLEGE OF ENGINEERING

Electrical Engineering and Computer Science

REAL-TIME SPEECH WORKLOAD ESTIMATION

Julian Fortune
Dr. Jamison Heard †
Dr. Julie A. Adams

PROBLEM STATEMENT

Designing, developing, and validating an algorithm for estimating speech workload with the following attributes:

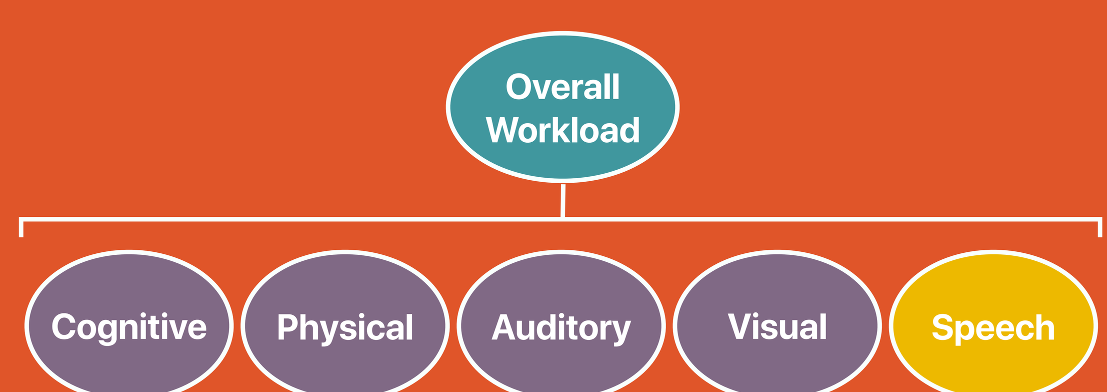
- Employs objective measures (e.g., pitch).
- Calculates a continuous speech-workload value.
- Invariant to individuals, human-robot teaming paradigms, and task environments.
- Sufficiently computationally efficient to be used in real-time.

Existing speech workload algorithms have limitations, and do not operate in real-time.

SPEECH WORKLOAD

Workload is the ratio of resources demanded by a task to the resources a human has available to allocate to the task.

- Comprised of multiple components, including speech workload.
- **Speech workload** occurs when an individual is required to use their voice to produce speech.
- Performance decreases in underload and overload conditions.



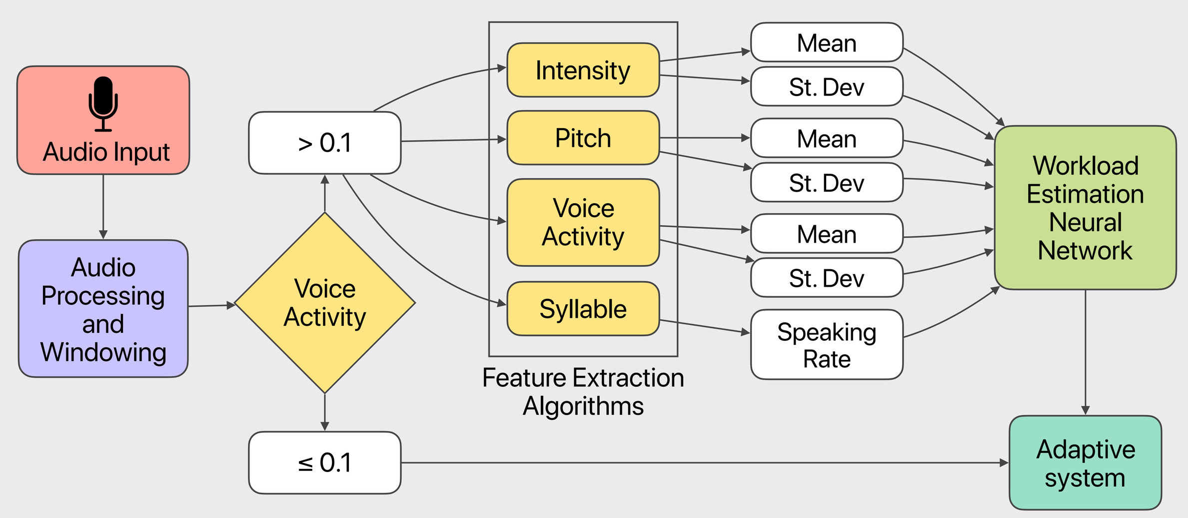
OBJECTIVE SPEECH WORKLOAD METRICS

Metric	Response to increasing speech workload	Extraction method
Intensity	Increases	Root-mean square (RMS)
Intensity Variation	Increases	RMS St. Dev
Pitch	Increases	Auto-correlation
Pitch Variation	Increases	Auto-correlation St. Dev
Speaking rate	Increases	Voiced Peaks

Oregon State University

† RIT Rochester Institute of Technology

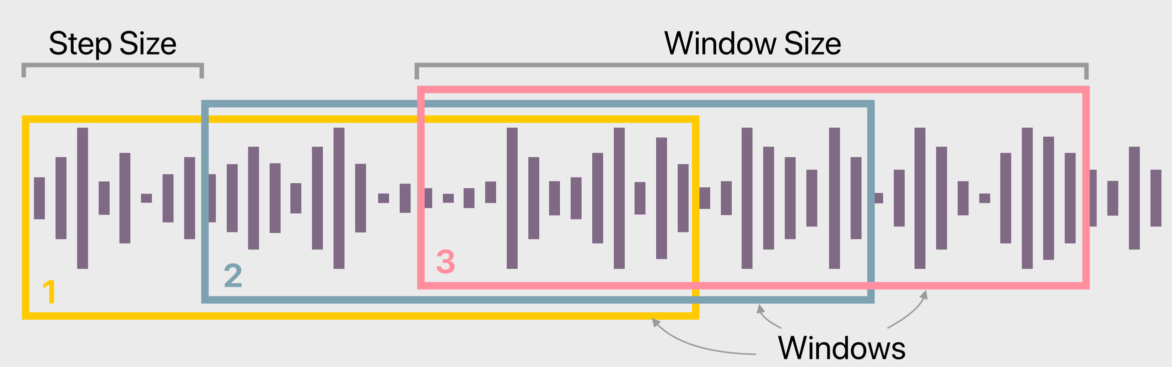
ALGORITHM



Audio Processing and Windowing

Feature extraction and speech workload estimation require audio *windows* at regular steps, looking back over a set duration.

- A step size of 1 second (s)
- A window size of 5s

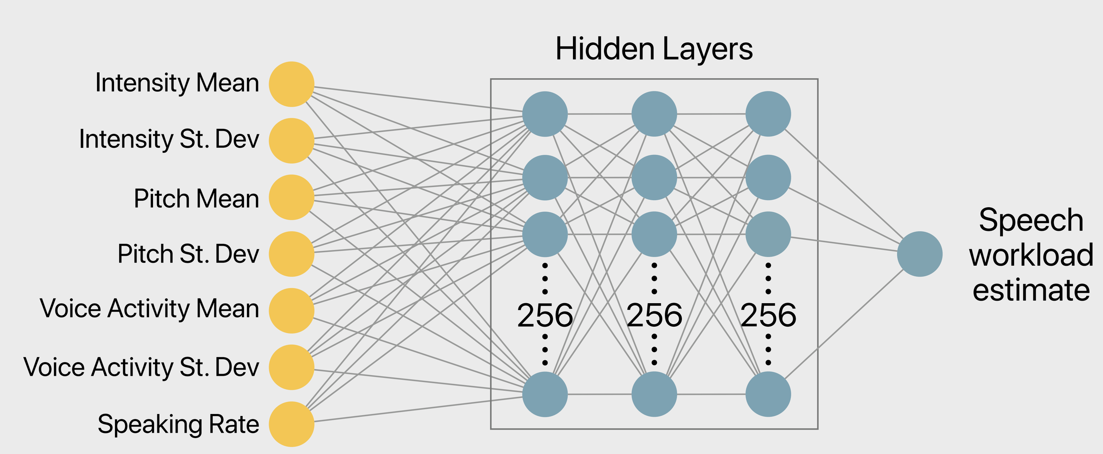


Machine Learning

A neural network with three fully-connected hidden layers and ReLU activation functions.

Trained using Adam optimizer, with a rate of 0.001 and batch size of 64.

The ground-truth labels were produced by IMPRINT Pro, and ranged from 0–4.



HUMAN SUBJECTS EVALUATIONS

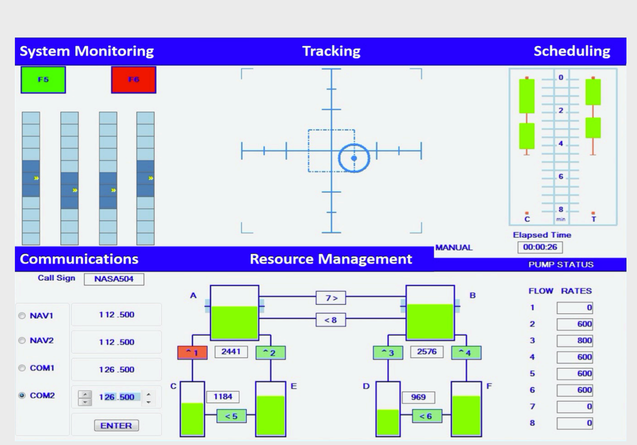


Figure 1: The interface of the NASA MATB-II, the task environment employed for the Supervisory Evaluation.

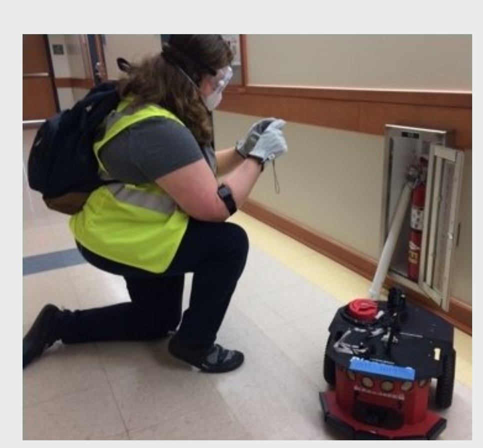


Figure 2: A participant and robot assistant performing a task in the Peer-Based Evaluation.

EMULATED REAL-WORLD CONDITIONS EXPERIMENT

The algorithm was trained on the first Supervisory Evaluation day, and tested on the second day.

The Supervisory Evaluation involved overseeing processes and performing adjustments via computer interface (Figure 1).

- First day: three trials at each workload level: underload, normal load, and overload.
- Second day: one trial where the workload condition transitioned every 5 minutes.

Table 1: The Pearson's correlation coefficient and Root-mean square error (RMSE) between the algorithm's estimates and the IMPRINT Pro model predictions by speech only restriction and workload condition.

Speech only?	Condition	Correlation	RMSE
No	Underload	0.144**	1.238
	Normal Load	0.046**	1.819
	Overload	0.008	2.494
	All	0.088**	1.859
Yes	Underload	1.000**	1.295
	Normal Load	1.000**	0.005
	Overload	1.000**	0.006
	All	0.929**	0.812

Lessons Learned

The speech workload estimation algorithm is capable of estimating speed workload accurately.

HUMAN-ROBOT TEAMING PARADIGM GENERALIZABILITY EXPERIMENT

The algorithm was trained on the Peer-Based Evaluation (Figure 2) and tested on the Supervisory Evaluation.

Peer-Based evaluation: search and collect samples in multiple environments aided by a robot assistant.

- Four consecutive tasks, each randomly assigned a workload condition: Low or High.

Table 2: The Pearson's correlation coefficient and RMSE between the algorithm's estimates and the IMPRINT Pro model predictions by speech only restriction and workload condition.

Speech only?	Condition	Correlation	RMSE
No	Underload	0.117**	2.330
	Normal Load	0.005	1.992
	Overload	-0.015*	2.155
	All	0.073**	0.701
Yes	Underload	0.987**	1.022
	Normal Load	0.989**	0.892
	Overload	0.983**	2.330
	All	0.937**	1.992

Lessons Learned

The speech workload estimation algorithm is invariant to the human-robot teaming paradigm and task environment.

REAL-TIME WINDOW SIZE EXPERIMENT

The algorithm's accuracy was assessed via leave-one-participant-out cross-validation using the Real-Time Evaluation.

The run-times of the feature extraction algorithms were recorded.

Table 3: The correlation between the algorithm's estimates and the IMPRINT Pro model speech workload predictions by window size used, condition, and speech only restriction. **Note**: ** represents $p < 0.0001$.

Speech only?	Condition	Window Size					
		1s	5s	10s	15s	30s	60s
No	Underload	0.145**	0.18**	0.215**	0.231**	0.232**	0.198**
	Normal Load	0.188**	0.258**	0.331**	0.387**	0.451**	0.43**
	Overload	0.146**	0.213**	0.31**	0.355**	0.445**	0.475**
	All	0.233**	0.33**	0.432**	0.485**	0.536**	0.531**
Yes	Underload	0.869**	0.879**	0.908**	0.92**	0.92**	0.894**
	Normal Load	0.82**	0.82**	0.831**	0.84**	0.847**	0.85**
	Overload	0.696**	0.692**	0.706**	0.716**	0.752**	0.762**
	All	0.847**	0.851**	0.863**	0.87**	0.885**	0.891**

Table 4: The mean (St. Dev.) run-times for each feature and window size, measured in seconds. Mean run-times > .5 are highlighted in yellow, and mean run-times > 1 are highlighted in red.

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)

Lessons Learned

Overall, a window size of 15 seconds is the most feasible size for real-time applications.

A 30s window size is the most reliable for offline estimation.

The algorithm is invariant to individuals.

CONTRIBUTIONS

Developed a speech workload algorithm that is invariant to individuals, human-robot teaming paradigms, and task environments during real-time task executions.

Secondary contributions

- Algorithm can be used offline, post hoc.
- Determination of appropriate window sizes.
- Analyzed physiological metrics' (e.g., respiration rate, filler utterances) relative to improved algorithm performance.

PUBLICATIONS

J. Fortune, J. Heard, and J. A. Adams, "Speech workload estimation for human- machine interaction," 2020. Submitted to the *Human Factors and Ergonomics Society Annual Meeting*.

J. Heard, J. Fortune, and J. A. Adams, "SAHRTA: A supervisory-based adaptive human-robot teaming architecture." *IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, 2020. *arXiv:2003.05823*.

J. Heard, J. Fortune, and J. A. Adams, "Speech workload estimation for human- machine interaction," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1, pp. 277–281, 2019.