Improved Multimodal Emotion Recognition for Better Game-Based Learning

Improved Multimodal Emotion Recognition for Better Game-Based Learning

Kiavash Bahreini

Dr. Rob Nadolski

Prof. dr. Wim Westera

For OULU Team from Finland, December 9, 2014,

Heerlen, The Netherlands

Welten Institute
Research Centre for Learning, Teaching and Technology

Open Universiteit
welten-institute.org

Agenda

1. What is this research about?
2. What is the target group?
3. Why this research?
4. How to do this research?
5. What have been done so far?
    1. Framework, participants, and software
    2. Experiment
    3. Results
6. Future direction

What is this research about?

1. Multimodal emotion recognition from two sources:
    1. Face and voice
        1. Happy, sad, fear, disgust, surprise, angry, and neutral

2. Learner support in serious games:
    1. Enhancing online communication skills training
    2. Self-awareness of own behavior
    3. Provide timely and adequate feedback

What is the target group?

1. Life long learners

(learners who are interested in improving communication skills using emerging technologies)

Why are we doing this research?

1. The importance of communication skills in knowledge society nowadays
2. Market demands
3. Communication-skills are high priority at EU level[1, 2, 3, 4]
4. Provide an attractive environment with regard to game-based learning

1. http://www.euca.eu/eu-project-erasmus-modes
2. http://www.epc.eu/documents/uploads/pub_1160_skills_and_education.pdf
3. http://softskillsproject.com/
4. http://www.fas.ie/en/pubdocs/SoftSkillsDevelopment.pdf

How to do this research?

1. Using devices and modern equipment
   - Webcams and microphones
2. Some developments: The overarching framework and software
3. Gather facial and vocal emotions in real-time
4. Integrate the software artifacts with a game-based engine:
   1. EMERGO (a game-based toolkit for delivery of multimedia cases)

What have been done so far?

Layers of FILTWAM

Learner

Web Interface

EMERGO Web Service Client

uses

Affective Computing Tool

Emotion Recognition from Facial Features

Face Detection → Face Feature Extraction → Face Emotion Classification

uses → Webcam

Emotion Recognition from Vocal Features

Voice Detection → Voice Feature Extraction → Voice Emotion Classification

uses → Microphone

Feedback /Content

Device

Facial Emotion Dataset

Vocal Emotion Dataset

Data

Live Stream

Internet

Network

Live Stream

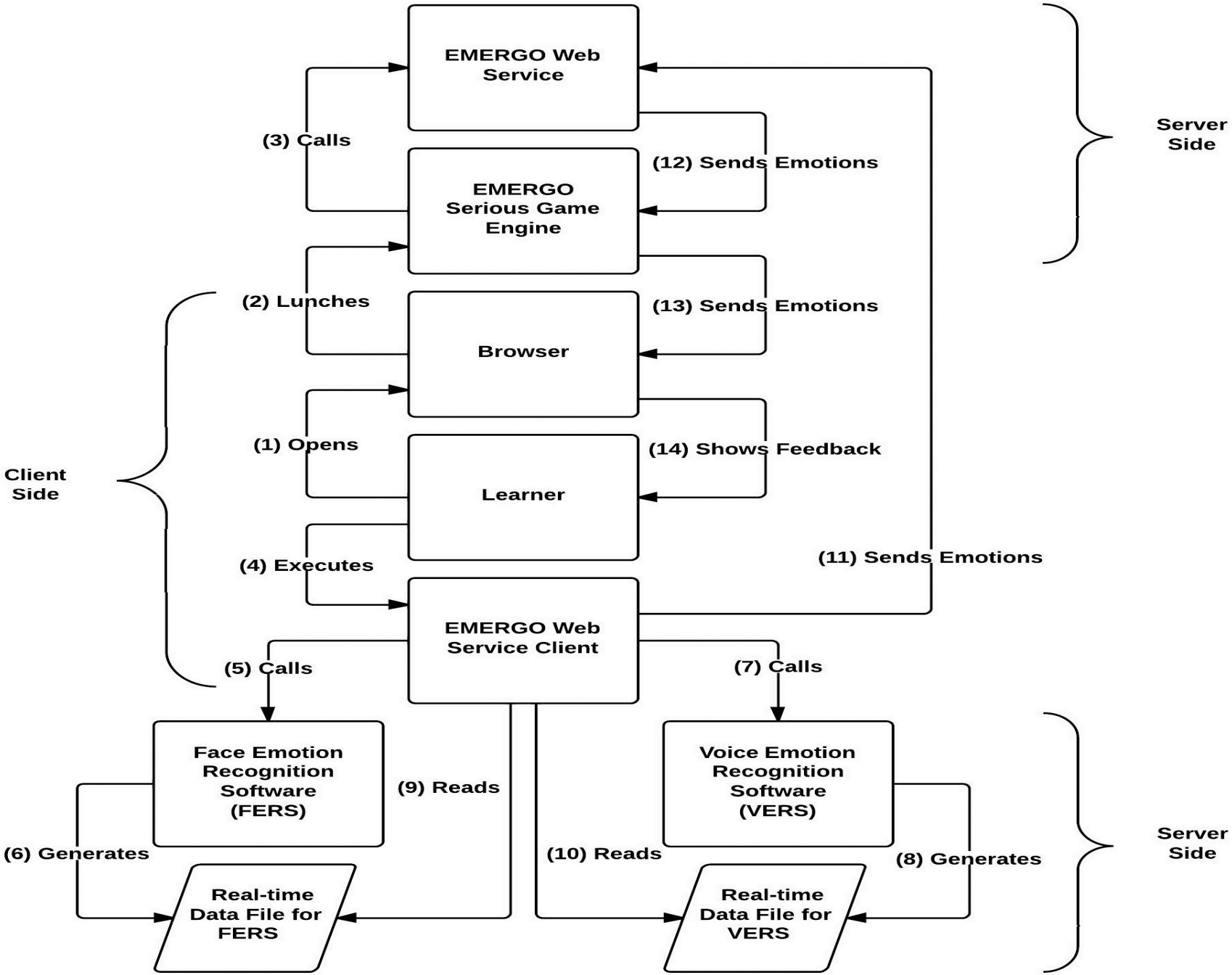
Feedback/Content

Game-based Learning Environment

EMERGO Web Service uses EMERGO Rule Engine

Application

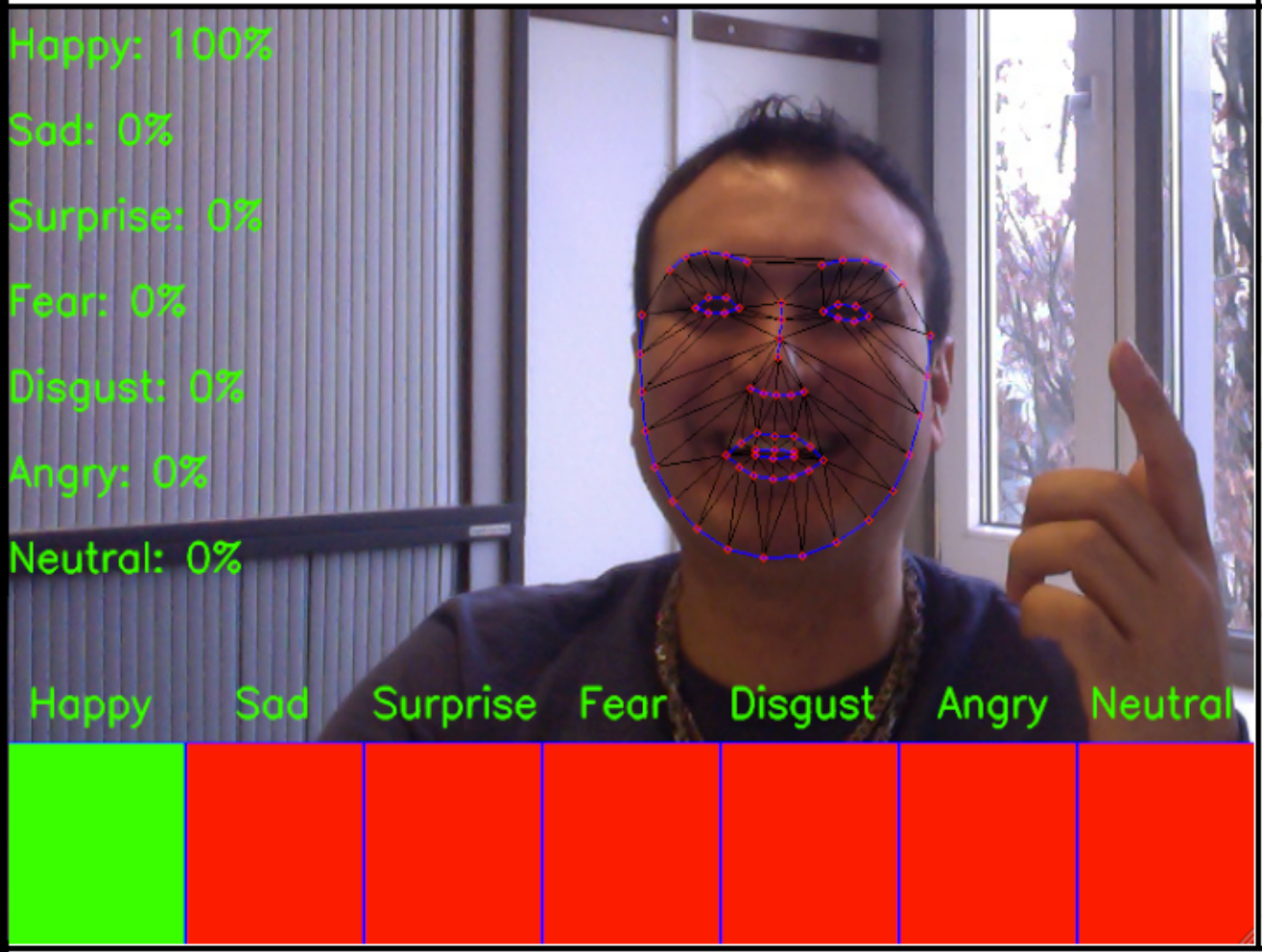
Data flow of the architecture



Participants and tasks

- Twelve participants (7 male, 5 female; age M=42, SD=10)
- Four consecutive tasks were given to the participants:
  - Mimic and repeat loudly the emotion that was presented through PowerPoint slides,
  - Mimic and repeat loudly the seven voice expressions,
  - Slides presented a text transcript (both sender and receiver) taken from a good-news conversation, participants were requested to read and speak aloud the sender 'slides' of transcript,
  - as in task 3, but in this case the text transcript was taken from a bad-news conversation

Face Emotion Recognition



Voice Emotion Recognition

```
    1        1:?  2:Neutral       0.923


=== Predictions on test data ===

 inst#     actual  predicted error prediction ()
    1        1:?  2:Neutral       0.677


=== Predictions on test data ===

 inst#     actual  predicted error prediction ()
    1        1:?  2:Neutral       0.824


=== Predictions on test data ===

 inst#     actual  predicted error prediction ()
    1        1:?  2:Neutral       0.677


=== Predictions on test data ===

 inst#     actual  predicted error prediction ()
    1        1:?  1:Happy         0.923
```

Please read the following sentence once you click on OK

Cancel OK

PowerPoint Slides

Sender Emotions: Happy

Actually, I have some really good news for you.

Participants' opinions

1. All tasks were moderately difficult and interesting to do
2. Instructions were clear
3. The feedback was pretty helpful
4. The self-assurance factor was not high among the participants
5. There was no distraction during the performance
6. The participants did not regard themselves as actors

Validation results of the face emotion recognition software

Validation of the Recognized Emotion by the Face Emotion Recognition Software Module

Happy	Sad	Surprise	Fear	Disgust	Angry	Neutral	Total
0.84	0.66	0.69	0.67	0.66	0.77	0.8	0.76

The Kappa value for the validation results of the face emotion recognition module for all the seven emotion for task 1, task 2, task3, and task 4.

Validation results of the voice emotion recognition software

Validation of the Recognized Emotion by the Voice Emotion Recognition Software Module

Happy	Sad	Surprise	Fear	Disgust	Angry	Neutral	Total
0.63	0.50	0.51	0.48	0.41	0.50	0.71	0.58

The Kappa value for the validation results of the voice emotion recognition module for all the seven emotion for task 1, task 2, task3, and task 4.

Validation results of the integration of the two software artifacts

Validation of the Recognized Emotion by the Face and the Voice Emotion Recognition Software Modules

Happy	Sad	Surprise	Fear	Disgust	Angry	Neutral	Total
0.68	0.50	0.53	0.50	0.43	0.55	0.73	0.61

The overall Kappa value for the validation results of the face and the voice emotion recognition software modules for all the seven emotion for task 1, task 2, task3, and task 4.

Multimodal fusion of the two software modules

1. Linear weighted fusion method, which is a type of rule-based fusion

$$V_m, 1 \le m \le n$$

$$V = \frac{1}{n} \sum_{m=1}^{n} w_m \times V_m$$

The overall accuracy of the multimodal fusion

Voice Emotion Recognition	Face Emotion Recognition						
	Happy	Sad	Surprise	Fear	Disgust	Angry	Neutral
Happy	89.7%	56%f	79%f	58%f	75%f	83%f	90%f
Sad	80%f	74.4%	79%f	66%v	75%f	83%f	90%f
Surprise	80%f	56%f	67.8%	58%f	75%f	83%f	90%f
Fear	80%f	56%f	79%f	50.3%	75%f	83%f	90%f
Disgust	80%f	56%f	79%f	58%f	63%	83%f	90%f
Angry	80%f	63%v	79%f	63%v	75%f	100%	90%f
Neutral	80%f	61%v	79%f	61%v	75%f	83%f	100%

Results

1. The overall accuracy of our face emotion recognition software based on the requested emotions and the recognized emotions is 75%.
2. The overall accuracy of our voice emotion recognition software based on the requested emotions and the recognized emotions is 52%.
3. Compare to our previous study, the accuracy of our voice emotion recognition dataset improved from 22.2% to 50%.
4. There are two reasons for the obtained false results:
    1. The malfunctioning of the software
    2. The participants were unable to mimic the requested emotions accurately

Future research

Creating a serious game in EMERGO

Tasks

Score Counter: 5/100

Completed Snippets: 1/20

A recorded video of the conversation partner

Reward

Detected Emotion from Face:

Happy

Detected Emotion from Voice:

Happy

The selected sentence and the emotion are:

- I'm pleased to announce that you've been selected! --- Neutral
- I'm pleased to announce that you've been selected! --- Happy
- I'm pleased to announce that you've been selected! --- Surprise
- Human resources will contact you to review the details of your new contract. --- Neutral
- Human resources will contact you to review the details of your new contract. --- Happy
- Human resources will contact you to review the details of your new contract. --- Surprise

You can now continue by pressing Next button.

Redo

Next

Thank you

Kiavash Bahreini

PhD Candidate at Welten Institute

Research Center for Learning, Teaching and Technology

Open University of the Netherlands


kiavash.bahreini@ou.nl


http://nl.linkedin.com/pub/kiavash-bahreini/29/9b6/93


http://twitter.com/KiavashBahreini


kbsocialnetwork@gmail.com


+31 (0) 61 41 28 48 5