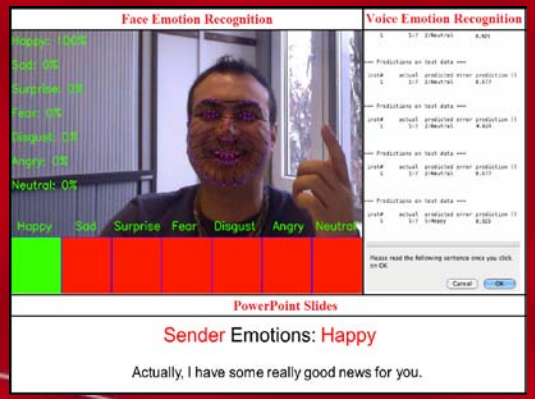
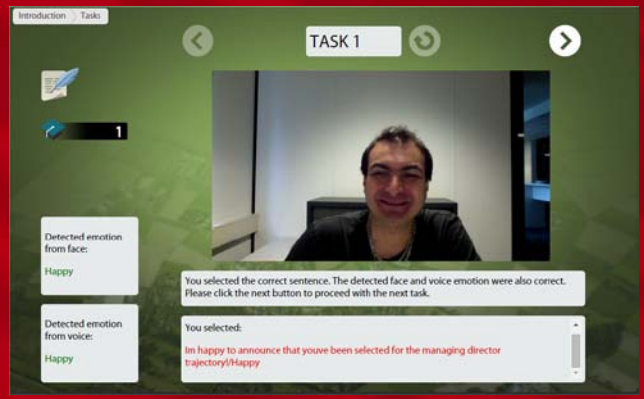
Multimodal Emotion Recognition as Assessment for Learning in a Game-Based Communication Skills Training



Kiavash Bahreini

Rob Nadolski

Wim Westera



ECTEL Conference, September 16-19, 2014

Graz, Austria

Welten Institute
Research Centre for Learning, Teaching and Technology

Open Universiteit
welten-institute.org

CC BY

Overview

1. Research area
2. The need for more and better communication skills
3. The problem with current communication trainings
4. Our – contribution to the - solution: *Communication Advisor*
5. Our R&D on assessment FOR learning with real-time multimodal emotion recognition
6. How to research the “evidence” of our solution?
7. Questions-discussion

1. Research area

1. Learner support (e.g., feedback) in serious games

   Goal: Provide **timely** and **adequate** feedback
   (feedback is based upon: cognitive states & affective states)

2. Learner support = Assessment *FOR* Learning (gather data + analysis)
   - multimodal emotion recognition from two sources (Face and Voice)
   - user choices (branching-stories, click-stream data)
   [unobtrusive in-game assessment]

3. Application area PhD project Kiavash
   - Communication skills
   [align verbal and non-verbal: learn to express the right emotion at the right time]
   Learner support towards
   - Self-awareness of own behavior
   - Enhancement of communication skills

   (multimodal emotion recognition can ALSO detect progress (cognitive)])

2. The need

More and better communication skills for life long learners/knowledge society

Communication-skills are given high priority at the EU level[1, 2, 3, 4]

align verbal and non-verbal

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

3. The problem

- “teacher bandwidth” impedes timely feedback

- prolonged practice (tiresome)

4. Our solution: *Communication Advisor(CA)*

Digital *game-based* communication training

- guided practice based upon multimodal emotion recognition **and** performance

Aims to result into:

- timely feedback, more personalized, and more motivating training for learners

no replacement of teachers, but more efficient for teachers

[*CA* as front-end for labor-intensive f2f-training with teacher support]

5. Our R&D on assessment FOR learning [with real-time Multimodal Emotion Recognition] ***geared towards the Solution***

1. Webcams and microphones (consumer equipment) for gathering facial and vocal emotions
2. An overarching framework and ***validated*** software development of artifacts for analyzing facial and vocal emotions in real-time (FILTWAM) [Research (R) and Development (D)]
3. Integration of the FILTWAM-software artifacts with a game-based engine EMERGO (a methodology and an open source game-based toolkit for the development and delivery of scenario-based serious games) [Development (D)]

Layers of FILTWAM

GALA
Games and Learning Alliance
Network of Excellence for Serious Games

Learner

Web Interface

EMERGO Web Service Client

uses

Feedback /Content

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

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

5. Research of software validation for Multimodal Emotion Recognition - Study

- Participants (adults: 7 male, 5 female) had to do four consecutive tasks:

  [1] Mimic (face) and speak loudly (voice) the requested emotion on the shown PowerPoint slide

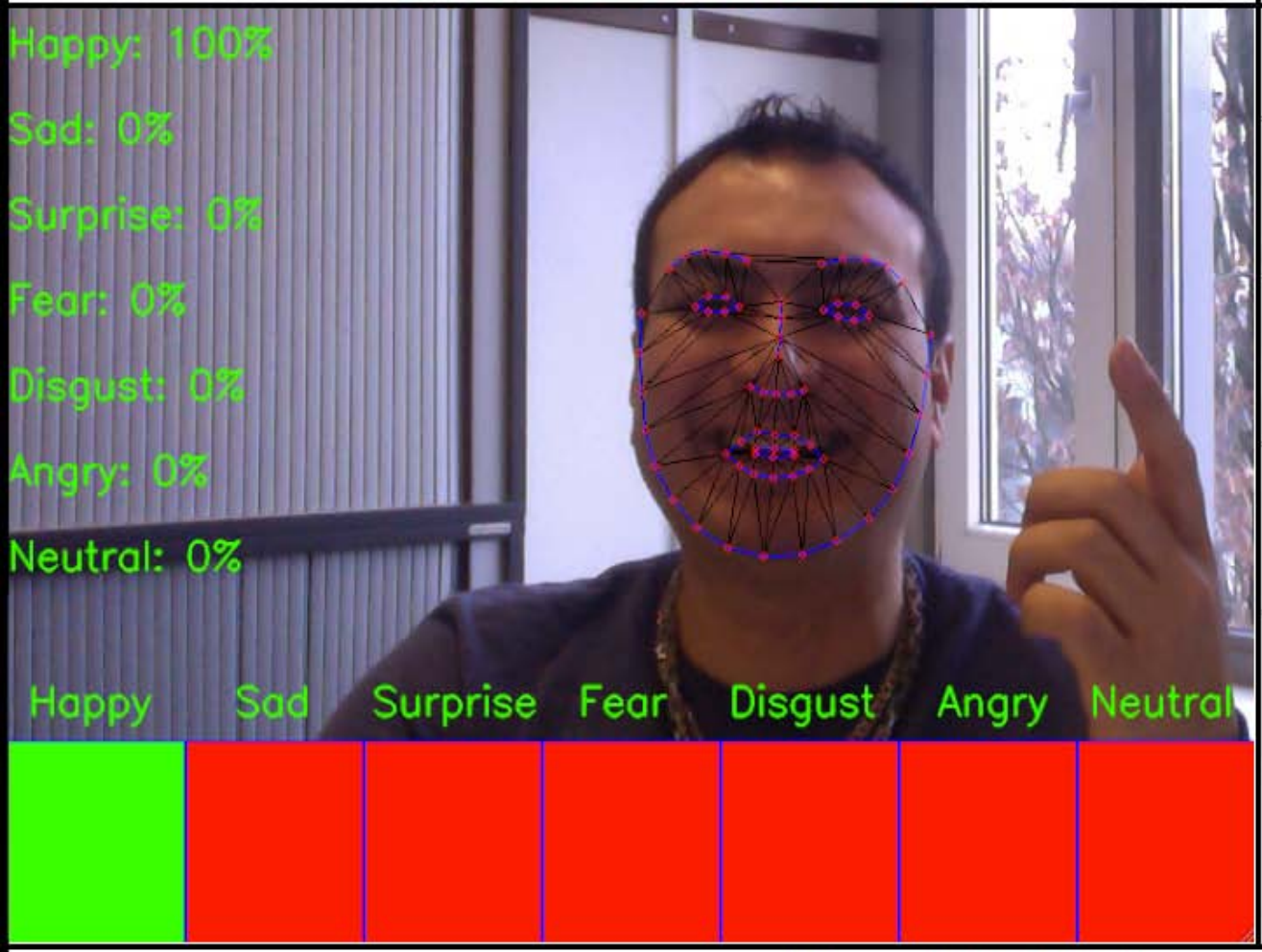
  [2] Mimic (face) and speak loudly (voice) the seven basic voice emotions on the shown PowerPoint slide (happy, sad, fear, disgust, surprise, angry, and neutral)

  [3] Mimic and speak aloud the sender slides of the presented text transcript (both sender and receiver) taken from a good-news conversation

  [4] As in task 3, but in this case the text transcript was taken from a bad-news conversation

- (in all cases, the recognized emotion was presented as Knowledge of Response-feedback: 'label of recognized emotion' was shown in green if it was the same as the requested emotion, in red if it differed from the requested emotion)

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.

5. Results of software validation for Multimodal Emotion Recognition

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.

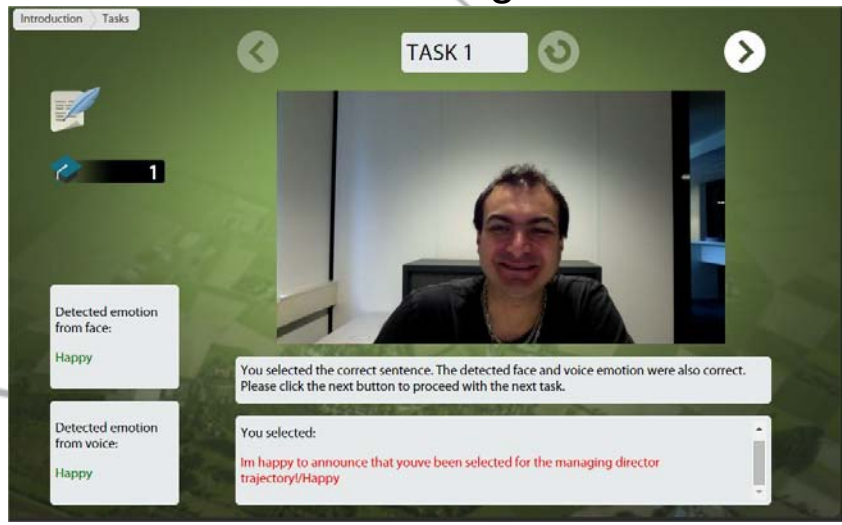
Validation of the Recognized Emotion by the Face and the Voice Emotion Recognition Software Modules

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.

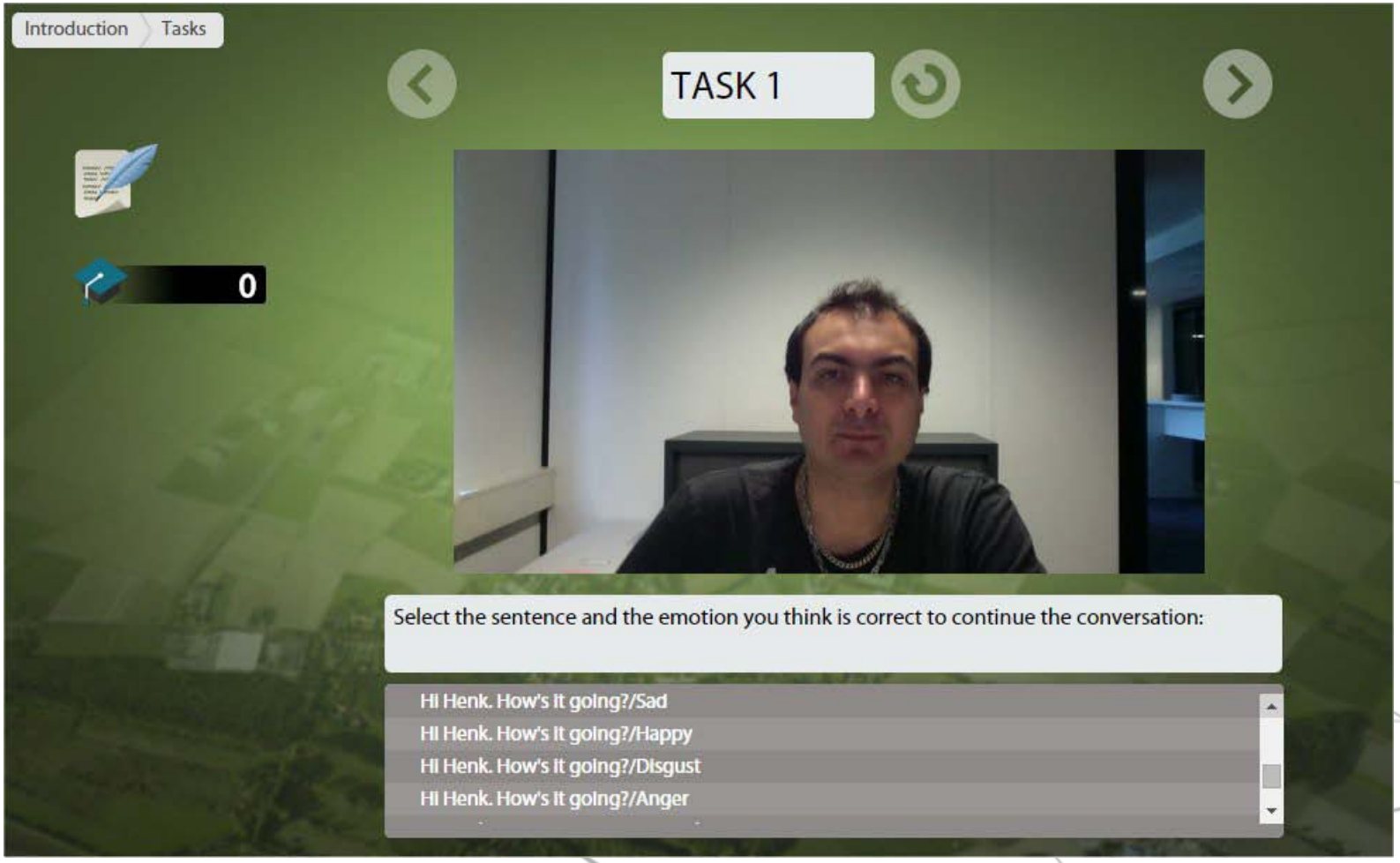
6. How to research the “evidence” of our Solution: proposal for Experimental design

- Five versions of *Communication Advisor* with immediate (timely) feedback
  - 1. [Cognitive] FB [on choosen alternatives])
  - 2. [Cognitive + one_sensorAffective] FB (face emotion detection)
  - 3. [Cognitive + one_sensorAffective] FB (voice emotion detection)
  - 4. [Cognitive + two_sensorAffective] FB (face & voice emotion detection)
  - 5. Control (= **no** FB)
- Twenty participants for each version (adults (female, male)) - randomized
- All versions contain ten micro stories/tasks + same reward mechanism
- Pre-test & Post-test on communication skills within all versions, using micro stories without feedback [→effectiveness] (scores)
- Motivation [→enjoyability] (questionnaire)
- Time investment [→efficiency] (time stamps)

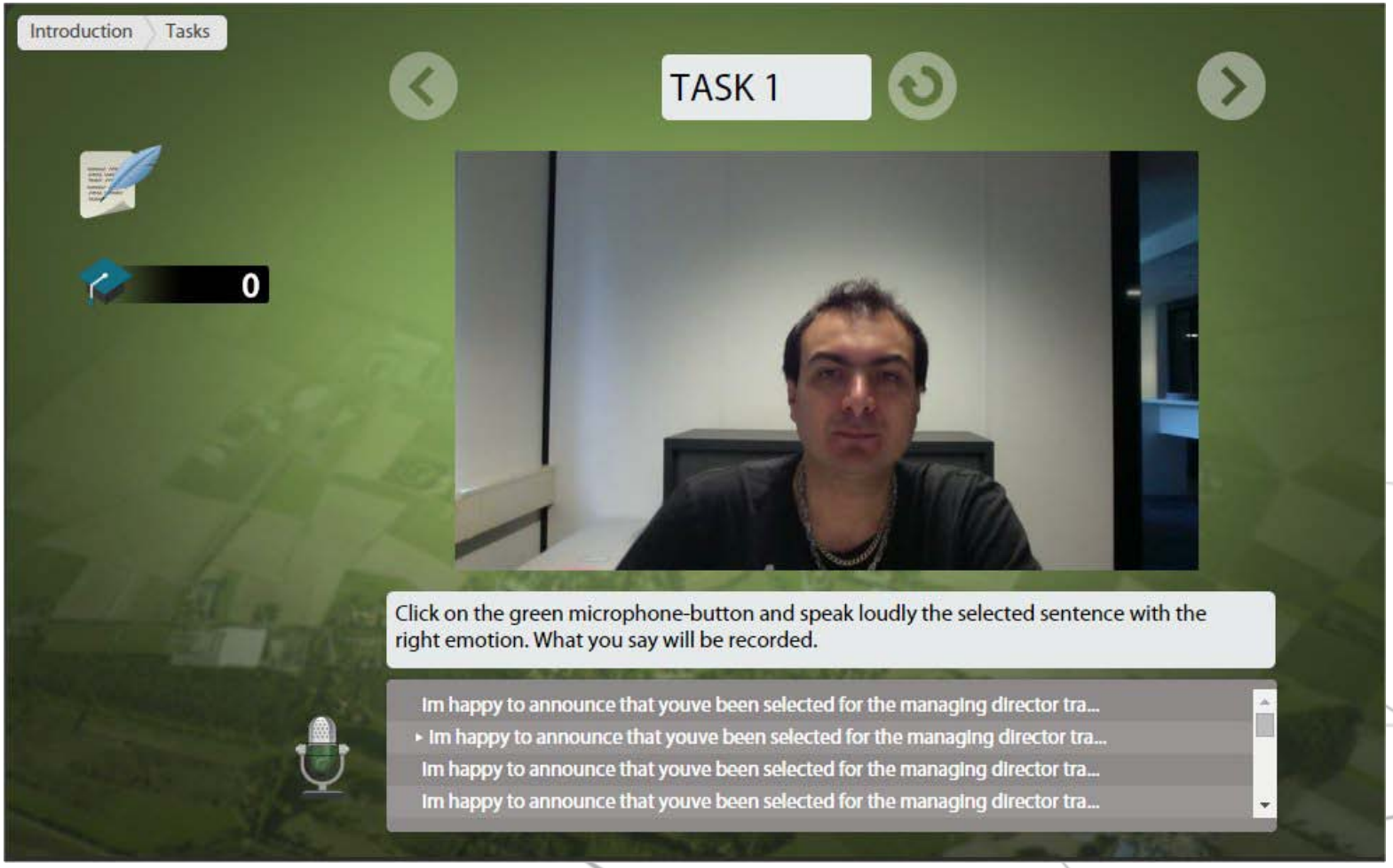


Welten Institute
Research Centre for Learning, Teaching and Technology

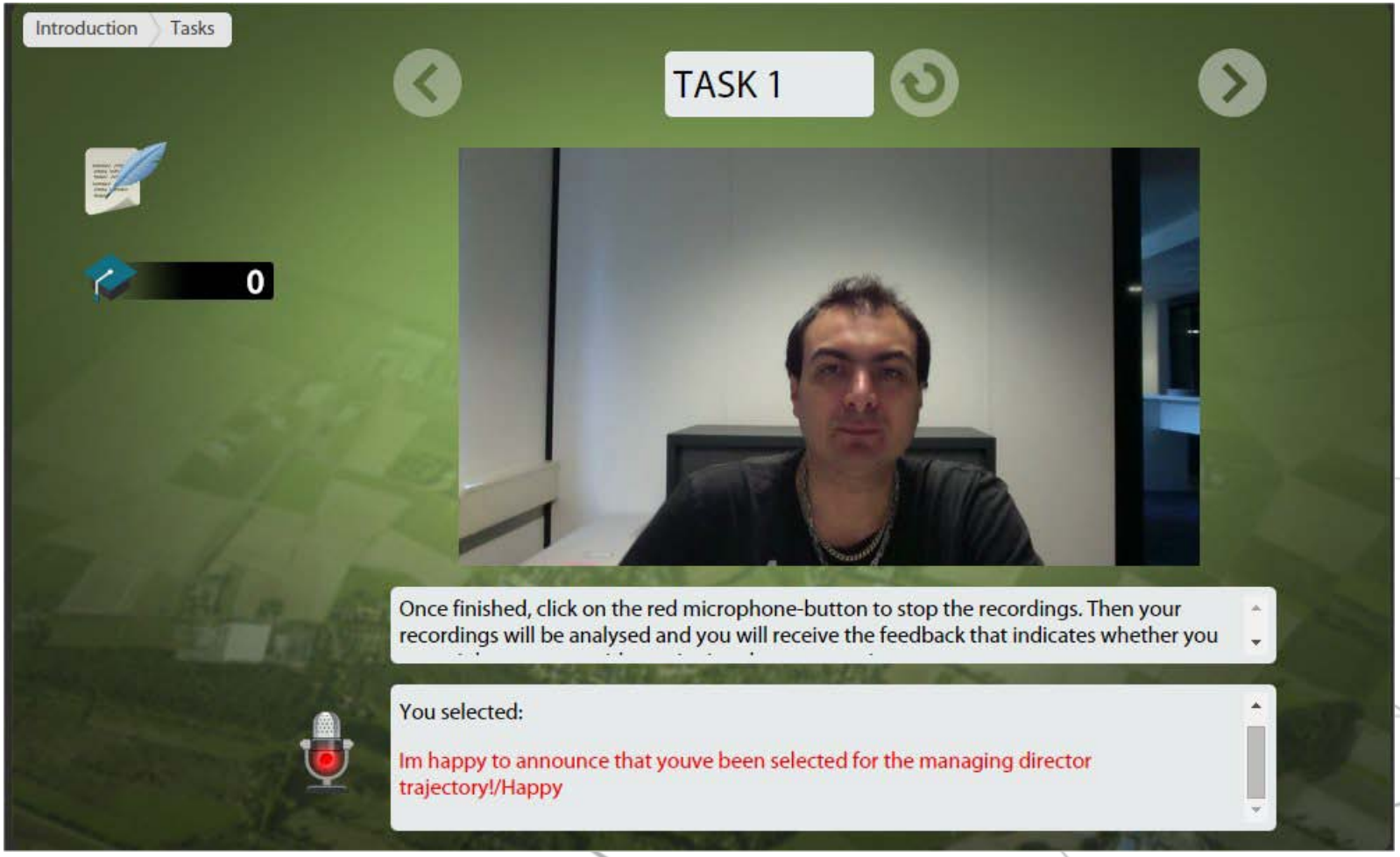
6. Prototype Communication Advisor - 1



6. Prototype Communication Advisor - 2



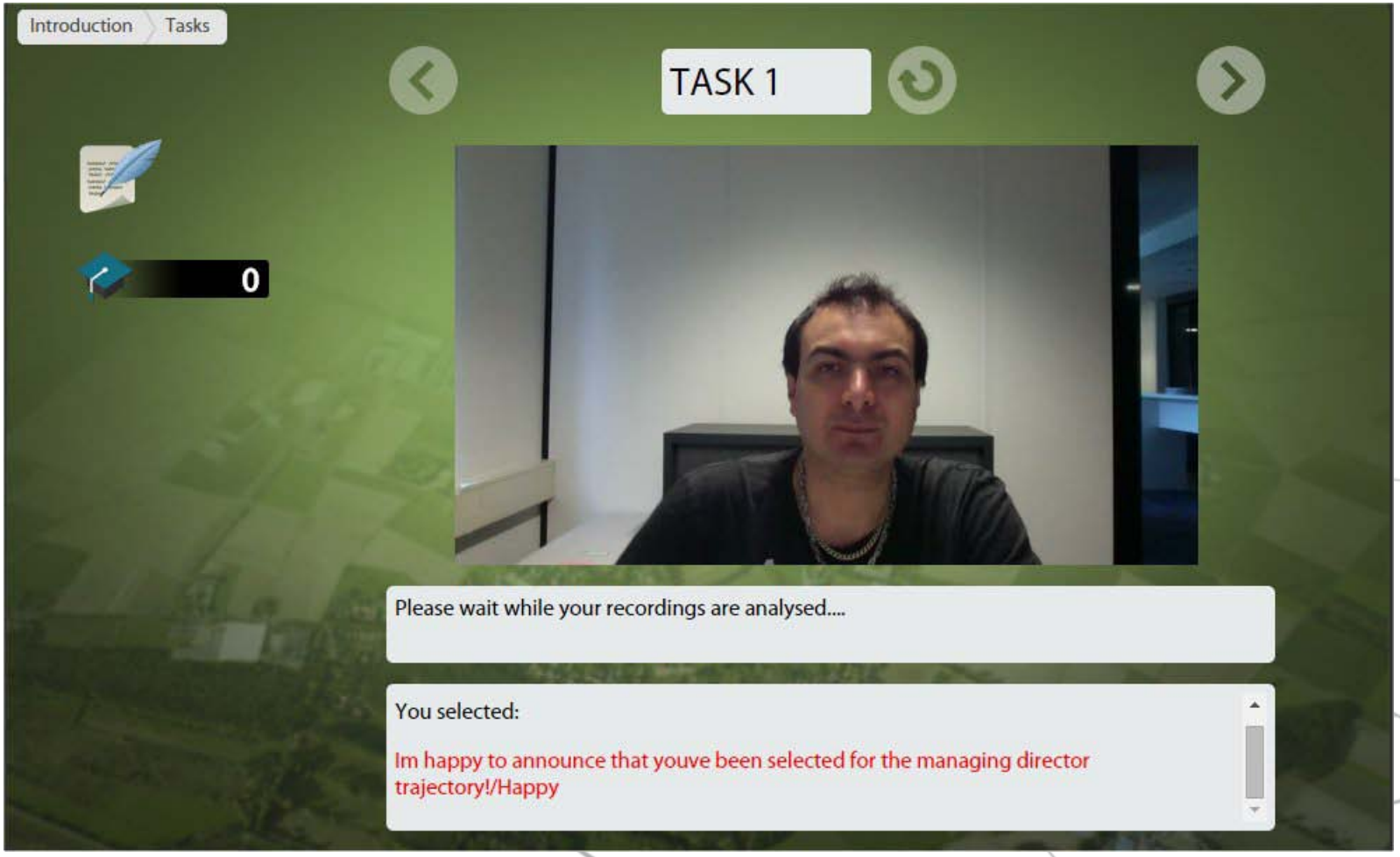
6. Prototype Communication Advisor - 3



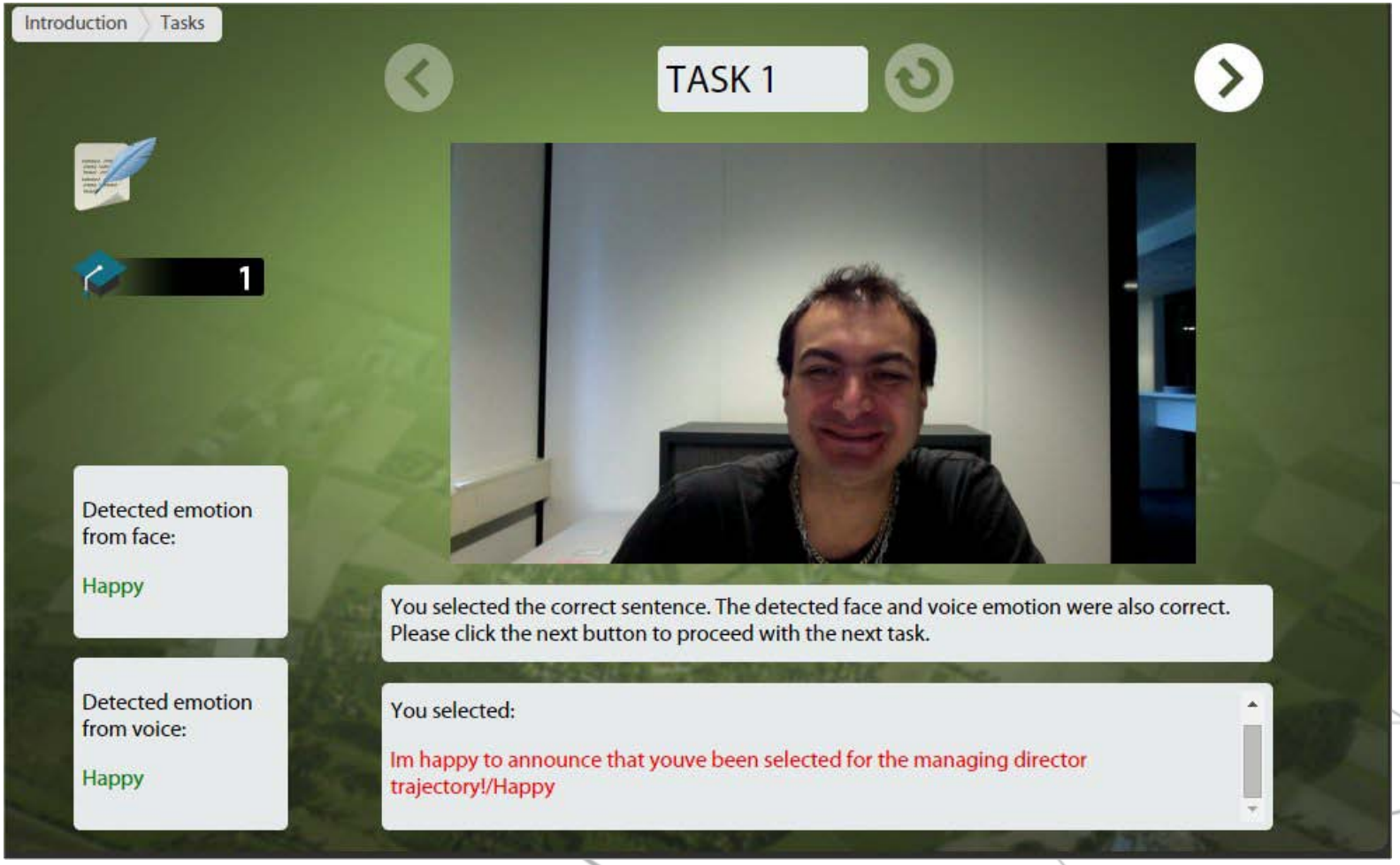
Welten Institute
Research Centre for Learning, Teaching and Technology

Open Universiteit
welten-institute.org

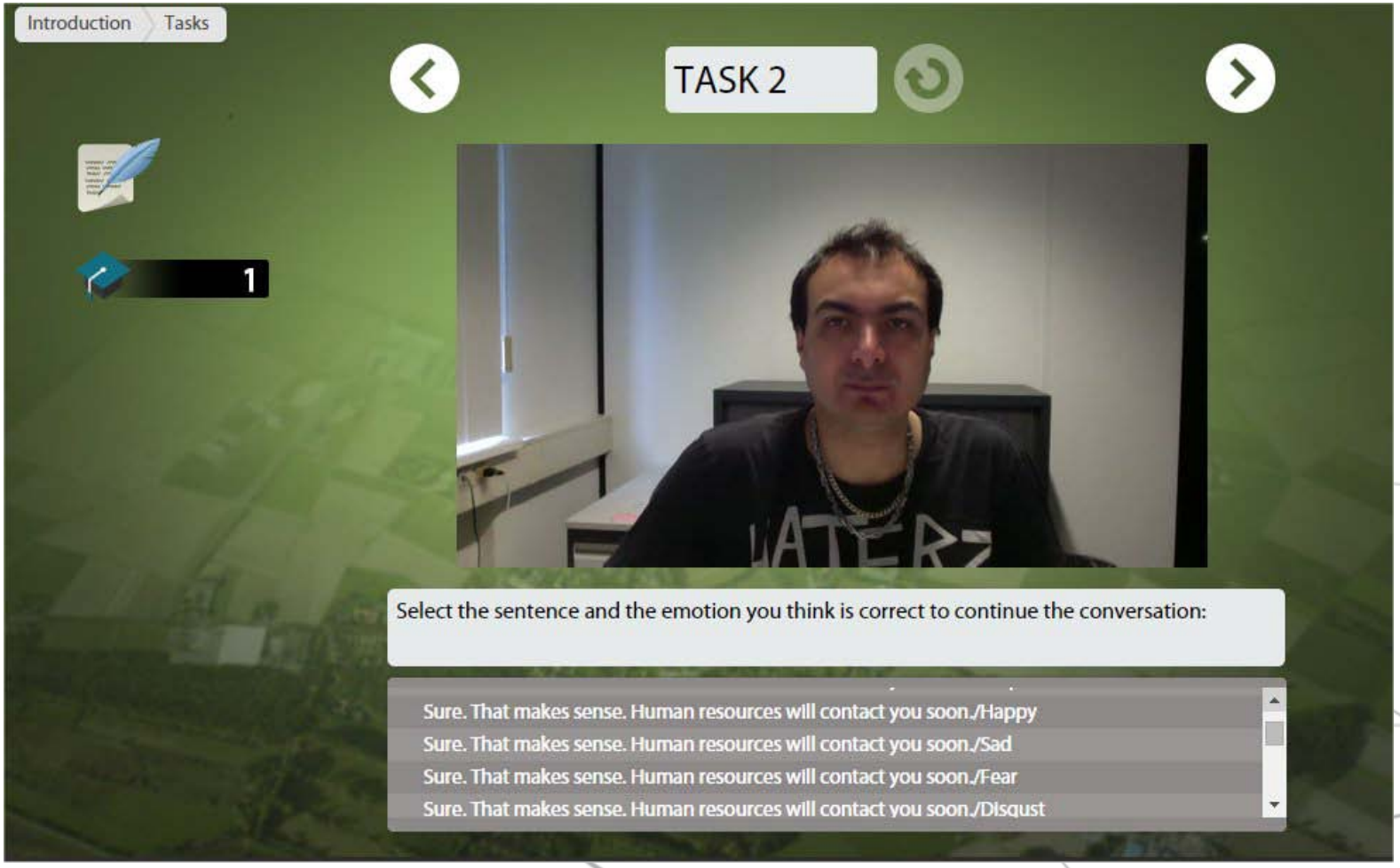
6. Prototype Communication Advisor - 4



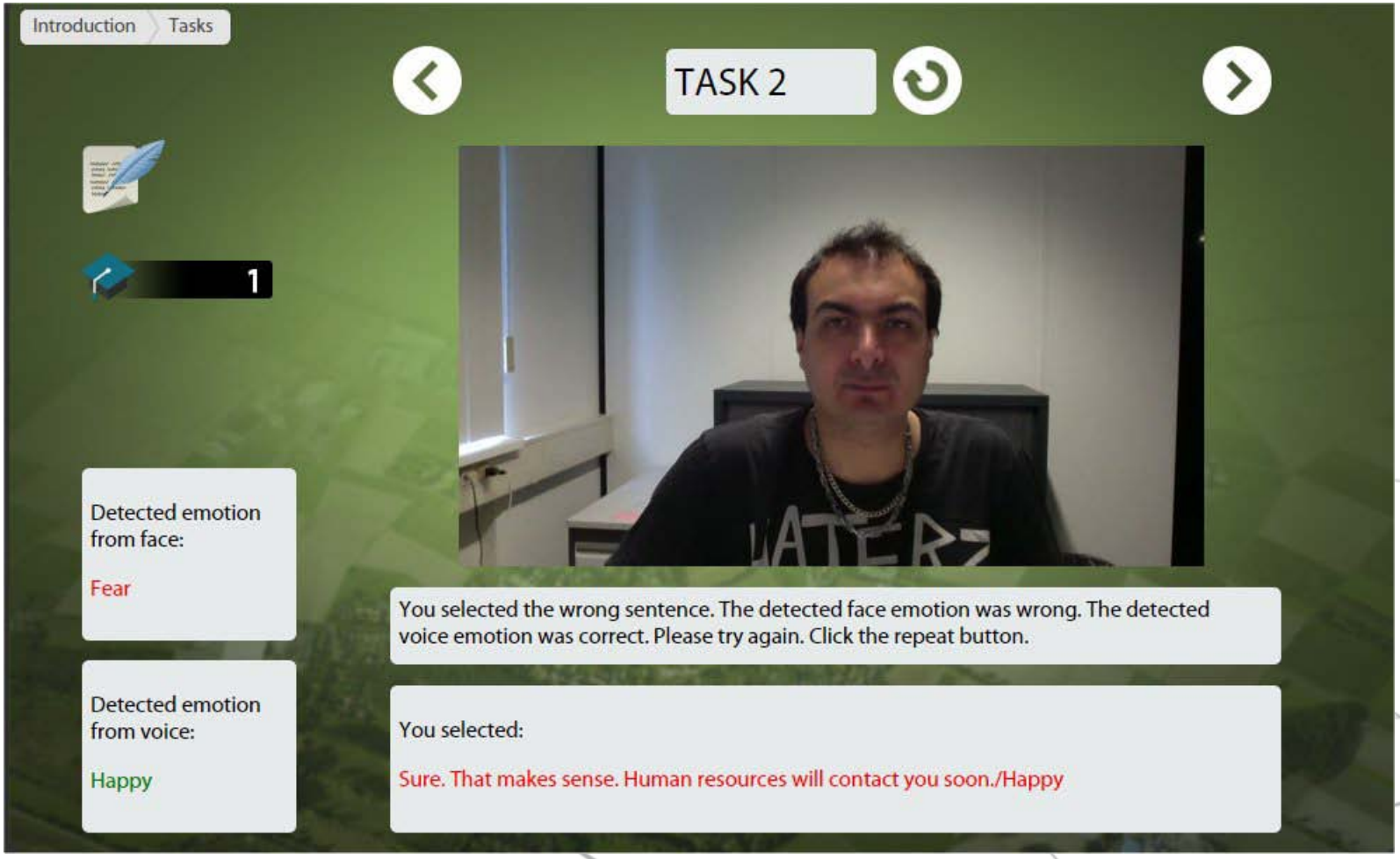
6. Prototype Communication Advisor - 5



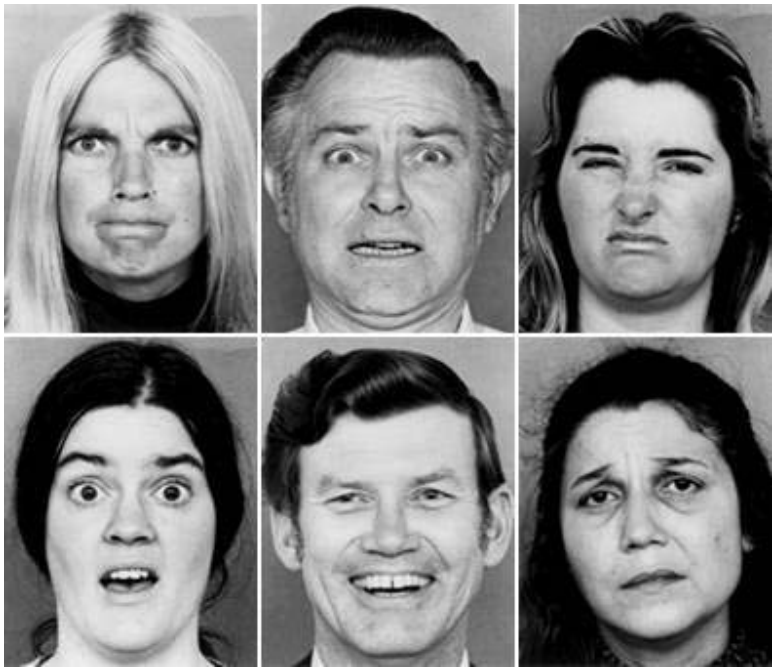
6. Prototype Communication Advisor - 6



6. Prototype Communication Advisor - 7



7. Questions-Discussion



Thank you

Rob Nadolski rob.nadolski@ou.nl

on behalf of

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