EmotionSense:  
A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research  
Rachuri K., Rentfrow P., Musolesi M., Longsworth C., Mascolo C., Aucinas A.

Mike Shaw

Computer Science Dept.  
Worcester Polytechnic Institute (WPI)
OUTLINE

- Motivation
- Related work
- Goal
- Assumptions & limitations
- Methodology
- Benchmarking
- Results
- Future work
MOTIVATION

● Study emotions and the relationship to environment

● Provide mental health and social science experts with –
  ● Emotional factors with respect to interpersonal relationships
  ● Identify locations and emotional responses
  ● Evaluate activity vs. emotions

● Smartphones allow field study w/o specialized equipment
  ● Past: In-home cameras, attached mics & diaries = Biased results
  ● Today: Ubiquity of smartphones desensitizes users from monitoring activities
RELATED WORK

- **Location and activity correlation**
  - BeTelGeuse [2] open source framework to gather situational information
  - CenseMe [4] detects activity at a location (e.g., dancing w/friends) and reports activity to social media

- **Social science experimentation**
  - Environmental activated recorder (EAR) to evaluate sociability contexts [3]

- **Self-reporting**
  - Use smartphone to report moods throughout the day to suggest therapy options [5]
GOAL

- “The overarching goal of EmotionSense is to exploit mobile sensing technology to study human social behavior.”
- Evaluate people’s emotions using smartphone sensors and speech-recognition tools to observe behavior patterns in social situations.
ASSUMPTIONS & LIMITATIONS

● Assumptions
  ● Participants will have smartphone with them majority of the time
  ● Microphone is unobstructed
  ● Participants gather frequently
  ● HTK produces correct results (before and after porting to Symbian)

● Limitations
  ● How well the participants represent persons who exhibit a wide range of detectable emotions
  ● How well the training data represents emotional signatures
METHODOLOGY

- Information flow

- Speaker recognition
  - Based on Gaussian Mixture Model (GMM) & Maximum A Posteriori (MAP) adaptation
  - Windows/Linux toolkit ported to Symbian OS

- Emotion recognition
  - Also based on Gaussian Mixture Model
  - Narrow emotional types are clustered into a broad classifications

- Adaptation framework
  - Generate rules to govern sensor sampling rates
Information Flow

- **Sensor Monitors/Classifier**
  - Movement detection
  - Bluetooth proximity detection
  - GPS monitor

- **Knowledge Base**
  - Converts sensor data into facts
  - `fact(<fact_name>, <value>)`
  - Ex: `fact(Activity, 1)`

- **Interference Engine**
  - Sensor sampling rate adaptation
  - Preservation of battery
  - Sample thresholds to minimize lossiness

- **Action Base**
  - Stores actionable events
  - `fact('action', <action_name>, <value>)`
  - Ex: `fact('action', 'ActivitySampling Interval', 10)`

**EmotionSense Manager**
- Starts all sensor monitor threads
- Instantiates Knowledge Base
- Invokes Inference Engine for fact collection
Speaker Recognition

Audio Data Collection & Parameterization

Apply GMM to distinguish study participants from others

MAP is applied to derive user-specific GMMs

Audio sequences are assigned user probabilities at run-time
Emotion Recognition

● Similar method as speaker recognition
  ● GMM trained on Emotional Prosody Speech and Transcripts library to classify emotions
  ● MAP adaptation is used to generate user specific models
  ● Emotional characteristics are assigned to audio sequences

● Emotion clustering
  ● Emotion grouping used by social psychologists
  ● Narrow emotion classification difficult even for humans

<table>
<thead>
<tr>
<th>Broad emotion</th>
<th>Narrow emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>Elation, Interest, Happy</td>
</tr>
<tr>
<td>Sad</td>
<td>Sadness</td>
</tr>
<tr>
<td>Fear</td>
<td>Panic</td>
</tr>
<tr>
<td>Anger</td>
<td>Disgust, Dominant, Hot anger</td>
</tr>
<tr>
<td>Neutral</td>
<td>Neutral normal, Neutral conversation, Neutral distant, Neutral tete, Boredom, Passive</td>
</tr>
</tbody>
</table>
BENCHMARKS

- Micro-benchmarks to evaluated system performance
  - Adaptation rules were collected from 12 users in a 24hr period
  - Tuned framework based on the Nokia’s 6210 sensor data captures

- Speaker recognition
  - 10min of training data from 10 users
  - Sample lengths varied from 1 to 15 seconds
  - 90% accuracy with sample lengths greater than 4 seconds

- Emotion recognition
  - Use pre-existing test and training library
  - 350 test samples per-sample length second
  - ~70% accuracy with sample lengths greater than 5 seconds
Benchmarks

Recognition accuracy & latency

Speaker recognition accuracy vs. audio sample length

Convergence ~90% > 4 seconds

Speaker recognition latency vs. audio sample length

Local benchmark based on 369MHz ARM 11 µP
Benchmarks

Power Consumption

Energy consumption vs. audio sample length

Energy consumption vs. maximum sampling interval
Benchmarks

Confusion Matrix

<table>
<thead>
<tr>
<th>Emotion [%]</th>
<th>Happy</th>
<th>Sad</th>
<th>Fear</th>
<th>Anger</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>58.67</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>29.33</td>
</tr>
<tr>
<td>Sad</td>
<td>4</td>
<td>60</td>
<td>0</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>Fear</td>
<td>8</td>
<td>4</td>
<td>60</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Anger</td>
<td>6.66</td>
<td>2.66</td>
<td>9.34</td>
<td>64</td>
<td>17.33</td>
</tr>
<tr>
<td>Neutral</td>
<td>6</td>
<td>5.33</td>
<td>0</td>
<td>4</td>
<td>84.66</td>
</tr>
</tbody>
</table>
RESULTS

● Trial conducted for 10 days with 18 participants

● Participant daily diaries
  ○ Activities
  ○ Who was present
  ○ Mood
  ○ Location

● Emotion Distribution
  ○ Neutral emotions are the most prevalent
  ○ Fear is the least prevent
Results

Emotion Distribution

Distribution of detected broad emotions

Most social activity exhibits neutral emotions

Distribution of detected broad emotions with respect to time of day

Emotions are more prevalent as the day progresses
Results

Emotion Distribution

Distribution of detected broad emotions within physical state

Non-neutral emotions are more prevalent in the idle state

Distribution of detected broad emotions with respect to number of co-located participants

Why is sadness experience in groups?
CONCLUSIONS

- Demonstrated smartphones are a viable tool for social science research
- Able to identify (to some degree) participant's emotions through speech recognition
- A majority of speech is categorized as neutral
- Emotion categorization algorithm produced underwhelming results
FUTURE WORK

- Galvanic skin response sensor
- Continue optimizing emotional recognition model
- Addition of more realistic noise models
- Real-time feedback, daily monitoring and user interaction options