TapSense: Combining Self-Report Patterns and Typing Characteristics for Smartphone based Emotion Detection

Surjya Ghosh, Niloy Ganguly, Bivas Mitra
Indian Institute of Technology Kharagpur

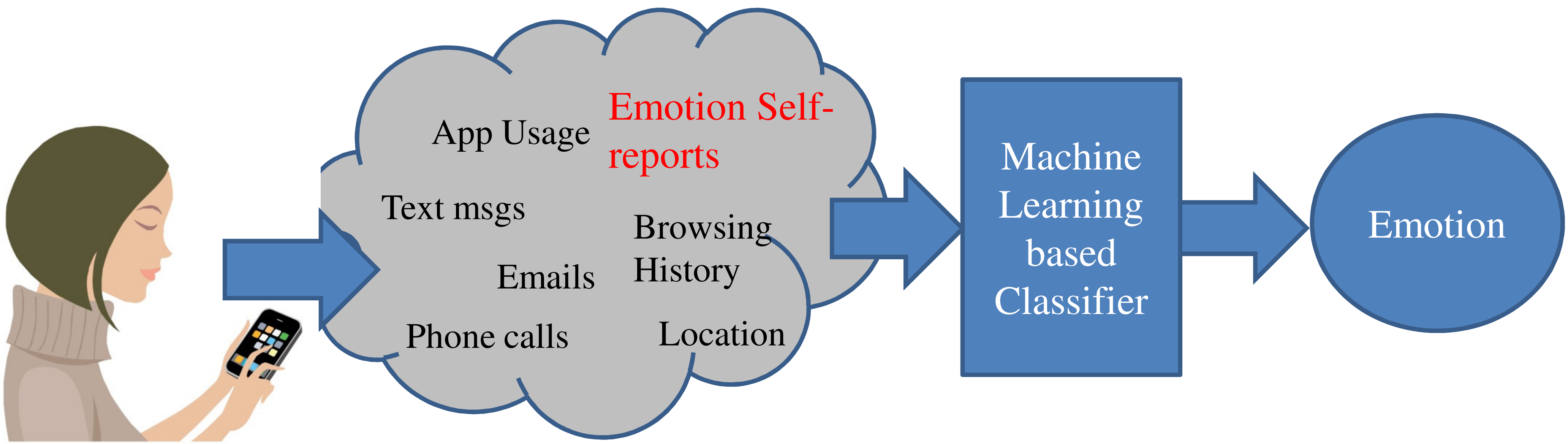
Pradipta De
Georgia Southern University

MobileHCI 2017
4th - 7th September, Vienna

contact: surjya.ghosh@iitkgp.ac.in

Smartphone-based Emotion Detection

- Smartphones
  - Integral part of our daily life
  - Easy to track activities, location details, call history etc.
  - Opportunity to determine emotion states
    - Moodscope [Mobisys 13], Boredom detection [UbiComp 15]



Objective

- Possibility of designing
  - Light-weight, non-intrusive emotion detection application

- *Typing activity* in smartphone
  - Non-intrusive
  - Low resource consumption
  - Prevents monitoring overhead of multiple sensors
  - Privacy preserving (if content not looked at)

  - Suitable for emotion detection ?

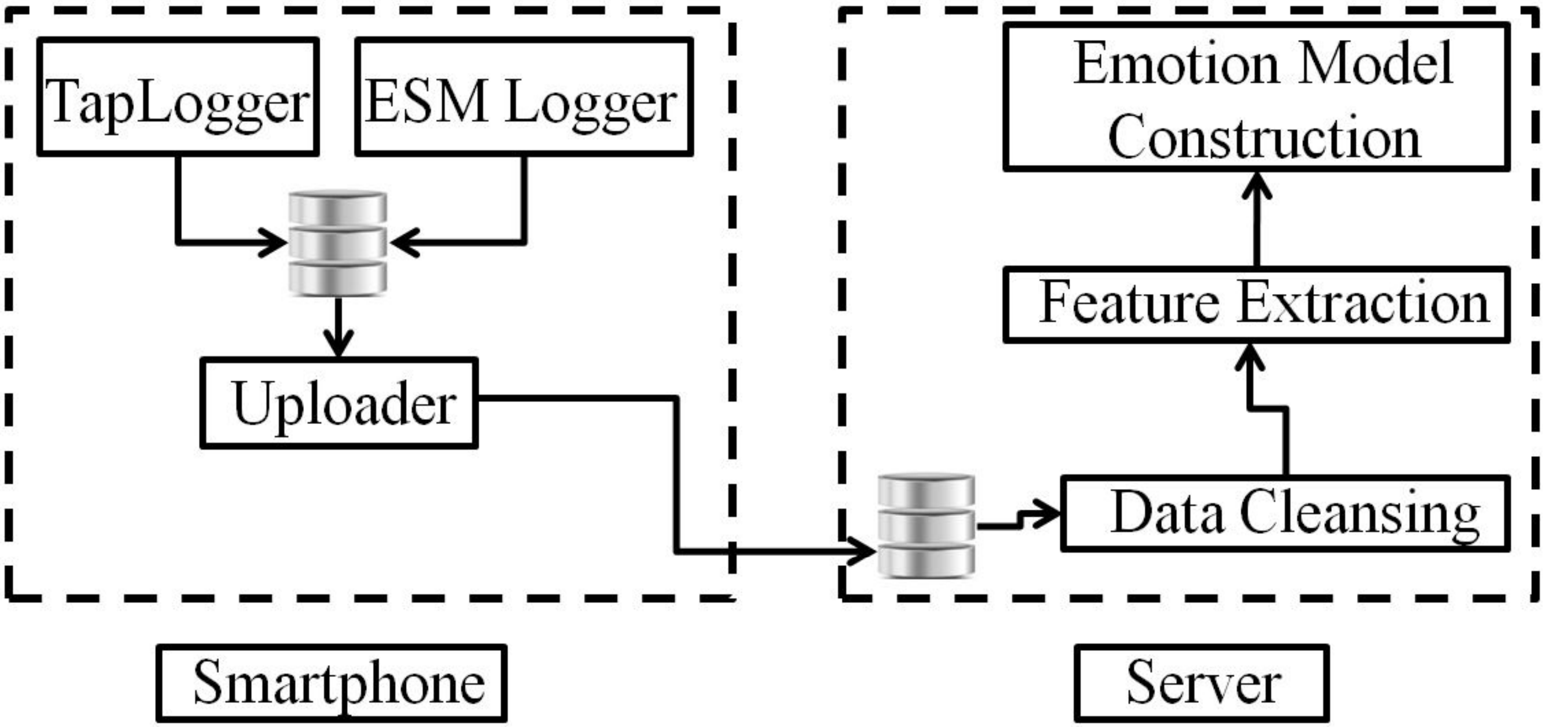
Objective

- Limitation of Typing
  - May not work for every individual
  - Typing cues may not be able detect emotions with high accuracy
  - Make use of other information source ?
    - Would make the system heavy
    - Need to record multiple sensor details
    - Users may not agree to provide other details
  - What's the way out ?
    - No need for additional details
    - Observe the self-reporting patterns
- Typing and Self-report patterns together for emotion prediction

Outline

- TapSense Architecture
  - Challenges
  - Design Principles
- Feature Identification
  - Keystroke Features
  - Self-reporting Pattern
- User Study
- Evaluation
- Take-home Points

TapSense Architecture

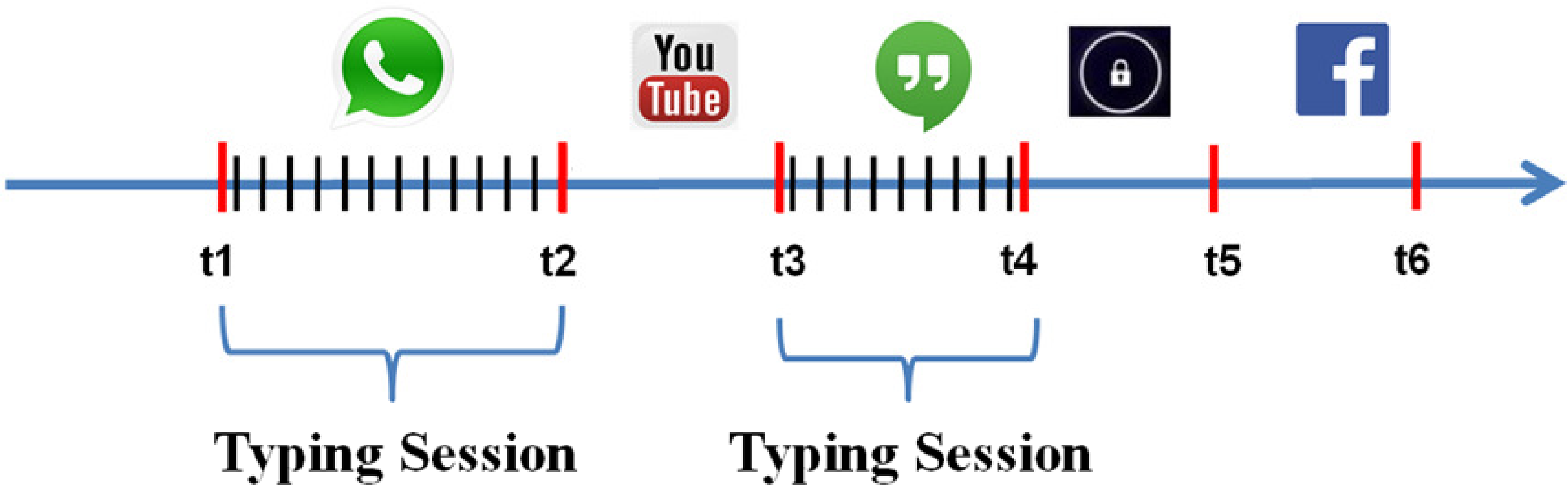


- *TapLogger*
  - Traces typing activity
- *ESMLogger*
  - Collects emotion self-reports
- *Feature Extraction*
  - Identify features
- *Model Construction*
  - Personalized, RF based

Challenges

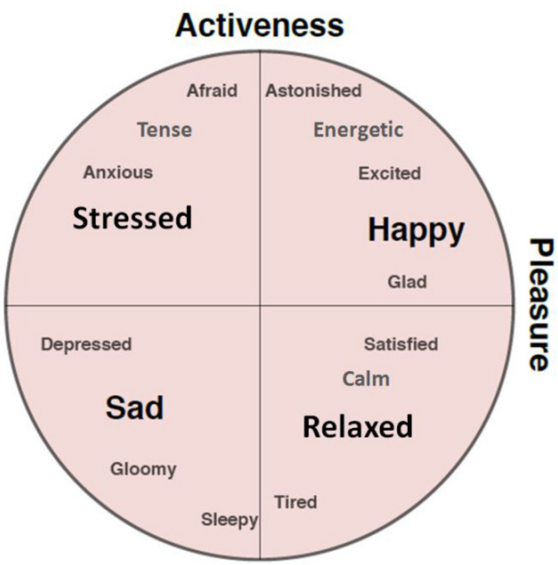
- Extract Typing details
  - Granularity of typing data collection
- Collect Self-reports
  - Manual → *survey fatigue*
  - Psycho-physical sensor based → *intrusive setup*
- Emotion detection model
  - Personalized → Training for every user

Typing Session Identification

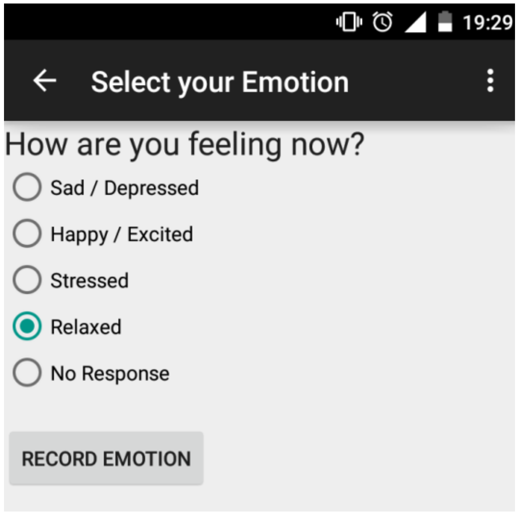


- Typing details are extracted session-wise
- Typing session
  - Collection of tap events within an app without changing it

Emotion Self-report Collection



(a) Emotion circumplex model

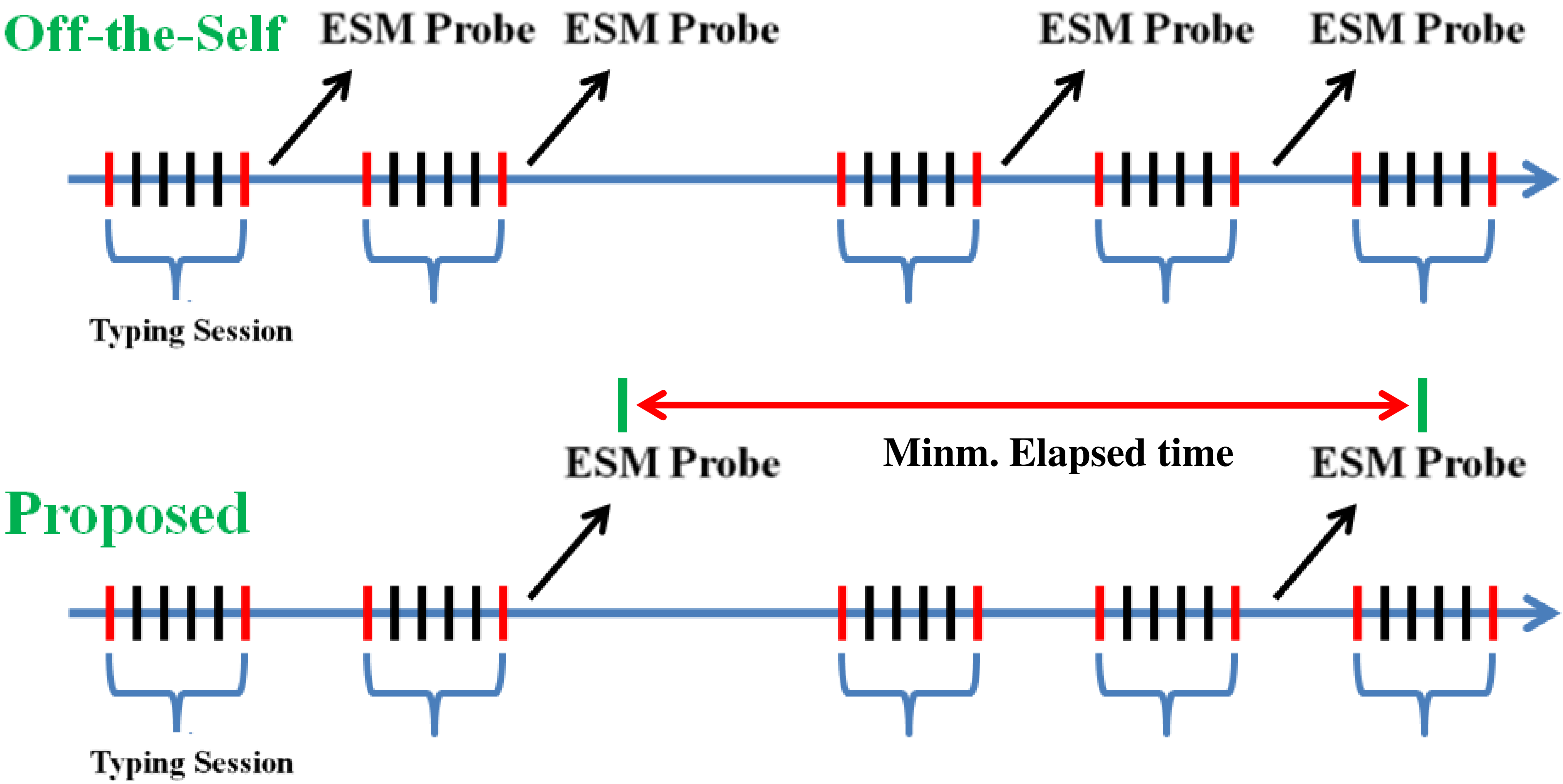


(b) Emotion collection UI

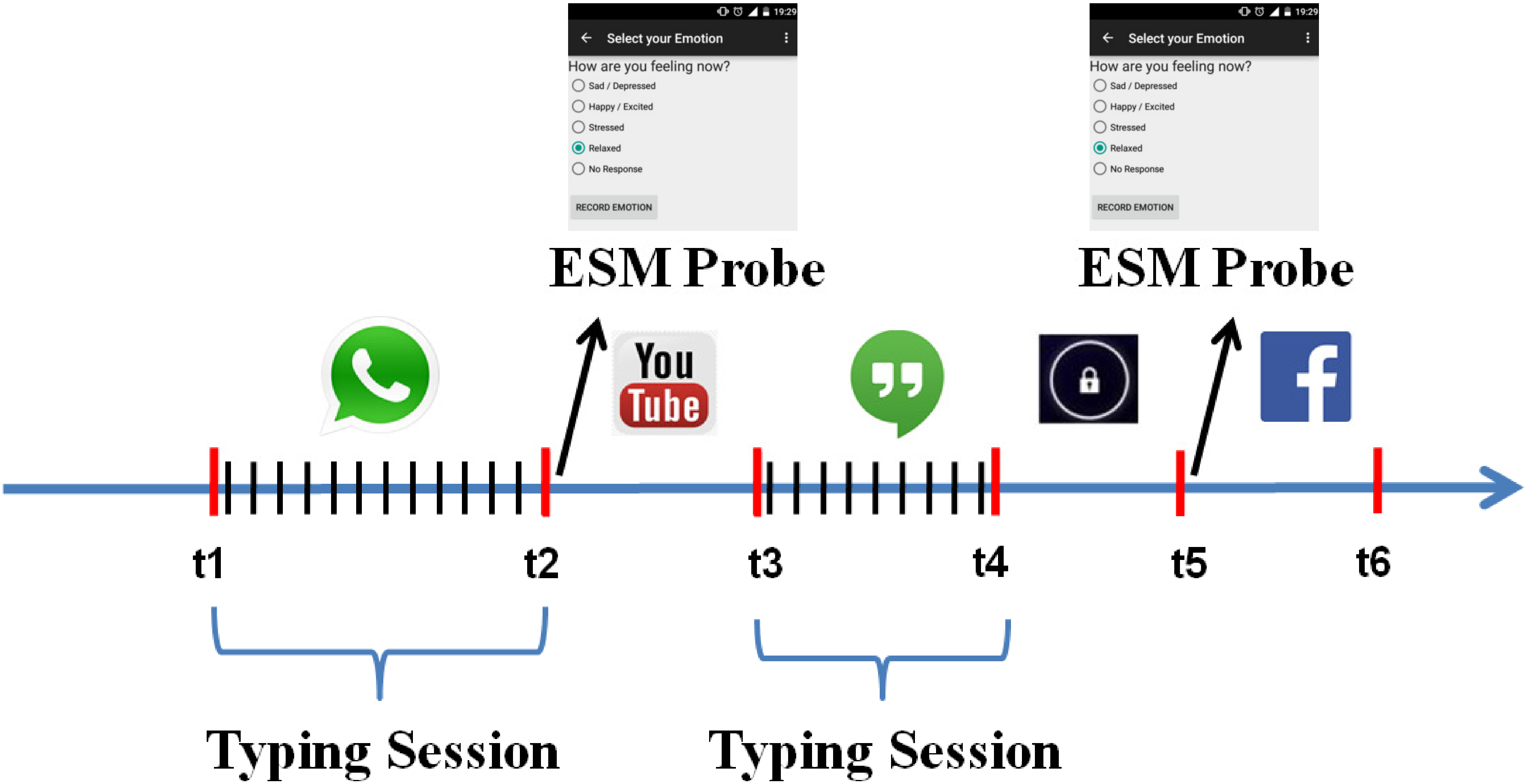
- Self-report collection
  - Manual, by using Experience Sampling Method (ESM)
  - Based on circumplex emotion model
  - Selected dominant emotion from each quadrant so that they are distinctly different
  - Emotion recording can be skipped by selecting *No Response*

Emotion Self-report Collection

- Self-report collection
  - *Survey fatigue* to be kept low

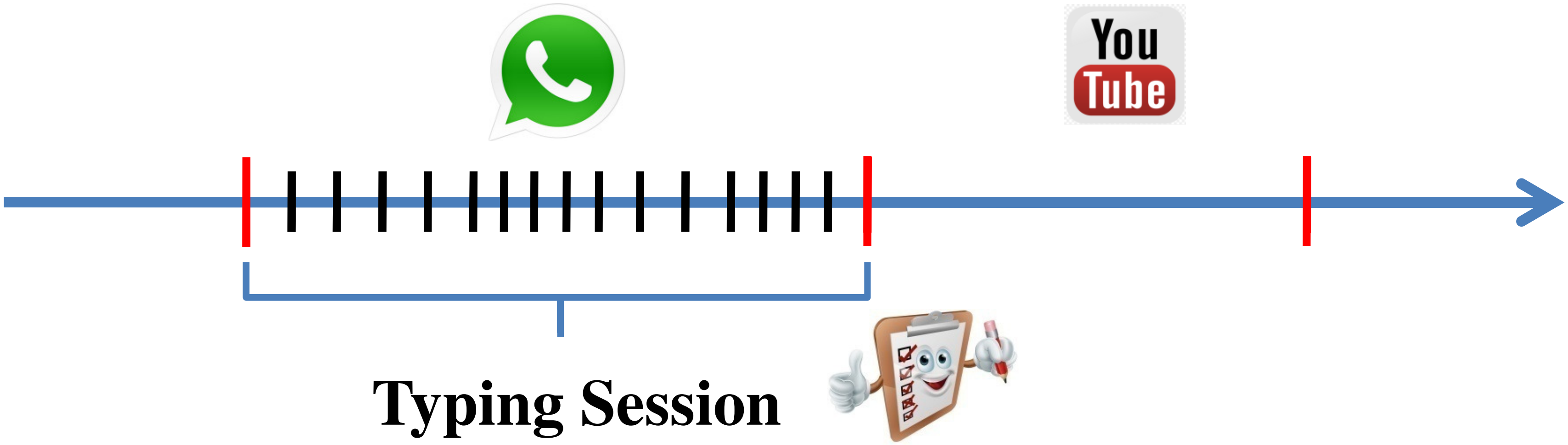


Attach Self-reports to Typing Session



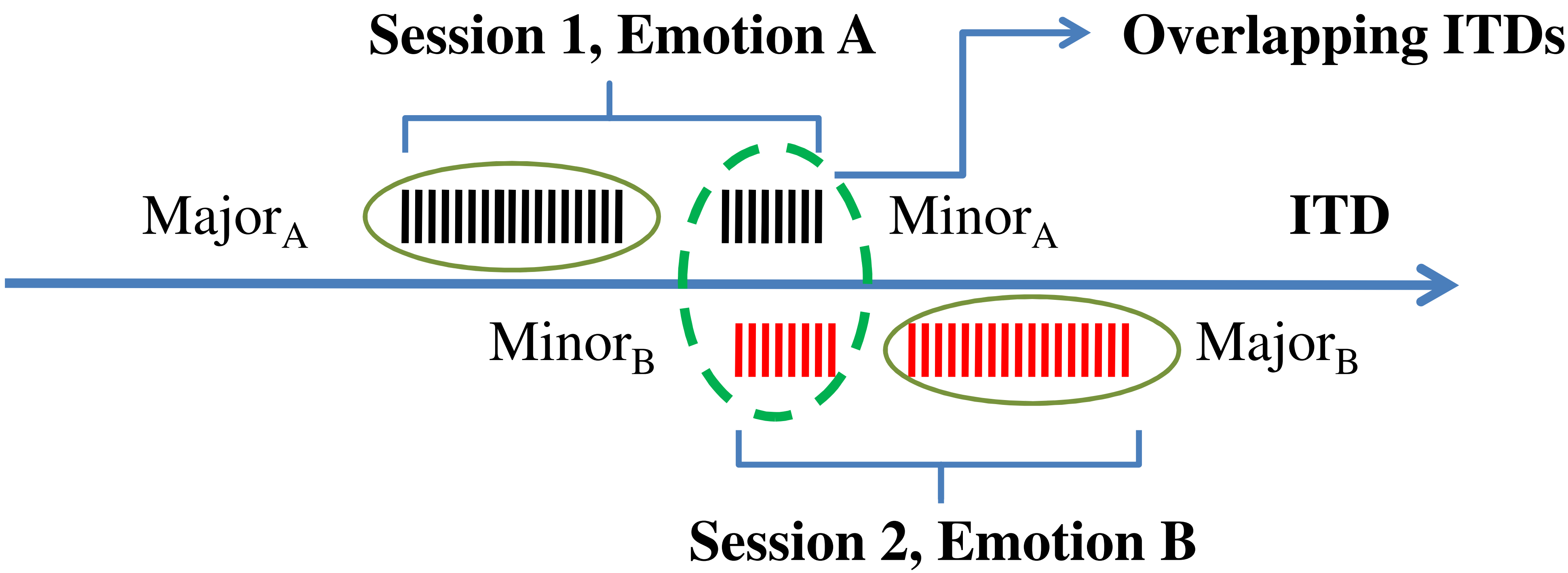
- Collected self-report is tagged with previous typing session

Feature Identification – Typing Speed



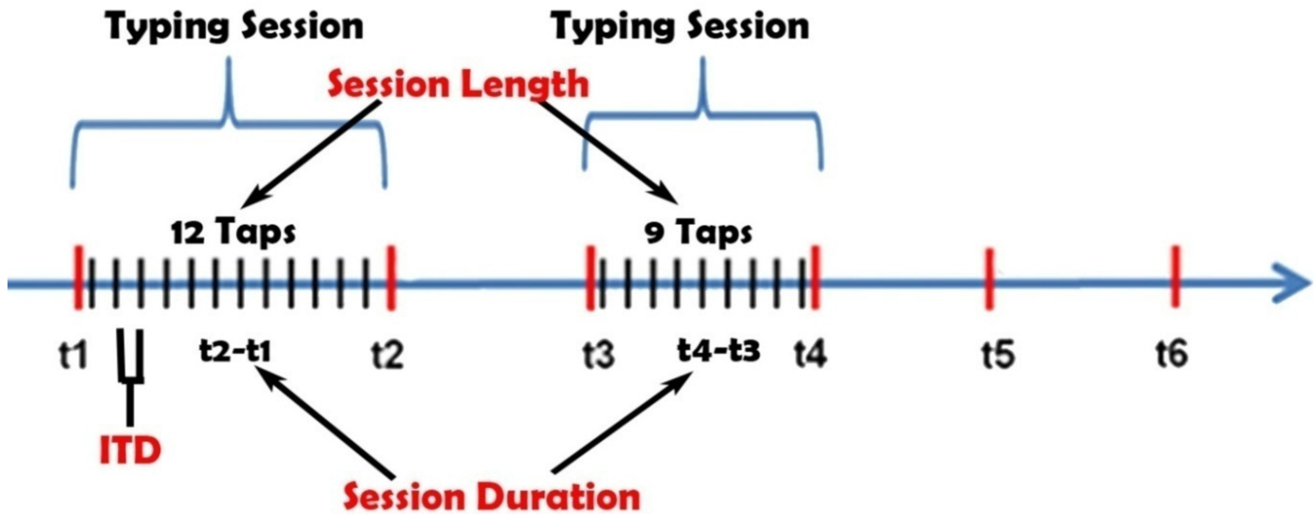
- Inter-Tap Distance (ITD)
  - Elapsed time between entering two character is *ITD*
- *Mean Session ITD*
  - Compute mean of all *ITD*s in a session, which is known as *Mean Session ITD*
  - Representation of typing speed

Refined Mean Session ITD (*RMSI*)



- *Mean Session ITD (MSI)*
  - Overlapping ITDs, not distinguishable enough
- *Refined Mean Session ITD (RMSI)*
  - Identify major cluster using K-means
  - Compute mean of ITDs present in that cluster

Keystroke Features



- *Session Length*
- *Session Duration*
- *Percentage of Backspaces in a Session*
  – To trace the typing mistakes performed in any given emotion
- *Percentage of Special Characters in a Session*
  – To trace usage of special chars in an emotion state

Self-reporting Pattern

- Can we estimate current self-report based on previous self-reports ?
- We define this as *Persistent Emotion (PRE)*
  - Model the same using discrete-time Markov Chain

$$\boxed{e_n = e_{n-1} . P}$$

- Need to construct the transition matrix *(P)*
  - Set of probability values (p_{xy})

$$\underbrace{\begin{pmatrix} H & S & T & R \end{pmatrix}_n}_{e_n} = \underbrace{\begin{pmatrix} H & S & T & R \end{pmatrix}_{n-1}}_{e_{n-1}} \times \underbrace{\begin{pmatrix} p_{hh} & p_{hs} & p_{ht} & p_{hr} \\ p_{sh} & p_{ss} & p_{st} & p_{sr} \\ p_{th} & p_{ts} & p_{tt} & p_{tr} \\ p_{rh} & p_{rs} & p_{rt} & p_{rr} \end{pmatrix}}_{P}$$

Symbol	Description
e_n	*PRE* of session 'n'
e_{n-1}	Self-report of session 'n-1'
P	Transition matrix
p_{xy}	Transition probability from state ***x*** to state ***y***

Persistent Emotion (*PRE*)

$$e_n = e_{n-1}.P$$

Session No	Self-report (Emotion Ground Truth)	Inter-session Gap (in Hr.)	PRE
1	Happy	-	
2	Relaxed	2	
3	Stressed	8	P
….	….	….	….
N-1	Relaxed (e_{n-1})	1	
N	Relaxed	1	Relaxed (e_n)

- *Persistent Emotion (PRE)*
  - Computed for every session
  - Used as a feature in the emotion prediction model

User Study

- Study duration – 3 Weeks (on-the-wild)
- Total number of participants – 30
  - University students
  - 24 males, 6 females, aged between (24 – 33) years
- Installed *TapSense* in participant mobile phones
- Excluded participants
  - 3 participants left in between
  - 5 participants recorded less than 40 labels
- Final participants – 22 (20 male, 2 female)

Dataset

- Keystroke details

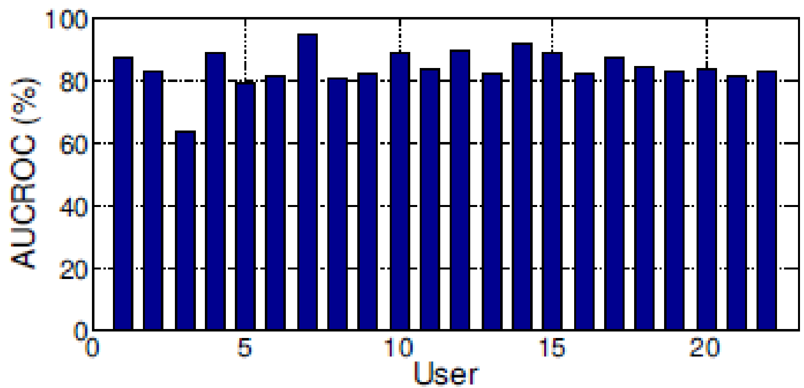
Parameter	Value
# of Typing events	529,698
Total typing duration	~135 Hours
Total typing sessions	2705 (mean=123, std dev = 25.6)
Session duration	Mean=180 sec. std dev = 34.2 sec.

- Self-report details
  - Applied SMOTE [Chawla et al.] to overcome emotion imbalance

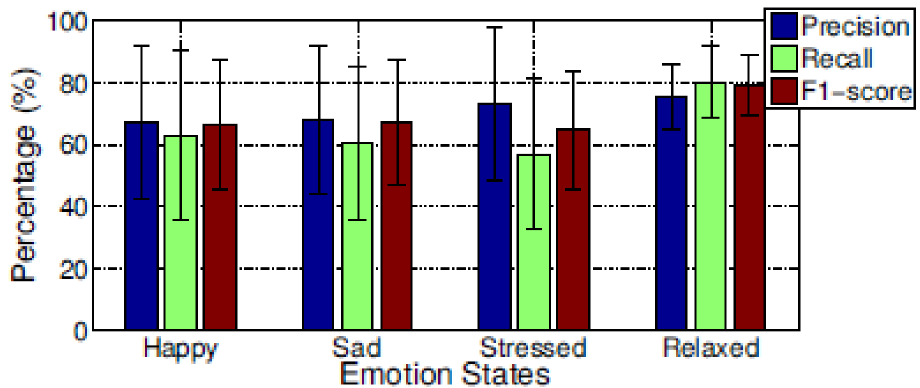
Emotion	Distribution
Happy	19%
Sad	9%
Stressed	23%
Relaxed	49%

Evaluation

- How accurate is the emotion prediction model ?



(a) Accuracy (AUCROC) of predicting different emotions across all users



(b) Mean value of Precision, Recall and F1-score for each emotion state. Error bar indicates standard deviation.

- Avg. accuracy (AUCROC) – 84% (std dev 6%)
- Relaxed state is identified with highest precision and recall

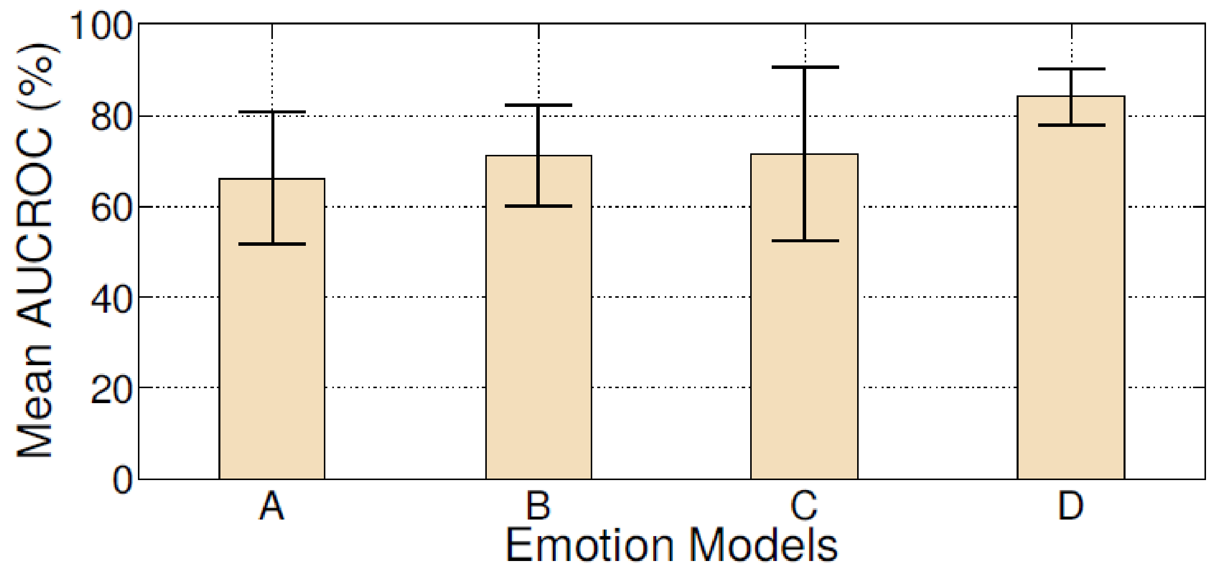
Baseline Models

Model Name	Description
Persistent Emotion (PRE) Model	Personalized model using persistent emotion (PRE) as the only feature
Keystroke only Model	Personalized model using only keystroke features
Aggregate Model	-Leave-one-participant-out cross validation - to reduce training overhead

Evaluation

- How does the proposed model perform w.r.t baseline models ?

Model	Description
A	*PRE* only model
B	Keystroke only model
C	Aggregate model
D	Proposed model



Proposed model outperforms all other models.

Evaluation

- Which features are most important ?

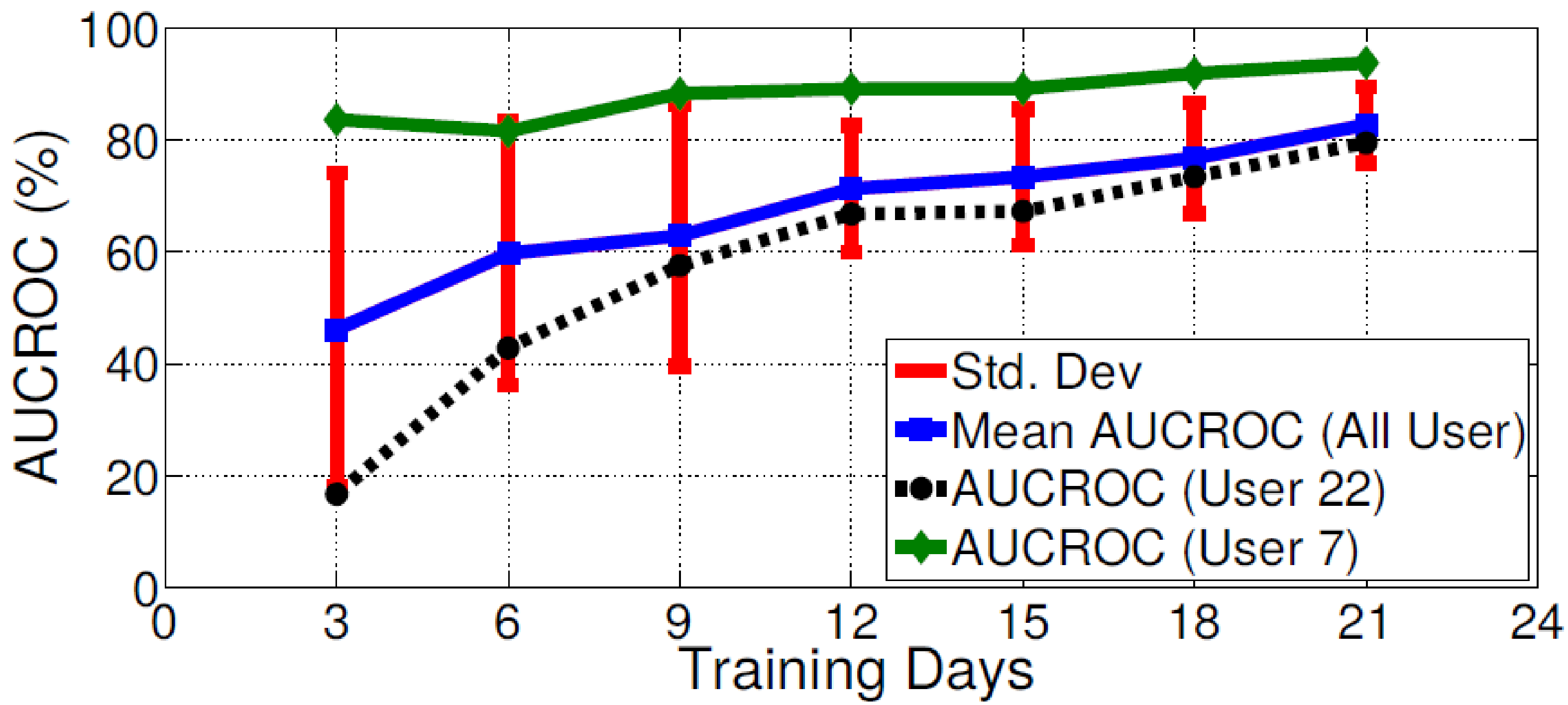
Feature Name	Rank	Average IG
PRE	1	0.4226
RMSI	2	0.2324
Working hour indicator	3	0.1368
MSI	4	0.1257
Backspace percentage	5	0.0529
Session duration	6	0.0270
Special char percentage	7	0.0226

Table 6: Ranking features based on Information Gain

- *PRE* is found to be the most discriminative feature followed by *RMSI*

Evaluation

- How much training is required ?

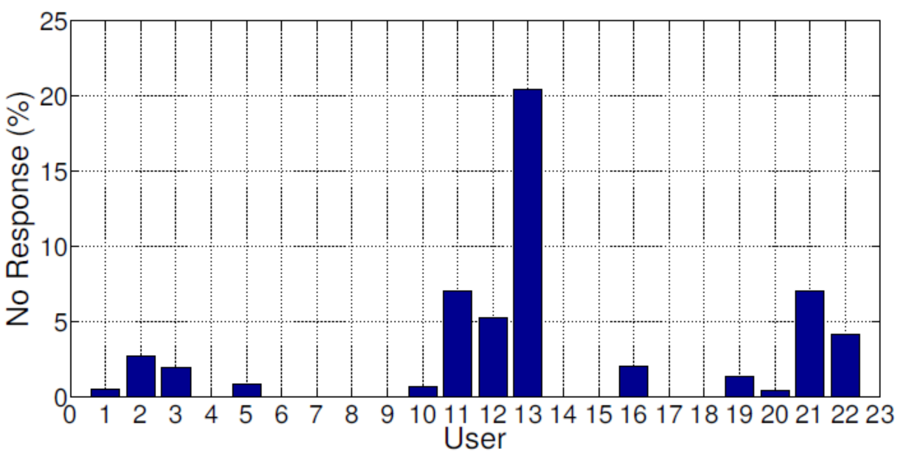


Within 12 days, average accuracy (AUCROC) of 71% is obtained, which touches 77% after 18 days.

Evaluation

- How effective is the proposed ESM in collecting self-reports?

Parameter	Value
Number of ESM probes	Avg. 4.6 per day for every user
Number of No Response sessions	Only 2.5% of all sessions



No Response distribution

- Proposed ESM is non-intrusive
  - Few probes per user
  - Less *No Response* in comparison to off-the-shelf ESM

Take-home Points

- Light-weight, non-intrusive emotion detection system by jointly modeling
  - Typing patterns
  - Self-reporting patterns
- Average accuracy (AUCROC) of 84% in a 3-week study involving 22 participants
- Personalized models appear to be superior

Last but not the least..

- We thank
  - all the participants, who took part in the survey
  - all the reviewers for their suggestion and feedback
  - Microsoft Research India for the travel grant

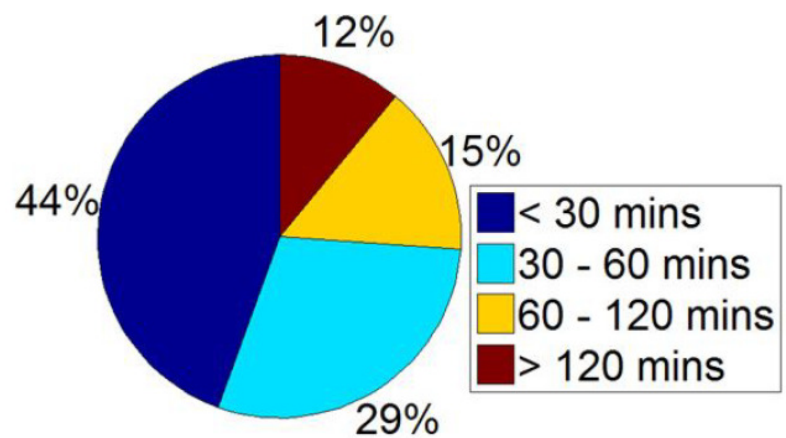
- Project website
  - http://cse.iitkgp.ac.in/~surjya.ghosh/projects.html

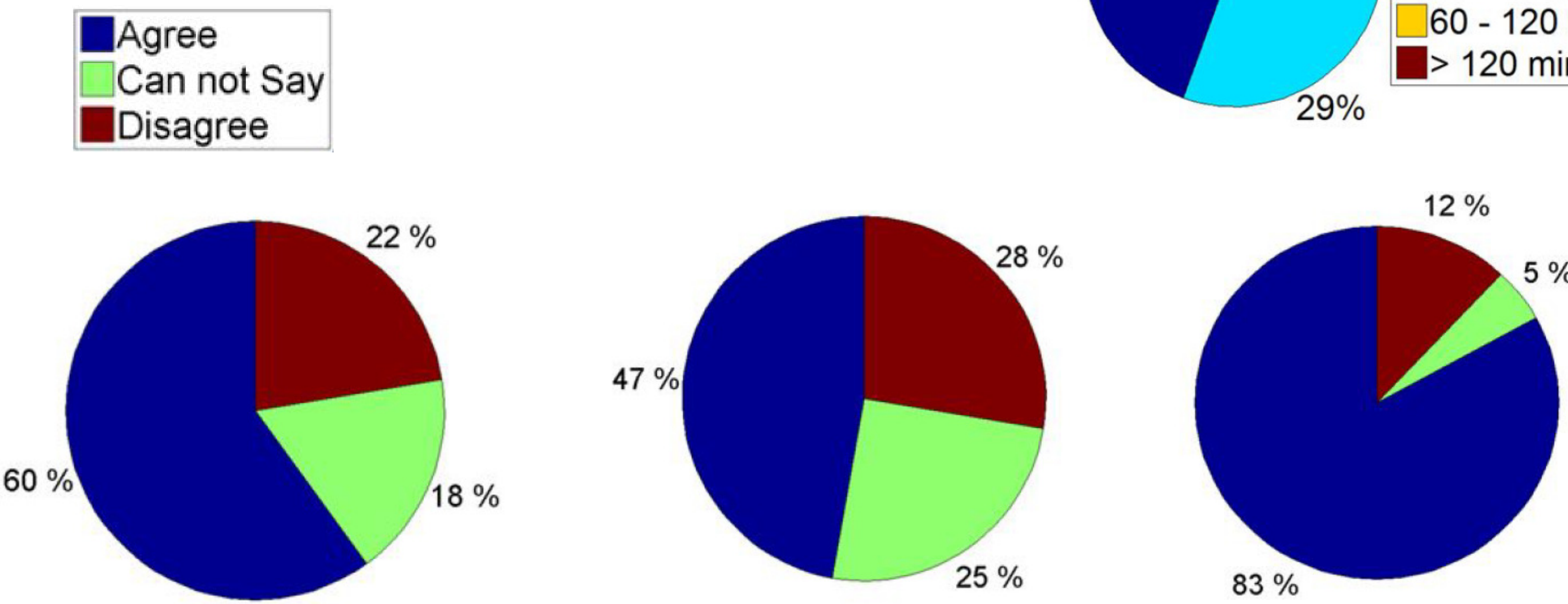
- Our research group (CNeRG, IIT Kharagpur, India)
  - http://www.cnergres.iitkgp.ac.in/

Thank You!!

Applicability of Typing in Emotion Detection

- Online survey involving 120+ participants
  - 56% participants indicated they spent at least half an hour typing daily
  - Measurable cues …





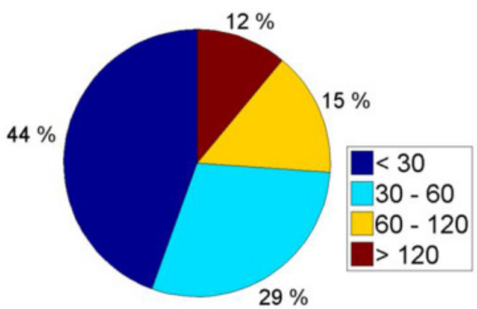
(c) Typing speed

(d) Typing mistake

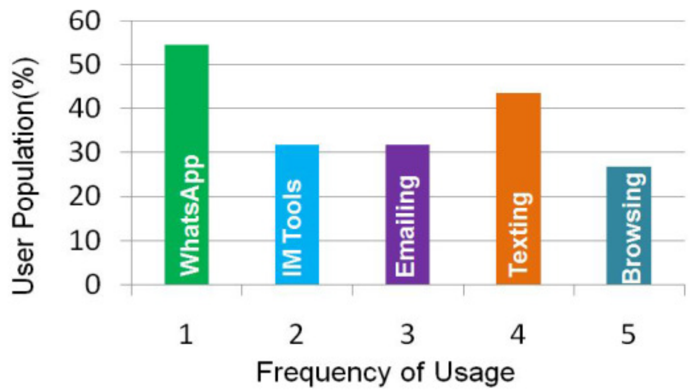
(e) Special character

Typing based Emotion Detection

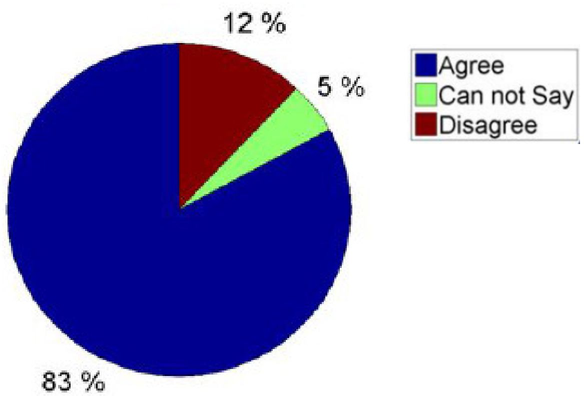
- Why Typing?
  - Keystroke dynamics shown to be an effective modality for desktop computers [CHI 2011]
  - Conducted an online survey involving 120+ participants
  - 56% participants spent at least half an hour daily in typing



(a) Daily typing duration



(b) App usage frequency

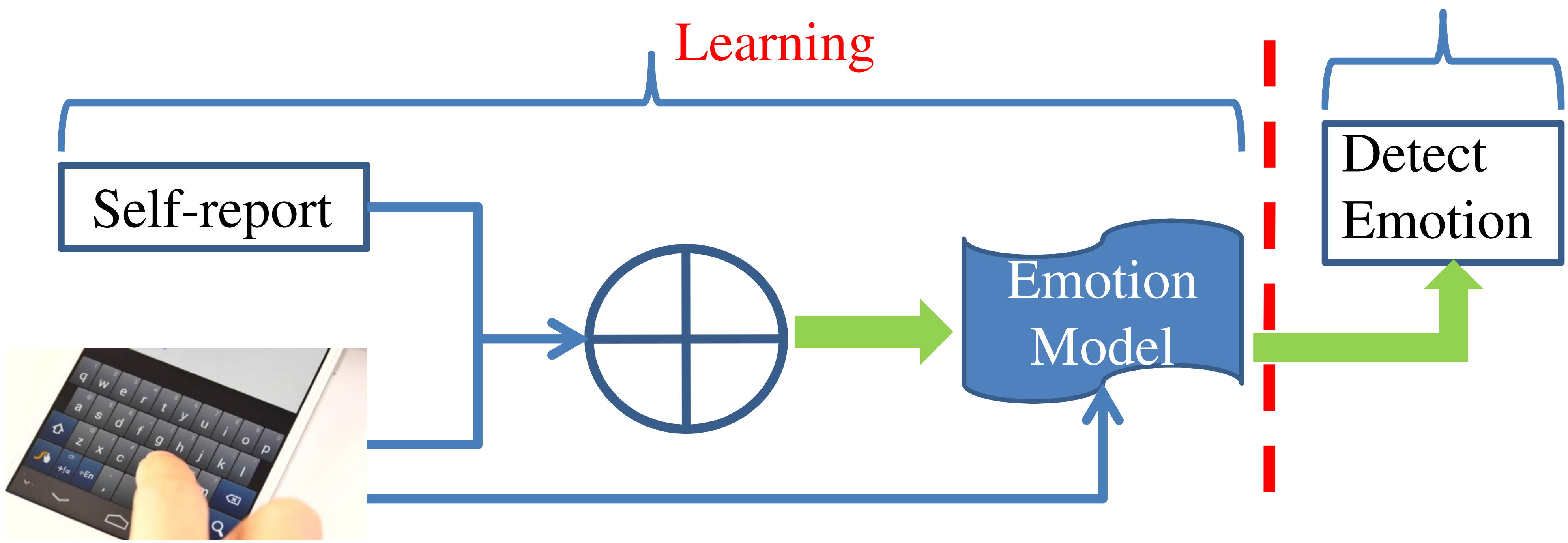


Emoticon Usage

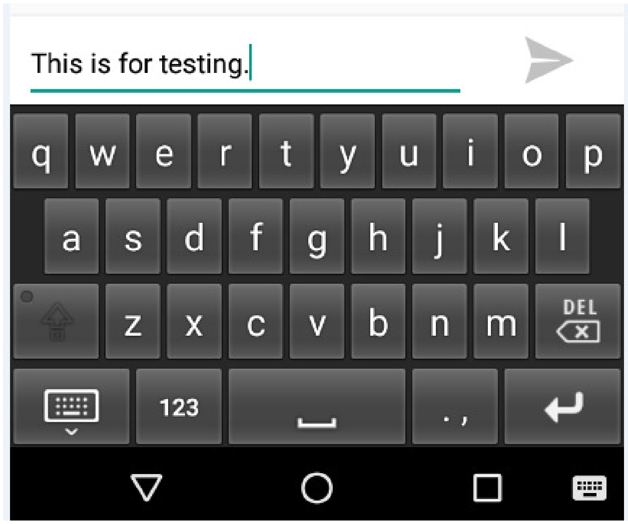
  - Variation in different measurable parameters (e.g. emoticon as indicated by 83% participants) with emotion
  - Prevents monitoring overhead of multiple sensors

Typing based Emotion Detection Application

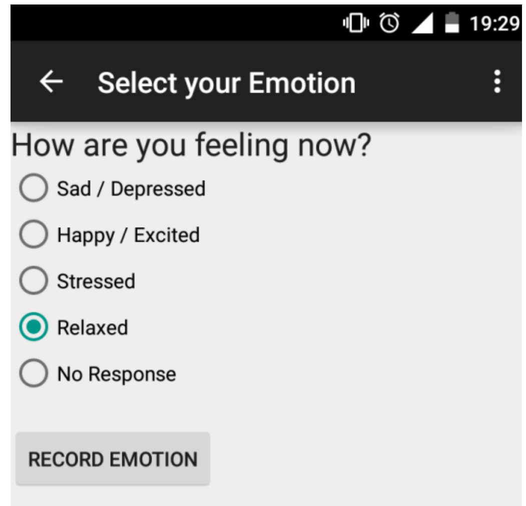
- Challenges
  - Extract Typing session
  - Collect Self-reports
    - Manual → *survey fatigue*
    - Psycho-physical sensor based → *intrusive setup*
  - Personalized training for every user



Experiment Apparatus



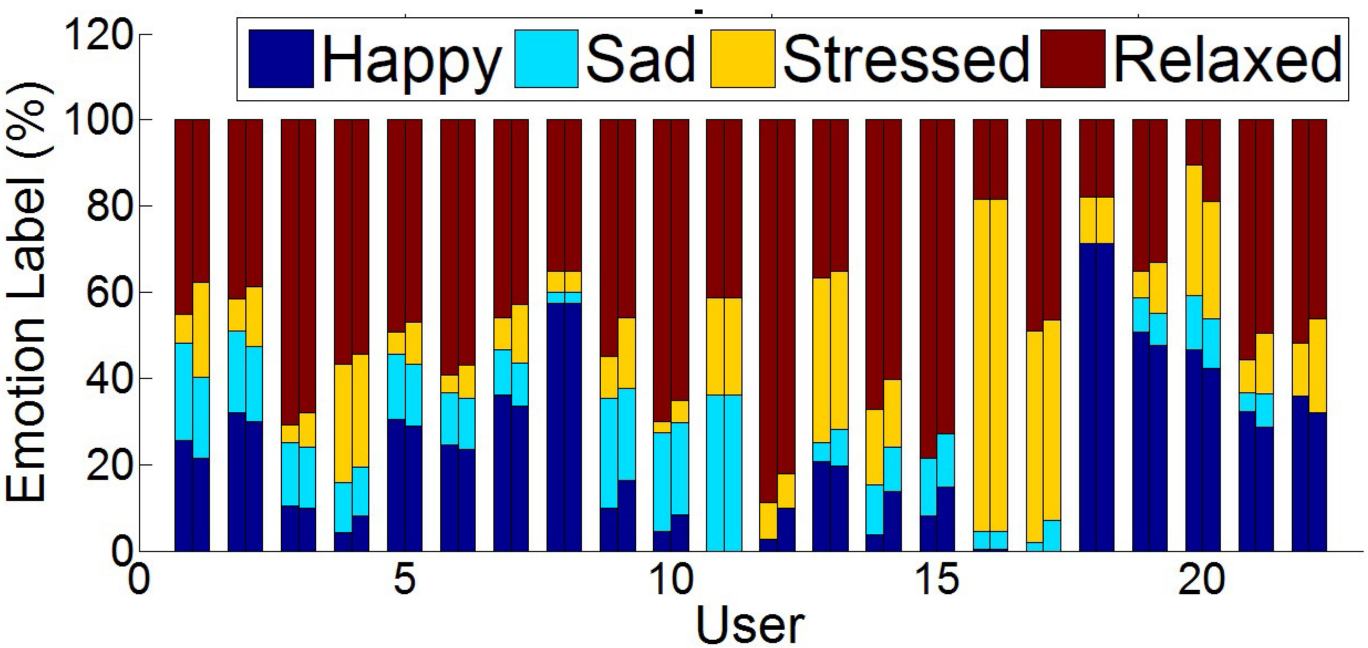
(a) *TapSense* Keyboard



(b) Emotion collection UI

- Custom keyboard to track typing
- Emotion recording can be skipped by selecting *No Response* in the self-report UI

Dataset

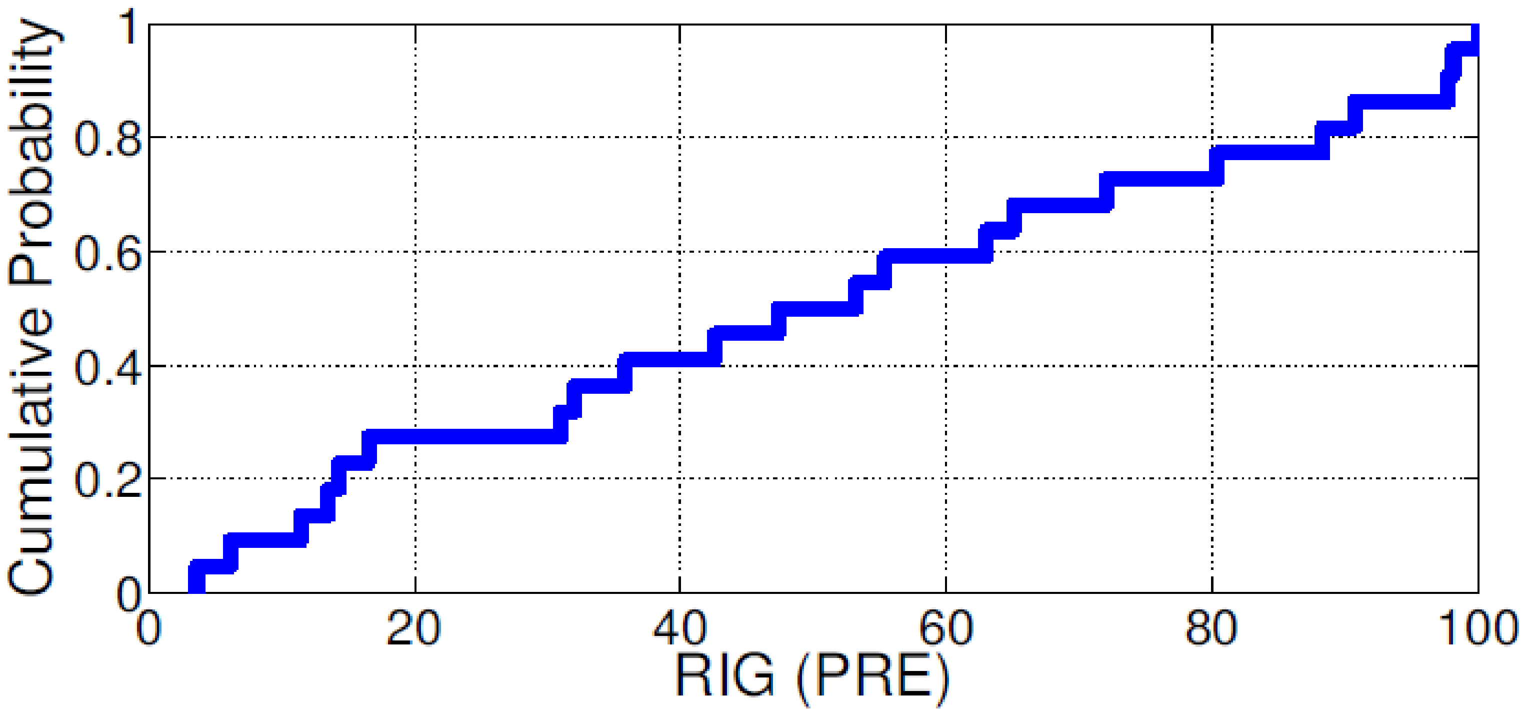


Emotion	Distribution
Happy	19%
Sad	9%
Stressed	23%
Relaxed	49%

- Applied SMOTE to overcome emotion imbalance

Evaluation

- Influence of *PRE*
  - Relative Information Gain (RIG) $RIG(f_i) = \frac{IG(f_i)}{\sum_{j=1}^{7} IG(f_j)}$



- Approximately 72% users are having *RIG (PRE)* > 30%

Evaluation

- Influence of *RMSI*

- One-way ANOVA
  - For every user, we form a group for every emotion state
  - Find users having at least one emotion state have significantly ($p < .05$) different *RMSI* than other

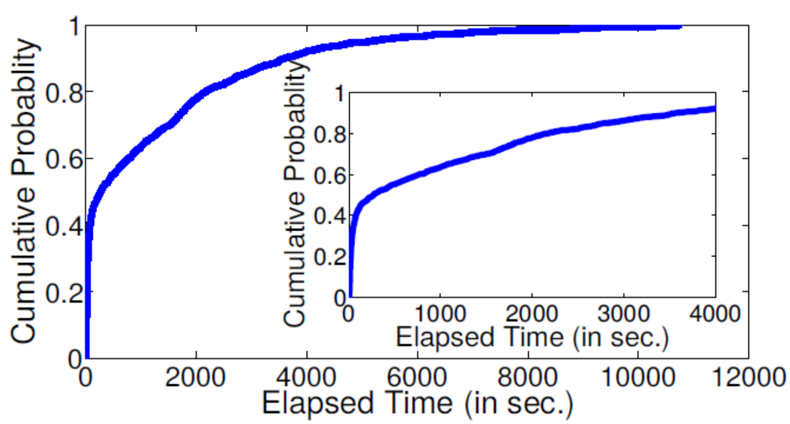
R - H	S - H	T - H	S - R	T - R	S - T
36%	78%	45%	78%	36%	56%

- *RMSI* alone can distinguish at least one emotion state for 50% of the population

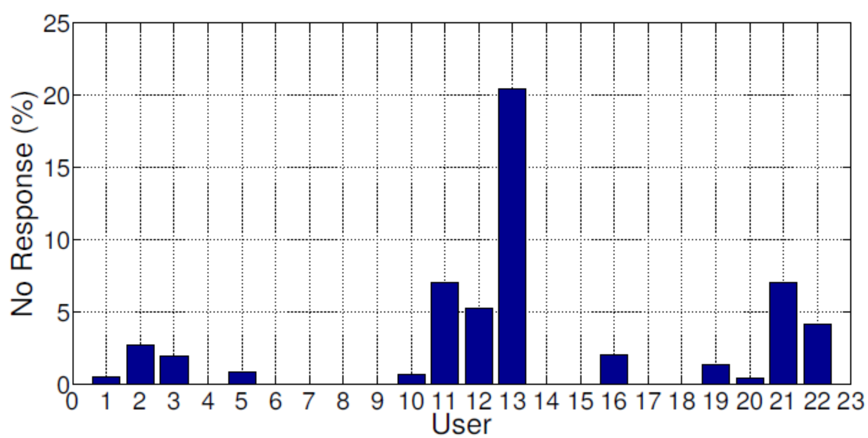
Evaluation

- How effective is the proposed ESM in collecting self-reports?

Parameter	Value
(i) Number of ESM probes	Avg. 4.6 (std. dev) per day for every user
(ii) Elapsed time between typing and emotion recording	Median elapsed time < 5 minutes
(iii) Number of No Response sessions	Only 2.5% of all sessions



(a) *Elapsed time* distribution



(b) *No Response* distribution

Self-reports are collected close to typing and does not cause major inconvenience.

Evaluation

- What is the amount of energy overhead?

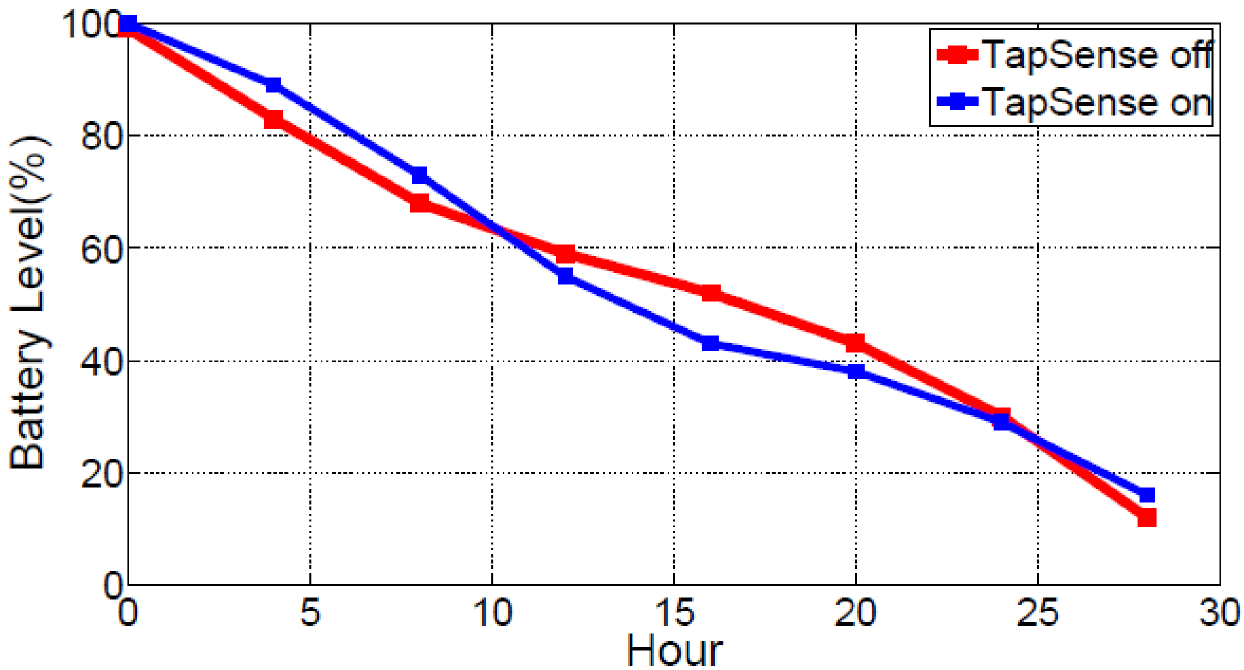


Figure 13: Battery depletion rate by keeping *TapSense* on and off

No significant variation in energy consumption with *TapSense*.

Evaluation

- Post-study Participant Feedback

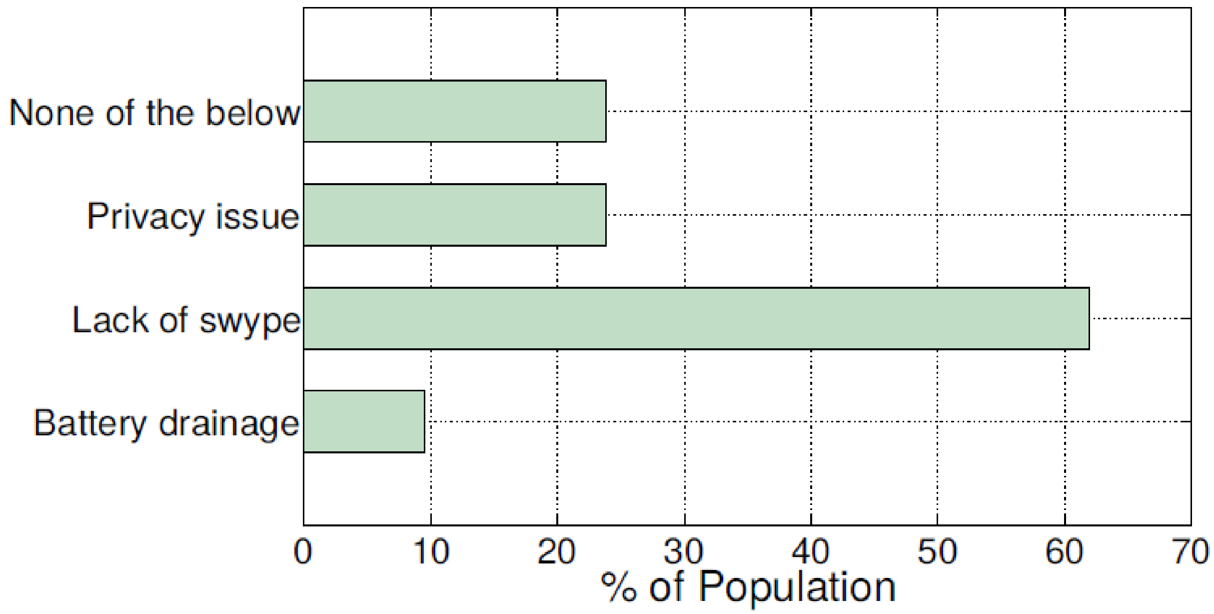


Figure 19: User preferences in *TapSense* usability survey

- 75% of the participants indicated as non-intrusive on a scale of 1 to 3
- Users were mainly concerned about absence of swipe facility