Indian Institute of Technology Ropar

Rupnagar, Punjab – 140 001

Mukesh Saini

(Assistant Professor, CSE)

Automatic Movie Making

Demo Video

Input Videos **Output Mashup**

How to synchronize videos?

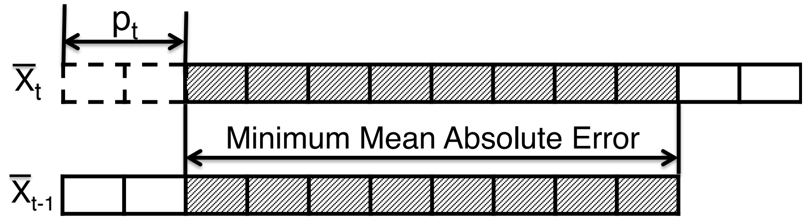
Obtain fingerprint of the audio and match the fingerprints!

How do human editors choose video?

1. Remove videos with bad view quality!
   – Shakiness, occlusion, tilt
2. Create a composition of videos taken from multiple perspectives!
   – Angle, distance

Shakiness

Step 1: Camera pan-tilt



Step 2: Filtered motion

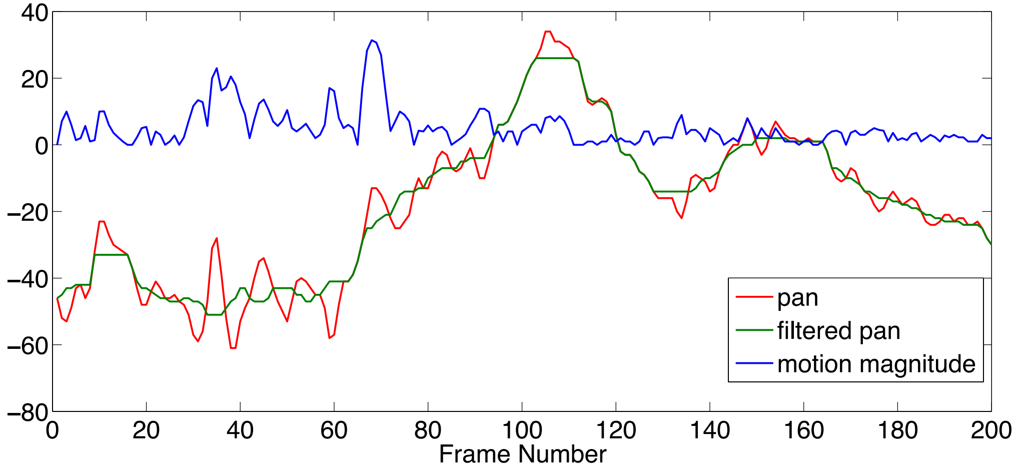
$$\nu_i = \sqrt{(p_i^a - p_i^f)^2 + (\tau_i^a - \tau_i^f)^2}$$

Step 3: Moves

n^p = clusters of shaky frames with $\nu_i > \theta_p$

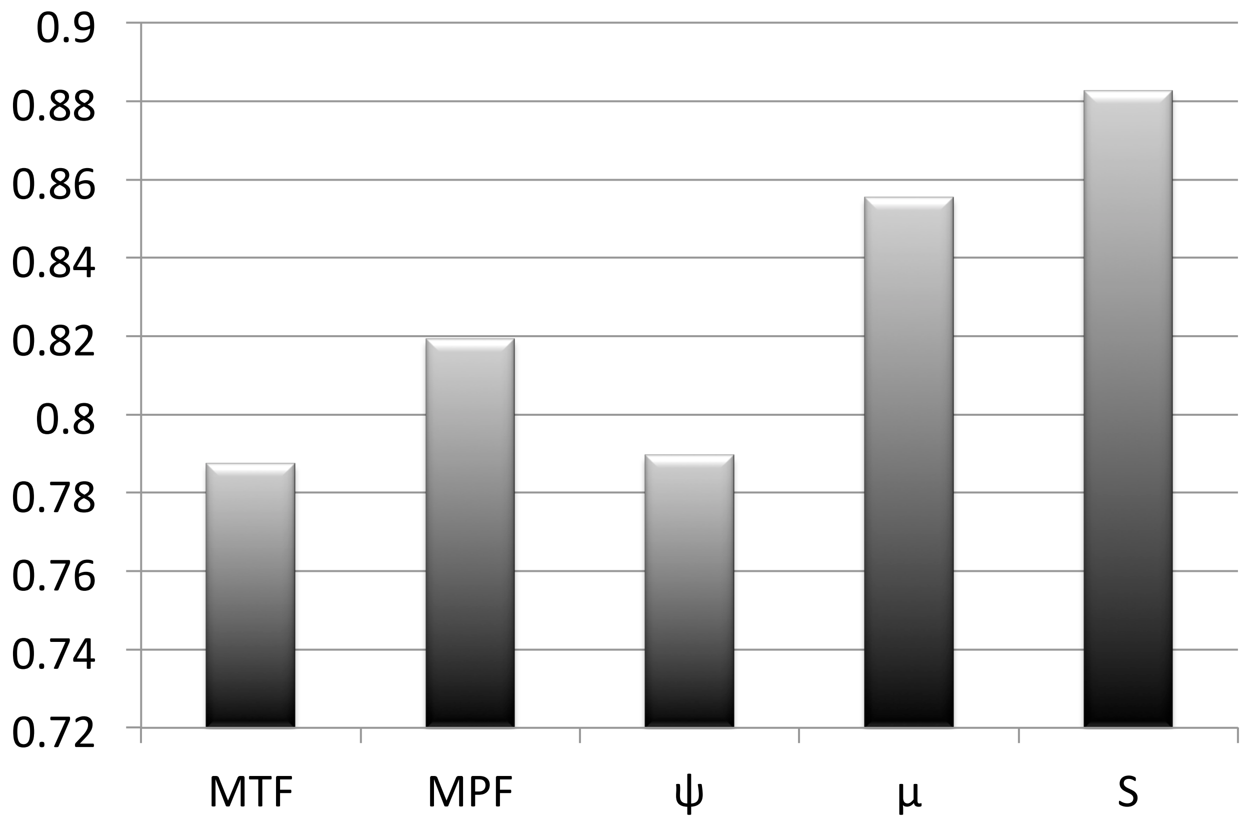
Step 4: Normalization

$$\mu = \frac{n^p}{n}$$



Shakiness Results

$$\mathcal{S} = c + \beta_p * MPF + \beta_t * MTF + \beta_f * \psi + \beta_m * \mu$$



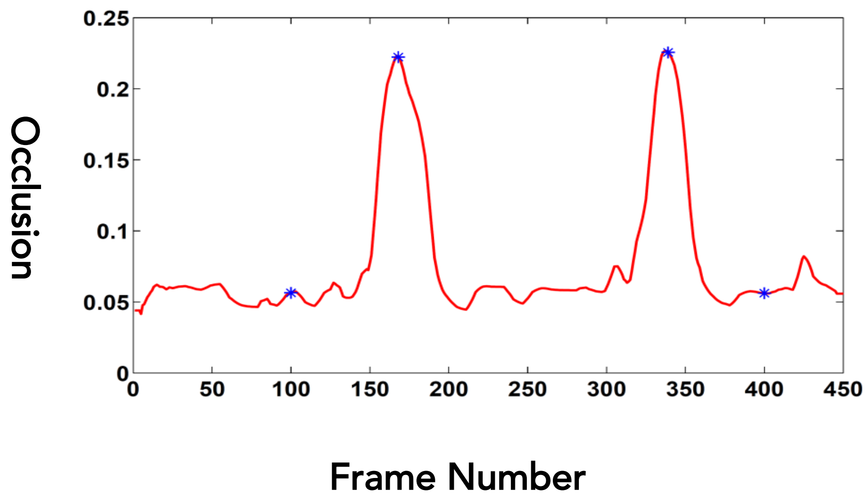
Occlusion

$$I^e(x,y) = \begin{cases} 1 & \text{if edge is detected at pixel } I(x,y) \\ 0 & \text{otherwise} \end{cases}$$

$$I^d = I^e \odot W$$

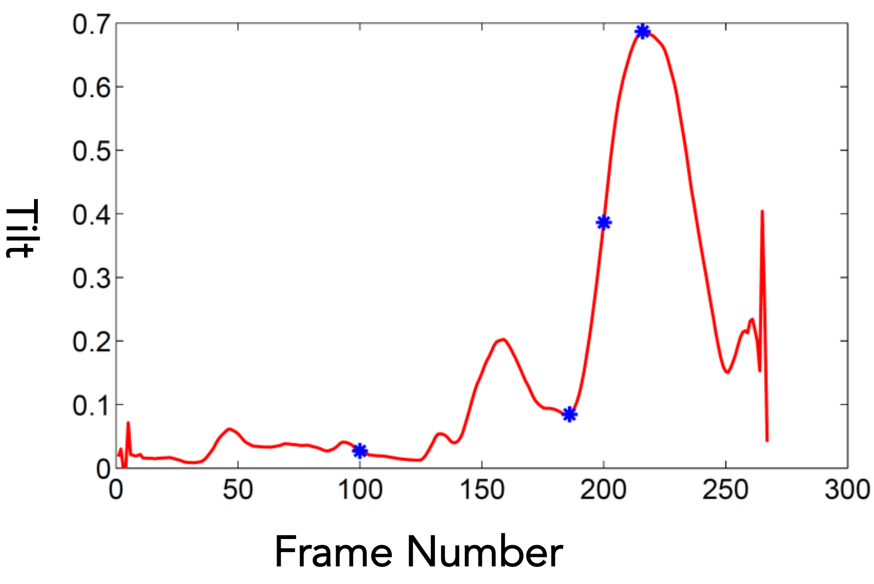
$$I^p(x',y') = \begin{cases} 1 & \text{if the sum of edge densities in the patch } (x',y') \text{ is greater than threshold} \\ 0 & \text{otherwise.} \end{cases}$$

$$f = \frac{\text{No of 1 patches}}{\text{Total number of patches}}$$



Tilt/Rotation

$$S^t = \frac{abs\left(\frac{1}{N^l}\sum_{i=1}^{N^l} o_i * l_i\right)}{\pi/4}$$



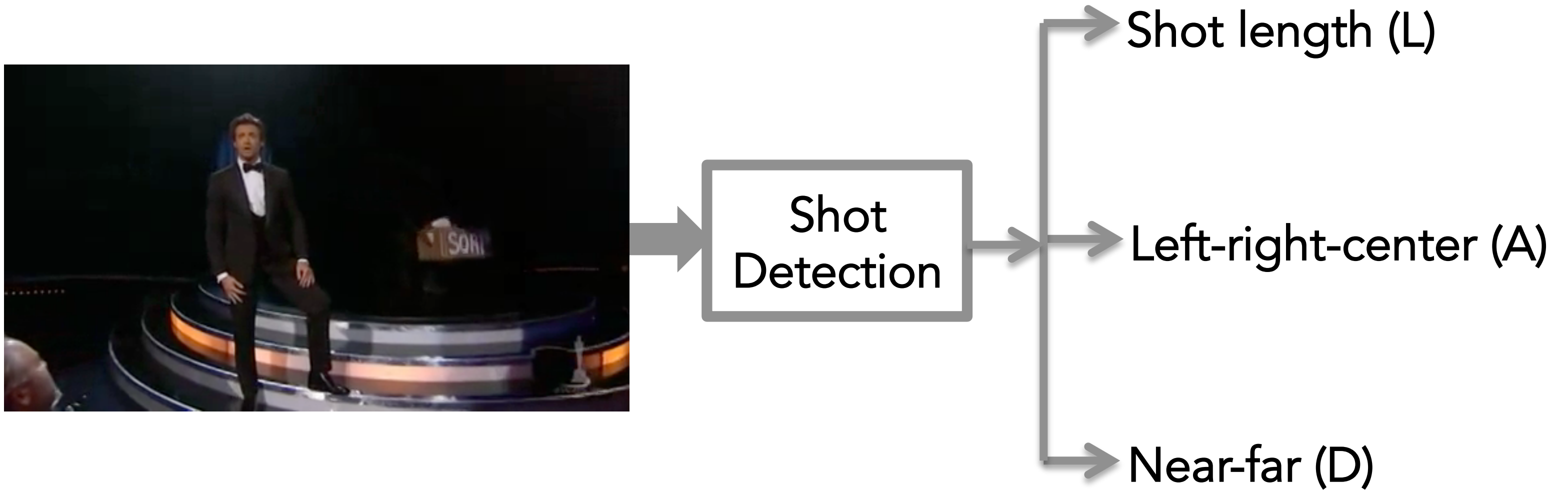
Frame 100

Frame 186

Frame 200

Frame 216

Learning from Human Editors



- Goal: Learn the rules along with uncertainty
- Probabilistic framework
- Hidden Markov Model
  - State variables: D, A
  - Observable: δ

HMM

Transition Matrix

	$\mathcal{CN}$	$\mathcal{CF}$	$\mathcal{RN}$	$\mathcal{RF}$	$\mathcal{LN}$	$\mathcal{LF}$
$\mathcal{CN}$	0	0.4	0.2	0.1	0.2	0.1
$\mathcal{CF}$	0.6	0	0.1	0.1	0.1	0.1
$\mathcal{RN}$	0.5	0.1	0	0.1	0.2	0.1
$\mathcal{RF}$	0.2	0.2	0.4	0	0.1	0.1
$\mathcal{LN}$	0.4	0.2	0.2	0.1	0	0.1
$\mathcal{LF}$	0.2	0.2	0.1	0.1	0.4	0

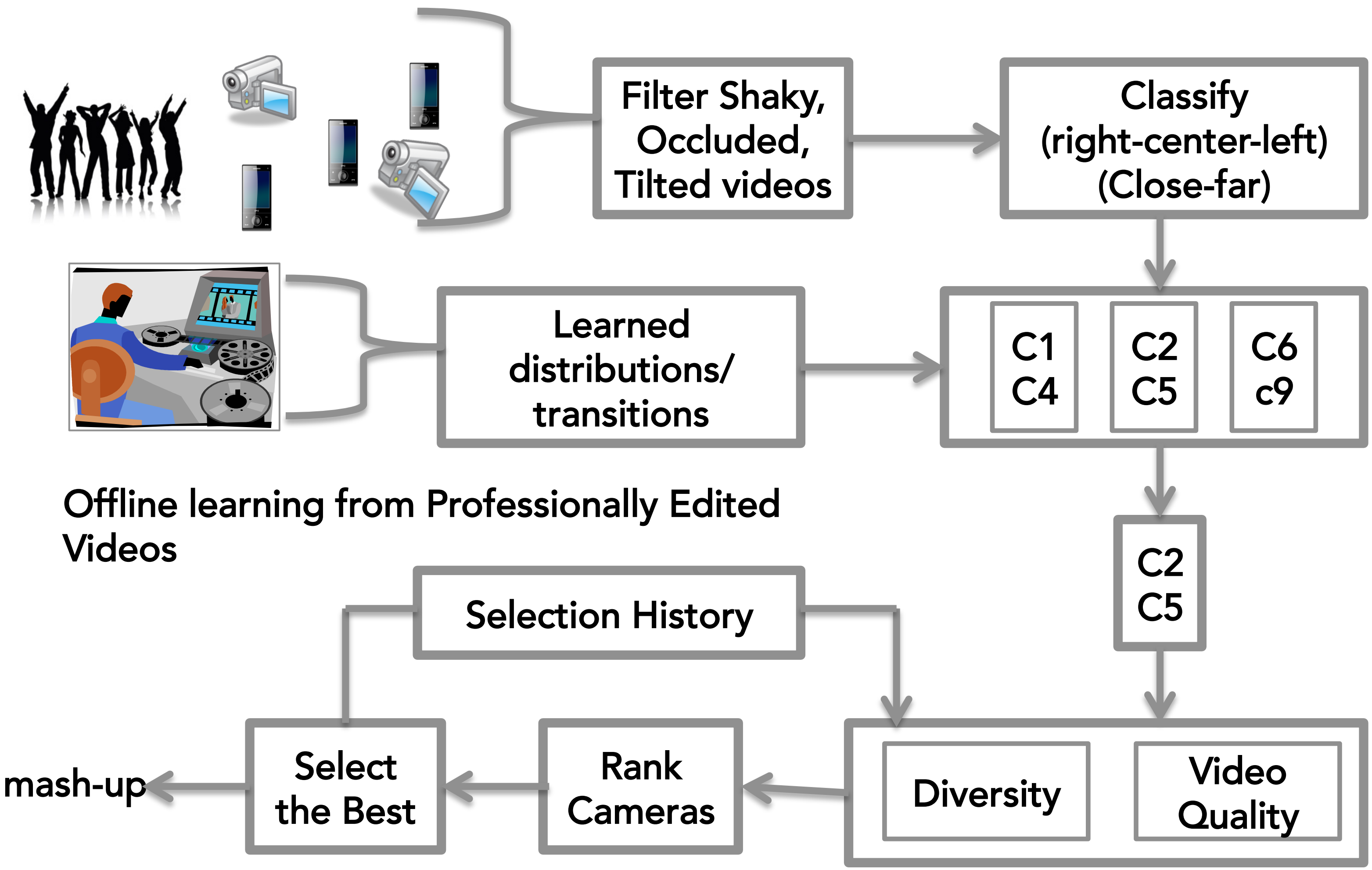
Emission Matrix

	1	2	3	4	5	6	7
$\mathcal{CN}$	1/31	2/31	4/31	7/31	7/31	6/31	4/31
$\mathcal{CF}$	3/12	4/12	2/12	1/12	1/12	1/12	0
$\mathcal{RN}$	2/15	3/15	4/15	3/15	2/15	1/15	0
$\mathcal{RF}$	3/10	4/10	2/10	1/10	0	0	0
$\mathcal{LN}$	2/15	3/15	4/15	3/15	2/15	1/15	0
$\mathcal{LF}$	3/10	4/10	2/10	1/10	0	0	0

1. Start $S_0 = CN$
2. Choose δ by sampling $P(\delta/S_t)$
3. Choose S_{t+1} by sampling $P(S_{t+1}/S_t)$

4. Repeat 2-3 forever

MASHUP Framework



Evaluation Dataset

P1: Group dance
(12 videos)

P2: Group dance
(12 videos)

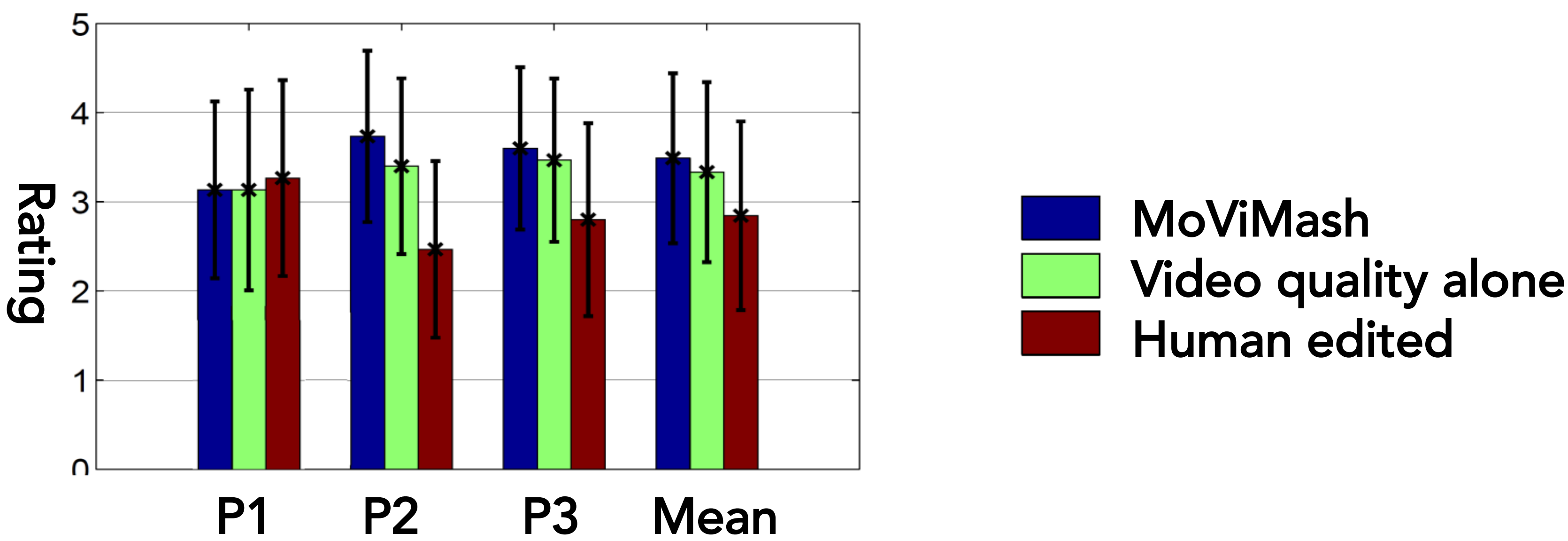
P3: Solo song
(5 videos)

Link - http://www.jiku.org/datasets.html

*[1] The Jiku mobile video dataset. **ACM MMSy**s,2013.*

Results

" How likely will you recommend the video to a friend?"



[*] MoViMash: Online Mobile Video Mashup, **ACM MM** 2012, (IF=1.22).

Multimedia Fatigue Detection for Adaptive Infotainment User Interface

Sultan Alhazmi[1], **Mukesh Saini**[12], Abdulmotaleb El Saddik[12]

[1]School of Electrical Engineering and Computer Science
University of Ottawa, Canada
[2]New York University Abu Dhabi

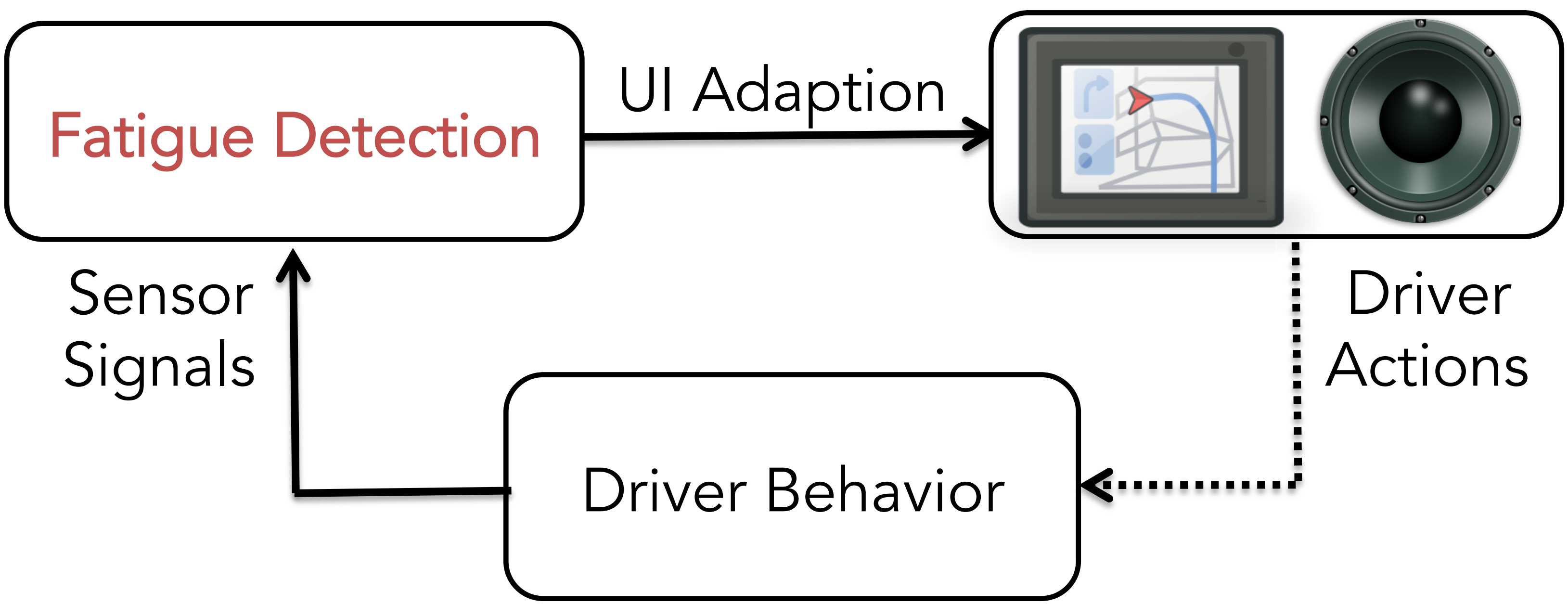
User Interfaces are designed to facilitate interaction between human and machine!

I **nudge** you and **speak louder** when you are **drowsy**!

Why same UI irrespective of different user states?

Around 30% of accidents are due to driver fatigue!

Purpose: Modulating the infotainment system user interface according to the fatigue level of the driver.



Fatigue Detection Approaches

1- Computer vision
2- Physiological signs analysis
3- Driving performance

Four performance cues and four context parameters!

Performance cues

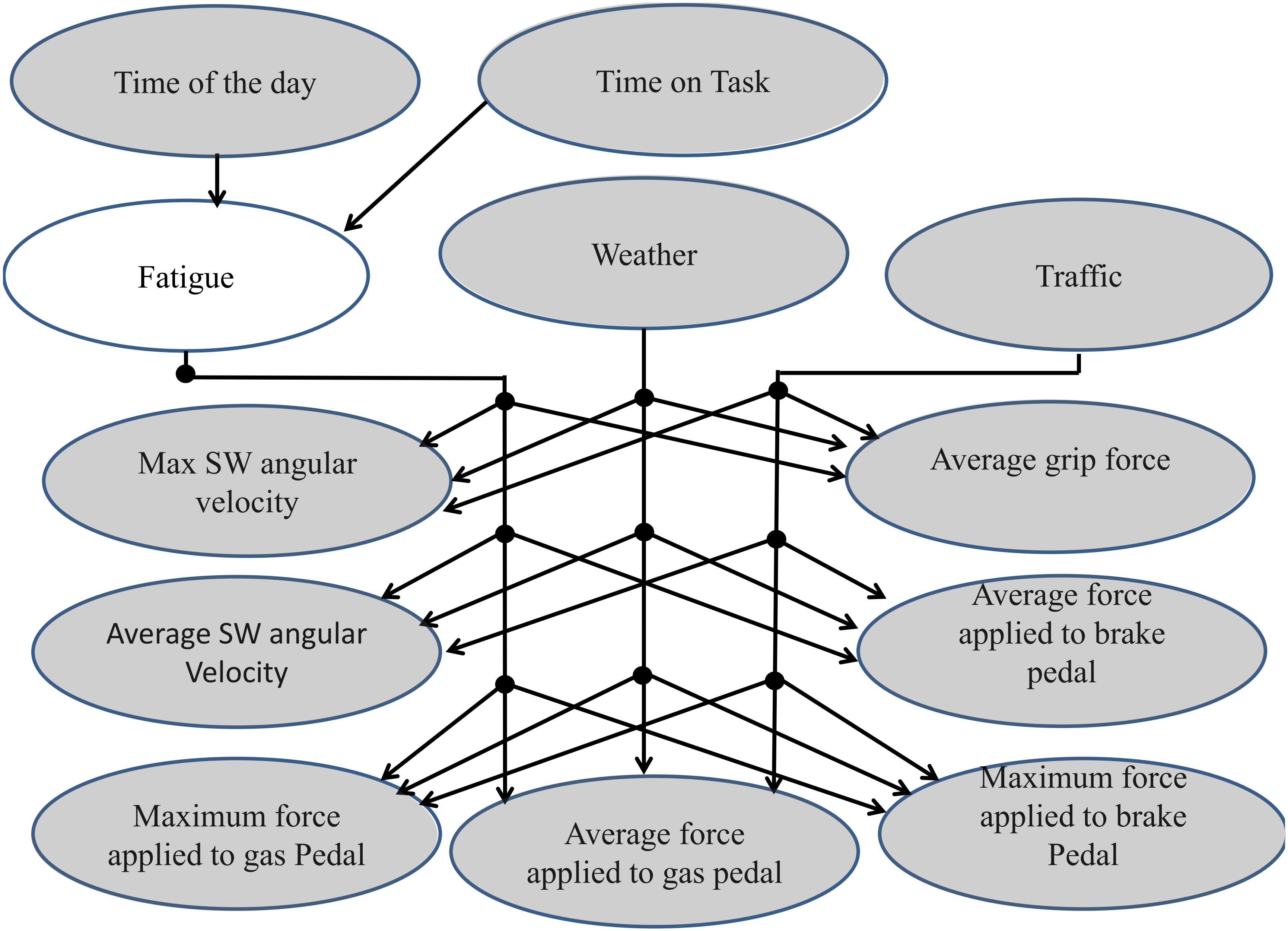
1. Steering wheel angular velocity
2. Grip force
3. Brake pedal
4. Gas pedal

Contextual cues

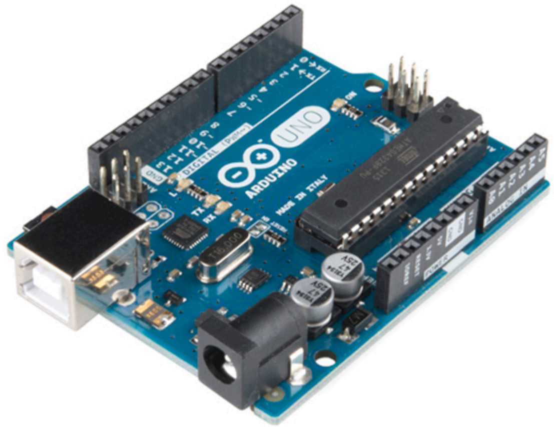
1. Time of the day
2. Traffic
3. Weather
4. Time on task

Context and performance cues are combined using a **Bayesian Network**

1. works with limited features and training data
2. works with partial information

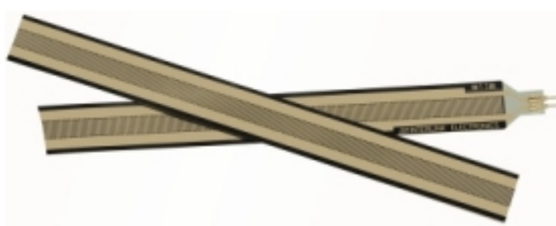

Time of the day
Time on Task
Fatigue
Weather
Traffic
Max SW angular velocity
Average grip force
Average SW angular Velocity
Average force applied to brake pedal
Maximum force applied to gas Pedal
Average force applied to gas pedal
Maximum force applied to brake Pedal

Experiment: Apparatus



(a) Ardino microcontroller

(b) Steering Wheel and pedals

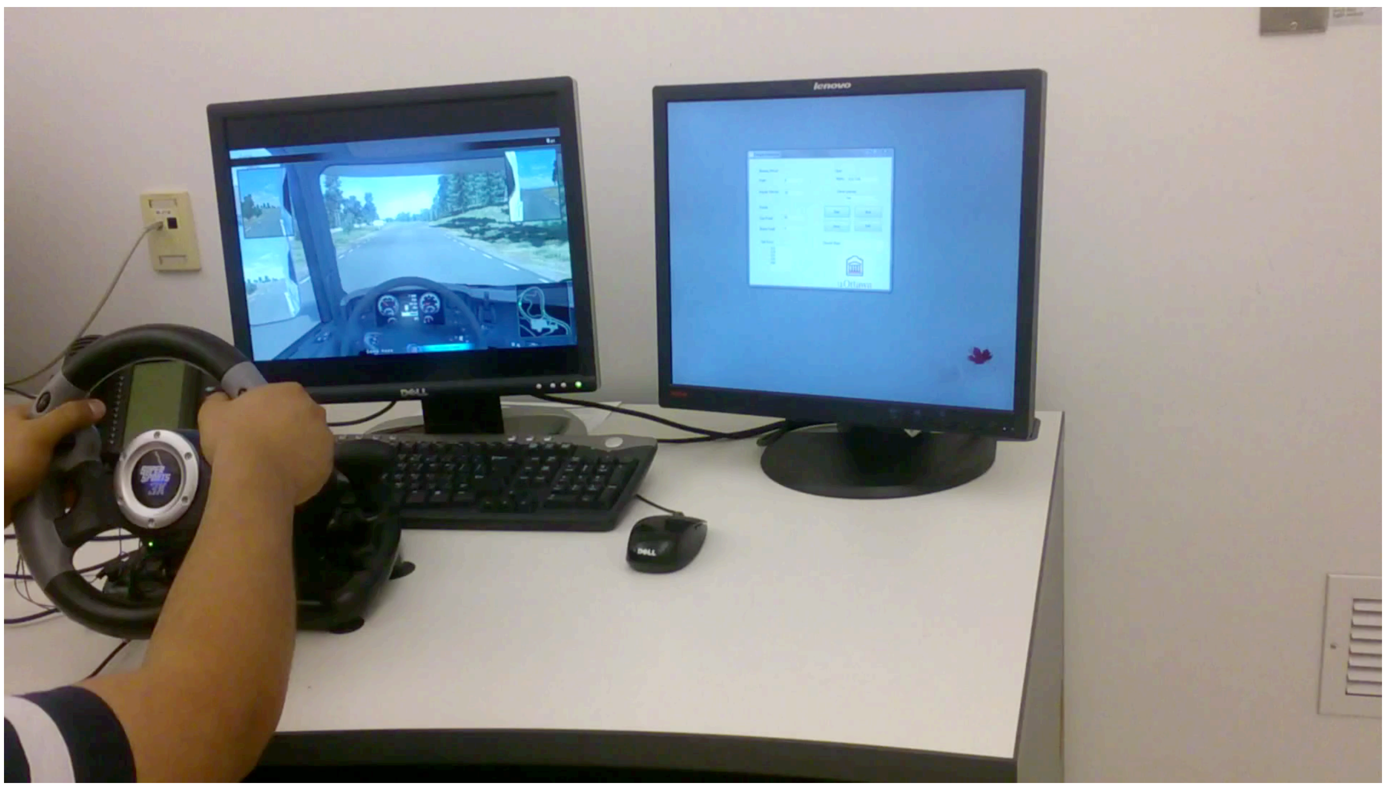


(c) Force Sensing Resistor

(d) Zephyr HxM BT

Experiment Protocol

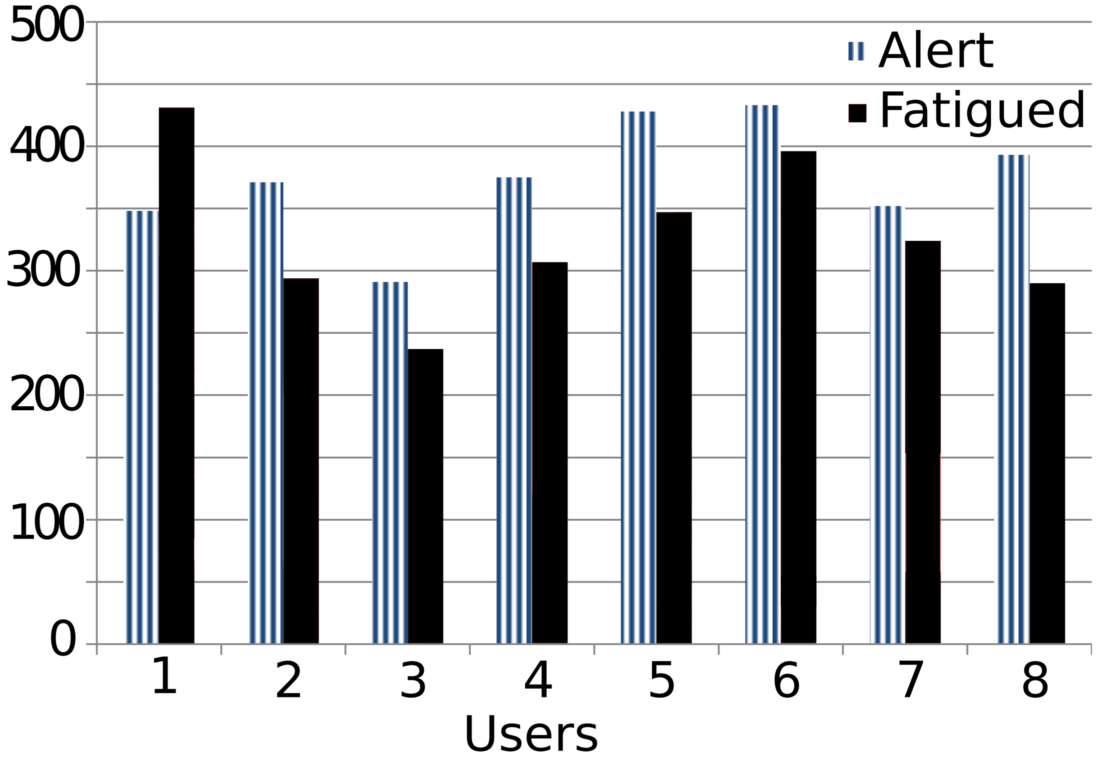
Users drove for one hour and rated fatigue levels on KSS!



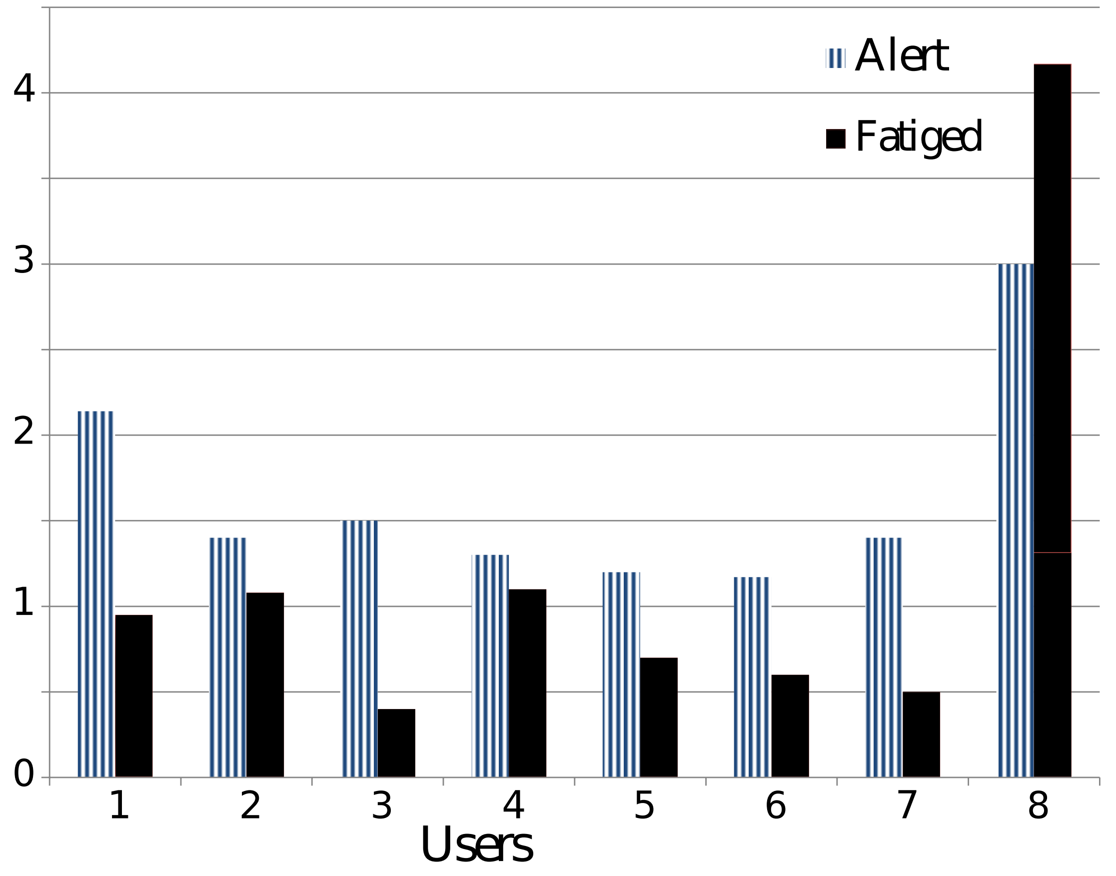
Karolinska Sleepiness Scale (KSS)

Level	Description	☑
1	Extremely alert	☐
2	Very alert	☐
3	Alert	☐
4	Rather alert	☐
5	Neither alert or sleepy	☐
6	Some signs of sleepiness	☐
7	Sleepy but no effort to stay awake	☐
8	Sleepy, some effort to stay awake	☐
9	Very sleepy, great effort to stay awake	☐

The average of grip force for the users



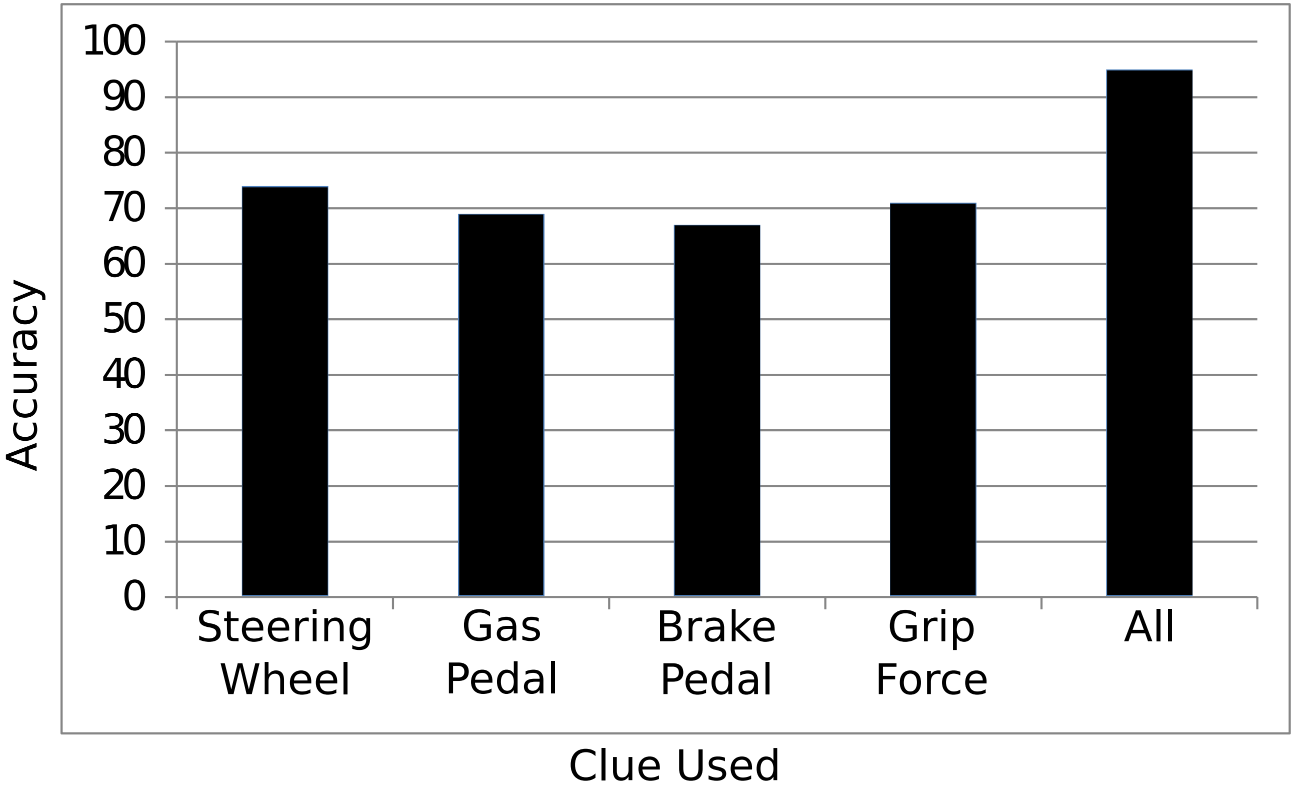
Average power on brake pedal



Accuracy for different cue combinations

Single cue		Two cues		Three cues	
Media streams	Result	Media streams	Result	Media streams	Result
Steering wheel	74	Steering wheel and Gas pedal	81	Steering wheel and Gas pedal and Brake pedal	87
Gas pedal	69	Steering wheel and Brake pedal	~~84~~	Steering wheel and Gas pedal and Grip force	~~90~~
Brake pedal	67	Steering wheel and Grip force	89	Steering wheel and Brake pedal and Grip force	93
Grip force	71	Brake pedal and Gas pedal	77	Brake pedal and Gas pedal and Grip force	78
		Grip force and Gas pedal	67		
		Grip force and Brake pedal	88		

Individual Vs Combined Accuracy



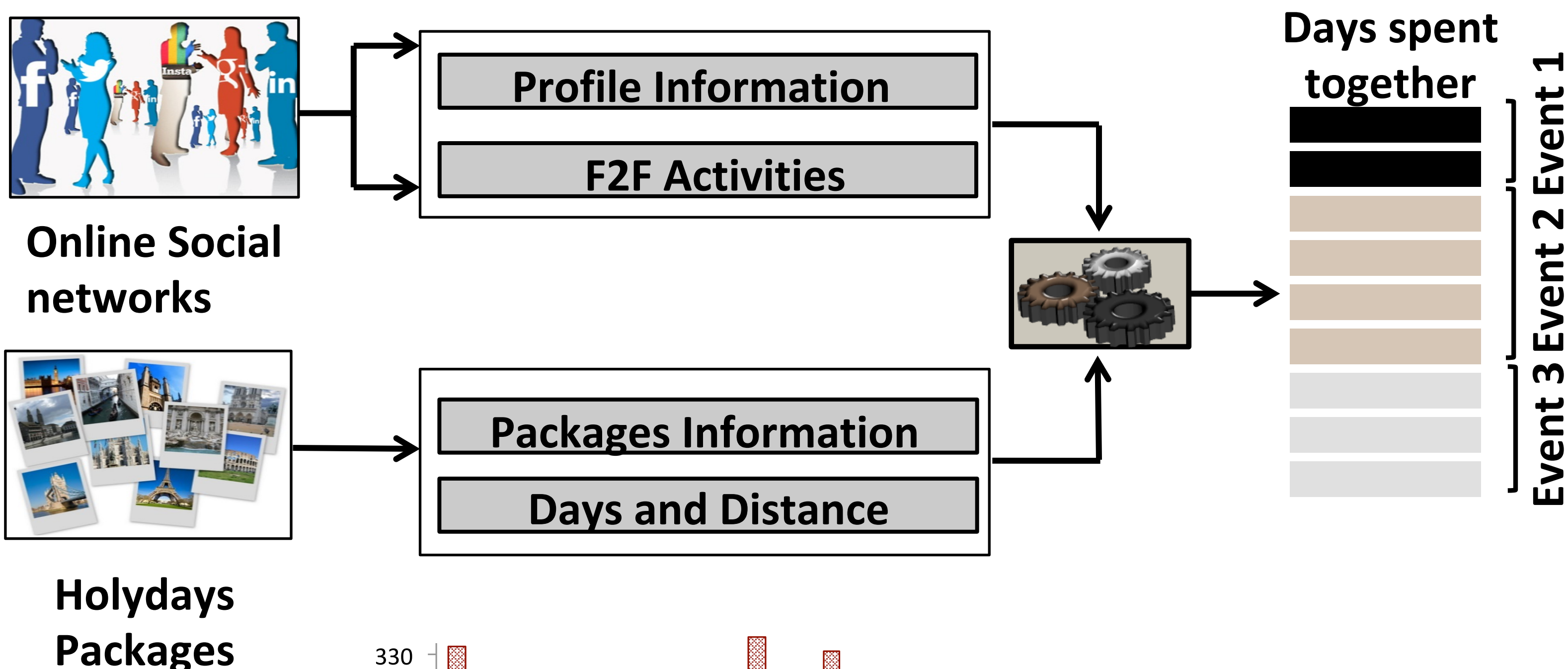
UI Adaption

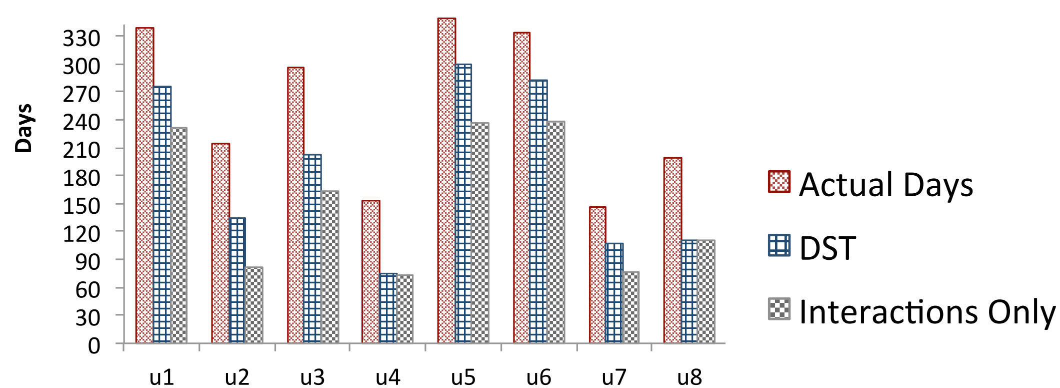
Property	Type	Intrusion Level	Control
Text	Visual	Low	Discrete
Colour	Visual	Low	Continuous
Illumination	Visual	Low	Continuous
Sound	Auditory	Medium	Continuous
Orientation	Visual	High	Discrete
Layout	Visual	Medium	Discrete
Haptic	Physical	High	Discrete
Tactile	Physical	High	Discrete

Conclusions

- Infotainment user interfaces should be adapted according to the driver fatigue level!
- Multiple cues improve fatigue detection accuracy!
- Bayesian networks are appropriate for fatigue detection and UI modulation!

Video Storytelling with OSN Data





[*] Personality Detection Using Multiple Online Social Networks**. Springer Multimedia Tools and Applications (MTAP), 2014. [Impact Factor 1.05]**.

[*] Towards Storytelling by Extracting Social Information from OSN, **ACM MM** 2014 (IF = 1.22)