Raja, Muneeba; Sigg, Stephan

RFexpress! - RF emotion recognition in the wild

Published in:
2017 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2017

DOI:
10.1109/PERCOMW.2017.7917516

Published: 02/05/2017

Document Version
Peer reviewed version

Please cite the original version:
RFexpress! – RF Emotion Recognition in the wild

Muneeba Raja  
Email: muneeba.raja@aalto.fi

Stephan Sigg  
Email: stephan.sigg@aalto.fi

Abstract—We present RFexpress! the first-ever system to recognize emotion from body movements and gestures via Device-Free Activity Recognition (DFAR). We focus on the distinction between neutral and agitated states in realistic environments. In particular, the system is able to detect risky driving behaviour in a vehicular setting as well as spotting angry conversations in an indoor environment. In case studies with 8 and 5 subjects the system could achieve recognition accuracies of 82.9% and 64%.

We study the effectiveness of DFAR emotion and activity recognition systems in real environments such as cafes, malls, outdoor and office spaces. We measure radio characteristics in these environments at different days and times and analyse the impact of variations in the Signal to Noise Ratio (SNR) on the accuracy of DFAR emotion and activity recognition. In a case study with 5 subjects, we then find critical SNR values under which activity and emotion recognition results are no longer reliable.

I. INTRODUCTION

Device free activity recognition (DFAR) leveraging radio frequency (RF) signal fluctuation has been explored intensively in recent research [1], [2]. However, most existing work has been carried out in controlled indoor environments, thereby ignoring many factors that affect radio wave propagation. For instance, the impact (in dB) that blocking of a signal path can have on the received signal is constrained by the strength of the signal itself. If the induced change lies below the Signal-to-Noise Ratio (SNR), the respective movement is unrecognizable for the DFAR system. We propose RFexpress! illustrated in Figure 1, an emotion recognition system that can be instrumented on standard off-the-shelf consumer equipment such as mobile phones, laptop computers or WiFi routers. RFexpress! exploits signal-fluctuation patterns recognizable at the RF-interface of a device to distinguish emotion-indicating movement, gesture and pose. This information is then exploited for the prediction of emotion and corresponding feedback. The recognition of emotion from DFAR systems, first suggested in [3], has been presented in [4] by reading pulse and respiration rate from RF-reflections for the recognition of emotional states. In contrast to our approach, such systems require ambitious measurement equipment and a controlled environment in which disturbance through movement in the environment is prohibited. The contributions of this paper are

1) RFexpress!: The first-ever RF-based DFAR emotion recognition system to exploit body movement, gesture and pose
2) Identification of critical SNR levels for DFAR-based recognition of relevant emotion-indicating gestures (case study with 5 subjects)
3) Exploitation of RFexpress! concepts for the first-ever RF-DFAR-based driver assistant system to detect risky driving behaviour (in-car case study with 8 subjects using a driving-simulation)
4) Exploitation of RFexpress! concepts for human motion-based emotion recognition (non-scripted case study with 5 subjects)

II. LITERATURE REVIEW

This section highlights previous research on (1) emotion recognition (2) activity recognition from RF signals.

1) Body Movements and Gestures for Emotion Sensing: The most well known emotion sensing methods exploit modalities such as facial expression, speech, text, physiological signals and body movements [5]. Modalities like facial expressions come with challenges like privacy intrusion, require high image quality and computationally intensive algorithms. Body movements, due to low cost and increasingly reliable body sensing technologies are suitable when emotions are to be detected from larger distances, for discriminating strong emotions and for emotions which can be deceived [6].

2) Body Movements and Gesture Recognition from RF: Activity and gesture recognition from RF is an active research topic. The fluctuation in the multipath propagation observed at a receive antenna indicates human movements or gestures [7], [8]. DFAR exploits existing infrastructure and overcomes device boundedness and limited range of traditional sensing.

Body movements map to human intentions, attention and emotions [5]. DFAR can, by detecting gestures or body movements, indicate respective emotions. However, in order to incorporate body movement detection with emotion sensing, we perform an SNR study to determine how wave propagation losses in real environments may limit the use of RF for activity detection and in turn emotion sensing.
RFexpress! follows a modular structure (cf. Figure 2). We will briefly overview the respective modules and leave the technical details for section V. The RF stimuli of activities are captured by the RF sensing module. We exploit RSSI information in our measurements in section IV. This continuous stream of data is then processed. The windowing, feature computation and data labelling is performed on processed data and features are forwarded to the classification learner. The classification learner is responsible for achieving a model for classifying gestures. The trained classification module is exploited to predict meaningful activity classes. These classes are then mapped to emotional states. In our case, in section V, we distinguish between angry and normal emotional states. This emotion-information is then used for the the generation of application-dependent feedback. We utilise and discuss these modules in the following sections.

III. RFexpress! Architecture

RFexpress! follows a modular structure (cf. Figure 2). We will briefly overview the respective modules and leave the technical details for section V. The RF stimuli of activities are captured by the RF sensing module. We exploit RSSI information in our measurements in section IV. This continuous stream of data is then processed. The windowing, feature computation and data labelling is performed on processed data and features are forwarded to the classification learner. The classification learner is responsible for achieving a model for classifying gestures. The trained classification module is exploited to predict meaningful activity classes. These classes are then mapped to emotional states. In our case, in section V, we distinguish between angry and normal emotional states. This emotion-information is then used for the the generation of application-dependent feedback. We utilise and discuss these modules in the following sections.

IV. Impact of SNR on Gesture Recognition

We determine the accuracy of DFAR systems in real environments. We first measure the radio characteristics of real environments where the SNR is the primary parameter. Then, we model these SNR values, and perform case studies with 5 subjects to identify critical SNR values for DFAR.

1) Measuring SNR Values in Real Environments: We consider office room, corridor of a building, mall, cafe and outdoor for the SNR study. We measure the signal strength and noise between a wifi router and a laptop at different distances. The experiment for each environment is repeated 5 times on different days and times of the day. Table I shows the SNR with changing distances in different environments.

Table I: Average SNR observed in all environments for various distances between the transmitting WiFi router and the laptop.

<table>
<thead>
<tr>
<th>Environment</th>
<th>0m</th>
<th>8m</th>
<th>17m</th>
<th>25m</th>
<th>30m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cafe</td>
<td>77.3</td>
<td>42</td>
<td>20.6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>outdoor</td>
<td>60.7</td>
<td>45.1</td>
<td>44.9</td>
<td>36.7</td>
<td>37.3</td>
</tr>
<tr>
<td>office</td>
<td>74.3</td>
<td>52.1</td>
<td>41.6</td>
<td>30.6</td>
<td></td>
</tr>
<tr>
<td>building corridor</td>
<td>76.09</td>
<td>49</td>
<td>56.4</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>mall</td>
<td>71.3</td>
<td>46.49</td>
<td>40.2</td>
<td>36.51</td>
<td>25.9</td>
</tr>
</tbody>
</table>

Table II: Impact of SNR on accuracy of gesture/activity detection. distance=2m & 4m.

<table>
<thead>
<tr>
<th>SNR Value (db)</th>
<th>Hands up/down, d=2m (%)</th>
<th>Hands up/down/claps, d=2m (%)</th>
<th>Hands up/down, d=4m (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
<td>90</td>
<td>83</td>
<td>84</td>
</tr>
<tr>
<td>42</td>
<td>88</td>
<td>81</td>
<td>83</td>
</tr>
<tr>
<td>22</td>
<td>79</td>
<td>68</td>
<td>69</td>
</tr>
<tr>
<td>12</td>
<td>65</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>53.5</td>
<td>47</td>
<td>49</td>
</tr>
<tr>
<td>0</td>
<td>52</td>
<td>41</td>
<td>42</td>
</tr>
</tbody>
</table>

2) Modelling SNR values in Lab Environment: The SNR vs. distance values obtained from different environments provide us with the range of SNR values to model in our lab environment. We set up USRP devices (2m apart) (Figure 3) and omni-directional antennas. In order to obtain the signal strength, we measure the power level at the receiver using a spectrum analyser during transmission, and subtract the added noise value. The transmitting frequency is 100kHz with a sampling rate of 1MHz and centre frequency of 2.48GHz. We can then tune power and noise values at the transmitter to obtain the desired SNR values at the receiver.

3) Modelling Real Environments and Analysing Effects of SNR on Activity Detection: Activity detection at 6 different modelled SNR values has been tested with 5 subjects. For each SNR, the person repeats the experiment 5 times with a system trained on 3 activities. We choose gestures for this experiment which map to certain emotional states: hands up, hands down and clapping.

4) Data Pre-processing and classification: We process the collected raw data using signal denoising, as in Abdelnasser et al. [1]. Then we calculate statistical and frequency-domain features (mean, standard deviation, entropy, zero crossing and average derivative) for activity labelling. We use the non overlapping window size to be 100,000 samples with 1MHz sampling frequency.

5) Analysis: The Table II shows that higher SNR fosters better accuracy but also depends on the complexity of the activity (cf. Figure 4). SNR values of about 30dB and higher can be considered for robust activity recognition of different activities. 20dB and below is not feasible for DFAR.
Fig. 4: The impact of SNR on gesture recognition. Each graph shows the gesture recognition at different SNR value. The gestures are hands down vs. clapping.

(a) neutral  (b) angry  (c) neutral  (d) angry

Fig. 5: Images captured during the driving car and conversation experiments. a, b indicate neutral vs. angry driving. c, d indicates the neutral vs. angry conversation.

V. REAL WORLD APPLICATIONS FOR EMOTION RECOGNITION

We exploit RFexpress! concepts for the distinction between different emotional states in two realistic cases: (1) Detection of risky, agitated driving behaviour and (2) detection of angry argument in an indoor setting.

A. Scenario 1: Detection of risky driving behaviour

Based on our results at different SNR levels that we presented, we venture forward to evaluate the effectiveness of DFAR systems in real world scenarios that may suffer from SNR degradation. In particular, we distinguish normal and angry driving behaviour.

1) Experiment: We performed a driving experiment with 8 subjects (5 males), aged 18-40 and belonging to 4 different countries. Each subject was assigned 30 minutes for performing the experiment (Figure 5 a, b). A driving simulator video \[1\] was played on a tablet and fixed on the windscreen. For normal driving, the subjects were asked to drive through the simulator video while performing normal driving. For the angry driving case, we created a cover story considering previous research [9] for emotion elicitation. The subjects act and respond accordingly.

2) Data Collection: The set-up for data collection can be seen in Figure 6. We used mean and variance features with non overlapping windows and a k-Nearest Neighbour (k-NN) classifier with k=6 neighbours and chose inverse weighted distance for computation. We trained the models for individual and aggregated data from all subjects.

3) Results and discussion: An overall accuracy of 98% for individual model and 82.9% for the inter-subject combined data model is achieved. Classification results after 5-fold cross

\[1\]https://www.youtube.com/watch?v=lyVctzSBArO

Fig. 6: Experiment set up for gestures activity detection and emotion recognition in car.

Fig. 7: Graphs of driving experiment for neutral (green) and angry case (red).

(a) scatterplot of one subject  (b) scatterplot of all subjects

Fig. 8: Scatterplots showing neutral (blue) and angry (red) case.

<table>
<thead>
<tr>
<th></th>
<th>Individual Training</th>
<th>Aggregated Training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neutral</td>
<td>Angry</td>
</tr>
<tr>
<td>Neutral</td>
<td>97%</td>
<td>3%</td>
</tr>
<tr>
<td>Angry</td>
<td>1%</td>
<td>99%</td>
</tr>
<tr>
<td>Precision</td>
<td>0.94</td>
<td>0.96</td>
</tr>
</tbody>
</table>

TABLE III: Confusion matrices from individual and aggregated trained model.
Fig. 9: Graphs of conversation experiment for neutral (green) and angry case (red).

<table>
<thead>
<tr>
<th></th>
<th>Distance= 2m</th>
<th></th>
<th>Distance= 5m</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neutral</td>
<td>Angry</td>
<td>Recall</td>
<td>Neutral</td>
</tr>
<tr>
<td><strong>Neutral</strong></td>
<td>89%</td>
<td>11%</td>
<td>0.86</td>
<td>96%</td>
</tr>
<tr>
<td><strong>Angry</strong></td>
<td>22%</td>
<td>78%</td>
<td>0.83</td>
<td>29%</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td>0.9</td>
<td>0.76</td>
<td></td>
<td>0.8</td>
</tr>
</tbody>
</table>

TABLE IV: Confusion matrices from a single subject’s trained model. Left: 2m distance bw receiver and subject; Right: 5m.

Validation are shown in Table III. If the angry state is being spotted for a long interval, then the driver needs to provided with safety feedback.

B. Scenario 2: Angry Behaviour Detection in a Conversation

In this scenario, a subject carries out a conversation, with another subject either on the phone or in person in an office environment. The main subject is standing and free to move during the conversation which makes it more flexible and challenging than the car environment.

1) Experiment: 5 subjects, with 3 nationalities participated in the study. Images from the experiment are shown in Figure 5(c, d). The space is a large meeting room of about 12x18 sqft. The distance between the transmitter and receiver is 8m, SNR is 42dB and all the equipment, configuration, emotion induction, data collection, pre-processing and feature calculation are the same as for the driving experiment in Section V-A. For each subject, the first neutral vs. angry data is first captured by keeping 2m distance between receiver and subject and then at 5m.

2) Results and Discussion: The results of the classification are shown in Table IV. As the SNR value is high enough for gesture recognition, it is clear that the body gestures are promising indicators of anger. The overall accuracy has, however, decreased as compared to results for driving experiments. The graphs in Figure 9 represent the difference of signals tracked between neutral case and angry case. Inter-subject classification for all subjects reduces the accuracy down to 64% for this scenario.

VI. CONCLUSION

We have presented RFexpress!, the first-ever Device-free motion and gesture-based emotion sensing system. The system has been exploited in a vehicular scenario and an indoor scenario, to detect risky driving behaviour and to distinguish between neutral and angry human communication. The experiment on the risky driving detection is performed by 8 drivers. The angry emotion is induced using well established emotion elicitation techniques. We observed that a car has well suited conditions for DFAR-based emotion recognition as could achieve an accuracy of 98%. In the office scenario, we considered the distinction of angry versus neutral conversation with increasing distance between subject and receiver. In this experiment we could achieve an accuracy of up to 82.9% for individually trained models and 64% for inter-subject models. The aggregated learning is challenging due to variation in body expressions with changing emotions in persons.

In addition, we studied the accuracy of DFAR, in real environments as compared to controlled environments. We measure the radio characteristics in different environments. Then we model them in our lab and perform gesture recognition experiment to analyse the variation in its accuracy with changing modelled radio characteristics. With SNR as our indicator, we consider 6 different SNR values and 3 gestures are detected at each SNR value. This experiment is performed by 5 participants 5 times each for every activity. The results show that accuracy above 80% can be achieved at SNR higher than 30dB. At SNR 20db and below, the accuracy of gesture recognition drops significantly. This accuracy also varies with the complexity of gesture performed.

For future work, we plan to extend our emotions to neutral, angry, happy, tired and sad and achieve high accuracy for car driving on road with more subjects varying in age, height and regions. We also need to have a clear methodology to distinguish between emotions for which the body language can be similar.

REFERENCES


