Machine Learning for Cognitive Load Inference from Physiological Signals

Veljko Pejović and Martin Gjoreski

Neuroergonomics Conference 2021

Veljko Pejović
Faculty of Computer and Information Science
University of Ljubljana

Martin Gjoreski
Faculty of Informatics
Università della Svizzera italiana

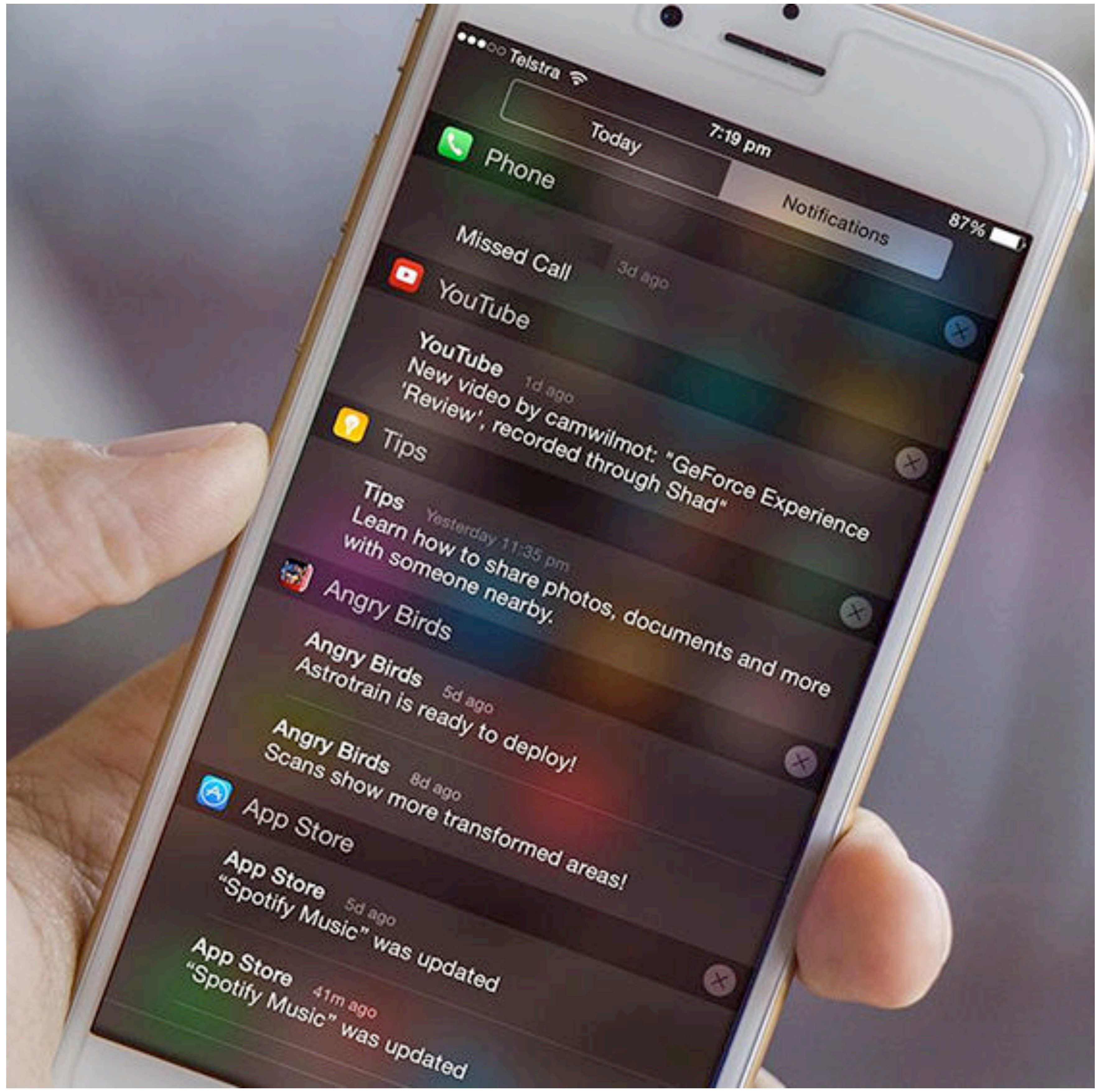
“There is more information available at our fingertips during a walk in the woods than in any computer system, yet people find a walk among trees relaxing and computers frustrating. Machines that fit the human environment instead of forcing humans to enter theirs will make using a computer as refreshing as taking a walk in the woods.”

Mark Weiser 1991

Mobile Interruptiblity Inference

Identifying opportune moments for notification delivery [Ubicomp2014]

- Hypothesis: context reveals interruptibility
- Path: mobile sensors reflect the context
- Implementation: InterruptMe Android library for notification management
    - Sensing location, movement, time
    - Personalized machine learning models of interruptibility



https://bitbucket.org/veljkop/intelligenttrigger

Sensed Context is not Everything

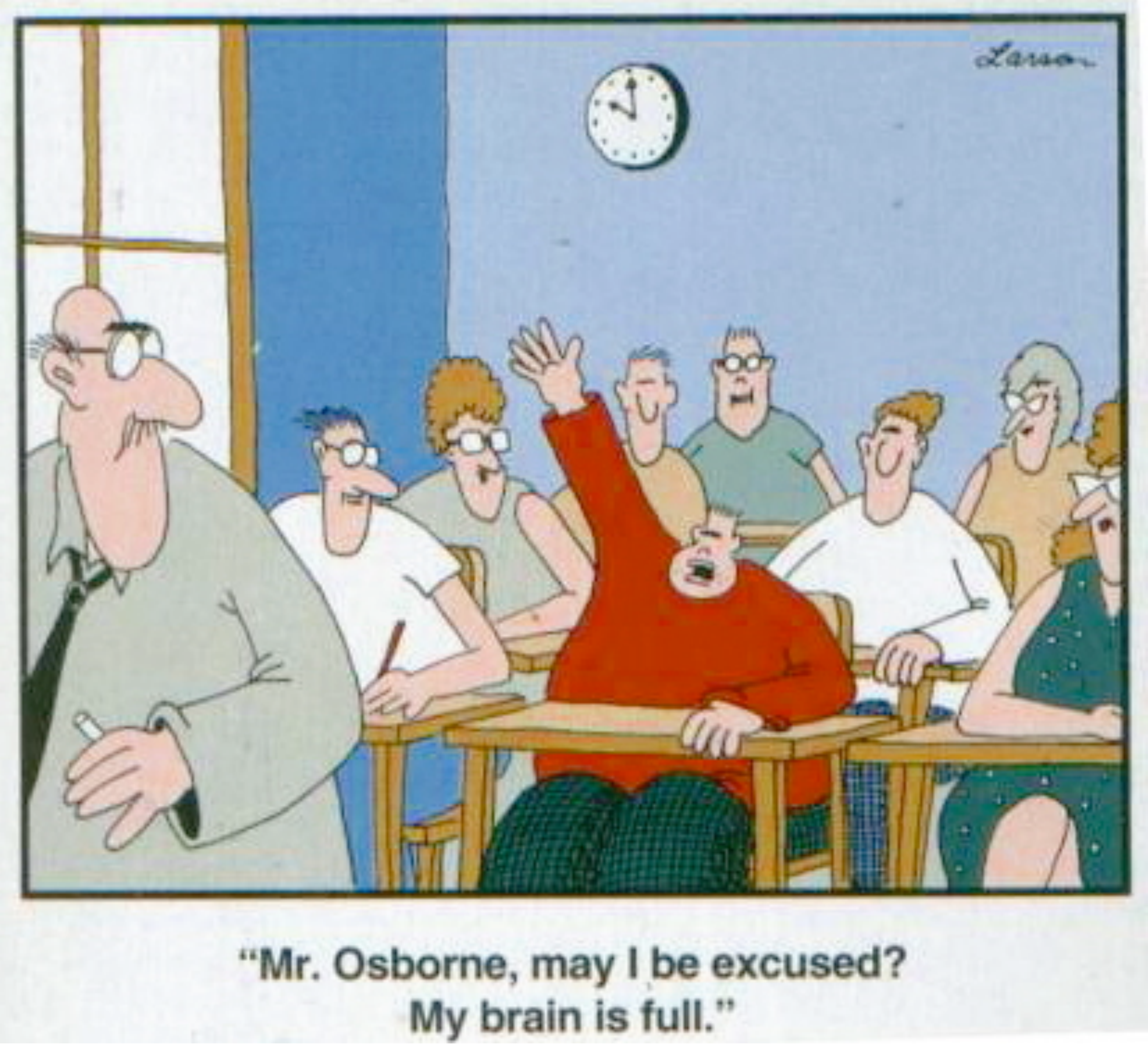
Identifying factors determining interruptibility [CHI2016]

- Hypothesis: factors, beyond those immediately sensed, reflect interruptibility
- Path: experience sampling study of notification handling
- Implementation: two-month study of 20 users, examining app usage/notifications, sender-receiver relationship, a user's task engagement

Role of Current Task Engagement

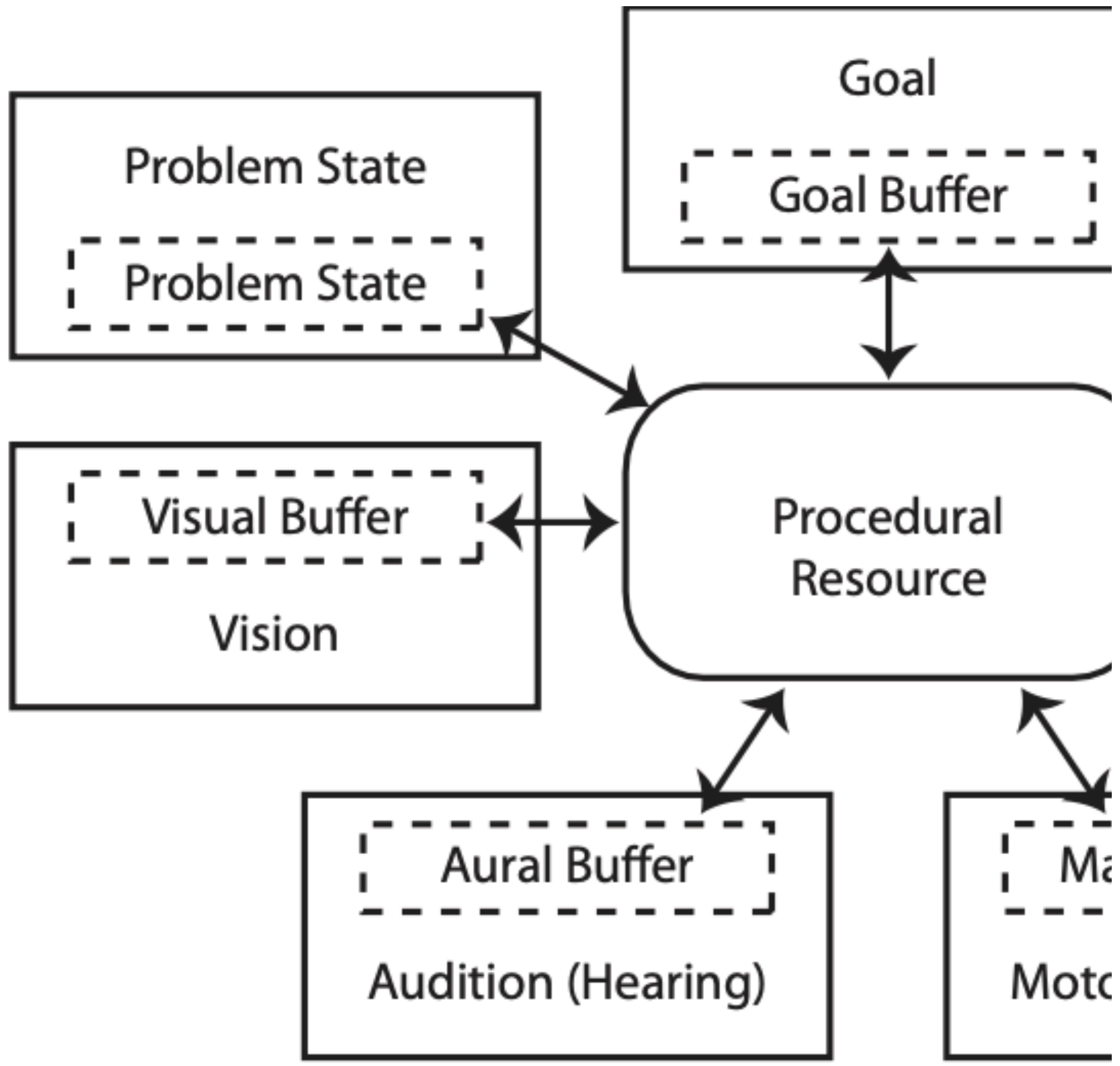
Identifying factors determining interruptibility [CHI2016]

- Notifications more disruptive if arriving when the user is in the middle of or finishing a task
- Perceived disruption increases with the complexity of an ongoing task
- Faster to react if engaged in a complex task



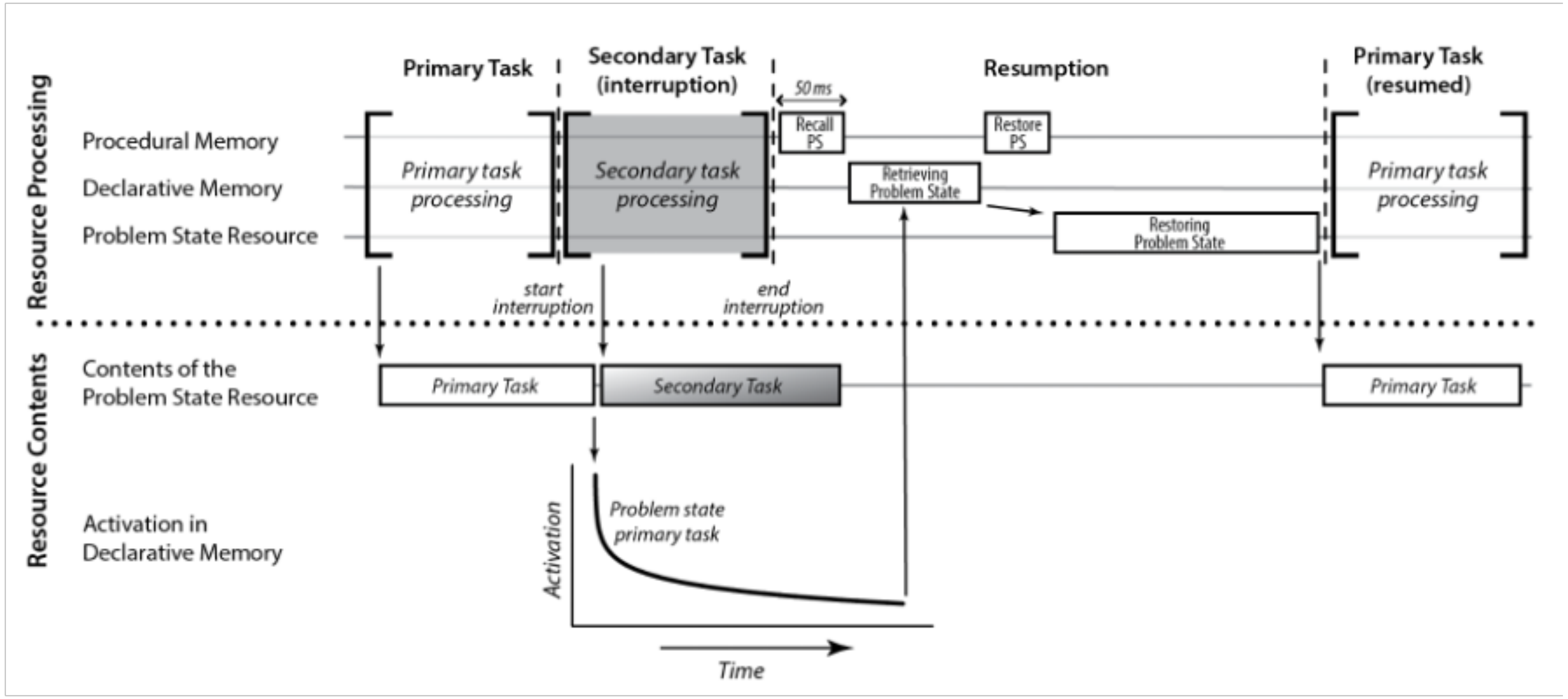
How Does a Thought Get Disrupted?

- Perceptual and motor resources
- Cognitive resources (e.g. procedural memory)
- Resource use is exclusive – one task at a time per resource
- Multiple problem threads run in parallel, but processing is still serial



How Does a Thought Get Disrupted?

- Interference when two or more threads ask for the same resource at a time
- Complex tasks require problem state saving/retrieving



Cognitive load - digging deeper

“Cognitive load is a multidimensional construct representing the load that performing a particular task imposes on the learner’s cognitive system”

Paas and Van Merriënboer, 1994

Inferring Cognitive Load

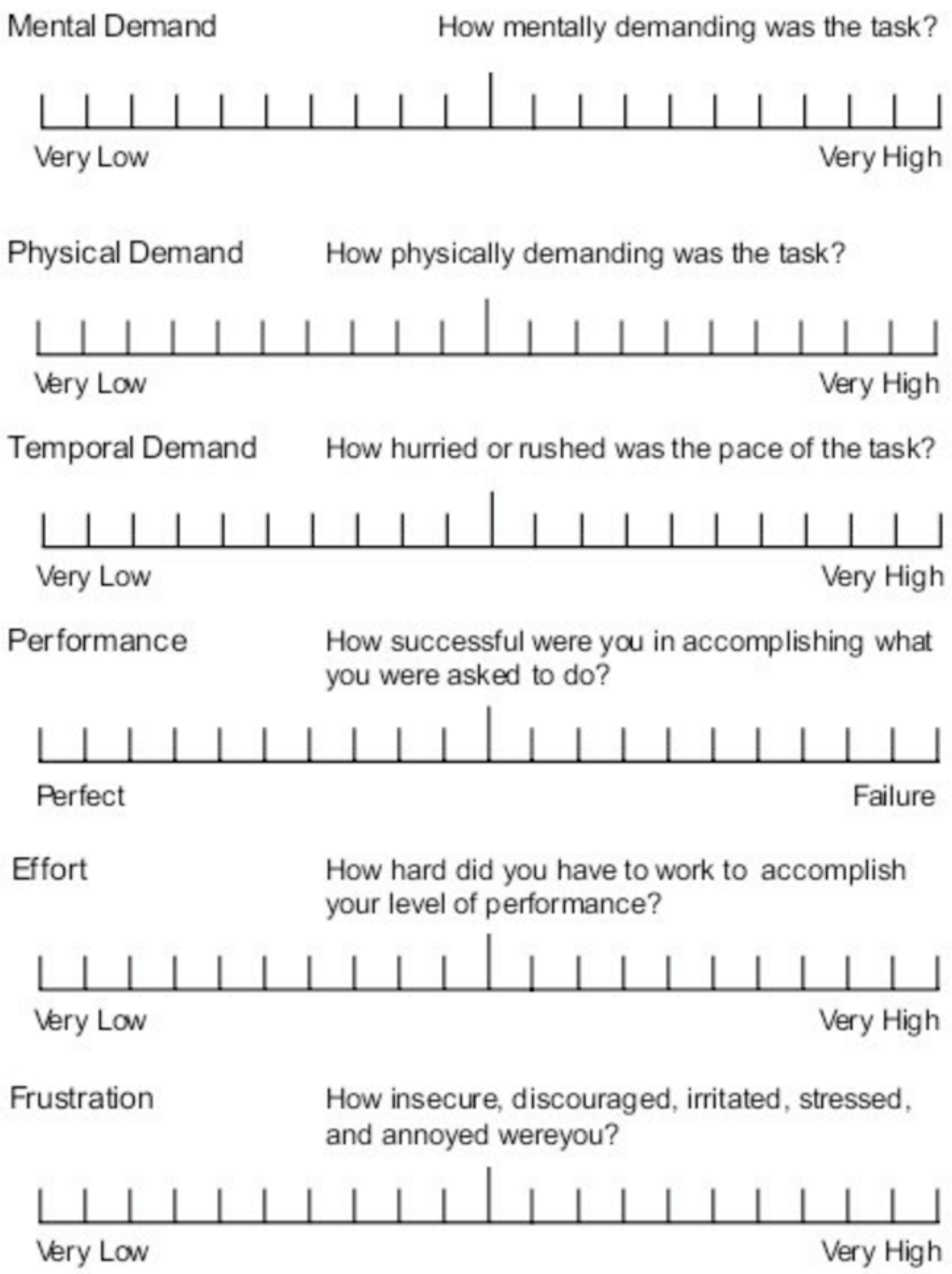
From abstract concepts to measurable phenomena

- Cognitive load dimensions:
  - Intrinsic
  - Extraneous
  - Germane

- Cognitive load assessment through:
  - Mental load
  - Mental effort
  - Performance

Measuring Cognitive Load

NASA-TLX


Mental Demand
How mentally demanding was the task?
Very Low
Very High

Physical Demand
How physically demanding was the task?
Very Low
Very High

Temporal Demand
How hurried or rushed was the pace of the task?
Very Low
Very High

Performance
How successful were you in accomplishing what you were asked to do?
Perfect
Failure

Effort
How hard did you have to work to accomplish your level of performance?
Very Low
Very High

Frustration
How insecure, discouraged, irritated, stressed, and annoyed wereyou?
Very Low
Very High

Physiological Reaction to Cognitive Load

- Pupil dilation
- Heart rate (variability)
- Breathing
- Heat flux

Cognitive Load Inference at Scale

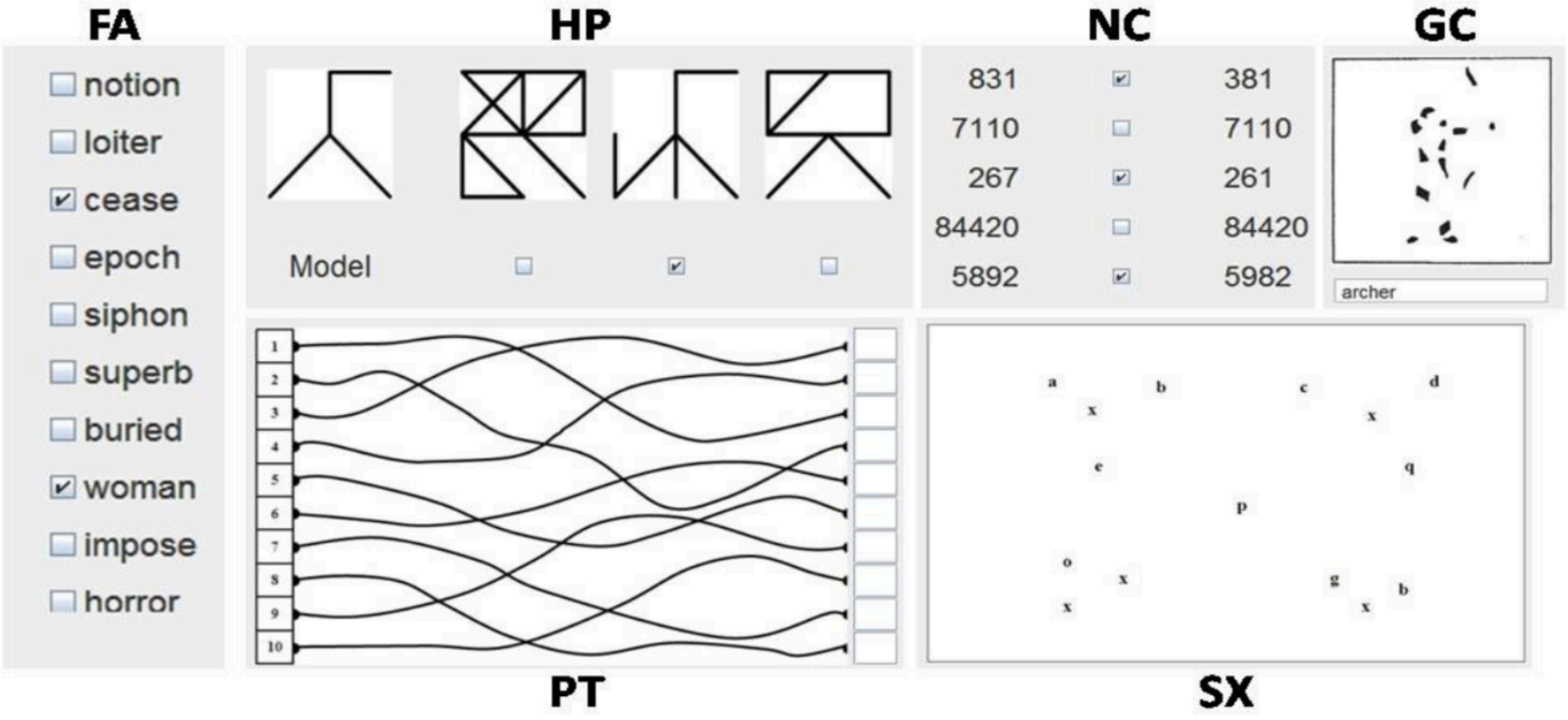
Ubiquitous computing for physiological signal sampling

- Cheap wearables can already capture:
  - PPG - heart rate (variability)
  - Skin temperature
  - Electrodermal activity
- Thermal cameras
- Wireless ranging

Data Collection Experiment

Ubiquitous computing for physiological signal sampling

- Elementary cognitive tasks (ECTs)



Data Collection Experiment

Ubiquitous computing for physiological signal sampling

Part 1	Demographic Questionnaire	2-back task	3 minutes Rest	3-back task	3 minutes Rest	Personality Questionnaire

Part 2	P-task n Intensity x	Task load Quest. + Rest	P-task n Intensity x	Task load Quest. + Rest	P-task n Intensity x	Task load Quest. + Rest	3 minutes Rest	6 cycles
	S-task		S-task		S-task			

- Different ECTs flavors (Easy, Medium, Hard)
- NASA-TLX
- Microsoft Band 2 wristband for physiological signal sampling
  - HR(V), ST, EDA

Can we automatically infer task engagement/cognitive load?

Inferring Cognitive Load - Machine Learning Challenge

Ubittention 2020 workshop

- Dataset split into:
  - Training+validation set
  - Test set
- 13 competition entries - ML pipelines
- Success metrics:
  - Overall accuracy, per subject, per task difficulty, recall, precission, etc.

Inferring Cognitive Load - Machine Learning Challenge

Details of submitted methods

Method	Preprocessing	Features	Feature Selection	Proposed Classifier
I	-	Handcrafted: general	Sequential backward floating search	Ensemble of 7 Gradient boosting decision trees
II	-	Handcrafted: features	-	Support vector machine
III	Standardization (subjectwise)	Handcrafted: general and domain-specific	Sequential forward floating search	Ensemble of support vector machines
IV	Standardization	Handcrafted: general and domain-specific	-	Logistic regression
V	Min-max normalization (overall and subjectwise)	Handcrafted: general and domain-specific (partially automated)	Feature discovery platform	Random forest
VI	Standardization	Handcrafted: general and domain-specific	-	Weighted sum of individualized and global logistic regression models
VII	Min-max normalization	Handcrafted: general and domain-specific	Maximal information coefficient	Multilayer perceptron
VIII	-	Handcrafted: general and domain-specific	-	Logistic regression
IX	Standardization (subjectwise)	Handcrafted: general and domain-specific	Gini impurity	Support vector machine
X	-	Handcrafted: correlation dimension	-	XGBoost classifier
XI	Standardization	Automated: pretrained on a larger dataset	-	CNN (6 layers per sensor data, each layer with batch normalization and ReLU activation)
XII	Min-max normalization (subjectwise)	Automated	-	Recurrent neural networ
XIII	Standardization (subjectwise)	Handcrafted: general and domain-specific	Gini impurity	Logistic regression

Inferring Cognitive Load - Machine Learning Challenge

Overall, per task difficulty, per experiment period performance

Method/Rank	Accuracy
I	0.694
II	0.679
III	0.674
IV	0.663
V	0.653
VI	0.653
VII	0.648
VIII	0.648
IX	0.627
X	0.580
XI	0.560
XII	0.554
XIII	0.503

Method	Designed Task Difficulty			
	Rest	Easy	Medium	Hard
I	0.632	0.679	0.714	0.857
II	0.600	0.714	0.771	0.771
III	0.695	0.679	0.686	0.600
IV	0.632	0.607	0.657	0.800
V	0.663	0.571	0.657	0.686
VI	0.579	0.571	0.686	0.886
VII	0.611	0.643	0.600	0.800
VIII	0.642	0.714	0.629	0.629
IX	0.621	0.571	0.714	0.600
X	0.674	0.429	0.457	0.571
XI	0.453	0.607	0.657	0.714
XII	0.632	0.643	0.457	0.371
XIII	0.516	0.607	0.343	0.543

	Experiment Period	
Method	First half	Second half
I	0.682	0.716
II	0.671	0.682
III	0.706	0.636
IV	0.694	0.625
V	0.671	0.659
VI	0.671	0.602
VII	0.647	0.659
VIII	0.635	0.659
IX	0.647	0.625
X	0.624	0.523
XI	0.588	0.489
XII	0.541	0.534
XIII	0.518	0.500

Inferring Cognitive Load - Machine Learning Challenge

Takeaway points

Dataset augmentation	No significant role
Preprocessing	Subjectwise standardization
ML approach	Ensemble of ML models
Features	Time-frequency domain, statistics, HRV
Feature selection	Sequential backward floating search
Hyperparameters	Bayesian optimization
Evaluation	Stratified subject-aware cross-validation
Other	- High-ranked methods have lower inter-subject accuracy difference. - High-ranked methods perform better for the instances that have a higher designed task difficulty. - Low-ranked methods are more sensitive to the different experiment periods.

References