# SEWA

## “Automatic Sentiment Analysis in the Wild”

## Innovation Action

## Horizon2020

## Grant Agreement no. 645094

# Deliverable D9.2

# Annual Report 2

| | |
|---|---|
| Deliverable Type: | R (Report) |
| Dissemination level: | CO (Consortium) |
| Month: | M24 |
| Contractual delivery date: | Jan 31, 2017 |
| Actual delivery date: | Feb 13, 2017 |
| Version: | 1.0 |
| Total number of pages: | |

## Document Information

| | | | |
|---|---|---|---|
| **Grant Agreement no.** | 645094 | **Acronym** | SEWA |
| **Full Title** | Automatic Sentiment Analysis in the Wild | | |
| **Project URL** | http://www.sewaproject.eu/ | | |
| **Document URL** | http://www.sewaproject.eu/deliverables/ | | |
| **EU Project Officer** | Philippe Gelin | | |

| | | | | |
|---|---|---|---|---|
| **Deliverable** | **Number** | D9.2 | **Title** | Annual Report 2 |
| **Work Package** | **Number** | WP9 | **Title** | Project Coordination and Management |

| | | | | |
|---|---|---|---|---|
| **Authors (Partner)** | Dionysia Kordopati (ICL), Maja Pantic (ICL), all SEWA partners | | | |
| **Responsible Author** | **Name** | Maja Pantic | **E-mail** | m.pantic@imperial.ac.uk |
| | **Partner** | Imperial College London | **Phone** | +44 207 594 8300 |

| | | | |
|---|---|---|---|
| **Version Log** | | | |
| **Issue Date** | **Rev. No.** | **Author** | **Change** |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

This project has received funding from the European Union's Horizon2020 Programme under grant agreement no 645094.

## Table of Contents

# PART A

## 1. Summary for publication

### 1.1. Summary of the context and overall objectives of the project

The overall aim of the SEWA project is to enable computational models for machine analysis of facial, vocal, and verbal behaviour in the wild. This is to be achieved by capitalising on the state-of-the-art methodologies, adjusting them, and combining them to be applicable to naturalistic human-centric human-computer interaction (HCI) and computer-mediated face-to-face interaction (FF-HCI). The target technology uses data recorded by a device as cheap as a web-cam and in almost arbitrary recording conditions including semi-dark, dark and noisy rooms with dynamic change of room impulse response and distance to sensors. It represents a set of audio and visual spatiotemporal methods for automatic analysis of human spontaneous (as opposed to posed and exaggerated) patterns of behavioural cues including analysis of rapport, mimicry, and sentiment such as liking and disliking.

In summary, the objectives of the SEWA project are:

(1) Development of technology comprising a set of models and algorithms for machine analysis of facial, vocal and verbal behaviour in the wild,

(2) Collection of the SEWA database being a publicly available benchmark multilingual dataset of annotated facial, vocal and verbal behaviour recordings made in-the-wild representing a benchmark for efforts in automatic analysis of audio-visual behaviour in the wild,

(3) Deployment of the SEWA results in both mass-market analysis tools based on automatic behaviour-based sentiment analysis of users towards marketed products and a sentiment-driven recommendation engine, and

(4) Deployment of the SEWA results in a novel social-network-based FF-HCI application – sentiment-driven Chat Social Game.

The SEWA project is expected to have many benefits. Technologies that can robustly and accurately analyse human facial, vocal and verbal behaviour and interactions in the wild, as observed by webcams in digital devices, would have profound impact on both basic sciences and the industrial sector. They could open up tremendous potential to measure behaviour indicators that heretofore resisted measurement because they were too subtle or fleeting to be measured by the human eye and ear. They would effectively lead to development of the next generation of efficient, seamless and user-centric human-computer interaction (affective multimodal interfaces, interactive multi-party games, and online services). They would have profound impact on business (automatic market research analysis would become possible, recruitment would become green as travels would be reduced drastically), and they could enable next generation healthcare technologies (remote monitoring of conditions like pain, anxiety and depression), to mention but a few examples.

## 1.2. Work performed from the beginning of the project to the end of the period covered by the report and main results achieved so far

### *1.2.1. WP1 – SEWA DB collection, annotation and release*

- Obtained ethical approval for the SEWA experiment.
- Designed the SEWA experiment protocol and implemented the data collection website.
- Conducted 204 successful data recording sessions using the aforementioned website. A total of 408 participants from 6 different cultural backgrounds (British, German, Hungarian, Serbian, Greek and Chinese) were recorded, resulting in more than 44 hours of audio-visual corpus covering a wide range of spontaneous expressions of emotions and sentiment during both video-watching and computer-mediated face-to-face communication sessions.
- Extracted the low-level acoustic features (ComParE and GeMAPSv01a) from all SEWA recordings.
- Automatically tracked the 49 facial landmarks in all SEWA recordings. These results were further refined through semi-automatic correction.
- Identified a total of 538 (~90 from each culture group) representative segments (high/low arousal, high/low valence, and liking/disliking) from the SEWA corpus. These segments were annotated fully in terms of facial landmarks, vocal and verbal cues, facial action units (FAUs), continuously valued emotion dimensions (valence and arousal), mimicry, sentiment, rapport, and template behaviours to form the core SEWA dataset.
- Released the SEWA database version 1.0 publicly as according to the data management plan.

### *1.2.2. WP2 - Low-level Feature Extraction*

- Implementation, evaluation, and publication of the toolkit *openXBOW* (formerly called *openWord*) to generate Bag-of-Words (BoW) representations from text, acoustic (and potentially visual) low-level descriptors (LLDs) in order to obtain a compact and robust representation of features.
- Implemented the incremental in-the-wild face alignment method for automatic facial landmark localisation. The tracker is capable of accurately tracking the 49 facial landmarks

in real-time and is robust against illumination change, partial occlusion and head movements.

- Feature enhancement by deep neural networks (LSTM) to improve acoustic features computed from noisy speech signals.
- Cross-corpus/cross-lingual emotion analysis, i.e., testing models for emotion analysis on languages which are not included in the training data.
- Generation of multi-lingual dictionaries for BoW representations with multi-databases in different languages.
- Application of the state-of-the-art of linguistic features employed in text retrieval to the sentiment analysis task.
- Investigation of acoustic landmarks as robust linguistic features for emotion recognition.
- Implementation of the incremental in-the-wild face alignment method for automatic facial landmark localisation.

### *1.2.3 WP3 – Mid-level feature extraction*

- Implementation and evaluation of robust mid-level feature extractor for facial action units.
- Development of Hidden Markov Model (HMM)-based method for head nod / shake detection.
- Implementation and evaluation of robust mid-level feature extractor for head and hand gestures and trained on the basic SEWA dataset.
- Annotated 100 sequences from the SEWA data for AU detection. Sequences were used for training and evaluation of the mid-level feature extractor.
- Implementation of the feature extractor for head nod/shake and hand gestures into standalone module and integrated into the SEWA back-end emotion recognition server using the HCI2
- Implementation of the Dynamic Ordinal Regression Toolbox for AU mid-level feature extraction.

### *1.2.4 WP4 – Continuous Affect and Sentiment Sensing in the Wild*

- Implementation, evaluation of the end-to-end learning based multimodality automatic affect predictor, by using audio and video recordings in-the-wild directly and avoid using tradition feature extraction.

- Implementation, evaluation of the bag-of-audio/video word method based multimodality affect predictor, by generating bag-of-words representations from audio-visual data recorded in the wild and then feeding these representations into regressors for continuous affect recognition.

### *1.2.5 WP5 – Behaviour Similarity in the Wild*

- Development, implementation, and evaluation of a novel methodology (i.e., the Temporal Archetypal Analysis) for unsupervised temporal segmentation of behaviour based on multimodal data.
- Development, implementation, and evaluation of an unsupervised representation learning method for extracting temporal and view/modality invariant data representatives which will be used in construction of behaviour templates.
- Development of a novel framework novel framework for dynamic behaviour modelling, analysis, and prediction. The framework resorts to a set of novel algorithms for learning dynamical system under real-world conditions, namely in the presence of noisy behavioural cues descriptors and possibly unreliable annotations.

### *1.2.6 WP6 – Temporal Behaviour-Patterning and Interpersonal Sentiment in the Wild*

- Selection of patterns for mimicry and rapport from the SEWA database.
- Initial experiments on inter-cultural prediction of behaviour based on state-of-the-art methods
- Data collected during experiments have been annotated for mimicry episodes.
- An audiovisual fusion method based on cross-prediction of each modality has been modified in order to be applied for mimicry detection.
- Preliminary experiments have been conducted on the MAHNOB mimicry database using the above method.

### *1.2.7 WP7 – Integration, Applications and Evaluation*

- Further refinement of the integrated application design, mapping between the capabilities of the SEWA emotion analysis and the interview skills trainer game, as outlined in D7.3.
- Completion of the development of the initial version of interview skills trainer game.
- Extensive user-testing of the initial version of the game with the target audience.

- Exploration of commercial opportunities through multiple meetings with a variety of recruitment technology providers.

### *1.2.8 WP8 – Dissemination, Ethics, Communication and Exploitation*

For detailed list of activities please see section 6 and Part B, section 1.2.8, of this report.

- Dissemination: 7 journal papers, 20 conference papers, 7 invited talks
- Organisation of the following challenges and workshops:
  - ComParE (Computational Paralinguistics challengE) at INTERSPEECH Conference 2016, San Francisco, CA, USA
  - AV+EC challenge at ACM MM 2016, Amsterdam, The Netherlands
  - CBAR 2016 workshop (Context-based Affect Recognition workshop) at IEEE CVPR 2016, Las Vegas, Nevada
- Communication:

  Efforts towards both General Public Dissemination and Industrial Dissemination have been intensified resulting in multiple TV coverage of the work done in SEWA, public speeches on the results of the SEWA project, and press coverage of efforts done in SEWA. RealEyes has won the Innovation Radar Prize 2016 and has engaged in a series of public talks promoting the SEWA results.
- Ethics:

  SEWA consortium had arranged for an **Ethical Advisory Board**, which consists of experts on various fields of ethics that concern the SEWA project.

  The members of the Ethical Advisory Board are Prof. Laurence Devillers of the Paris-Sorbonne IV University in France and Prof. Jean-Gabriel Ganascia of the University Pierre et Marie Curie in France. The Ethical Advisory Board meets at most once a year with the PMC. The first meeting was held in conjunction with the SEWA kick-off meeting on 12-13 February 2015, in London, UK. The recommendations made by the Ethical Advisory Board have been discussed by the PMC, adopted by the project, and are forwarded to the Commission as part of deliverable D8.2. The Ethical Advisory Board will be consulted in all ethical issues as they arise in the course of the work in the various research lines.

### *1.2.9 WP9 – Project co-ordination and management*

- Overall strategic and operational management and steering of the project, ensuring the accuracy, quality and timeliness of deliverables. Conduction of the financial and administrative management of the project. Management of liaison with the European Commission; management of public face of the project and networking with other related projects. Co-ordination of coherence of all developments between Work Packages.

- List of managed and submitted deliverables during this period:

D2.2, Robust Visual Feature Extractor

D2.3, Improved acoustic-linguistic feature extractor

D3.1, Component / Demonstrator for mid-level visual features extraction

D1.1, SEWA Database

D7.2, Initial version of the Ad Recommendation Engine

D7.3, Initial Version of SEWA Chat Game

D3.2, Audio-visual detector of nonverbal vocalisations

D4.1, Multi-modal affect recognizer

D5.1, Visual behaviour similarity estimator

D9.2, Annual Report 2

- List of project meetings during this period:

Phone meetings: 08/02/2016, 07/03/2016, 18/04/2016, 21/06/2016, 12/07/2016, 02/09/2016, 02/11/2016, 28/11/2016, 09/01/2016, 27/01/2016

Plenary meeting – ICL, London, UK : 17/02/2016 & 22/09/2016

Review meeting – Luxembourg: 15-17/05/2016

Valorisation meeting – London, UK : 23/09/2016

## 1.3 Progress beyond the state of the art and expected potential impact (including the socio-economic impact and the wider societal implications of the project so far)

### *1.3.1 WP1 – SEWA DB collection, annotation and release*

- We released the SEWA database (SEWA DB), a multilingual dataset of annotated facial, vocal and verbal behaviour recordings made in-the-wild. This database will be not only an extremely valuable resource for researchers both in Europe and internationally but it will also push forward the research in automatic human behavioural analysis and user-centric HCI and FF-HCI in a similar manner as PASCAL pushed forward the field of object detection. SEWA DB will be used for a number of challenges and benchmarking efforts and will have more than 200 active users worldwide by the end of the project. The SEWA DB can be accessed online at http://db.sewaproject.eu/.

### *1.3.2 WP2 – Low-level Feature Extraction*

- It has been shown that the Bag-of-Audio-Words (BoAW) approach is able to predict emotions in terms of arousal and valence in a better way than all other known approaches and published results (proven on the RECOLA database used ,e.g., in the AV+EC challenges 2015 & 2016).
- The deteriorating effect of noise on acoustic features is overcome using de-noising auto encoders (feature enhancement).
- Development of a hybrid system combining BoAW (acoustic features) and BoW (Bag-of-Words, linguistic features) and also BoVW (Bag-of-Visual-Words) with different feature fusing schemes. The toolbox *openXBOW* has been released and has already been used on different tasks. It has a high chance of becoming a default tool in the research community.
- Implemented the incremental in-the-wild face alignment method for automatic facial landmark localisation. The tracker is capable of accurately tracking the 49 facial landmarks in real-time and is robust against illumination change, partial occlusion and head movements.

### *1.3.3 WP3 – Mid-level feature extraction*

- The Variable-State Latent Conditional Random Field (VSL-CRF) model was developed for the task of AU detection in-the-wild. This method achieves better generalization performance compared to traditional CRFs and other related state-of-the-art models.
- The model has been released on GitHub and on the iBug website. It is currently used for different computer vision and machine learning tasks.

### *1.3.4 WP4 – Continuous Affect and Sentiment Sensing in the Wild*

- Realisation of a fully automatic continuously-valued sentiment and affect dimensions predictor from audio-visual data recorded in the wild, which gets performance competitive or better than other state-of-the-art approaches.

### *1.3.5 WP5 – Behaviour Similarity in the Wild*

- Extensive experimental results on three publicly available datasets demonstrate that the developed unsupervised video segmentation method (i.e., the Temporal Archetypal

Analysis) outperforms the compared state of the art methods in temporal human behaviour segmentation.

- Experimental evidence indicates that the developed method for learning linear dynamical systems in the presence of gross noise is more robust compared to the state of the art.
- The developed dynamic behaviour analysis framework has been applied to vision-based conflict intensity prediction, valence and arousal prediction, and tracklet matching. Extensive experiments on real-world data drawn from these application domains demonstrate the robustness and the effectiveness of the proposed framework.

### *1.3.6 WP6 – Temporal Behaviour-Patterning and Interpersonal Sentiment in the Wild*

- We have developed one of the first approaches that attempts continuous detection of behavioural mimicry, using data of natural interaction between minimally constrained subjects.
- We have modified a prediction-based approach for audiovisual fusion in order to detect mimicry without explicit knowledge of what action has been mimicked.

### *1.3.7 WP7 – Integration, Applications and Evaluation*

Further refining the integrated application as an interview skills training game called #interviewSkillz. In the game, one player takes on the role of an interviewer and the other the role of a candidate. The purpose is to provide job seekers with a platform to develop the skills necessary to be successful in the emotional and social aspects of the interview process.

The potential socio-economic impact of the application addresses the largely ignored skill-gap of young people attending their first job interview, which in Europe constitutes 10 to 16 million young people each year.  The end-users expected to benefit includes; young people studying in education institutions or undertaking work-based training, education institutions offering communication skills development courses and services, apprenticeship training providers, employers running graduate training schemes, employment agencies offering value-added recruitment services to recruiters and jobseekers and welfare to work providers offering training and support to jobseekers.

With respect to progress beyond the state of art, the job interview skills trainer is the first of its kind, offering trainees automated interview training feedback using advanced emotional analytics. The trials using the initial version of the application were received very well by the end-user groups.

### *1.3.8 WP8 – Dissemination, Ethics, Communication and Exploitation*

SEWA partners have increased the interest of general public and the industry in the field of automatic emotion recognition and the SEWA applications. There are several evidences for this: a major article on SEWA in German national press (Passauer Neue Presse), three TV reports on the work done n SEWA filmed at Imperial College in the lab of the SEWA coordinator, SEWA coordinator’s TEDx talk at EC Digital Assembly in September 2016, invitation to the SEWA coordinator to speak on SEWA and related technologies at UNI Global Union Summit in mid-November 2016 and at Global Innovation Summit in late November 2016, RealEyes winning the Innovation Radar Prize 2016 and speaking about SEWA results at public events (see section 6 and Part B, section 1.2.8, for details).

The awareness of the scientific community about the importance of research focus on automatic analysis of human behaviour observed in the wild and automatic audio-visual sentiment analysis has been raised by means of both the International Workshop on Context-based Affect Recognition (CBAR’16), which has been organised by and sponsored by SEWA, and a large number of Keynotes given by the SEWA partners at which SEWA project and its aims have been explained (for the complete list, see section Part B of this reposrt, section 1.2.8).

Further impact in the scientific community is ensured by the organisation of AV+EC, and ComParE challenges with a considerable number of participants each, and by the high number of peer-reviewed publications.

## 2. Deliverables

| Deliverable Number | Deliverable name | Work package number | Lead beneficiary | Type | Dissemination level | Delivery date from Annex 1 | Actual delivery date | If deliverable not submitted on time: forecast delivery date | Status |
|---|---|---|---|---|---|---|---|---|---|
| D2.2 | Robust Visual Feature Extractor | 2 | ICL | DEM | CO | M13 (Feb 16) | 29/02/2016 & revised 09/08/2016 | | complete |
| D2.3 | Improved acoustic-linguistic feature extractor | 2 | UP | DEM | CO | M13 (Feb 16) | 29/02/2016 | | Complete |
| D3.1 | Component / Demonstrator for mid-level visual features extraction | 3 | ICL | DEM | CO | M15 (Apr 16) | 06/05/2016 | | Complete |
| D1.1 | SEWA Database | 1 | ICL | Dataset | CO | M18 (July 16) | 17/08/2016 | | Complete |
| D7.2 | Initial version of the Ad Recommendation Engine | 7 | RealEyes | DEM | CO | M18 (July 16) | 24/08/2016 | | Complete |
| D7.3 | Initial Version of SEWA Chat Game | 7 | PlayGen | other | CO | M18 (July 16) | 09/08/2016 | | complete |
| D3.2 | Audio-visual detector of nonverbal vocalisations | 3 | ICL | DEM | CO/PU | M24 (Jan 17) | 15/02/2017 | | |
| D4.1 | Multi-modal affect recognizer | 4 | PU | DEM | CO/PU | M24 (Jan 17) | 06/02/2017 | | |
| D5.1 | Visual behaviour similarity estimator | 4 | ICL | DEM | CO/PU | M24 (Jan 17) | 10/02/2017 | | |
| D9.2 | Annual Report 2 | 9 | ICL | R | PU | M24 (Jan 17) | 13/02/2017 | | |

## 3. Milestones

| Milestone No | Milestone title | Related WP(s) no. | Lead beneficiary | Delivery date from Annex 1 | Means of verification | Achieved | If not achieved forecast achievement date | comments |
|---|---|---|---|---|---|---|---|---|
| M2 | SEWA Applications V1 | WP2 – WP7 |  | M18 | 1st version of the SEWA applications implemented and tested | Yes |  |  |

## 4. Ethical Issues

The SEWA consortium had arranged for an Ethical Advisory Board, which consists of experts on various fields of ethics that concern the SEWA project. The members of the Ethical Advisory Board are Prof. Laurence Devillers of the Paris-Sorbonne IV University in France and Prof. Jean-Gabriel Ganascia of the University Pierre et Marie Curie in France.

The Ethical Advisory Board meets at most once a year with the PMC. The first meeting was held in conjunction with the SEWA kick-off meeting on 12-13 February 2015, in London, UK. The recommendations made by the Ethical Advisory Board have been discussed by the PMC, adopted by the project, and are forwarded to the Commission as part of deliverable D8.2. The Ethical Advisory Board is consulted in all ethical issues as they arise in the course of the work in the various research lines.

# 5. Critical implementation risks and mitigation actions

## 5.1 Foreseen Risks

| Risk Number | Description of Risk | Work Packages | Proposed risk-mitigation measures |
|---|---|---|---|
| R1 | Integration of WP1 with WP3 | WP1, WP3 | See description in WP1 |
| R2 | Audio feature extraction is too slow | WP2 | See description in WP2 |
| R3 | Linguistic feature extraction performs poorly in adverse acoustic environments | WP2 | See description in WP2 |
| R4 | Optimisation of dynamic texture descriptors fails | WP3 | See description in WP3 |
| R5 | Underperformance of affect recognition | WP4 | See description in WP4 |
| R6 | Integration of WP1 with WP4 | WP1, WP4 | See description in WP4 |
| R7 | Underperformance of behaviour similarity matching | WP5 | See description in WP5 |
| R8 | Usability of the tools developed in WP2-WP5 for deployment in the SEWA applications | WP7 | See description in WP7 |
| R9 | Reliability of the hardware/software ecosystem of the SEWA applications | WP7 | See description in WP7 |
| R10 | Users‘ acceptance of audio-visual measurements | WP7 | See description in WP7 |
| R11 | Delay in SEWA technology and / or applications development | WP7, WP8 | See description in WP8 |
| R12 | Competing technology emerges from another academic or industrial institution | WP8 | See description in WP7 |
| R13 | Loss of key technology / application partner | WP8, WP9 | See description in WP8 |

## 5.2 Unforessen Risks

| Risk Number | Description of Risk | Work Packages Concerned | Proposed risk-mitigation measures |
|---|---|---|---|
| 1 | | | |

## 5.3 State of the Play for Risk Mitigation

| Risk number | Period number | Did you apply risk mitigation measures – YES / NO | Did your risk materialise YES / NO | Comments |
|---|---|---|---|---|
| **R1** | **M1-12** | **YES** | **YES** | As described in Deliverable D9.1. Annual Report 1 |
| **R2** | **M1-9** | **YES** | **NO** | The extraction of low and mid-level acoustic / visual features is running in real-time. |
| **R3** | **M1-13** | **YES** | **Partially** | **Task 2.1** Our efforts to create noise robust feature representations for real-life emotion recognition from speech, succeeded.<br>**Task 2.3** ASR has problems with strong dialect, noise and cross-talk, which is found throughout the SEWA database. The influence on the final performance of emotion recognition is low, as linguistic information is only used to augment the |

| | | | | |
|---|---|---|---|---|
| | | | | acoustic/visual information. We are further investigating acoustic landmarks to get linguistic-like features as they are more robust. |
| **R4** | **M3-15** | **YES** | **YES** | Task 3.2 The work in T3.2 is currently based on facial landmark location and their trajectories in time. Appearance-based features are not used. Our implementation of the landmark tracker is highly efficient, being able to track 8 video streams in parallel at 50fps on our data processing machine (CPU: Intel Core i7-5960X, memory: 32GB). The tracker is highly accurate and produces accurate results for AU detection (T3.2), making the use of costly appearance features unnecessary. |
| **R5** | **M15-27** | **YES** | **NO** | Affect recognition yields (almost) state-of-the-art results on the SEWA database (in the wild). |
| **R6** | **M15-27** | **YES** | **NO** | The core SEWA dataset (540 representative short segments) has now been annotated. This data-set is now beeing used to develop and test the methods in WP4. |
| **R7** | **M12-30** | **NO** | **NO** | The work on behaviour similarity matching is ongoing and the first results are promising (see Part B, section 1.2.5). However, as explained in the project description, the completion of the project does not depend on the development of accurate and fast behaviour similarity matching. |
| **R8** | **M12-42** | **NO** | **NO** | So far, tools from WP2-WP4 are such that real-time performance of the SEWA applications is feasible. |
| **R9** | **M12-42** | **NO** | **NO** | Actual user testing of the integrated applications has not started yet. |
| **R10** | **M12-42** | **YES** | **Potentially** | User's stated concern with their video/audio being seen by others. Possible solutions include replacement of face with simple avatar, or closed systems. |
| **R11** | **M12-42** | **NO** | **NO** | So far, all developments have been completed in a timely fashion. |
| **R12** | **M1-42** | **NO** | **NO** | No competing technology has emerged from another academic or industrial institution. |
| **R13** | **M1-42** | **NO** | **NO** | No loss of key technology / application partner has occured. |

## 6. Dissemination and exploitation of results

### 6.1. Scientific publications

| **Type of scientific publication** | **Title of the scientific publication** | **DOI** | **ISSN or eSSN** | **Authors** | **Title of the journal or equivalent** | **Number, date** | **Publisher** | **Place of publication** | **Year of publication** | **Relevant pages** | **Public & private publication** [4] | **Peer-review** | **Is/Will open access provided to this publication** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *[Article in journal]*<br>*[*Publication in conference proceeding/ workshop*]*<br>*[Books/Monographs]*<br>*[*Chapters in books*]*<br>*[Thesis/dissertation]* | [insert title of the publication] | [insert DOI reference] | [insert ISSN or eSSN number] | [insert authors' name(s)] | [insert title of the journal] | [insert number of the journal]<br>[insert month of the publication]<br>[insert year of the publication] | [insert name of the publisher] | [insert place of publication] | [insert year of the publication] | [insert first page of the publication] - [insert last page of the publication] | *[YES]*<br>*[NO]* | *[YES]*<br>*[NO]* | *[*Yes - Green OA [insert the length of embargo if any]*]*<br>*[*Yes - Gold OA [insert the amount of processing charges in EUR if any]*]*<br>*[NO]* |
| *Journal article* | Discrimination Between Native and Non-Native Speech Using Visual Features Only | 10.1109/TCYB.2015.2488592 | 2168-2275 | C. Georgakis, S. Petridis, M. Pantic | IEEE Transactions on Cybernetics | December 2016 | IEEE | | 2016 | pp. 2758 - 2771 | | | *YES-Green OA* |
| *Publication in conference proceedings* | Multi-Modal Neural Conditional Ordinal Random Fields for Agreement Level Estimation | 10.1109/ACII.2015.7344679 | 2156-8111 | N. Rakicevic, O. Rudovic, S. Petridis, M. Pantic | Proceedings of the Int'l Conference on Pattern Recognition | December 2016 | IEEE | Cancun, Mexico | 2016 | N/A | | | *YES – Green OA* |
| *Journal article* | Joint Facial Action Unit Detection and Feature Fusion: A Multi-conditional Learning Approach | 10.1109/TIP.2016.2615288 | Online ISSN:1941-0042 | S. Eleftheriadis, O. Rudovic, M. Pantic | IEEE Transactions on Image Processing. | December 2016 | IEEE | | 2016 | pp. 5727 - 5742 | | | *YES-Green OA* |
| *Publication in workshop proceedings* | TensorLy: Tensor Learning in Python | N/A | N/A | J. Kossaifi, Y. Panagakis, M. Pantic | NIPS Tensor-Learn Workshop | December 2016 | | Barcelona, Spain | 2016 | N/A | | | |
| *Publication in conference proceedings* | Variational Gaussian Process Auto-Encoder for Ordinal Prediction of Facial Action Units | N/A | 0302-9743 | S. Eleftheriadis, O. Rudovic, M.P. Deisenroth, M. Pantic | Asian Conference on Computer Vision | November 2016 | Springer | Taipei, Taiwan | 2016 | N/A | | | *YES-Green OA* |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Journal article* | Robust Correlated and Individual Component Analysis | DOI: 10.1109/ TPAMI.2015.2497700 | ISSN 0162-8828 | Y. Panagakis, M. A. Nicolaou, S. Zafeiriou and M. Pantic | IEEE Transactions on Pattern Analysis and Machine Intelligence | November 2015 | IEEE | Piscataway, USA | 2016 | pp. 1665-1678 | | *YES* | |
| *Publication in conference proceedings* | The University of Passau Open Emotion Recognition System for the Multimodal Emotion Challenge | DOI: 10.1007/ 978-981-10-3005-5_54 | ISSN 1865-0929 | Jun Deng, Nicholas Cummins, Jing Han, Xinzhou Xu, Zhao Ren, Vedhas Pandit, Zixing Zhang, and Bjorn Schuller | Proceedings of the 7th Chinese Conference on Pattern Recognition System for the Multimodal Emotion Challenge | November 2016 | Springer | Chengdu, China | 2016 | pp. 652-666 | | *YES* | *YES – Green OA* |
| *Publication in conference proceedings* | A Bag-of-Audio-Words Approach for Snore Sounds' Excitation Localisation | No DOI | ISSN 0932-6022 | M. Schmitt, C. Janott, V. Pandit, K. Qian, C. Heiser, W. Hemmert and B. Schuller | Proceedings of 14th ITG Conference on Speech Communication | October 2016 | VDE, IEEE | Paderborn, Germany | 2016 | pp. 230-234 | | *YES* | *YES – Green OA* |
| *Publication in conference proceedings* | Towards Cross-lingual Automatic Diagnosis of Autism Spectrum Condition in Children's Voices | No DOI | ISSN 0932-6022 | M. Schmitt, E. Marchi, Fabien Ringeval and B. Schuller | Proceedings of 14th ITG Conference on Speech Communication | October 2016 | VDE, IEEE | Paderborn, Germany | 2016 | pp. 264-268 | | *YES* | *YES – Green OA* |
| *Publication in conference proceedings* | At the Border of Acoustics and Linguistics: Bag-of-Audio-Words for the Recognition of Emotions in Speech | DOI: 10.21437 /Interspeech.2016-1124 | ISSN 1990-9770 | M. Schmitt, F. Ringeval and B. Schuller | Proceedings of 17th Annual Conference of the International Speech Communication Association (INTERSPEECH) | September 2016 | ISCA | San Francisco, USA | 2016 | pp. 495-499 | | *YES* | *YES – Green OA* |
| *Publication in conference proceedings* | Discriminatively trained recurrent neural networks for continuous dimensional emotion recognition from audio | No DOI | No ISSN | F. Weninger, F. Ringeval, E. Marchi, and B. Schuller | In Proceedings of the 25th International Joint Conference on Artificial Intelligence | July 2016 | IJCAI/ AAAI | New York, USA | 2016 | pp. 2196–2202 | | *YES* | *YES – Green OA* |
| *Publication in workshop proceedings* | 7 Essential Principles to Make Multimodal Sentiment Analysis Work in the Wild | No DOI | No ISSN | B. Schuller | In Proceedings of the 4th Workshop on Sentiment Analysis where AI meets Psychology | July 2016 | IJCAI/ AAAI | New York, USA | 2016 | 1 page | | *YES* | *YES – Green OA* |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Publication in conference proceedings* | Copula Ordinal Regression for Joint Estimation of Facial Action Unit Intensity | DOI: 10.1109/CVPR.2016.530 | eISSN 1063-6919 | R. Walecki, O. Rudovic, M. Pantic and V. Pavlovic | Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'16) | June 2016 | IEEE | Las Vegas, USA | 2016 | pp. 4902-4910 | | *YES* | |
| *Publication in conference proceedings* | A Framework for Joint Estimation and Guided Annotation of Facial Action Unit Intensity | DOI: 10.1109/CVPRW.2016.183 | eISSN 2160-7516 | R. Walecki, E. Coutinho, O. Rudovic, M. Pantic, V. Pavlovic | Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPRW'16). 4th Workshop on Context Based Affect Recognition | June 2016 | IEEE | Las Vegas, USA | 2016 | pp. 1460-1468 | | *YES* | |
| *Publication in conference proceedings* | Gaussian Process Domain Experts for Model Adaptation in Facial Behavior Analysis | DOI: 10.1109/CVPRW.2016.184 | eISSN 2160-7516 | S. Eleftheriadis, O. Rudovic, M. P. Deisenroth and M. Pantic | Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPRW'16). 4th Workshop on Context Based Affect Recognition | June 2016 | IEEE | Las Vegas, USA | 2016 | pp. 1469-1477 | | *YES* | *YES – Green OA* |
| *Publication in conference proceedings* | Towards a Common Linked Data Model for Sentiment and Emotion Analysis | No DOI | No ISSN | J. Sanchez-Rada, B. Schuller, V. Patti, P. Buitelaar, G. Vulcu, F. Burkhardt, C. Clavel, M. Petychakis, and C. A. Iglesias | In Proceedings of the 6th International Workshop on Emotion and Sentiment Analysis (ESA 2016) | May 2016 | ELRA | Piscataway, USA | 2016 | pp. 48-54 | | *YES* | *YES – Green OA* |
| *Journal article preprint* | openXBOW – Introducing the Passau Open-Source Crossmodal Bag-of-Words Toolkit | No DOI | No ISSN | M. Schmitt and B. Schuller | arxiv.org | arxiv.org: 1605.06778, May 2016 | arxiv.org | Portoroz, Slovenia | 2016 | 9 pages | | *NO* | *YES – on arxiv.org* |
| *Journal article* | Probabilistic Slow Features for Behavior Analysis | DOI: 10.1109/TNNLS.2015.2435653 | ISSN 2162-237X | L. Zafeiriou, M. A. Nicolaou, S. Zafeiriou, S. Nikitidis, M. Pantic | IEEE Transactions on Neural Networks and Learning Systems | vol. 27, no. 5, May 2016 | IEEE | Piscataway, USA | 2016 | pp. 1034-1048 | | *YES* | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Journal article* | Discriminant Incoherent Component Analysis | DOI: 10.1109/TIP.2016.2539502 | ISSN: 1057-7149 | C. Georgakis, Y. Panagakis and M. Pantic | IEEE Transactions on Image Processing | vol. 25, no. 5, May 2016 | IEEE | Piscataway, USA | 2016 | pp. 2021-2034 | | *YES* | |
| *Publication in workshop proceedings* | Empirical Mode Decomposition: A Data-Enrichment Perspective on Speech Emotion Recognition | No DOI | No ISSN | B. Dong, Z. Zhang, and B. Schuller | In Proceedings of the 6th International Workshop on Emotion and Sentiment Analysis, satellite of the 10th Language Resources and Evaluation Conference (LREC) | May 2016 | ELRA | Portoroz, Slovenia | 2016 | pp. 71-75 | | *YES* | *YES – Green OA* |
| *Publication in workshop proceedings* | Multimodal Sentiment Analysis in the Wild: Ethical considerations on Data Collection, Annotation, and Exploitation | No DOI | No ISSN | B. Schuller, J. Ganascia, and L. Devillers | In Proceedings of the Workshop on ETHics in Corpus Collection, Annotation, and Application, satellite of the 10th Language Resources and Evaluation Conference (LREC) | May 2016 | ELRA | Portoroz, Slovenia | 2016 | pp. 29-34 | | *YES* | *YES – Green OA* |
| *Publication in conference proceedings* | Cross Lingual Speech Emotion Recognition Using Canonical Correlation Analysis on Principal Component Subspace | DOI: 10.1109/ICASSP.2016.7472789 | eISSN 2379-190X | H. Sagha, J. Deng, M. Gavryukova, J. Han, and B. Schuller | In Proceedings 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) | March 2016 | IEEE | Shanghai, China | 2016 | pp. 5800-5804 | | *YES* | *YES – Green OA* |
| *Publication in conference proceedings* | Enhanced Semi-supervised Learning for Multimodal Emotion Recognition | DOI: 10.1109/ICASSP.2016.7472666 | eISSN 2379-190X | Z. Zhang, F. Ringeval, B. Dong, E. Coutinho, E. Marchi, and B. Schuller | In Proceedings 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) | March 2016 | IEEE | Shanghai, China | 2016 | pp. 5185-5189 | | *YES* | *YES – Green OA* |
| *Publication in conference proceedings* | ADIEU Features? End-to-end Speech Emotion Recognition Using a Deep Convolutional Recurrent Network | DOI: 10.1109/ICASSP.2016.7472669 | eISSN 2379-190X | G. Trigeorgis, F. Ringeval, R. Brückner, E. Marchi, M. Nicolaou, B. Schuller, and S. Zafeiriou | In Proceedings 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) | March 2016 | IEEE | Shanghai, China | 2016 | pp. 5200-5204 | | *YES* | *YES – Green OA* |
| *Publication in* | Deep Complementary Bottleneck Features | DOI: 10.1109/ICASSP.2 | eISSN 2379-190X | M. Pantic and S. Petridis | In Proceedings 41st IEEE International Conference on | March 2016 | IEEE | Shanghai, China | 2016 | pp. 2304 - 2308 | | *YES* | *YES – Green OA* |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *conference proceedings* | for Visual Speech Recognition | 016.7472088 | | | Acoustics, Speech, and Signal Processing (ICASSP) | | | | | | | | |
| *Journal article* | Continuous Estimation of Emotions in Speech by Dynamic Cooperative Speaker Models | DOI 10.1109/TAFFC.2016.2531664 | ISSN 1949-3045 | A. Mencattini, E. Martinelli, F. Ringeval, B. Schuller, and C. Di Natale | IEEE Transactions on Affective Computing | February 2016 | IEEE | Piscataway, USA | 2016 | 14 pages | | *YES* | |
| | | | | | | | | | | | | | |

## 6.2. Dissemination and communication activities

| **Type of dissemination and communication activities** | **Number** |
|---|---|
| *Organisation of a Conference* | 0 |
| *Organisation of a workshop* | 2 |
| *Press release* | 2 |
| Non-scientific and non-peer reviewed publications (popularised publications) | 5 |
| *Exhibition* | 4 |
| *Flyers* | 0 |
| *Training* | 0 |
| *Social media* | 3 |
| Web-site | 1 |
| *Communication campaign (e.g radio, TV)* | 4 |
| *Participation to a conference* | 16 |
| *Participation to a workshop* | 5 |
| Participation to an event other than a conference or workshop | 11 |
| *Video/film* | 2 |
| *Brokerage event* | 0 |

| | |
|---|---|
| *Pitch event*<br>*Trade fair*<br>Participation in activities organised jointly with other H2020 project(s)<br>*Other* | 1<br>1<br>5 |
| **Total funding amount** | €44,450.29 |

| **Type of audience reached**<br>**In the context of all dissemination & communication activities**<br>('multiple choices' is possible) | **Estimated Number of persons reached** |
|---|---|
| *[Scientific Community (higher education, Research)] >1500*<br>*[Industry] >5500 (at UNI Global Union Summit, Global Innovation Summit, EU Digital Assembly, and Royal Society and Science Exhibitions)*<br>*[Civil Society] >10,000 (also via TV broadcasts)*<br>*[General Public] >10,000 (also via TV broadcasts)*<br>*[Policy makers] > 1000 (at UNI Global Union Summit, Global Innovation Summit, and EU Digital Assembly)*<br>*[Medias] > 1000 (via websites, Facebook, and popular press coverage excluding TV broadcasts)*<br>*[Investors] > 20 (via Valorisation Board meeting, and Global Innovation Summit)*<br>*[Customers] > 20 (via Valorisation Board meeting, SEWA website, and personal contacts)*<br>*[Other]* | |

## 6.3. Intellectual property rights resulting from the project

PlayGen has not sought to formally protect IP rights, as Patents are not granted in the EU for software, Trademark of experimental software is not sought, the registered design or utility models do not apply.

| **Type of IP Rights** | **Application reference** | **Date of the application** | **Official title of the application** | **Applicant(s)** | **Has the IPR protection been awarded?** | **If available, official publication number of award of protection** |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| *Patent* | *GB 1620476.0* | 02/12/2016 | **COMPUTER-IMPLEMENTED METHOD OF PREDICTING PERFORMANCE DATA** | Realeyes OU | *PENDING* | |

## 6.4 Innovation

Does the project include the following activities and if so how many of each?

| **Activities developed within the project** | **Number** |
|---|---|
| Prototypes | PLAYGEN:  3 Prototypes (1 Sumobate, 2 InterviewSkills games)<br>REALEYES: 1x baseline prediction model for sales lift data<br>1x baseline prediction model for social media data<br>1x 2nd version of the prediction model for sales lift data |
| Testing activities (feasibility/demo) | PLAYGEN: 30 (testing sessions)<br>REALEYES: Detailed cross validation tests for all prediction models<br>Detailed generalisation tests for all prediction models<br>Recommendation feasibility demonstration via ideal sub-panel selection<br>Recommendation feasibility demonstration via similarity search |
| Clinical trials | 0 |

**Will the project lead to launching one of the following into the market (several possible):**

| | |
|---|---|
| New product (good or service) | *YES* |
| New process | *NO* |
| New method | *NO* |

**How many private companies in your project have introduced or are planning to introduce innovations (within the project lifetime or 3 years thereafter):**

Mars and Marketcast have already introduced recommendations based on predictive modelling into their companies in a limited experimental capacity. Over the next 3 years as the service matures we are expecting this introduction to grow significantly.

IPSOS, our industrial partner in SEWA project and world’s leading Market Research Company, is also interested in utilisation of predictive modelling across their products. IPSOS is conducting market research on behalf of many worlds' most well-known brands and over the next 3 years we are expecting to see introduction of Realeyes predictive modelling based recommendations to many companies via this cooperation with IPSOS.

## 7. Open Research Data

| Identifier, DOI (if available) | Title/Identifier (if no DOI available) | Is this dataset Openly accessible[5]? | Is this dataset re-usable[6] | If the dataset is linked to a publication, specify the DOI of the publication |
|---|---|---|---|---|
| [insert DOI reference] | The SEWA Database (accessible at: http://db.sewaproject.eu/) | *Yes* | *Yes* | A journal paper introducing the SEWA database is in preparation. |

## 8. Gender

Gender of researchers and other workforce involved in the project.

| Beneficiaries | Number Women researchers[8] (all levels, incl. postdocs and PhD students) | Number Men researchers[8] (all levels, incl. postdocs and PhD students) | Number Women in the workforce other than researchers | Number Men in the workforce other than researchers |
|---|---|---|---|---|
| ICL | 2 | 11 | 2 | 0 |
| PU | 1 | 3 | 0 | 0 |
| PlayGen | 2 | 5 | 1 | 2 |

| | | | | |
|---|---|---|---|---|
| RealEyes | 0 | 8 | 10 | 2 |

Gender dimension in the project

Does the project include a gender dimension in research content? NO

# PART B

## 1. Explanation of the work carried out by the beneficiaries and Overview of the progress

The overall aim of the SEWA project is to enable computational models for machine analysis of facial, vocal, and verbal behaviour in the wild. This is to be achieved by capitalising on the state-of-the-art methodologies, adjusting them, and combining them to be applicable to naturalistic human-centric human-computer interaction (HCI) and computer-mediated face-to-face interaction (FF-HCI). The target technology uses data recorded by a device as cheap as a web-cam and in almost arbitrary recording conditions including semi-dark, dark and noisy rooms with dynamic change of room impulse response and distance to sensors. It represents a set of audio and visual spatiotemporal methods for automatic analysis of human spontaneous (as opposed to posed and exaggerated) patterns of behavioural cues including analysis of rapport, mimicry, and sentiment such as liking and disliking.

### 1.1 Objectives

In summary, the objectives of the SEWA project are:

- Development of technology comprising a set of models and algorithms for machine analysis of facial, vocal and verbal behaviour in the wild,
- Collection of the SEWA database being a publicly available benchmark multilingual dataset of annotated facial, vocal and verbal behaviour recordings made in-the-wild representing a benchmark for efforts in automatic analysis of audio-visual behaviour in the wild,
- Deployment of the SEWA results in both mass-market analysis tools based on automatic behaviour-based sentiment analysis of users towards marketed products and a sentiment-driven recommendation engine, and
- Deployment of the SEWA results in a novel social-network-based FF-HCI application – sentiment-driven Chat Social Game.

The SEWA project is expected to have many benefits. Technologies that can robustly and accurately analyse human facial, vocal and verbal behaviour and interactions in the wild, as observed by webcams in digital devices, would have profound impact on both basic sciences and the industrial sector. They could open up tremendous potential to measure behaviour indicators that heretofore resisted measurement because they were too subtle or fleeting to be measured by the human eye and ear. They would effectively lead to development of the next generation of efficient, seamless and user-centric human-computer interaction (affective multimodal interfaces, interactive multi-party games, and online services). They would have profound impact on business (automatic market research analysis would become possible, recruitment would become green as travels would be reduced drastically), and they could enable next generation healthcare technologies (remote monitoring of conditions like pain, anxiety and depression), to mention but a few examples.

## 1.2 Explanation of the work carried per WP

### *1.2.1. Work Package 1 (WP1) - SEWA DB collection, annotation and release*

**Task 1.1: Ethical Approval**

The ethical approval of the SEWA experiment has been obtained in M1.

**Task 1.2: SEWA data acquisition**

Task 1.2 has been completed in M12. In the SEWA data-collection experiment, recruited participants have been divided into pairs based on their cultural background, age and gender. Each pair of the subjects participated in two parts of the experiment: watching a total of 4 advertisements, and then discuss about the last advertisement though video-chat. The entire watching of adverts and the subsequent conversation between the volunteers were recorded using web-cameras and microphones integrated into the laptops/PCs of the volunteers.

During the experiment, we recorded 6 groups of volunteers (around 30 persons per group) from six different cultural backgrounds: British, German, Hungarian, Greek, Serbian, and Chinese. The volunteers in each group have a broad distribution in gender and age. Specifically, there are at least three pairs of native speakers in each age group (18~29, 30~39, 40~49, 50~59, and 60+) for each culture. The resulting database contains a total of 204 sessions of experiment recordings, with 1525 minutes of audio-visual data of people's reaction to adverts from 408 individuals and 568 minutes of recorded computer-mediated face-to-face interactions between pairs of subjects.

**Task 1.3: SEWA data annotation**

The SEWA database contains annotations for facial landmarks, LLD features, hand gestures, head gestures, facial action units (AUs), verbal and vocal cues, continuously-valued valence, arousal and liking / disliking (toward the advertisement), template behaviours, episodes of agreement / disagreement, and mimicry episodes.

Due to the large amount of raw data acquired from the experiment, the SEWA database has been annotated iteratively, starting with sufficient number of examples to be annotated in a semi-automated manner and used to train various feature extraction algorithms developed in SEWA. Specifically, 538 short (10~30s) video-chat recording segments were manually selected to form the fully-annotated basic SEWA dataset. These segments were selected based on the subjects' emotional state of low / high valance, low / high arousal, and liking / disliking. All 6 cultures were evenly represented in the basic SEWA dataset, with approximately 90 segments selected from each culture based on the consensus of at least 3 annotators from the same culture.

Details about the annotations included in the SEWA database are as follows:

1. Facial landmarks: The 49 facial landmarks were annotated for all segments included in the basic SEWA dataset. The annotation was performed semi-automatically [Chrysos et al, 2015]. We first applied an automatic facial landmark tracker [Asthana et al, 2014] on all video segments and checked the tracking results to identify the frames with tracking errors. We then manually corrected 1/8 of these frames and used the annotation result to train a set of person-specific trackers. These person-specific trackers were applied to the rest of the frames to obtain more accurate tracking results. Afterward, the updated landmark locations were manually verified, and if necessary, corrected, to form the final annotation result. An example of the facial landmark annotation obtained from this process is show in Figure 1.

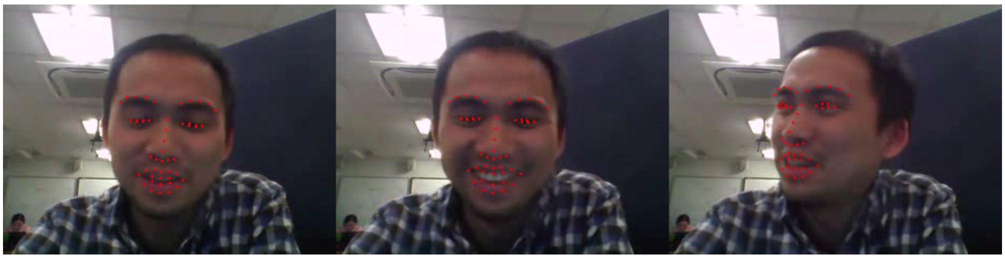
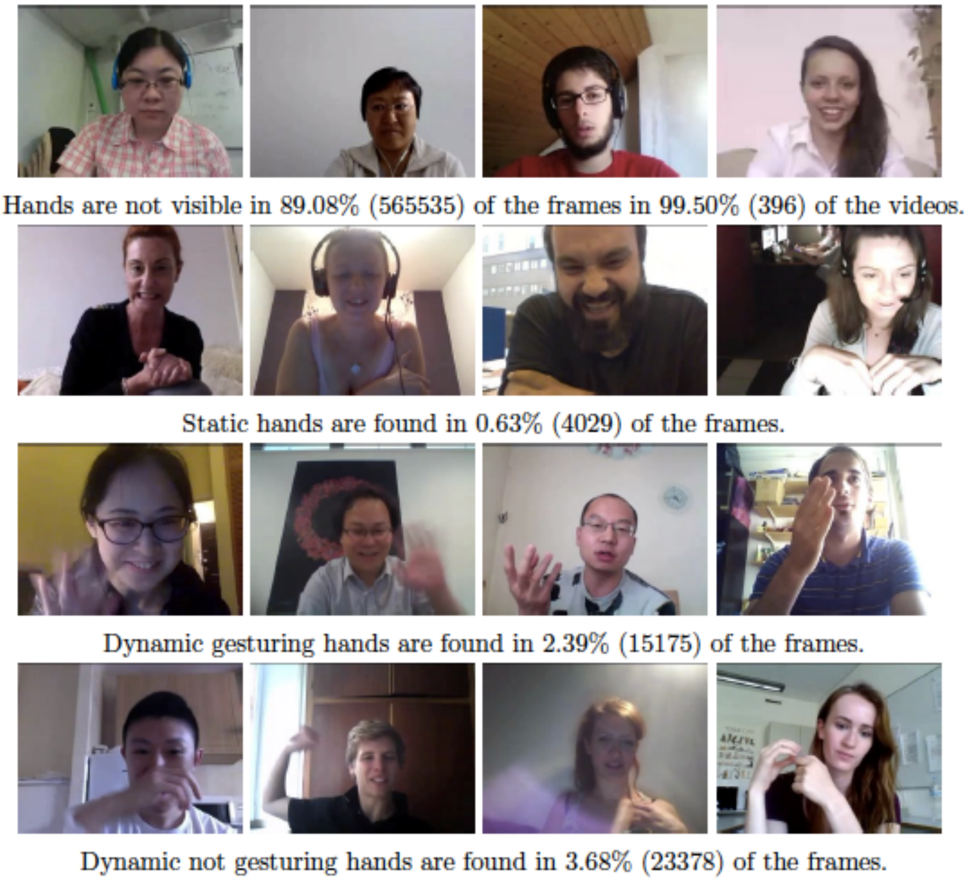


*Figure 1: Examples of facial landmark annotation.*

2. Audio low-level descriptors: We provide two sets of low-level audio descriptor (LLD) features for all recordings included in the SEWA database: the 65 dimension ComPareELLD and the more compact 18 dimension GeMAPSv01aLLD [Schuller et al, 2013]. The features were extracted automatically in 10ms steps.
3. Hand gestures: We annotated hand gestures for all video-chat recordings in 5 frame steps. Five types of hand gestures were labelled: hand not visible (89.08%), hand touching head (3.32%), hand in static position (0.63%), display of hand gestures

(2.39%), and other hand movements (3.68%). Some examples of the labelled frames are shown in Figure 2.

4. Head gestures: We annotated head gestures in terms of nod and shake for all segments in the basic SEWA dataset. The annotation was performed manually on a frame-by-frame basis. We emphasised specifically on high precision during the annotation process. Specifically, only un-ambiguous displays of head nod / shake were labelled. In the end, a total of 282 head nod sequences and 122 head shake sequences were identified. Examples of the labelled head nod / shake sequences are shown in Figure 3.
5. Facial action units: We extracted examples of 5 facial action units (AU) from the basic SEWA dataset: inner eyebrow raiser (AU1, 109 examples), outer eyebrow raiser (AU2, 79 examples), eyebrow lowerer (AU4, 94 examples), lip corner puller (AU12, 104 examples), and chin raiser (AU17, 61 examples). Similar to the case of facial landmarks, the AU examples were identified in a semi-automatic manner. Specifically, we first applied automatic AU detectors to the video segments and manually removed all false-positives from the detection results. Consequently, the AU annotation is not exhaustive, meaning that some AU activations may be missed. Examples of the annotated action units are shown in Figure 4.



Hands are not visible in 89.08% (565535) of the frames in 99.50% (396) of the videos.

Static hands are found in 0.63% (4029) of the frames.

Dynamic gesturing hands are found in 2.39% (15175) of the frames.

Dynamic not gesturing hands are found in 3.68% (23378) of the frames.

*Figure 2: Examples of hand gesture annotation.*

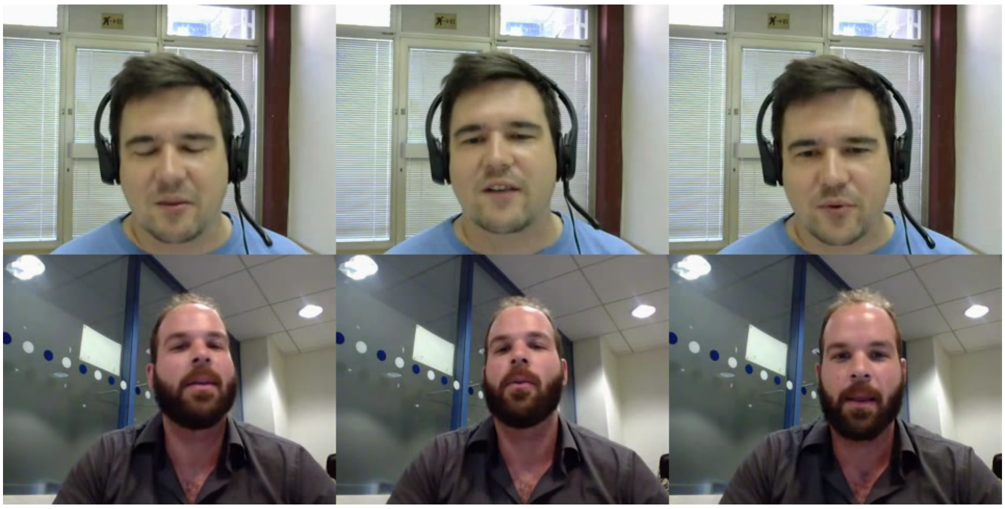


*Figure 3: Examples of head nod (top row) and head shake (bottom row) sequences.*

6. Audio transcript: We provide the audio transcript of all video-chat recordings. In addition to the verbal content, the transcript also contains labels of certain non-verbal cues, such as sighing, coughing, laughter, and so on.

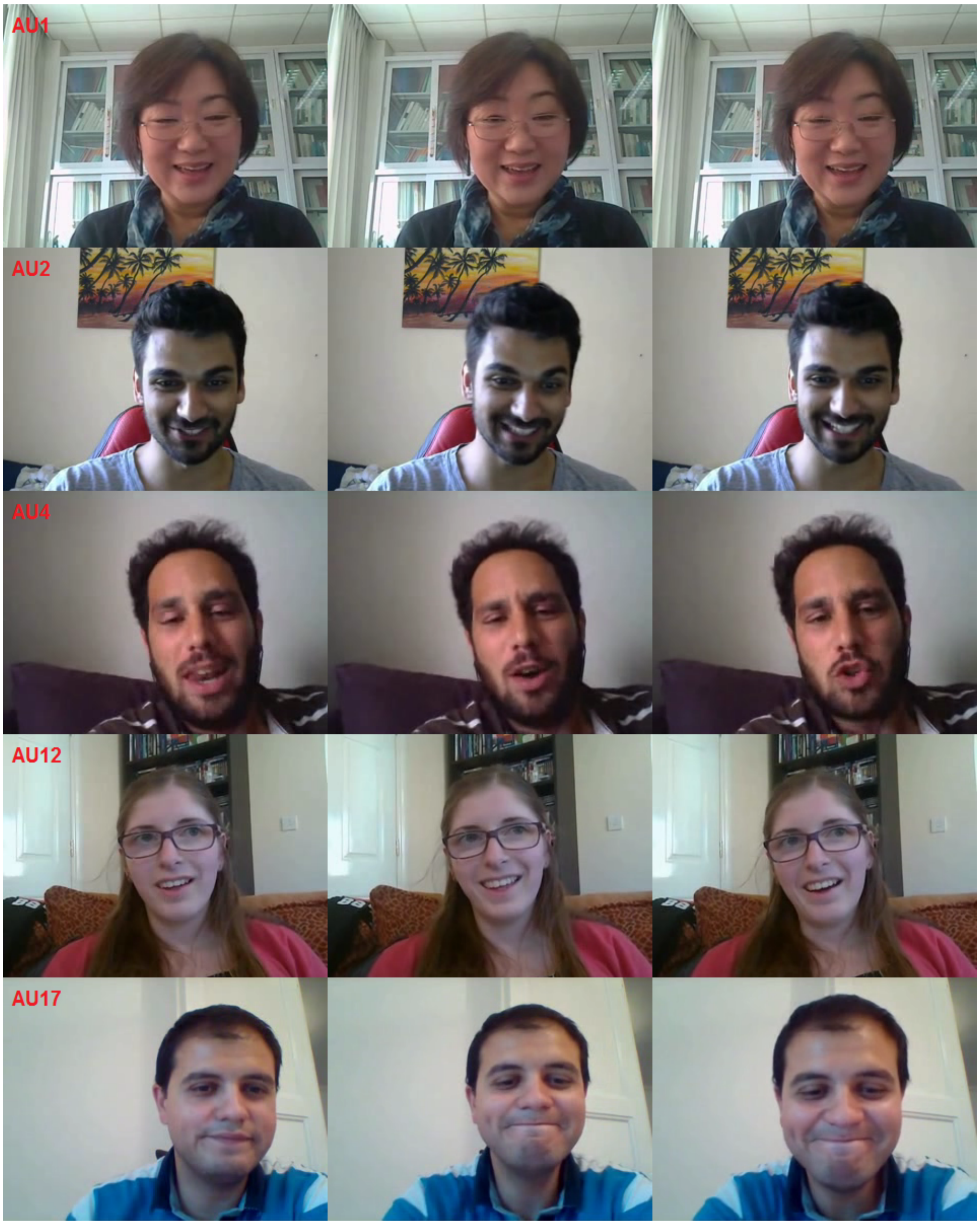


*Figure 4: Examples of the facial action units annotation.*

7. Valence, arousal and liking / disliking: Continuously-valued valance, arousal and liking / disliking (toward the advertisement) were annotated for all segments in the basic SEWA dataset. In order to identify the subtle changes in the subjects' emotional state, annotators were always hired from the same cultural background of the recorded subjects. In addition, to reduce the effect of the annotator bias, 5 annotators were recruited for each culture. The annotation was performed in real-time using a joystick with a sample rate of 66 Hz. To avoid cognitive overload on the annotators, the three dimensions (valence, arousal and liking / disliking) were annotated separately in three passes. Furthermore, for each dimension, the segments were annotated three times, first based on audio data only, then based on video data only and finally based on audio-

visual data. An example of the end result of this annotation process is illustrated in Figure 5.

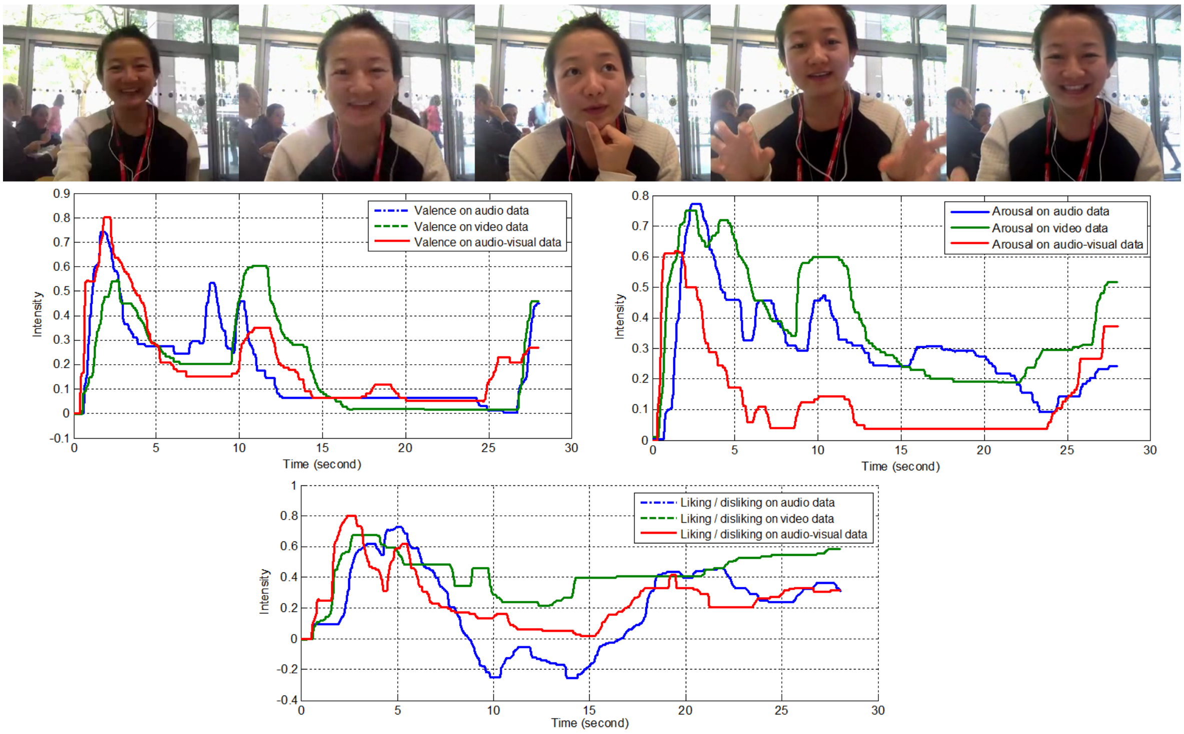


*Figure 5: Example of the continuously-valued annotation of valence, arousal and liking / disliking.*

8. Behaviour templates: We identified behaviour templates for each culture when the subjects are in the emotional state of low / high valence, low / high arousal or showing liking / disliking toward the advertisement. For each category, at least two examples were included. Table 1 shows the exact distribution of the templates found in the basic SEWA dataset.

*Table 1: Templates Behaviours Identified in the Basic SEWA Dataset*

| Culture | Low Valence | High Valence | Low Arousal | High Arousal | Liking | Disliking |
|---|---|---|---|---|---|---|
| British | 2 | 2 | 2 | 2 | 2 | 2 |
| German | 4 | 4 | 3 | 3 | 4 | 4 |
| Hungarian | 2 | 2 | 2 | 2 | 2 | 2 |
| Serbian | 6 | 5 | 2 | 6 | 6 | 6 |
| Greek | 2 | 2 | 2 | 2 | 2 | 2 |
| Chinese | 3 | 4 | 2 | 4 | 5 | 4 |

9. Agreement / disagreement episodes: We extracted a number of episodes from the video-chat recordings in which the pair of subjects were in low, mid or high level of agreement / disagreement with each other. The selections were based on the consensus of at least

3 annotators from the same culture of the recorded subjects. The exact numbers of agreement / disagreement episodes are shown in Table 2.

10. Mimicry episodes: A total of 197 mimicry episodes (48 British, 31 German, 39 Hungarian, 20 Serbian, 41 Greek and 17 Chinese), in which one subject mimicked the facial expression and / or head gesture of the other subject, were identified from the video-chat recordings.

**Table 2: Agreement / Disagreement Episodes Identified in the Video-Chat Recordings**

| Culture | Strong Agreement | Moderate Agreement | Weak Agreement | Weak Disagreement | Moderate Disagreement | Strong Disagreement |
|---|---|---|---|---|---|---|
| British | 12 | 26 | 29 | 7 | 3 | 3 |
| German | 7 | 7 | 7 | 6 | 9 | 6 |
| Hungarian | 7 | 6 | 6 | 5 | 5 | 5 |
| Serbian | 7 | 7 | 7 | 4 | 6 | 4 |
| Greek | 5 | 5 | 5 | 5 | 5 | 5 |
| Chinese | 5 | 6 | 6 | 4 | 5 | 3 |

References:

- G. S. Chrysos, E. Antonakos, S. Zafeiriou and P. Snape. Offline deformable face tracking in arbitrary videos. In IEEE International Conference on Computer Vision Workshops (ICCVW), 2015. IEEE, 2015.
- Asthana, S. Zafeiriou, G. Tzimiropoulos, S. Cheng, and M. Pantic. From pixels to response maps: Discriminative image filtering for face alignment in the wild. IEEE PAMI, 37(6):1941–1954, 2015.
- Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, S. Kim, The Interspeech 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism, Proc. Interspeech 2013, ISCA, Lyon, France, 2013.

**Task 1.4: SEWA database design and release**

The SEWA database has been released online at: http://db.sewaproject.eu/. The web-portal provides a comprehensive search filter allowing users to search for specific recordings based on various criteria, such as demographic data (gender, age, cultural background, etc.), availability of certain types of annotation, and so on. This will facilitate investigations during and beyond the project in the field of machine analysis of facial behaviour as well as in other research fields.

The SEWA database is made available to researchers for academic-use only. To comply with clauses stated in the Informed Consent signed by the recorded participants, all non-academic/commercial uses of the data are prohibited. To enforce this retraction, an end-user licence agreement (EULA) is prepared. Only researchers who signed the EULA will be granted access to the database. In order to ensure secure transfer of data from the database to an authorised user's PC, the data are protected by SSL (Secure Sockets Layer) with an encryption key.

### *1.2.2. Work Package 2 (WP2) - Low-level Feature Extraction*

**Task 2.1: Environmentally robust acoustic features**

Task 2.1 has been finished in M9. The methods developed during Task 2.1, especially bag-of-audio-words (BoAW) and openXBOW have been advanced in later tasks and work packages, especially in Task 2.3 and in work package 4.

**Task 2.2: Environmentally robust visual features**

This task addressed the problem of joint detection of faces and facial landmarks in input videos. Our main achievement is to improve the robustness in case of large and steady changes in head pose, illumination, occlusion, and the expression of the face.

We developed a facial landmark tracker which detects the user's face in video recordings on a frame by frame basis and accurately tracks a set of 49 facial landmarks. Even though very good performance for the facial landmark localisation has been shown by many recently proposed discriminative techniques, accurate tracking in 'in-the-wild' scenarios (such as that of SEWA applications) remains challenging. One way to increase tracking accuracy in this context is to automatically construct person-specific models through incremental updating of the generic model. Specifically, we proposed efficient strategies to update the discriminative model trained by a cascade of regressors. Experimental evaluation on LFPW and Helen datasets shows our method outperforms state-of-the-art generic face alignment strategies. Tracking experiment using SEWA dataset also shows promising results. The method has been implemented as a standalone module, which is further integrated into the SEWA back-end emotion recognition server using HCI^2 Framework. Our implementation is also highly efficient, being able to track 8 video streams in parallel at 50fps on our data processing machine (CPU: Intel Core i7-5960X, memory: 32GB).

**Task 2.3: Cross-lingual language related features**

Back in Task 2.1, the openXBOW toolkit has been initially implemented (then called openWord) to generate bag-of-audio-words (BoAW) representations of acoustic features. During work on Task 2.3, the standard bag-of-words (BoW) method known from natural language processing has been added to openXBOW. Thus, the toolkit makes it possible to generate generic bag-of-words representations of both symbolic (e.g., text) and numeric (e.g., acoustic or visual) descriptors.

OpenXBOW [Schmitt and Schuller, 2016] has been released as a GitHub repository (https://github.com/openXBOW/openXBOW) to make it available for the research community. Though initially targeting emotion recognition [Schmitt et al., INTERSPEECH, 2016], it has already been employed in further tasks [Schmitt et al., ITG, 2016].

In the scope of this work package, openXBOW was used to generate multi-lingual BoW features for multi-language sentiment and emotion analysis. As mentioned above, we provided to combine BoAW and BoW, fusing these two methodologies on two different levels, namely merging the feature vectors or combining the decision output of each system. All default extensions to BoW have been implemented, such as n-grams, n-character-grams, TF/IDF-weighting, histogram normalization and stop words. The success of the implementation has been proven on the Thinknook dataset for sentiment analysis of Tweets achieving state-of-the-art results.

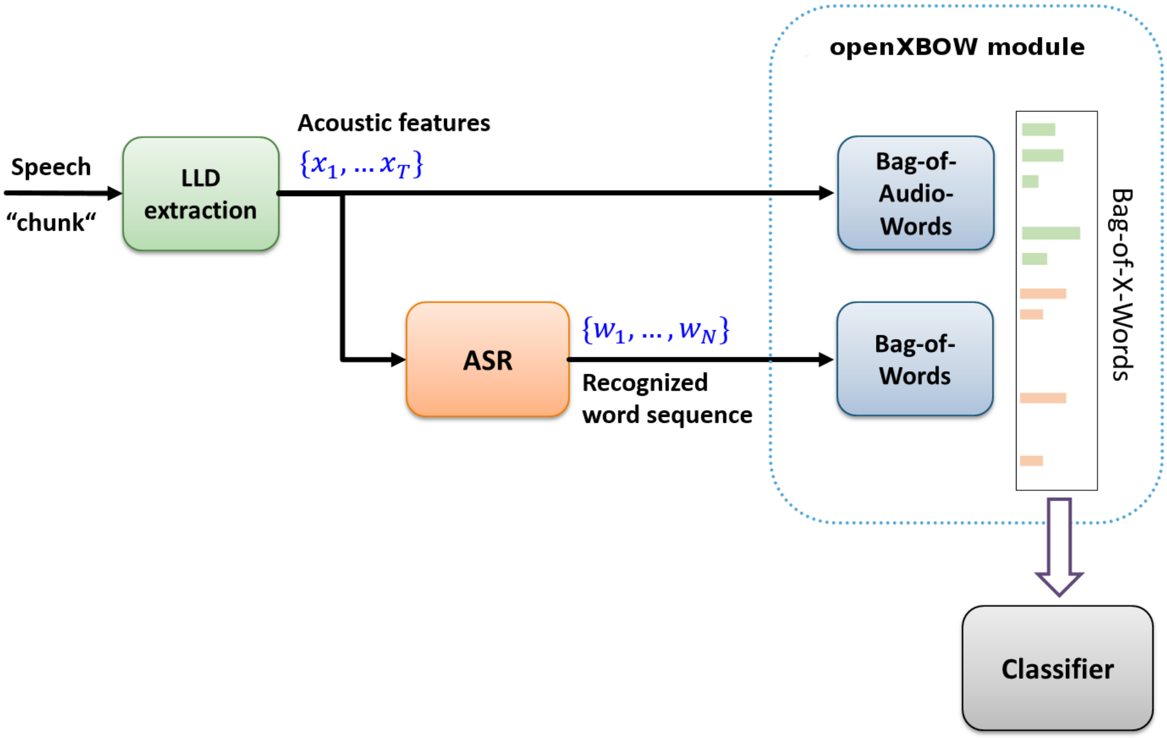


*Figure **Error! No sequence specified.**.2.2.1: SEWA acoustic/linguistic bag-of-words processing chain*

To obtain a textual representation (i.e., the transcription) of the conversation, a system for automatic speech recognition (ASR) is necessary, which outputs a sequence of words as shown in Figure 1.2.2.1. Usually, the ASR system takes Mel frequency cepstral coefficients (MFCC) as an input, which are also meaningful and commonly used low-level descriptors for emotion recognition [Schmitt et al., INTERSPEECH, 2016]. Thus some computational effort can be saved in the architecture of the system.

ASR modules usually consist of an acoustic model (AM) and a language model (LM) part (see Figure 1.2.2.2). While the acoustic model recognizes phonemes, syllables, or words, the LM chooses the most probable sequence of words from a set of candidates, based on knowledge about the target language. Without a LM, an optimum decoding of words is not possible as different words are usually vocalized in a similar way and the actual meaning reveals only from the context.

As the LM and, depending of the family of the language, also the AM are specific for each language, the ASR module must be trained for each target language. This is difficult especially for rare languages, such as Hungarian, Serbian, or Greek, which are part of the SEWA database. Another problem is that the presence of strong dialect (e.g., Bavarian dialect in the German subset of the database) decreases the performance dramatically. As for training, a huge amount of data is necessary, the SEWA database is not suitable for that purpose. Moreover, data from video chat conversations contain a large amount of cross-talk, noise and signal loss, which further deteriorate the performance of ASR.

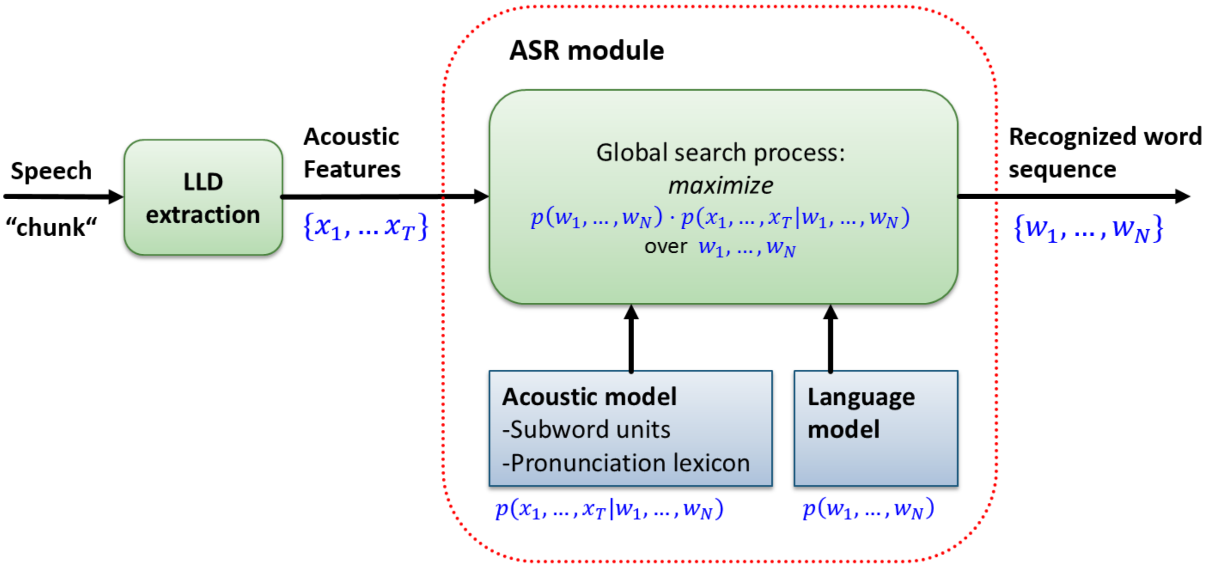


*Figure 1.2.2.**Error! No sequence specified.**: ASR module used in SEWA, based on the Kaldi toolkit*

The ASR module used within the scope of SEWA is based on the Kaldi toolkit (http://kaldi-asr.org/). The following table presents results obtained on the LibriSpeech corpus of audio books. Word error rates (WERs) below 5 % show that our models keep up with state-of-the-art ASR.

| **Dataset (LibriSpeech)** | **WER (%)** |
|---|---|
| Development set (clean) | 4.96 |
| Test set (clean) | 5.30 |
| Development set (noisy) | 12.90 |
| Test set (noisy) | 13.68 |

A similar performance can, however, not be reached on the `in the wild' SEWA dataset due to the above mentioned reasons. For emotion recognition, it would be sufficient to recognize meaningful keywords related to emotion and sentiment. The impact of reduced ASR performance will be studied in later stages of the project, e.g., during WP4 and the first integration with the applications from the industrial partners.

We considered two ways to overcome the described problems and make use of linguistic information anyhow. To obtain BoAW representations more similar to linguistic words, they can also be created over short segments which are the output of automatic syllabification of the speech signal. A more promising method is, however, to use acoustic landmarks.

Acoustic landmarks are symbolic representations of speech units, often related to phonemes. The motivation behind using them is that they are more robust than ASR (due to the much more limited number of phonemes or speech units) while still containing linguistic information. They are not supposed to be completely language independent, but are assumed to have at least some similarities between languages. Initial results are reported in deliverable D2.3.

References:

- M. Schmitt and B. W. Schuller. "openXBOW-Introducing the Passau Open-Source Crossmodal Bag-of-Words Toolkit." *arXiv preprint arXiv:1605.06778* (2016).
- M. Schmitt et al. "A Bag-of-Audio-Words Approach for Snore Sounds' Excitation Localisation." *Proceedings of 12. ITG Symposium on Speech Communication,* VDE, 2016.
- M. Schmitt et al. "At the border of acoustics and linguistics: bag-of-audio-words for the recognition of emotions in speech." Proceedings of INTERSPEECH, San Francisco, USA, 2016.
- Z. Zhang et al. "Facing realism in spontaneous emotion recognition from speech: Feature enhancement by autoencoder with LSTM neural networks." Proceedings of INTERSPEECH, San Francisco, USA, 2016.

### *1.2.3. Work Package 3 (WP3) - Mid-level feature extraction*

We present the robust mid-level visual feature detection component developed for the SEWA project. In particular, the component detects facial action units (AUs) from the face images of the user in video recordings, and on a frame-by-frame basis. To obtain the tool that can be applied for the detection task, two steps have been employed: semi-automatic data annotation and training of the devised models for the AU detection. In the first step, we used the state-of-the- art sequence modelling discriminative method, trained on two publicly available datasets with AUs intensity annotations (0-5), to semi-automatically annotate the target SEWA videos. Then, a group of experts have manually checked such segments and classified them into active and non-active AUs, providing the starting and end frame of the target AU. Finally, in step two, the obtained annotations were used to train models for automatic detection of target AUs. These models are based on the proposed state-of-the-art sequence modelling method, Variable-state latent Conditional Random fields (VSL-CRFs [2]), for AU detection.

**Task 3.1: Automatic Detection of Head and Hand Gestures**

The WP3 envisioned using the head and hand gestures as the mid-level features, in addition to AUs estimated directly from the facial landmarks. Note, however, that the hand gestures will not be included in this set due to their scarce occurrences in the SEWA videos (see also section 3.1.1 and D9.1 for further explanation).

**Task 3.1.1 Head Gestures**

We developed a Hidden Markov Model (HMM)-based method for head nod / shake detection. Our method takes the face orientation vector (yaw, pitch and roll, as given by the Chehra face tracker [3] described in D2.2) as input. For each frame, we first calculate the heads angular velocity by comparing the mean face orientation vectors in the two short (0.1 second) windows before and after the current frame. The windows are used to reduce the methods sensitivity to small errors in the face orientation vectors estimated by the Chehra tracker. We then discretize the heads angular velocity into four directional code-words (upward, downward, leftward and rightward motion). Based on these code- words, we trained three HMMs to recognize nod (nodHMM), shake (shakeHMM) and other arbitrary (otherHMM) head movements. These models were trained on the head nod / shake annotations in the basic SEWA dataset, which consists of 538 short segments selected from the SEWA video-chat recordings. To recognize head gesture in each frame, we first feed the code-word sequence obtained from a 0.6 second window prior to the frame into the three models to calculate their likelihood given the observations. The head gesture (nod, shake or other) labelled by the most likely model is then given as the recognition result.

### Task 3.1.1 Hand Gestures

As a part of Task 3.1, building a touching-the-face hand gesture detector was initially envisioned. The automated extraction of touching the-face gestures have been attempted in two ways by analysis of dynamic hand movement and static face touching. After a careful analysis of the SEWA videos, we came to the conclusion that the face touching and dynamic gestures cannot be seen in many videos, and most of the events are very short. In summary, the classes of hand gestures are quite heterogeneous, and the real interpretation of each event will most likely be quite varying. Hand gesture trackers yield quite poor results, having many false positives and low true positives that will not be useful in commercial applications we are building. For this reason, we excluded these gestures from the set of mid-level features originally envisioned.

### Task 3.3: Automatic detection of Facial Action

For this task, we focus on automatic detection of five AUs (1,2,4,12 and 17). We use these 5 AUs as they are occurring most in naturalistic data, as given by the SEWA dataset, and are important for high-level reasoning about sentiment, as described in WP5.

We adopted a semi-automatic approach to annotation of AUs. In the first step, we automatically classified each image frame from SEWA videos into one of the intensity levels of target AU. This is performed by means of the CORF [1]. The output of this model was then manually inspected by expert annotators to identify the starting and end frame of target AUs from SEWA videos in question. Once the labels for presence/absence of target AUs in the inspected segments were verified, the modelling of each AU using SEWA annotations is performed by means of the VSL-CRF [2] model.

The proposed pipeline for AU detection is depicted in Fig. 1. As input, it receives input features (i.e., locations of fiducial facial points extracted using the Chachra tracker [3, 4].). these are then passed through several blocks for data pre-processing, including normalization and dimensionality reduction. Then classification of each target frame into active/non-active AU is performed using VSL-CRF classifiers trained for each AU independently.

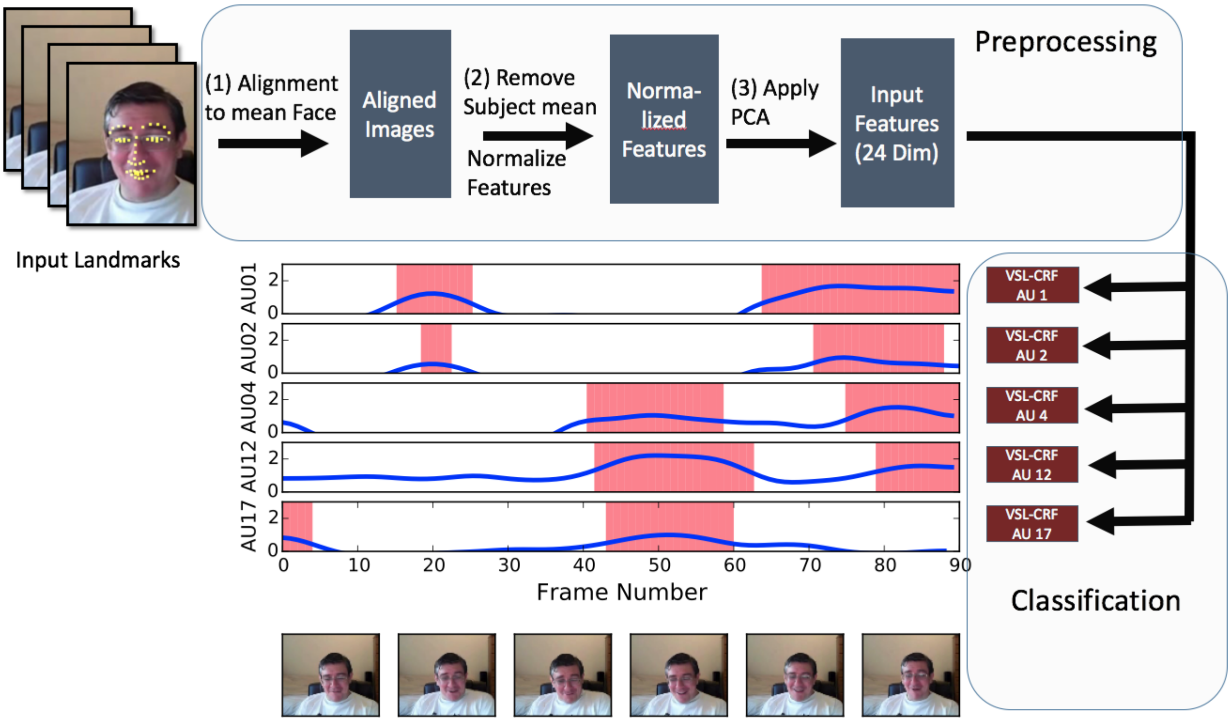


*Fig. 1: Pipeline for AU detection: The facial points coming from the proposed tracker are (1) aligned to the mean face, (2) the mean value is removed and the (3) the dimensionality is reduced using PCA. The resulting sequential data is then classified using the VSL-CRF model.*

We have implemented the proposed pipeline for feature extraction and the algorithms for AU detection in two separate modules that are directly linked. The module for feature extraction is named 'landmarks2feature' and can be applied on the landmarks coming from the proposed tracker. Steps 1-3 in Fig. 1 show the processes in this module. The only parameter that has to be specified is the number of principle components. To make the computation as fast as possible, we have pre-computed the facial mean shape from all the data from the SEWA databases, and stored it in this module. The affine transformation is the computationally most expensive process which is performed using the C++ implementation from OpenCV. The process is highly efficient, being able to process 100+ fps on a regular workstation (CPU: Intel Xeon 3.6GHz, memory: 32GB). The output of the feature extraction module is directly linked to the VSL-CRF model for AU detection. Therefore, we have used the Matlab implementation from the Dynamic Ordinal Classification (DOC) Toolbox. The inference can be performed for all AUs in parallel, and we achieve 65 fps on the workstation mentioned above.

[1] M. Kim and V. Pavlovic. Structured output ordinal regression for dynamic facial emotion intensity prediction. ECCV, pages 649–662, 2010.

[2] R. Walecki, O. Rudovic, V. Pavlovic, and M. Pantic. Variable-state latent conditional random fields for facial ex- pression recognition and action unit detection. *IEEE Inter- national Conference on Automatic Face and Gesture Recognition*, pages

1–8, 2015.

[3] Akshay Asthana, Stefanos Zafeiriou, Georgios Tzimiropoulos, Shiyang Cheng, and Maja Pantic. From pixels to response maps: Discriminative image filtering for face alignment in the wild. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 37(6):1312–1320, 2015.

[4] C. M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

### *1.2.4. Work Package 4 (WP4) - Continuous Affect and Sentiment Sensing in the Wild*

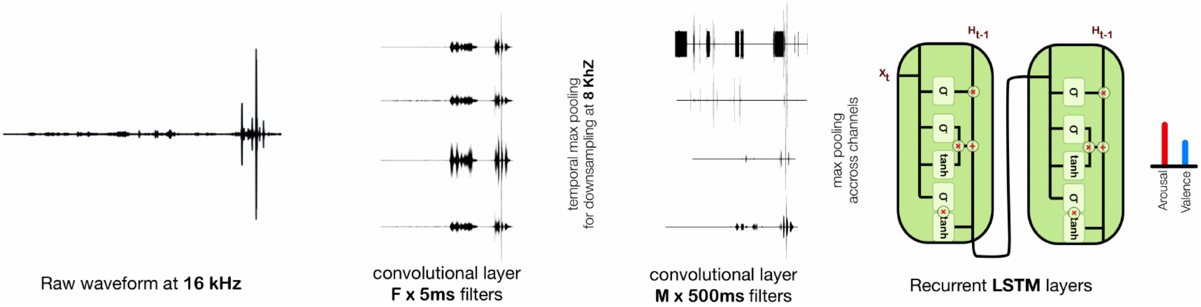
**Task 4.1: Multi-modal analysis of sentiment and affect-related states**

The objective of this task is to exploit technologies to address the problem of automatic continuous estimation of emotion in unconstrained audio and video recordings based on low- and mid-level features extracted in WP2. The ultimate goal is to develop an automatic audio-visual affect recognizer, supporting time-continuous, context-sensitive, and multidimensional affect predictions.

One state-of-the-art affect predictor is based on end-to-end learning, which exploits raw input signals instead of hand-engineered features to better fit the task at hand. Specifically, Convolutional Neural Networks (CNNs) are utilized to get higher level representations directly from raw input, and then the output of CNNs could be fed into recurrent neural networks for the application of continuous emotion prediction. We investigated the end-to-end learning on both audio and video raw input to build an end-to-end audio/video emotion recognizer.

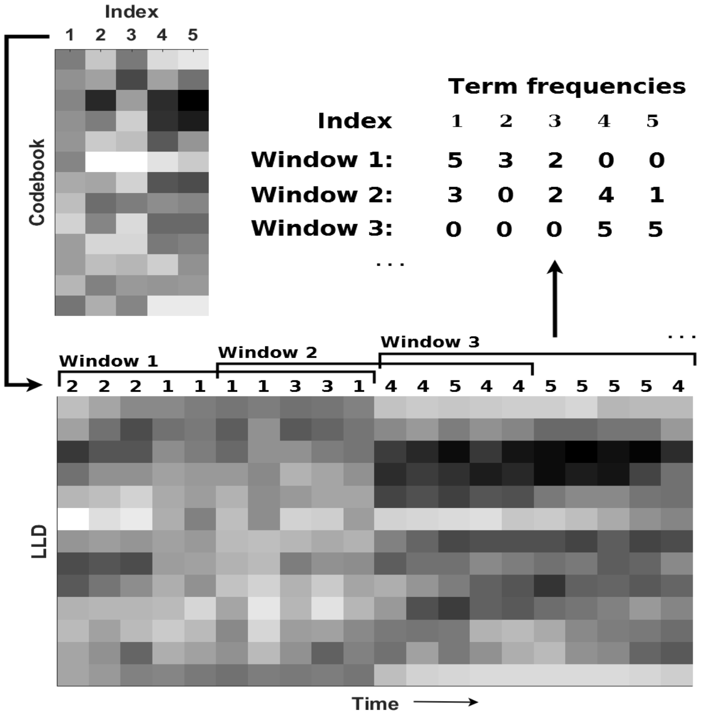
Furthermore, based on former investigation, the performance is better for recognition of both arousal and valence with Bag-of-Audio-Words (BoAW) compared to traditional acoustic Low-Level Descriptors (LLDs). In this task, therefore, we also investigated the possibility to combine both the BoAW features with Bag-of-Video-Words (BoVW) features as another affect predictor.

*For end-to-end learning topology*, it is summarised and depicted in figure below.



- *Input*. We segment the raw waveform to 6s long sequences after we pre-process the time-sequences to have zero mean and unit variance to account for variations in different levels of loudness between the speakers. At 16 kHz sampling rate, this corresponds to a 96000-dimensional input vector.
- *Temporal Convolution*. We use F = 40 space time finite impulse filters with a 5 ms window in order to extract fine-scale spectral information from the high sampling rate signal.
- *Pooling across time*. The impulse response of each filter is then passed through a half-wave rectifier (analogous to the cochlear transduction step in the human ear) and then down-sampled to 8kHz by pooling each impulse response with a pool size of 2.
- *Temporal Convolution*. We use M = 40 space time finite impulse filters of 500 ms window, These are used to extract more long-term characteristics of the speech and the roughness of the speech signal.
- *Max pooling across channels.* We perform max-pooling across the channel domain with a pool size of 20. This reduces the dimensionality of the signal while preserving the necessary statistics of the convolved signal.
- *Recurrent layers*. We segment the 6s sequences to 150 smaller sub-sequences to match the granularity of the annotation frequency of 40 ms. We use two bidirectional LSTM layers with 128 cells each, although we get similar performance with the uni-directional approach.

For the bag-of-X-Words approach, similar workflow as in WP 2 is taken, but on not only audio features, but also based on video features. Firstly, low-level features are extracted. Then the codebook is generated on the training partition by random sampling or k-means clustering. Once the codebook has been generated, features within a certain window are quantised. Then a histogram is created from the frequencies. This is also exemplified in figure below.



We have shown that both the end-to-end learning method and the bag-of-X-Words method are suitable for employment within the SEWA project, to build an automatic audio/video affect recognizer. The former method avoids using hand-crafted features and models the patterns better directly from raw signals with the usage of convolutional neural networks and recurrent neural networks. However, the latter method exploits the low-level features and generates higher-level features from them. Although it depends on the features extracted, the performance is competitive or even better than the former methods.

Once all annotations of the SEWA database are finished, we will evaluate both approaches on these data and then decide which one to be employed in SEWA. It would also be possible to integrate two methods in one by fusing their predictions on decision level to further boost the performance.

**Task 4.3: Confidence measures**
Work on Task 4.3 has not yet started.

### *1.2.5. Work Package 5 (WP5) - Behaviour Similarity in the Wild – start M12 end M30*

Objectives of the WP: WP5 formulates the problem of behaviour similarity in which the aim is to address the question if audio-visual recordings of two behaviours are similar instead of recognizing the actual behaviours (e.g., smile, interest etc). Therefore, instead of training machine learning algorithms to detect target classes, using large amounts of manually annotated data (usually unavailable or excessively expensive), here it is only required to extract suitable data representations that capture certain spatio-temporal dynamics unique to the target typical behaviours, referred to as "template" behaviours. The extracted representations from the "templates" are compared to the representations extracted from currently observed behaviour for similarity measurement. Hence, the objective of WP5 is to develop in-the-wild technology for spatiotemporal behaviour similarity measurement (i.e., how similar are two behavioural recordings in space and time). The approach adopted here is (i) to temporally segment behaviour by finding regions that undergo specific temporal changes, (ii) to extract suitable data representations that align in a spatio-temporal manner the segmented behaviour with the "template" gesture, and (iii) to measure the similarity of the behaviour shown in the template and the current video (e.g. positive sentiment, boredom, etc.). Finally, a similar set of processing steps will be attempted to attain spatiotemporal audio-visual behaviour similarity measurement.

**Task 5.1: Online Unsupervised Segmentation of Behaviour and latent feature extraction (ICL 6 PM) Start M12 – End M18**

Learning a temporally invariant subset of data points, referred to as representatives or archetypes, which can efficiently describe high-dimensional time series is an important visual data analysis problem. Temporally invariant archetypes are essentially the most informative slow-varying, data points of the time series and thus they can be used for summarization, representation, clustering, and segmentation of high-dimensional time series, such as videos. Representing time-varying data with a small number of archetypes has several advantages over working with long high-dimensional time series. First, archetypes facilitate the removal of outliers since they are not true representatives of the data. Moreover, the performance, the memory requirement, and the computational cost of clustering and segmentation algorithms is improved. The problem of learning temporally invariant archetypes becomes rather challenging when dealing with multiple time series arising from different modalities. For instance, human motion can be represented by multimodal time series of pixel intensities, depth maps, and motion capture data. Similarly, a particular human behaviour can be identified by certain vocal, gestural, and facial features extracted from both the audio and visual modalities. In this multimodal setting, the task is to find slow varying or temporally invariant prototypical data points efficiently describing the multiple time series with the additional property of being invariant across different modalities.

To this end, the **temporal archetypal analysis** is proposed in [1], enabling the discovery of slow-varying and modality invariant data representatives from multiple high-dimensional time series. In particular, we seek to express each data point in each time series as a convex combination of slowly-varying archetypes with the combination coefficients being shared among the different modalities. Moreover, the archetypes of each time series are also restricted to be convex combinations of the data. To find such invariant archetypes, a novel constrained optimization problem is solved by employing an iterative algorithm with guaranteed convergence.

The performance of the proposed method is assessed by conducting experiments in unsupervised action segmentation by employing three different datasets. In particular, the temporal archetypal analysis is tested on both single- and multimodal data. Experimental results indicate the effectiveness of the proposed approach on this application, outperforming state-of-the-art compared methods.

**Task 5.2 Spatio-temporal alignment of segmented visual behaviours and measuring similarity: (ICL 9PM) Start M15 – End M24**

To measure the similarity between two given segmented behaviours a novel framework has been devised [2]. The proposed framework is used to first estimate the parameters of a dynamical system. These parameters are used for temporal invariant data representations that capture the underlying dynamics of the behaviours. The similarity between the two behaviours will be measured by computing the distance of the systems' parameters in a suitable matrix manifold.

Concretely, analysis of human behaviour concerns detection, tracking, recognition, and prediction of complex human behaviours including affect and social behaviours such as agreement and conflict escalation/resolution from audio-visual data captured in naturalistic, real-world conditions. Representative machine learning models employed for automatic, continuous behaviour and emotion analysis include Hidden Markov Models (HMMs) for facial expression recognition, Dynamic Bayesian Networks (DBN) for human motion classification and tracking, Conditional Random Fields (CRFs) for prediction of visual backchannel cues (i.e., head nods), Long-Short Term Memory (LSTM) Neural Networks, and regression-based approaches. Despite their merits, these methods rely on large sets of training data, involve learning of a large number of parameters, they do not model dynamics of human behaviour and affect in an explicit way, and, more importantly, they are fragile in the presence of gross non-Gaussian noise and incomplete data, which is abundant in real-world (visual) data.

In this WP, we model and tackle the problem of dynamic behaviour analysis in the presence of gross, but sparse noise and incomplete visual data under a different perspective. The modelling assumption here is that for smoothly varying dynamic behaviour phenomena, such as conflict escalation and resolution, temporal evolution of human affect described in terms of valence and arousal, or motion of human crowds, among others, the observed data can be postulated to be trajectories (inputs and outputs) of a linear time-invariant (LTI) system. Recent advances in system theory indicate that such dynamics can be discovered by learning a low-complexity (i.e., low-order) LTI system based on its inputs and outputs via rank minimization of a Hankel matrix constructed from the observed data. Here, continuous-time annotations characterizing the temporal evolution of relevant behaviour or affect are considered as system outputs, while (visual) features describing behavioural cues are deemed system inputs. In practice, visual data are often contaminated by gross, non-Gaussian noise mainly due to pixel corruptions, partial image texture occlusions or feature extraction failure (e.g., incorrect object localization, tracking errors), and human assessments of behaviour or affect may be unreliable mainly due to annotator subjectivity or adversarial annotators. The existing structured rank minimization-based methods perform sub-optimally in the presence of gross corruptions. Therefore, to robustly learn a LTI system from grossly corrupted data, we formulate a novel q-norm regularized (Hankel) structured Schatten-p norm minimization problem.

The proposed model is the heart of a general and novel framework for dynamic behaviour modelling and analysis. A common practice in behavioural and affective computing is to train machine learning algorithms by employing large sets of training data that comprehensively cover different subjects, contexts, interaction scenarios and recording conditions. The proposed approach allows us to depart from this practice. Specifically, we demonstrate for the first time that complex human behaviour and affect, manifested by a single person or group of interactants, can be learned and predicted based on a small amount of person(s)-specific observations, amounting to a duration of just a few seconds. The developed framework is summarized in Fig. 1.

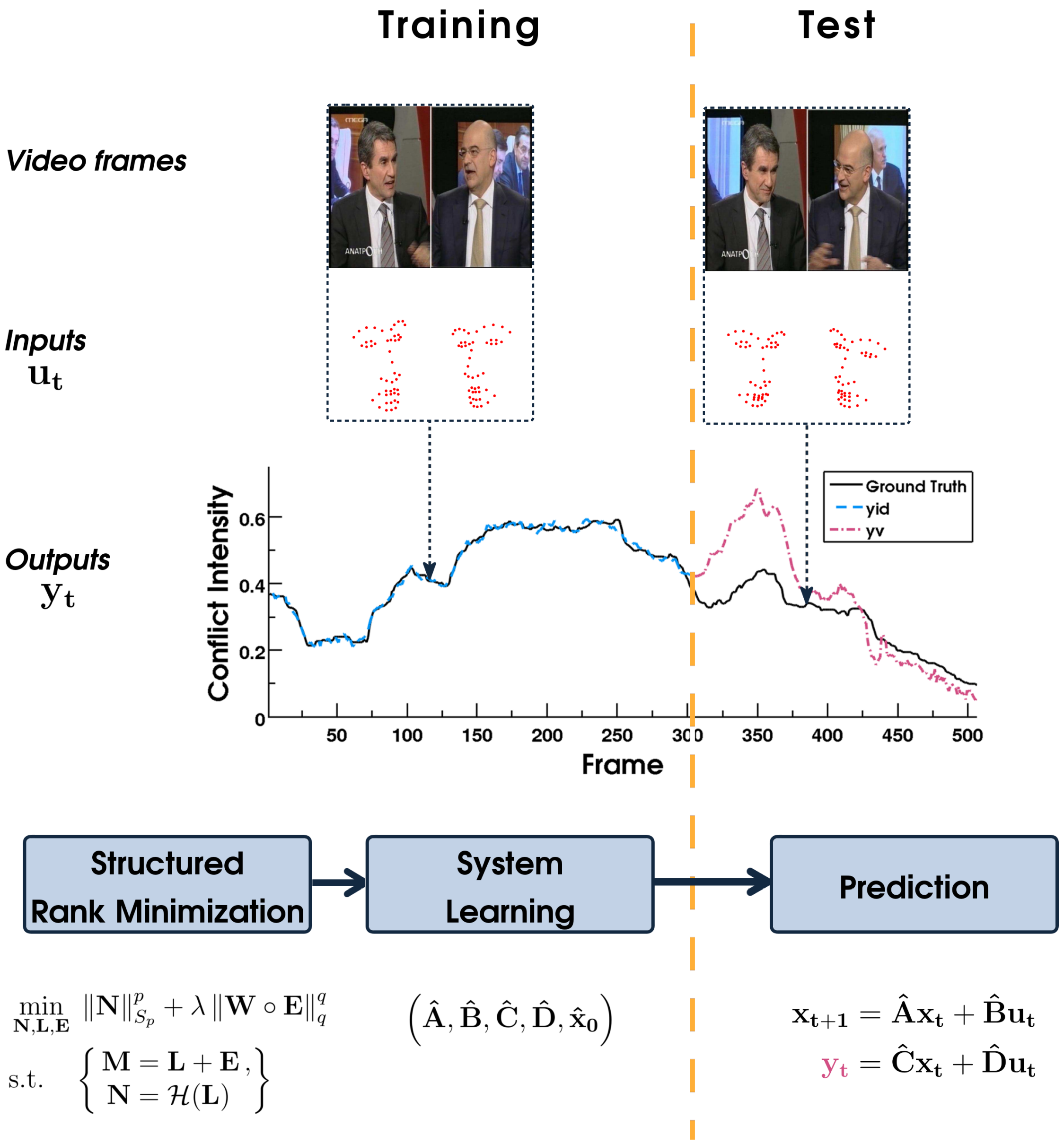

Training
Test
Video frames
Inputs
$\mathbf{u_t}$
Outputs
$\mathbf{y_t}$
Conflict Intensity
0
0.2
0.4
0.6
50
100
150
200
250
300
350
400
450
500
Frame
Ground Truth
yid
yv
Structured Rank Minimization
System Learning
Prediction
$\min_{\mathbf{N},\mathbf{L},\mathbf{E}} \|\mathbf{N}\|_{S_p}^p + \lambda \|\mathbf{W} \circ \mathbf{E}\|_q^q$
s.t. $\left\{ \begin{matrix} \mathbf{M} = \mathbf{L} + \mathbf{E}, \\ \mathbf{N} = \mathcal{H}(\mathbf{L}) \end{matrix} \right\}$
$\left(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}, \hat{\mathbf{x}}_0\right)$
$\mathbf{x_{t+1}} = \hat{\mathbf{A}}\mathbf{x_t} + \hat{\mathbf{B}}\mathbf{u_t}$
$\mathbf{y_t} = \hat{\mathbf{C}}\mathbf{x_t} + \hat{\mathbf{D}}\mathbf{u_t}$

*Fig. 1 Illustration of the proposed dynamic behaviour analysis framework, as applied on the task of conflict intensity prediction for a sequence from CONFER dataset. A portion of the sequence frames is used for LTI system learning through the proposed structured rank minimization method (training), while the remaining frames are used for prediction (test).*

The effectiveness and the generalizability of the proposed model is corroborated by means of experiments on synthetic and real-world data. In particular, the generalizability of the proposed framework is demonstrated by conducting experiments on 3 distinct dynamic behaviour analysis tasks, namely (i) conflict intensity prediction, (ii) prediction of valence and arousal, and (iii) tracklet matching. The attained results outperform those achieved by other state-of-the-art methods on both synthetic and real-world data and, hence, evidence the robustness and effectiveness of the proposed approach.

**Task 5.3: Audiovisual behaviour similarity measurement.**
Task started in January 2017.

References

[1] E. Fotiadou, Y. Panagakis, and M. Pantic, “Temporal Archetypal Analysis”, in Proc. 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017), to appear.

[2] C. Georgakis, Y. Panagakis, and M. Pantic, Dynamic Behavior Analysis via Structured Rank Minimization, International Journal of Computer Vision. 2017.

### *1.2.6. Work Package 6 (WP6) - Temporal Behaviour-Patterning and Interpersonal Sentiment in the Wild*

**Task 6.1: Prediction-based approach to detection of temporal behaviour-patterning**

Work on this WP began in January 2016. The focus of the research conducted within the scope of this work package is on the prediction of mimicry and rapport. Mimicry means that a person imitates the voice, language, facial expressions, or gestures of his/her conversational partner. Rapport means that one dialogue partner fully understands and appreciates the position and the attitudes of the partner. To facilitate the definition of rapport, in the SEWA project, it was agreed to predict rather agreement and disagreement as these are defined more clearly.

Most of our efforts have been focused on annotating the data collected during the SEWA experiments. As basis for the research, short segments from the video chats contained in the SEWA database were selected. For each of the six cultures (Chinese, Hungarian, German, British, Serbian, and Greek), at least 42 audio-visual chunks were extracted from the video chats in 6 different categories of agreement/ disagreement (strong/ medium/ weak agreement, weak/ medium/ strong disagreement). Preliminary results based on acoustic features and support vector machine provide an unweighted average recall (UAR) between 18 % and 39 % (chance level: 16.7 %).

We have also modified a prediction-based approach for audio-visual fusion in order to detect mimicry without explicit knowledge of what action has been mimicked. In order to get some initial results (while the annotation procedure was taking place) we have used a relatively new multimodal database, containing mimicry episodes as they occur in naturalistic dyadic interactions, the MAHNOB Mimicry Database [1].

Database description: The corpus is recorded using ambient and individual close-talk fixed microphones, individual cameras from 6 frontal and 1 overhead view(s), and a profile-view wide-angle camera. All output signals were exactly synchronized using external triggers. Video data was recorded at 58 frames/second, and audio was sampled at 48kHz. The dataset consists of 54 recordings of dyadic face- to-face interactions: 34 are discussions on a political topic, and 20 are conversations situated in a role-playing game. Each session is between 5 and 20 minutes long. The subjects consist of 40 participants and 3 confederates, across a range of ethnic backgrounds and first languages. This data has been partially annotated for multiple behaviours, including dialogue acts, head gestures, hand gestures, body movement and facial expression, and mimicry episodes. Due to only partial availability of annotations, we used 10 sessions, with a session length median of 14 minutes. Table 1 presents statistics about the subjects used in this work. We define a mimicry episode as the occurrence of a behaviour shown by a dyad participant as a result of the other dyad participant's prior display of that signal. The episode onset is taken to be the onset of the dyad participant's action subsequently manifested by the mimicker (i.e. the other dyad participant), whilst the offset is taken to be the offset of the mimicker's display of that action. The upper bound on the time lag between mimicked action offset and mimicker's action onset is set at 4 seconds - a longer delay would be unlikely to be a mimicry episode [1]. Mimicry behaviours may occur multiple times within the same episode, either due to overlapping occurrences (i.e. the onset of a behaviour to be mimicked occurs before the offset of a previously mimicked behaviour), or "reflective" mimicry, i.e. subject 2 mimicking an action of subject 1, which is subsequently mimicked by subject 1, as in contagious laughter.

Audio features: Cepstral features, such as MFCCs, have been widely used in speech recognition, language-identification, and discrimination between linguistic/non-linguistic vocalizations. We use the first 6 MFCCs, computed every 10ms, over a window of 40ms, giving a frame rate of 100 frames/second.

Visual features: Changes in facial expression are captured using the point tracker described in [2], which uses an online appearance model to track rigid head movements and non- rigid facial motion, using 113 landmark facial points. It also decouples this movements to output MPEG-4 facial animation parameter (FAP) estimates, corresponding to mouth width, mouth height, eyebrow pose etc.

*Table 1: Session statistics (class size and episode length reported as number of samples, session length as time in seconds)*

| Session # | Class size | | Episode length mean/var | | Session length |
|---|---|---|---|---|---|
| | Mimicry | Non-mimicry | Mimicry | Non-mimicry | |
| 3 | 1273 | 24532 | 254/60 | 4583/6021 | 7m24s |
| 4 | 2714 | 52685 | 226/126 | 4777/4466 | 15m54s |
| 5 | 4040 | 52647 | 237/140 | 3284/7342 | 16m16s |
| 6 | 2146 | 54016 | 214/94 | 5369/5277 | 16m07s |
| 11 | 2350 | 52105 | 195/86 | 4299/4978 | 15m38s |
| 21 | 1967 | 33057 | 281/126 | 4696/2393 | 10m03s |
| 32 | 4800 | 32087 | 228/152 | 1515/1750 | 10m36s |
| 33 | 1072 | 54826 | 172/62 | 9137/6093 | 16m03s |
| 42 | 6009 | 36651 | 214/104 | 1307/1598 | 12m14s |
| 44 | 3833 | 14384 | 212/124 | 845/1125 | 5m13s |

Methodology: We adapt a method suggested in [3], where each feature vector is split into two disjoint subsets - one subset of features is reconstructed from the other using a class-specific regression model, and the model with minimum reconstruction error classifies the sample. In our case, our subsets are the subject-specific audio-visual features. For each of the two classes, mimicry and non-mimicry, we train a regression model from the first subject's features to the second subject's features, and vice versa. This is done for multiple time lags, both positive and negative, to account for subject reaction time, and directionality of mimicry. We use the long short-term memory network (LSTM) as our underlying regression model to account for sequential dependencies in our data, without resorting to concatenation of multiple samples from a window into one very large feature vector. The relationship between the subject 1 and subject 2's features for both mimicry $S_1^M, S_2^M$ and non-mimicry $S_{\mathbf{1}}^{\bar{M}}, S_{\mathbf{2}}^{\bar{M}}$ is modelled by  for mimicry and $f^{\bar{M}}$ $f^{\bar{M}}$ for non-mimicry as follows:

$$f^{M}_{S_2 \rightarrow S_1}(S_2^M) = \hat{S}_1^{\,M} \approx S_1^M \qquad (1)$$

$$f^{M}_{S_1 \rightarrow S_2}(S_1^M) = \hat{S}_2^{\,M} \approx S_2^M \qquad (2)$$

$$f^{\bar{M}}_{S_2 \rightarrow S_1}(S_2^{\bar{M}}) = \hat{S}_1^{\,\bar{M}} \approx S_1^{\bar{M}} \qquad (3)$$

$$f^{\bar{M}}_{S_1 \rightarrow S_2}(S_1^{\bar{M}}) = \hat{S}_2^{\,\bar{M}} \approx S_2^{\bar{M}} \qquad (4)$$

Once the model parameters are learnt, an unseen example is given the label of the pair of class-specific models that produce the lowest reconstruction error. When new samples are available (for both subjects), the audio and visual features are computed, and are then fed to the models from eq. 1, 2, 3, 4, and 4 error values are produced. We use mean squared error (MSE) to scalarize the vector of reconstruction errors.

$$e^{M}_{S_2 \rightarrow S_1} = MSE(\hat{S_1}^{M}, S_1^{M}) \qquad (5)$$
$$e^{M}_{S_1 \rightarrow S_2} = MSE(\hat{S_2}^{M}, S_2^{M}) \qquad (6)$$
$$e^{\bar{M}}_{S_2 \rightarrow S_1} = MSE(\hat{S_1}^{\bar{M}}, S_1^{\bar{M}}) \qquad (7)$$
$$e^{\bar{M}}_{S_1 \rightarrow S_2} = MSE(\hat{S_2}^{\bar{M}}, S_2^{\bar{M}}) \qquad (8)$$

We then compute a weighted mean of the MSE, for each class, as shown in eq. 9 and 10, where the weights are optimized using grid search during model selection on the validation set.

$$e^{M} = w_M \times e^{M}_{S_2 \rightarrow S_1} + (1 - w_M) \times e^{M}_{S_1 \rightarrow S_2} \qquad (9)$$
$$e^{\bar{M}} = w_{\bar{M}} \times e^{\bar{M}}_{S_2 \rightarrow S_1} + (1 - w_{\bar{M}}) \times e^{\bar{M}}_{S_1 \rightarrow S_2} \qquad (10)$$

A frame is classified as mimicry or non-mimicry depending on which pair of models (corresponding to a particular class) produced the best feature reconstruction, i.e. the pair with the lowest combined reconstruction error:

$$IF\ e^{M} > e^{\bar{M}}\ THEN\ \overline{\mathbf{M}}\ ELSE\ \mathbf{M} \qquad (11)$$

Pre-processing steps: We split our data into training, validation and test sets on a per-session basis, as mimicry behaviours vary considerably between different pairs of subjects. The training set consisted of the first contiguous block of the session such that it contained half of all the mimicry episodes. This was then split into individual sequences and used for training. The contiguous block containing the next quarter of all the mimicry episodes formed the validation set, and the remaining data was used for testing and performance evaluation. Before training, all features are z-normalized (per session) to zero mean and unit standard deviation, and smoothed using a Savitzky-Golay filter of window size 15 and degree 3.

Training: Mimicry and non-mimicry models are trained with sequences from their respective classes only. We use an ensemble of the classifiers detailed above, with lags of {-24,0,24} samples, corresponding to time lags of {-0.5,0,0.5} seconds. As many mimicry episodes were short, models with longer time lags would have had even less training data per session than currently available, due to the need to clip the ends of each training sequence after timeshifting one relative to the other (for example, when using a 150 sample length sequence to train a model with a lag of 58 frames ≡ 1s, clipping the sequence after time-shifting would lose 40% of the data for that sequence). Preliminary experiments also showed that including longer time lags had no meaningful effect on performance, for this model. We define lag relative to subject 1, hence a model with a negative lag implies that it models the relationship between data from subject 2 with earlier data from subject 1. So, for each class we train regression models to predict the audiovisual features at t in stream 2 based on the features at t−24 in stream 1 (and vice versa), and models to predict the audiovisual features at t in stream 1 based on the features at t−24 in stream 2 (and vice versa), as well as models to predict the features at t in stream 1 from t in stream 2 (and vice versa). Models with time-lags suffer from inevitable edge effects (e.g. when training with the first sample in a session, there are no prior samples to train a time-lagged model with); rather than zero-pad the sequence ends, we clip those samples that have no corresponding samples (at the correct time) to train with.

Labelling procedure: After the regression models for each class have produced a reconstruction of their complementary features, the error values from eq. 5 to 8 are smoothed using a Savitzky-Golay filter, with a window size of 29 frames, and degree 5. The reconstruction errors from each pair of regressors are then compared to generate a label prediction as per eq.11. As mentioned above, we use an ensemble of classifiers with different time lags, each of which produces a label for a given sample. Therefore each frame is labelled 3 times. These "votes" are then combined using a majority voting decision rule. The performance measures we use are precision and recall. Note that we are not classifying presegmented sequences, rather we are performing classification on individual frames along the entire length of the sequence. Training of each regression model is performed using presegmented sequences (as they are only trained using data from their respective classes), however labelling of new frames is done continuously, to take advantage of the stateful LSTM model.

Model selection: We trained networks using only one hidden layer. The number of hidden neurons was optimized using a line search across the range [25-75] in steps of 10, where the hidden layer size for networks in both classes was constrained to be equal. Networks were trained using resilient back-propagation, with a training epoch limit of 500. Our method also requires optimization of the weights from eq. 9 and 10, with respect to classification performance. This is performed using a single resolution grid-search in steps of 0.001.

*Table 2: Class-specific precision and recall measures for detection of mimicry of laughter, smiles, and linguistic vocalization*

| Session # | Non-mimicry | | Mimicry | |
|---|---|---|---|---|
| | precision | recall | precision | recall |
| 3 | 93.8 (1.1) | 56.2 (1.4) | 13.8 (1.2) | 65.4 (6.6) |
| 4 | 95.4 (1.5) | 63.4 (1.2) | 22.6 (1.3) | 77.8 (8.0) |
| 5 | 98.3 (1.0) | 56.1 (2.4) | 14.6 (1.5) | 88.4 (6.0) |
| 6 | 98.9 (3.6) | 61.3 (1.2) | 10.5 (6.8) | 86.8 (4.1) |
| 11 | 97.8 (4.8) | 53.5 (4.7) | 6.2 (0.5) | 71.4 (6) |
| 21 | 83.3 (4.2) | 69.9 (3.2) | 52.3 (5.4) | 70.2 (7.7) |
| 32 | 95.2 (0.8) | 63.9 (0.9) | 39.1 (0.7) | 87.7 (2.2) |
| 33 | 98.9 (4.9) | 49.2 (3.6) | 6.3 (0.5) | 84.4 (8.2) |
| 42 | 91.2 (1.9) | 63.3 (1.5) | 47.1 (1.9) | 84.2 (3.4) |
| 44 | 40.6 (2.0) | 63.9 (3.0) | 79 (1.8) | 59.2 (1.8) |

Table 2 shows the experimental results (mean and standard deviation of 5 runs) on 10 full sessions of the MAHNOB Mimicry database. We can see that the performance is highly session dependent, however the models have a bias towards labelling a frame as mimicry, as shown by the generally high positive recall performance. This may be due to the significant class imbalance in the data. Although our method is not directly discriminating between the two class distributions in the feature space, the abundance of non-mimicry data may allow the non-mimicry model to learn a smoother approximation between the two sets of features, allowing better generalization. Even after filtering, the high-frequency noise in the mimicry model error is more prominent than in the non-mimicry model error. This noise seems to cause the false positives when reconstruction error is low for both models, examples of which can be seen in Figure 1. However we can see that our method can successfully detect boundaries between mimicry and non-mimicry in some cases, as in Figure 1.

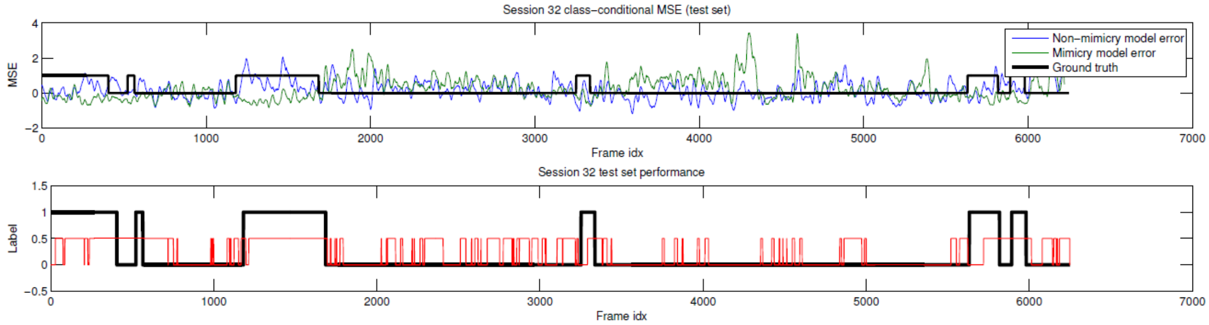


Figure 1: Model error and subsequent frame classification on session 32’s test set.

**References**

*[1] S. Bilakhia, S. Petridis, A. Nijholt, M. Pantic. Pattern Recognition Letters, The MAHNOB Mimicry Database - a database of naturalistic human interactions, pp. 52 - 61. November 2015.*
*[2] J. Orozco, O. Rudovic, J. Gonzlez, and M. Pantic, "Hierarchical on-line appearance-based tracking for 3D head pose, eyebrows, lips, eyelids and irises," Image and Vision Computing, February 2013.*
*[3] S. Petridis, M. Pantic , Prediction-based Audiovisual Fusion for Classification of Non-linguistic Vocalisations, IEEE Transactions on Affective Computing, December 2015 2015.*

**Task 6.2: Behaviour-similarity-based approach to detection of temporal behaviour-patterning**

Work on Task 6.2 will start in M27.

**Task 6.3: Automatic assessment of interpersonal sentiment**

Work on Task 6.2 will start in M33.

### *1.2.7. Work Package 7 (WP7) - Integration, Applications and Evaluation*

**Tasks 7.2-7.5: Chat Social Game by Playgen**

PlayGen has further refined the integrated application design with specific attention focused on the mapping between the capabilities of the SEWA emotion analysis and the interview skills trainer game. This includes identification and application of specific high level feature extractions afforded by the SEWA analytics algorithms to provide specific feedback for likeability, interest, happiness, sincerity, calmness dominance, stress and anger.

The initial version of the application has been developed using a combination of Unity game engine together with components from the H2020 RAGE project for social gamification and the OpenTok video streaming and messaging platform. PlayGen carried out extensive user-testing of the initial version of the game with the target audience, both onsite at the PlayGen offices and remotely at end-users own locations.

The appraisal and evaluation of V1 of the application lead to the conclusion that the WebGL platform is too unstable for a commercial deployment of the game, hence the second version will be developed from the ground up using HTML5.  Another potential bottleneck maybe reliance on Websocket technologies which appear to be blocked in some educational establishments, therefore alternative network transport mechanics including WebRTC are being evaluated.

**Tasks 7.2, 7.4, 7.5: Advert Recommender Engine by Realeyes**

RealEyes has focused on several major activities. First, they identified the similarity metrics of adverts and users. Second, they proved that sentiment-based recommender engine can work in real-life environment. Third, they continued building out further connections with different industrial players who could benefit from the SEWA results. In particular, they:

- Strengthened collaboration with key industrial partners in this project – Mars and Marketcast.
- This resulted in an increase of available advert performance datasets, which are being used to develop the core of the recommendation engine, both in the form of sales and social media performance.
- Interest and relevance of the developed service was further backed by one of the industrial partners with an additional investment into the project to enable RealEyes collect more behavioural data.
- Initiated collaboration with additional potential future partners to help define target groups and clarify their needs, such as eBay and Analect.
- Implemented the first version of similarity measurements among adverts and users.
- Confirmed that emotion profile based ad recommendation is feasible.
- Designed a scoring system that allows for an objective and useful measure of the matching quality.
- Designed and implemented a baseline model that will serve as a reference for the next development stage that will integrate advanced sentiment analysis tools provided by our partners in SEWA.
- Tested out opportunities in audio based behaviour signal processing through collaboration with industrial and academic partners and identified main challenges in this direction that would need to be tackled.
- Further progressed core technical functionality by improving the baseline model and set of predictive signals used and advanced technical integration discussions.
- Together with Imperial College London lead organization of the 2nd Valorisation Board meeting.

### *1.2.8. Work Package 8 (WP8) – Dissemination, Ethics, Communication and Exploitation*

Much effort has been put on dissemination of the SEWA project and the technologies and applications developed within the project. This includes the mass media (newspapers, journals, social media) and academic media (journals, conference contributions, organization of challenges). Multimodal emotion recognition is present on all major conferences in the field (such as ACM MM, INTERSPEECH, & ACII).

**Task 8.1: SEWA website and e-services**

The website has been updated regularly. The new webmaster is Jean Kossaifi, a PhD student at Imperial College London.

**Task 8.2: Valorisation Advisory Board**

The Valorisation board meeting took place on Friday 23rd September 2016 at The Gore Hotel in South Kensington, London. The Valorisation Board member who participated were: Elissa Moses (IPSOS), Simon Hughes (Jobatar), Jim Hodgkins (VisualDNA), Paul Martin (Xaxis), Zoe Ilic (FT), Bjorn Schuller (audEERING).

The agenda of the meeting was as follows:

- Brief introduction of all present.
- Summary presentation of all currently used audio-visual tracking capabilities developed by Imperial/Passau with a short peak into the future Presentation/demo of the Social Chat Game by Playgen.
- Presentation/demo of Realeyes' ad recommendation engine.
- Project Ethics, Communication, Dissemination and Exploitation plans.
- Open panel discussion with Valorisation Board members, chaired by Realeyes.

The final suggestions and summary of the valorisation board, following a series of presentations, demonstrations and discussions on crucial issues, are:

1. Overall the Valorisation Board was very impressed by the progress and content of the panel presentations. It is felt that there are impressive advancements in practical innovation which will have value to the marketplace and generate successful business applications.
2. "Let's Go!" At this point in time, the Valorisation Board recommends taking some of the ideas that have been developed to the next level quickly. It is unknown how long

these newly created competitive advantages will be without competition, so action should be taken to pre-empt the market.

3. There are solid “proofs of concept” that should be shared with the EU short term, industry and press. We need to mark our progress and generate business potential. The next step is to define 2 technologies that are priority for advancing to industry. The Advisory Board will help to guide the 1-2 available applications and provide counsel on how to proceed to market.
4. Ideally one of the new applications should have a broad humanitarian quality such as a health care or health diagnostic applications. The Board challenged the group to consider and innovate against new humanitarian applications to make a contribution to society.
5. There was a discussion on how best to publicize the output of the SEWA project. Used properly, publicity can help propel the adoption and success of SEWA developed technology. However, careful effort should be given to planning a media strategy once the priority technologies are selected for messaging. Special regard should be given for getting the message right, not inviting any controversy and leveraging the advancements in the best light.
6. More data appears to be needed for advancing a world class facial expression tracking tools. It is very important that sales data can be attributed to emotions and advisors could help with further developing this attribution. We need to define what it is, how much do we need and why, how might we be able to acquire the training data that is required to move forward. Ipsos might be able to help provide at least some of the desired data. Other data will need to be generated by Realeyes and the project team.
7. Data are also needed for the training initiative and we are encouraged to use creative thinking for finding ways on how to obtain data from various sources (3rd party databases), novel annotation techniques, smart annotation games, other approaches. The ultimate question we should spend some time thinking about is how can we come up with a solution for the training data challenge that solves the problem in a breakthrough way.
8. The project teams should stay aware of other companies doing business in the US, Israel, etc. around the world in a similar space. Our technology developments need to be understood in the context of other potential competitors so that we stay ahead with competitive advantages.

9. Miscellaneous general feedback: build on discoveries, make more use of the Valorisation Board, recruitment, intern market, hospitality, colleges, etc.
10. In order to keep momentum, industrial SEWA partners are encouraged to have calls in-between the annual meetings with the Valorisation Board partners when it is beneficial for joint development of SEWA products. For example, IPSOS is planning to meet with Realeyes to discuss the possibilities of collection of Chinese facial expressions data to be used to improve the Ad Recommendation Engine.

**Task 8.3: R&D output publications and conference participation - Dissemination**

In M3, an overall dissemination plan of SEWA project was delivered in D8.1 and it specifies our dissemination strategy in detail.

The full list of R&D output publications is given in Part A, section 6, of this report.

We have organised the following challenges and workshops:

- ComParE (Computational Paralinguistics challengE) at INTERSPEECH Conference 2016, San Francisco, CA, USA (http://emotion-research.net/sigs/speech-sig/is16-compare)
- AV+EC challenge at ACM MM 2016, Amsterdam, The Netherlands (http://sspnet.eu/avec2016/)
- CBAR 2016 workshop (Context-based Affect Recognition workshop) at IEEE CVPR 2016, Las Vegas, Nevada (http://cbar2016.blogspot.co.uk)

In addition, SEWA partners had a number of Keynote speeches:

- **Keynote** by the SEWA coordinator at Int'l Conf. Image Processing Theory, Tools and Applications (IPTA'16), Oulu, Finland, 15.12.2016
- Björn Schuller: **Keynote** “Engage to Empower: Emotionally Intelligent Computer Games & Robots for Autistic Children”, Conference on “The world innovations combining medicine, engineering and technology in autism diagnosis and therapy”, SOLIS RADIUS, Rzeszow, Poland, 29.09.2016.
- Björn Schuller: **Keynote** “Computational Paralinguistics in Everyday Environments”, The 4th International Workshop on Speech Processing in Everyday Environments (CHiME 2016 Workshop), San Francisco, CA, 13.09.2016.
- Björn Schuller: **Keynote** “7 Essential Principles to Make Multimodal Sentiment Analysis Work in the Wild”, 4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016), IJCAI 2016 Workshop, IJCAI/AAAI, New York, NY, 10.07.2016.
- Björn Schuller: **Keynote** “Say no more – the computer already deeply knows you?“, SWS 2016 Speech Signal Processing Workshop, ACL/ACLCLP, National Taiwan University, Taipei, Taiwan, 18.03.2016.

Furthermore, SEWA partners have a number of invitation for Keynote speeches:

- Björn Schuller: **Keynote** “Big Data, Deep Learning – At the Edge of X-Ray Speaker Analysis”, 19th Conference on Speech and Computer (SPECOM 2017), Hatfield, UK, 12.-16.09.2017.

- Björn Schuller: Opening Plenary "Artificial Emotional Intelligence – A Game Changer for AI and Society?", Annual Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB), AISB, Bath, UK, 19.-21.04.2017.
- Björn Schuller: **Keynote** "Reading the Author: A Holistic Approach on Assessing What is in one's Words", 18th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), Springer, Budapest, Hungary, 17.-23.04.2017.

### Task 8.4: Management of interactions with other EU projects

The SEWA partners are currently working on and collaborating with several EU-funded projects in the field of emotion recognition and intelligent behaviour analysis.

- In FP7 ERC Starting Grant iHEARu (www.ihearu.eu), methods and tools for holistic analysis of real-life speaker characteristics are developed. One major output of this research is the crowdsourcing game iHEARu-play. The goal of this software, which can be run on several software and hardware platforms, is to make the process of data annotation (audio, video & text) more pleasant for the raters. We are planning to enrich the SEWA database with further innovative labels than the predefined standard labels using this tool to reach a broader research community.
- In FP7 TERESA project (teresaproject.eu), an important aim was to achieve vision-based detection of face and facial landmarks in unconstrained indoor environments. The first versions of the Chehra facial landmark tracker, selected for further use in SEWA (see section 6.3), were originally developed for the TERESA project.
- In FP7 Marie Curie IEF ConfER project, automatic estimation of conflict escalation and resolution in human-human interactions was investigated. The output of this project, in the area of spatio-temporal alignment of two behaviour episodes (time series), is now being used in WP5. Pilot studies in this direction are now underway.
- In H2020 MixedEmotions project on Social Semantic Emotion Analysis for Innovative Multilingual Big Data Analytics Markets, the synergies are mainly in the emotion recognition from multilingual audio content.
- In H2020 SpeechXRays project, Realeyes' role is to develop face tracking, attention, sentiment and affect analysis capabilities and test the use of these capabilities in multi-channel biometric application. Hence, R&D in the area of sentiment and affect analysis and tracking in SEWA project is of interest also in SpeechXRays.

### Task 8.5: Engagement with the public

There have been a number of press releases in 2016, which have been making the general public aware of the need and the essence of the technology we aim to bring together with SEWA:

- The article ‘The gofer in the machine’ featured in bi-monthly cultural magazine ‘1843 (formerly Intelligent Life)’ published by prestigious ‘The Economist Group’ summarizes the evolution of the technology from the most simplistic telephones, to the smartphones with touch-screen interfaces, to today’s spoken interfaces, envisioning what all this means to our future. The view shared by UP PI Björn Schuller has been a key contributor in making these projections, the verbatim amounting to nearly one third of the article. The UP PI points out why making these system emotionally aware is the next logical step in the evolution of virtual assistants, advocating and citing the reasons for the same from different angles (technological, psychological and consequently, the competitive market point of view).
- The article ‘Machines That Talk to Us May Soon Sense Our Feelings, Too’ published in the popular science magazine ‘Scientific American’ features interview of UP PI Björn Schuller produced in conjunction with the World Economic Forum. In the interview, he explains how the technologies like Siri, Cortana work, the lack of paralinguistic dimensions to these products and its need, and how he envisions the future, the next game-changing breakthroughs to be.

Efforts towards both General Public Dissemination and Industrial Dissemination have been intensified resulting in multiple TV coverage of the work done in SEWA and public speeches on the results of the SEWA project.

TV and Radio coverage:

- **ITV News at 22:00**, 21-Dec-2016 (www.facebook.com/maja.pantic.758/videos/1165111813604730/)
- **TV Program** UNI Global Union TV, "The Future World of Work", 15-Nov-2016 (https://www.youtube.com/watch?v=VeUaM_GHO1Q&list=PLDzsZ-wrSPx_xOFNifCupS7zDpBjPym8H&index=3)
- **CBS 60 Minutes**, 9-Oct-2016 (www.cbsnews.com/news/60-minutes-artificial-intelligence-charlie-rose-robot-sophia/)
- **BBC World News,** 1-Jul-2016 (https://www.youtube.com/watch?v=Ddtf9kn0hv0&feature=youtu.be)
- **Inforadio, Rundfunk Germany ,** 25-Nov-2016**,** Radio interview broadcast 7:10-7:20 AM, “Verhältnis Mensch-Maschine im Jahr 2050”**,** (http://www.inforadio.de/dossier/2016/grossprojekte/das-vernetzte-ich-ii/woche-2/ersetzt-die-maschine-den-menschen/ersetzt-die-maschine-den-menschen.html)

Public Speeches and Events:

- **PechaKucha Talk** by SEWA coordinator at Global Innovation Summit, London, UK, November 2016
- **Talk** by SEWA coordinator at the UNI Global Union Summit, Nyon, Switzerland, November 2016
- **Inspirational Talks** by both the SEWA coordinator and the PI for RealEyes, Dr Elnar Hajiyev, at the Science Festival: State of Emotion, Berlin, Germany, November 2016
- **TEDx** Talk by the SEWA coordinator at European Commission's Digital Assembly 2016, Bratislava, Slovakia, September 2016
- **Royal Society's event:** "Driverless Cars – ask the experts", the SEWA coordinator was one of the three panellists at the event, London, UK, July 2016
- **Royal Society's Summer Science Exhibition**, Talk by the SEWA coordinator on Artificial Emotional Intelligence, London, UK, July 2016
- **Royal Society's event:** "Learning Machines & How Computers Got Smart", the SEWA coordinator was one of the three panellists at the event, London, UK, April 2016

RealEyes has won the **Innovation Radar Prize 2016** and has engaged in a series of public talks promoting the SEWA results, including the Science Festival in Berlin and a large number of events (12 in total), listed in section 4.

### Task 8.6: Data Management Plan and dissemination of software and datasets

- The SEWA Database is released at the following link: http://db.sewaproject.eu/ for research purposes. The SEWA database includes annotations of the recordings in terms of facial landmarks, facial action unit (FAU) intensities, various vocalisations, verbal cues, mirroring, and rapport, continuously valued valence, arousal, liking, and prototypic examples (templates) of (dis)liking and sentiment. The data has been annotated in an iterative fashion, starting with a sufficient amount of examples to be annotated in a semi-automated manner and used to train various feature extraction algorithms developed in SEWA, and ending with a large DB of annotated facial behaviour recorded in the wild.
- A toolkit called 'openXBOW - the Passau Open-Source Multimodal Bag-of-Words Toolkit' implemented in Java has now been made publicly available at the following link: https://github.com/openXBOW/openXBOW. It generates a bag-of-words representation from a sequence of numeric and/or textual features, e.g., acoustic LLDs, visual features and transcriptions of natural speech. The tool provides a multitude of options, e.g., different modes of vector quantisation, codebook generation, term frequency weighting and methods known from natural language processing.
- The code for 'The SEWA AU Detector tool' has been released at the SEWA official website: http://sewaproject.eu/resources .This application requires the 49 fiducial facial

points extracted using the SEWA facial point tracker as input. These are passed through several blocks for data pre-processing, including normalization, alignment and dimensionality reduction. The output is the classification of each target frame in terms of target AUs being active/non-active.

- A Tensor Learning toolbox called TensorLy has been released and is available at https://github.com/tensorly/tensorly. TensorLy is a state of the art general purpose library for tensor learning. Written in Python, it aims at following the same standards adopted by the main projects of the Python scientific community and fully integrating with these. It allows for fast and straightforward tensor decomposition and learning and comes with exhaustive tests, thorough documentation and minimal dependencies. It can be easily extended and its BSD licence makes it suitable for both academic and commercial applications.
- AFEW-VA is a new dataset collected in-the-wild composed of 600 challenging video clips extracted from feature film, along with highly accurate per-frame annotations of valence and arousal. Added to these are per-frame annotations of 68 facial landmarks. The dataset is made publicly available and released along with baseline and stat-of-the-art experiment as well as a thorough comparison of features demonstrating the need of such in-the-wild data. (https://ibug.doc.ic.ac.uk/resources/afew-va-database/)
- The Conflict Escalation Resolution (CONFER) Database is a collection of excerpts from audio–visual recordings of televised political debates where conflicts naturally arise, and as such, it is suitable for the investigation of conflict behaviour as well as other social attitudes and behaviours. The database contains 142 min of naturalistic, 'in-the-wild' conversations and is the first of its kind to have been annotated in terms of continuous (real-valued) conflict intensity on a frame-by-frame basis. The release of CONFER Database is accompanied by the first systematic study on continuous estimation of conflict intensity, where various audio and visual features and classifiers are examined (http://www.sciencedirect.com/science/article/pii/S0262885616302190).

**Task 8.7: Ethical Advisory Board**

SEWA involves recording and storing of data from adult volunteers, and then releasing them to the scientific community to facilitate investigations on the topic within and beyond the project. In order that the project remains compliant with ethical principles and applicable EU and national law, SEWA consortium arranged for an Ethical Advisory Board, which consists of two experts in the field of ethics that concern the SEWA project.

The members of the Ethical Advisory Board are Prof. Laurence Devillers of the Paris-Sorbonne IV University in France and Prof. Jean-Gabriel Ganascia of the University Pierre et Marie Curie in France. The Ethical Advisory Board meets at most once a year with the PMC. The first meeting was held in conjunction with the SEWA kick-off meeting on 12-13 February 2015, in London, UK.

The recommendations made by the Ethical Advisory Board concerned all:

1. Data Collection
2. Privacy and security
3. Dual Use
4. Payment and compensation
5. Identifying, excluding, and reporting potentially illegal material
6. Access to data by third parties
7. Minimizing potential misuse of the data or findings to stigmatize any groups or communities
8. Declaration on what SEWA project does NOT involve
9. The consent form.

The recommendations made by the Ethical Advisory Board have been discussed by the PMC, adopted by the project, and are forwarded to the Commission as part of deliverable D8.2. The Ethical Advisory Board will be consulted in all ethical issues as they arise in the course of the work in the various research lines.

**Task 8.8: Organisation of challenges and benchmarking**

A number of workshops and challenges have been organized in the year 2016 to best encourage the research community to improve and compare the performance of methods for automatic prediction of sentiment, arousal, and valence.

- The Sixth Audio/Visual Emotion Challenge and Workshop-AVEC 2016 (a satellite workshop of ACM-Multimedia 2016) was organized in Amsterdam, Netherlands.
- INTERSPEECH 2016 featured Computational Paralinguistics Challenge (ComParE) in San Francisco, USA.

**Task 8.9: Exploitation plans and industrial dissemination**

PlayGen further developed its business strategy with respect to exploitation of the project results. The strategy as outlined in further detail in the updated version of D8.1 seeks to establish the application as a ground-breaking new product the uses advanced emotional analysis offered by SEWA to substantially deliver on the promise of *increasing the chances of getting hired for a job through better interview skills.* Once the value of the application is established and demonstrated, through testing of V2 of the application, the next part of the strategy is to scale the business in a number of directions including, a self-service systems and a white label version that can be branded by direct clients such as university career groups and organisation operating in the apprenticeship and recruitment in general. Additionally, partners are being sought for the launch of the launch of V3 of the application.

PlayGen produced and distributed a short video to be used for dissemination of the existing SEWA application, this can be found on the SEWA website and directly at: https://www.youtube.com/watch?v=fM03h13J3mM . The video has been viewed over 5000 times thus far. Going forward, once the integrated version of the application is delivered in 2017, new marketing material will be developed for promotion, together with targeted events including specialist Career Guidance, Apprenticeship and Recruitment exhibitions.

Exploitation and dissemination plan of Realeyes with regard to ad recommendation engine is detailed out in the revised version of D8.1 report. It focuses around enhancement of existing and development of new Realeyes products, targeting of 3 major groups of clients (Agencies, Brands, and Media) and promotion of its services through industrial conferences.

While in its essence the plan has not changed, there are additional updates and clarifications to be mentioned. Enhancement of an existing Realeyes Creative Testing product with the predictive modelling capabilities has already started and this functionality is already being trialled with a number of selected clients, such as Mars and Marketcast. Feedback we have received so far is encouraging and we are progressing in this direction. Additionally, we are in the process of creating a plan for the development of an entirely new product line together with IPSOS that will also benefit from the recommendation properties of predictive modelling developed for ad recommendation engine in SEWA. The plan is to launch the 1st version of the product already in Q3 2016. Both products assume manual recommendations of video ads to users based on the results of predictive modelling. Complete automation and integration of a fully automated ad recommendation engine into programmatic advertising networks proved to be a more challenging task. Our plan is to first identify main opportunities for manual recommendation using various predictive modelling approaches with selected clients, then leverage the learnings towards automation of the recommendations in the programmatic advertising context.

PlayGen carried out a large number of bi-lateral meetings for exploitation of the interview skills game. The meetings included the following organisation as potential exploitation partnership deals: EasyRecrue, Lumesse, Shine, Meet & Engage, Predictive Hire, Cingo Recruitment, Social Talent, Hire Serve, Jobvite, Cut-E and Launchpad Recuit. The application and it's description was met with universal interest, however given that version 1 shown did not incorporate the SEWA analytics yet, the vendors and potential partners are awaiting the 2nd version of the application which will show the real differential and the true value of the SEWA platform.

Realeyes represented SEWA project on ICT Proposer day and won The Innovation Radar Prize in category of Horizon 2020 ICT innovator.
Blog posts on Realeyes website:

http://www.realeyesit.com/blog/detecting-more-complex-affective-states

http://www.realeyesit.com/blog/innovation-radar-award

http://www.sewaproject.eu/files/REALEYES.pdf

In addition to normally planned presence at industrial conferences, winning ICT Innovation Award allowed RealEyes to secure a booth at the upcoming CeBIT 2017 exhibition. We plan to use this opportunity to promote goals of SEWA project and ad recommendation engine within Marketing solutions sections of the fair.

Realeyes co-presented with Mars and Ehrenberg-Bass Institute for Marketing Science results of development of predictive model for ad recommendation engine in SEWA: https://www.realeyesit.com/i-com-global-forum-for-marketing-data-and-measurement

Realeyes held an industrial talk presentation on challenges of Scalable Emotion Intelligence at Spark Summit Europe 2016 in Brussels.

https://www.realeyesit.com/spark-summit-europe-2016

http://go.databricks.com/videos/spark-summit-eu-2016/scalable-emotion-intelligence-realeyes

Realeyes CEO gave a talk on how emotion recognition technology developed at Realeyes and in SEWA is becoming more sophisticated and can be used to generate real-time data to help marketers:

https://www.lsnglobal.com/seed/article/18997/backlash-culture-network-evening-mihkel-jaeaetma

Realeyes created promotional video content for SEWA project and leveraged it to raise awareness of the project and technology that it set as a goal to develop:

https://www.realeyesit.com/sewa

A list of additional industrial-dissemination-oriented events is given in section 4 of this report.

### 1.2.9 Work Package 9 (WP9) - Project co-ordination and management

**Task 9.1: Coordination of the consortium's activities**

This task comprised the management related communications and meetings. The Coordinator and the management team were in close contact with the European Commission's representatives which guaranteed a productive communication flow at all levels of the project. During the reporting period the consortium met on a regular basis by monthly conference telephone calls as well as face-to-face meetings in parallel to the plenary, review and valorisation meetings.

The project coordinator and project manager organized two (2) plenary meetings, ten (10) phone meetings, one (1) Valorisation Advisory Board meeting and one (1) review meeting. The project manager also assisted with organization of sub-team meetings.

**Task 9.2: Quality control and work plan monitoring**

The main objective of the project management was to carry out and to guarantee the effective coordination and management of the project: focusing especially on the day-to-day administration, coordination, and monitoring of the project's progress. The task implies the coordination of the overall project and related activities. The achievements of the project´s objectives, such as the deliverables, milestones, and periodic reports including their timely provision, was constantly reviewed and traced.

For each deliverable, the following process is adopted: (i) the WP leader prepares a draft of the deliverable with all partners working on the WP, (ii) two partners who did not work on the deliverable comment on the draft, (iii) the project coordinator provides final comments on the deliverable, and (iv) the project manager uploads the deliverable.

The project manager reviewed M13 to M24 financial transactions and usage of man months.

**Task 9.3: Reporting to the European Commission**

The Coordinator and his team were responsible for timely collection, review, consolidation, and preparation of the Annual report 2 (M13 – M24) according to the provisions of the Grant Agreement.

The status of the deliverables and milestones were constantly monitored by the Coordinator and the management team. During the second reporting period (M13 – M24) 10 deliverables in total were submitted to the European Commission via the ECAS system. The full list is below:

| | |
|---|---|
| D2.2 | Robust Visual Feature Extractor |
| D2.3 | Improved acoustic-linguistic feature extractor |
| D3.1 | Component / Demonstrator for mid-level visual features extraction |
| D1.1 | SEWA Database |
| D7.2 | Initial version of the Ad Recommendation Engine |
| D7.3 | Initial Version of SEWA Chat Game |
| D3.2 | Audio-visual detector of nonverbal vocalisations |
| D4.1 | Multi-modal affect recognizer |
| D5.1 | Visual behaviour similarity estimator |
| D9.2 | Annual Report 2 |

**Task 9.4: Legal and Contractual management**

ICL’s dedicated EU team and the project management team requested an amendment to the agreement due to an unforeseen linked third party member to join the SEWA project.

The project coordinator administrated the drafting of exploitation agreement, ensuring that partners have access to A/V sentiment analysis systems and other foreground produced during the project in a way that ensures the graceful execution of their business and commercialization plan.

## 1.3 Impact

**WP1**

As the one of the largest dataset of its kind, the SEWA database will be extremely valuable to researcher working in the field of automatic human behavioural analysis and user-centric HCI and FF-HCI. SEWA DB will be used for a number of challenges and benchmarking efforts and will have more than 200 active users worldwide by the end of the project.

**WP2**

The openXBOW toolkit has been published on GitHub and has already earned some attention from the community. It is supposed to have a high impact as it can be used for any audio/video/language recognition tasks and is generic in that sense.

**WP3**

The VSL-CRF is part of the DOC-Toolbox (https://github.com/RWalecki/DOC-Toolbox). This Toolbox contains different sequence classification methods including Conditional Ordinal Random Fields (CORF), Hidden Conditional Ordinal Random Fields (HCORF), Conditional Random Fields (CRF), Hidden Conditional Random Fields (HCRF) and the Variable State Latent Conditional Random Fields (VSL-CRF) that was developed for the mid-level feature extraction. It is easy to use and different stat-of-the-art methods can be quickly trained and evaluated. Many research groups and institutes are currently using the toolbox for tasks like facial expression recognition, AU-detection and gesture recognition. Further impact in the scientific community is ensured by maintaining and the toolbox keeping it up to date.

**WP4**

Crossmodal bag-of-words representations have a huge potential of becoming a state-of-the-art method for emotion recognition, as shown in deliverables D2.1 & D2.3 (D4.1). Moreover, they can be applied to many other tasks, such as scene classification and multimedia mining, and will thus have an impact to a broad scientific and industrial community.

End-to-end learning on the raw signal, which is a field of growing interest, has been applied successfully for emotion recognition in the SEWA project for the first time ([Trigeorgis et al., 2016] - honourable mention at ICASSP).

**WP5**

WP5 has a profound impact on the advancement of the state of the art in natural, multimodal human-computer interfaces. That is, we produced technology for behaviour similarity measurement, based on a fully unsupervised learning approach. This technology will answer the question “are these two multimodal inputs similar?” instead of “what is the conveyed meaning of the displayed behaviour?” The developed technology could represent the solution to the long-standing problem in machine analysis of human behaviour – the lack of annotated data to learn from. To wit, in the behaviour-similarity-matching paradigm, minimal annotation of training data is needed; it is only required to pinpoint “typical” example(s) of the target behaviour and “templates” of the target behaviour are compared to the currently observed behaviour for similarity measurement.

**WP6**

Work on WP6 aims to develop tools for both automatic analysis of interpersonal behaviour matching, that is, temporal patterning of the facial, vocal and verbal behaviour shown by two interacting people (like in mirroring / mimicry), and automatic assessment of the interpersonal sentiment based on the presence, frequency and duration of mirroring. SEWA will enhance our understanding of interaction patterns and will facilitate more natural and smooth computer-mediated dyadic interactions. Finally, new algorithms are being developed which aim to match similar behaviour without explicit knowledge of the type of behaviour. This is an important achievement since it will minimize the data annotation effort which is one of the main bottlenecks in machine learning research.

**WP7**

RealEyes in WP7 is working on an Ad Recommendation Engine based on emotional profile. This is in an early stage, but the partners have high interest in that project. RealEyes is working on finding a partner, who can use the prototype to validate the approach and prove better recommendations can be delivered by using emotional data. As the plan is to launch a product based on this research, it will have big impact on the business and in revenue as well. Due to nomination for Innovation Radar Prize, an increased interest in the company’s website has been observed.

As far as PlayGen is concerned, the estimated impact will be visible/ countable only after the incorporation of SEWA technologies into a new product which will happen in the next reporting period.

**WP 8**

Much effort has been put on dissemination of the SEWA project and the technologies and applications developed within the project. This includes the mass media (newspapers, journals, social media), academic media (journals, conference contributions, organization of challenges) and industry fairs and contacts. The SEWA tools and database have also been released and that produces a further deep and long-lasting impact of the work done in the SEWA project.

## 2. Update of the plan for exploitation and dissemination of result

There has been no deviation from the exploitation and dissemination plan (revised deliverable D8.1) submitted in M18.

## 3. Update of the data management plan

There has been no deviation from the data management plan (deliverable D8.2) submitted in M6.

## 4. Follow-up of recommendations and comments from previous review

Following the project review meeting in May and the letter from the Project Officer on the 1st of June 2016 titled*: Result of the Review of your H2020 project 645094 — SEWA*, the Commission recommendations to be implemented are as follows:

**Recommendation 1:** Prepare and deliver within six months from the reception of this report a draft exploitation agreement, ensuring that partners (notably RealEyes, PlayGen) have access to A/V sentiment analysis systems and other foreground produced during the project in a way that ensures the graceful execution of their business and commercialization plans. The agreement should be prepared jointly by the partners in order to ensure that IPR issues will not be a set-back to the industrial exploitation of the project' results.

**Recommendation 2:** Provide updated versions of deliverables D2.1, D2.2, D2.3 in order to reflect how technology development choices have been driven by industry user/application requirements, but also how developed systems serve the intended innovation. Minor updates to the existing documents will be required, mainly in terms of presenting a matching of requirements to technology development choices and results. Please provide the updated documents within two months after the reception of this report.

**Recommendation 3:** Update the dissemination plan (D8.1) in order to include planned activities towards the industry, including activities that will boost the commercialization plans

of the partners. Please be factual and realistic, by planning for targets that boost exploitation, while being achievable within the scope of the project. Include actions towards target customers. Provide the updated documents within two months after the reception of this report.
**Recommendation 4:** For the next reporting period, ensure a tighter integration between WP7 and technology workpackages in progress (WP3, WP4, WP5, WP6) making sure that technology development work is user-driven and focused on industrial requirements. Please make sure that a technical architecture for the innovative applications (in WP7) is produced and used to drive technology integration. Likewise taken into account suggestions and risks outlined by the Valorisation Board in the technical developments. Explore the impact of these suggestions on the project's risk management approach.
**Recommendation 5:** Complement excellent dissemination work towards the research community with activities that boost the commercial exploitation of the project's results. Main activities should be underlined in the updated version of the dissemination plan D8.1. Ensure a more active presence in social media (e.g., towards communicating the project's achievements and attracting followers from the sentiment analysis, facial analysis and acoustic analysis communities).

**Actions w.r.t. Recommendation 1**: Letters of understanding between the partners have been made as a further explanation as to how we intend to achieve a successful transfer of foreground knowledge developed during the project and facilitate the companies to achieve their part in the project and have a successful commercialisation of their work. As requested by the Project Officer, the letters are appended to this report (Appendix 1).

**Actions w.r.t. Recommendation 2**: Deliverables D2.1, D2.2, D2.3 have been reviewed to include a list of requirements driving the development of the components, based on work already undertaken in WP7. All three revised deliverables were submitted within the recommended period.

**Actions w.r.t. Recommendation 3**: The dissemination Plan was updated to include a list of dissemination activities towards industry, including potential customers of RealEyes and PlayGen regarding the innovative applications of the project. The revised Deliverable D8.1 was submitted within the recommended period.

**Actions w.r.t. Recommendation 4**: The valorisation meeting on 23rd of September 2016, was organised in order to get further advice on design and architecture of SEWA applications and to investigate commercialization potential of each aspect of the SEWA technology. The agenda of the meeting focused on Dissemination and Exploitation plans and therefore on the commercial exploitation of the project' results. Overall the Valorisation Board was very impressed by the progress and content of the panel presentations and it was felt that there are impressive advancements in practical innovation which will have value to the marketplace and generate successful business applications. All suggestions and risks outlined by the Valorisation Board have been taken into account in the technical developments.

**Actions w.r.t. Recommendation 5**: The updated Overall Dissemination Plan is underlined in the revised Deliverable D8.1, which was submitted within the recommended period. The revised Dissemination plan includes new activities towards industry in order to boost commercialisation plans and a list of upcoming relevant events that provide excellent opportunities to meet with new clients and potential partners. Furthermore, a Database paper is being written in order to be submitted to "IEEE Transactions on Pattern Analysis and Machine Intelligence Special Issue: The Computational Face (TPAMI)" in February 2017.
List of events attended during the reporting period:

| | | | |
|---|---|---|---|
| London Job Show | 2016 Sep 30th | Showcases some of the very best employment and training opportunities from the region's most respected International, National, Regional and Local Employers. | http://www.mkjobshow.co.uk/ |
| Recruitment Leaders Connect | 2016 Oct 20th | Recruitment Leaders Connect is the largest recruitment industry event series in the UK. | http://recruitmentleadersconnect.com/ |
| Reconverse | 2016 Dec 1st | How Recruitment Technology Can Improve Your Hires | https://reconverse.com/ |
| Skill 2016 | 2016 Dec 9/10th | Aimed at 15-24 year olds, this two day event will provide young people and their families with a rare chance to discover careers through interactive, inspirational activities and demonstrations | http://www.skillslondon2016.co.uk/ |
| ClickZ Shift | 2016 Aug 30/31st | Digital Marketing Conference - Inspiring and educating over 250,000 digital marketers and leaders for 18 years across North America, Europe & Asia. | https://www.clickzlive.com/ |
| Dmexco | 2016 Sept | Digital Marketing Exposition and | http://dmexco.de/ |

| | | | |
|---|---|---|---|
| | 14/15th | Conference in Cologne, Germany. Almost 43000 attendees are expected to join in this Trade Show, Fair and Exhibition. | |
| Advertising Week NY | 2016 Sept 26/30th | For one week, from September 26th to 30th, the brightest leaders from the marketing and entertainment industry join together in New York to share their visions, passions, and practices at Advertising Week. | http://newyork.advertisingweek.com/ |
| Festival of Marketing | 2016 Oct 5/6th | Festival of Marketing is an event with more than 200 speakers, workshops, awards, experience rooms and training, it is the only event that truly reflects the creative, strategic and tactical job that marketers do. | http://www.festivalofmarketing.com/ |
| Turnaround management Association UK | 2016 Oct 12th | The TMA UK brings together professionals from across the UK, Europe and worldwide to meet, network and hear the latest news within business recovery, corporate turnaround and restructuring. | http://www.tma-uk.org/ |
| IAB Engage | 2016 Oct 12/13th | The Internet Advertising Bureau (IAB) is the trade association for online and mobile advertising. It promotes growth and best practice for advertisers, agencies and media owners. | http://www.iabuk.net/ |
| Wired 2016 | 2016 Nov 3/4th | Wired is a conference event that joins the innovators, inventors and entrepreneurs defining the future, as they explore the big trends shaping tomorrow. | http://www.wiredevent.co.uk/wired2016 |
| Ad:tech | 2016 Nov 2/3rd | An exclusive gathering of 1000 elite brand, agency and media players in the heart of Shoreditch. The hottest start-ups pitching to become Unilever's Next Big Thing. A new content and networking-fuelled format dedicated to performance marketing & tech innovation. | http://ad-techlondon.co.uk/ |

## 5. Deviations from Annex 1 and Annex 2

### 5.1 Tasks

There were no deviations on tasks during this reporting period.

However, all tasks either involving machine learning (i.e., WP3, WP4, WP5, WP6) or involving data collection and data annotation (WP1), required a greater effort than what has been estimated at the project proposal writing. Hence, in all WPs we have an **"overspent" in terms of PMs**, though we have **stayed within the budget** and have enough budget to complete the project as originally envisioned (see section 5.2).

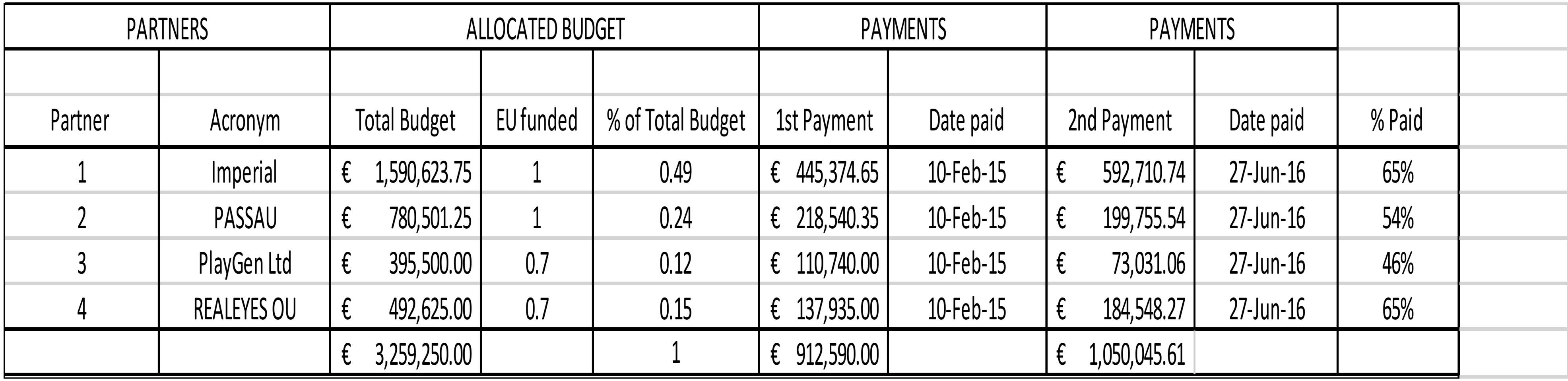
Specifically, the current trend in machine learning is deep learning and, each of our solutions needs to be compared to the solution provided by deep learning algorithms. Training deep learning algorithms and finding the right parameters costs a lot of time, much more than what we would usually spent if comparing to standard statistical machine learning or Bayesian learning algorithms. Due to this, we had a large "overspent" in terms of PMs on WP3-WP6. Similarly, and as already reported in Deliverable D9.1, we had a large "overspent" in terms of PMs on WP1. First, more time was spent on scheduling face-to-face interactions (sometimes multiple times due to unstable Internet connection) than was originally estimated. In the second reporting period, we have encountered various problems with data annotations. Some of the annotators proved to be less reliable than others and cleaning the annotations, re-annotating, and organizing all annotations to be stored in the online version of the database took much more time than what estimated originally.

The problems with annotation are also the main reason why we had an **overspent in the sub-contracting budget**, which will be covered from the unused equipment budget so that we remain **within the overall budget** (see section 5.2). As planned originally in the project proposal, Imperial has subcontracted a company for doing specific (and very tedious) annotation of the data. The company, Anaii, was our subcontractor. We have also asked them to take over SEWA-customisation and maintenance of the face-to-face chat platform and the population of the database and the SEWA website. As we explain in the first periodic report, deliverable D9.1, this work would cost a lot if done by Imperial's RAs, while the work is of pure engineering nature and very tedious and could be easily completed by Anaii. The work has now been completed but we have an overspent of GBP 13,814.92 in total. This amount will be moved from the equipment budget to cover the overspent in the sub-contracting (for further explanation, please see section 5.2).

## 5.2 Estimation of effort and budget

The second interim payment of EURO 1,050,045.61 was received by the co-ordinator (ICL) on 13th Jun 2016. Imperial transferred the contribution to the partners according to their % of contribution to the project within 45 days from date of receipt of payment from the commission.

Table: shows Beneficiary payments for second interim payment (Period 1 costs)



| PARTNERS | | ALLOCATED BUDGET | | | PAYMENTS | | PAYMENTS | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |
| Partner | Acronym | Total Budget | EU funded | % of Total Budget | 1st Payment | Date paid | 2nd Payment | Date paid | % Paid |
| 1 | Imperial | € 1,590,623.75 | 1 | 0.49 | € 445,374.65 | 10-Feb-15 | € 592,710.74 | 27-Jun-16 | 65% |
| 2 | PASSAU | € 780,501.25 | 1 | 0.24 | € 218,540.35 | 10-Feb-15 | € 199,755.54 | 27-Jun-16 | 54% |
| 3 | PlayGen Ltd | € 395,500.00 | 0.7 | 0.12 | € 110,740.00 | 10-Feb-15 | € 73,031.06 | 27-Jun-16 | 46% |
| 4 | REALEYES OU | € 492,625.00 | 0.7 | 0.15 | € 137,935.00 | 10-Feb-15 | € 184,548.27 | 27-Jun-16 | 65% |
| | | € 3,259,250.00 | | 1 | € 912,590.00 | | € 1,050,045.61 | | |

Due to this annual report being submitted by 31 Jan 2016 i.e. mid-way through the second reporting period, the co-ordinator is only able to collate financial and effort information from M13 up to M22. Full financial reporting will be submitted after M30 at the end of the second reporting period.

Table: shows Summary of Person Months per Work Package per Beneficiary for M1-M22 vs Planned PM / WP for M1-M42

| | Lead WP | Total Planned Person Months / PMs | ICL Planned PM M1-42 | ICL Actual PM M1-M22 | ICL Remaining PM | Passau Planned PM M1-42 | Passau Actual PM M1-M22 | Passau Remaining PM | PlayGen Planned PM M1-42 | PlayGen Actual PM M1-M22 | Playgen Remaining PM | RealEyes Planned PM M1-42 | RealEyes Actual PM M1-M22 | RealEyes Remaining PM | Total Planned PM M1-42 | Total Actual PM M1-M22 | Total Remaining PM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **WP1;** (Collection, annotation & release) | #1-IMP | 54 | 23 | 55.59 | (32.59) | 12 | 14.37 | (2.37) | 3 | 3.00 | 0.00 | 16 | 9.33 | 6.67 | 54 | 82.29 | (28.29) |
| **WP2;** (Low level feature extraction) | #2-PASSAU | 24 | 6 | 12.60 | (6.60) | 18 | 19.24 | (1.24) | 0 | 0.00 | 0.00 | 0 | - | 0.00 | 24 | 31.84 | (7.84) |
| **WP3;** (Mid-level feature extraction) | #1-IMP | 35 | 26 | 63.75 | (37.75) | 3 | 2.38 | 0.62 | 0 | 0.00 | 0.00 | 6 | 2.31 | 3.69 | 35 | 68.44 | (33.44) |
| **WP4;** (Continuous affect & Sentiment sensing in the Wild) | #2-PASSAU | 27 | 12 | 14.44 | (2.44) | 15 | 14.53 | 0.47 | 0 | 0.00 | 0.00 | 0 | - | 0.00 | 27 | 28.97 | (1.97) |
| **WP5;** (Behaviour similarity in the Wild) | #1-IMP | 27 | 21 | 21.33 | (0.33) | 6 | - | 6.00 | 0 | 0.00 | 0.00 | 0 | - | 0.00 | 27 | 21.33 | 5.67 |
| **WP6;** (Temporal Behvaiour - Patterning & Interpersonal Sentiment in the Wild) | #1-IMP | 42 | 21 | 4.83 | 16.17 | 21 | 6.00 | 15.00 | 0 | 0.00 | 0.00 | 0 | - | 0.00 | 42 | 10.83 | 31.17 |
| **WP7;** (Intergration, Applications & Evaluation) | #4-REALEYES OU | 104 | 9 | 0.00 | 9.00 | 7 | 1.06 | 5.94 | 44 | 25.79 | 18.21 | 44 | 27.59 | 16.41 | 104 | 54.44 | 49.56 |
| **WP8;** (Dissemination, Ethics, Communication & Exploitation) | #4-REALEYES OU | 30 | 6 | 1.28 | 4.72 | 4 | 3.73 | 0.27 | 8 | 2.88 | 5.12 | 12 | 6.95 | 5.05 | 30 | 14.84 | 15.16 |
| **WP9;** (Project co-ordination and Management) | #1-IMP | 24 | 18 | 8.49 | 9.51 | 2 | 0.78 | 1.22 | 2 | 1.14 | 0.86 | 2 | 1.61 | 0.39 | 24 | 12.02 | 11.98 |
| **TOTAL PERSON MONTHS** | | **367** | **142** | **182.31 *** | **(40.31)** | **88** | **62.09** | **25.91** | **57** | **32.81** | **24.19** | **80** | **47.79** | **32.21** | **367** | **325.00** | **42.00** |

Red figures in brackets denotes that more PM have been worked than the planned PM.

* Denotes that PM have not been included between M19-M22 for one person due to incomplete timesheets. All PM will be included at end of second reporting period M30

**Imperial - Explanation for budget and PM usage:**

Due to a combination of various factors, we are using and need to use a large number of PMs in different WPs, whilst remaining within the budget.

The factors in question are as follows:

- The effort needed for creation and annotation of the SEWA database has been underestimated in the initial estimation presented in the project proposal; as already explained in section 5.1, a variety of problems with annotation has been encountered and, as a result, many more working hours than originally planned were needed.
- The current trend in machine learning is deep learning and each of our solutions needs to be compared to the solution provided by deep learning algorithms. Training deep learning algorithms and finding the right parameters require more time and effort in comparison with using statistical machine or Bayesian learning. Hence, the larger number of PMs used.
- A number of our senior researchers have left our team in order to work for industries and junior researchers have been recruited. Since then, these new promising junior researchers have been trained and are working as part of the SEWA team. The training of the new team members used additional PMs and given that they are less experienced researchers, a larger number of PMs is needed to complete the tasks. Yet, we consistently remain within the initially estimated budget.

**Passau - Explanation for budget and PM usage:**

Overall, the excess from the initial estimation of PM in WP1 is about 2-3 PM and in WP2 is about 1-2 PM. This occurred mainly because Passau wanted to make fast progress during the first months of the project as a large part of the following work packages is based on the SEWA database. The acquisition of subjects for the database needed much more effort than foreseen. Especially people over 40 are very picky about privacy and do not like to be recorded. Due to technical problems with the network connections of the volunteers, a large number of scheduled recording sessions failed. In addition to what has been proposed, also Chinese people have been recorded and Passau was partly responsible for those volunteers. The annotation of the database required more efforts than assumed as we decided to make a precise manual transcription of all recordings, so that we are able to measure the performance of automatic speech recognition. We do not see any problems w.r.t. the timely delivery of all deliverables and milestones.

The excess of planned PM for work package 2 (Low-level feature extraction) has the following reasons:

Task 2.1 & 2.3: We developed a novel toolkit (openXBOW) from scratch (published on the SEWA website) and implemented more features than promised in the project proposal to be able to promote it in a better way in the research community.

Task 2.3: The development of automatic speech recognition (ASR) modules to be used 'in the wild' required much more efforts than expected (due to strong dialect, noise, poor recording conditions).

**PlayGen - Explanation for budget and PM usage:**

There is no overspend on budget or PM for PlayGen during the reporting period

**RealEyes - Explanation for budget and PM usage:**

There is no overspend on budget or PM for REALEYES during the reporting period.

### *5.2.1 Unforeseen sub-contracting*

Anaii Ltd has been used as the sub-contractor for the SEWA project to customise the SEWA Video Chat enabler, SEWA Database modelling, SEWA website implementation and maintenance, and provide annotations of the data in terms of facial points and gestures.

Due to the increased work on data annotation, as explained in section 5.1, we have an overspent on the sub-contracting. This overspent has been further increased by the new terms and conditions of H2020 projects, according to which VAT can be claimed for all actual costs. Imperial has automatically charged the VAT on subcontracting costs to the project, due to the HMRC tax rules for the UK (when an EU supplier invoices for their services sub-contracted on a H2020 project). Unfortunately this was not accounted for at the proposal stage of the project for the sub-contracting budget, and therefore the expenditure for the sub-contractor Anaii has been additionally overspent.

### *5.2.2 Unforeseen Financial Adjustment for Period 1 to be done at M30*

This is to notify the Commission that due to unforeseen changes in the terms and conditions of the financial administration in how the personnel salary is calculated in Reporting Period 1, Imperial will need to submit a Financial Adjustment. Salary was previously calculated using the Last Closed Financial Year methodology but since the P1 submission, Imperial’s methodology has reverted back to FP7 process using 1540 as the standard number of hours per annum following Imperial's financial year (01 Aug – 31 Jul).

Furthermore, due to the unforeseen overspent of the sub-contracting budget explained above, a deviation of the use of resources as described in the Grant Agreement for Imperial College is necessary. To cover the cost of this overspent, the Project Coordinator suggested a virement between "other direct cost" items budget and has obtained the approval of the Project Officer for this. Specifically, unused equipment budget will be used to cover the costs of the overspent on subcontracting. The equipment budget has not been fully utilised as the server to host the database has not been bought but, instead, we have used an existing server. Also, we have planned to buy processing machines but instead of that, we opted for a cheaper option and bought graphical cards, which boosted the performance of our existing machines. In that way we have an underspent on the equipment budget that could be used to cover the overspent on the sub-contracting. Overall, we remain with the budget.

### *5.2.3 Unforeseen use of in kind contribution from third party against payment or free of charges*

There was no unforeseen in–kind contribution from third party during reporting period two.

### *Unforeseen Linked Third Party*

Realeyes OÜ

There was a request for an amendment due to an unforeseen linked third party member to join the SEWA project.

Realeyes OÜ is currently the beneficiary of the Grant Agreement and to date the costs incurred related to the project have been related to employees of Realeyes OÜ based either in London or Budapest. Given the size of the team now in Budapest they have been required by local legislation to set up a local legal entity and will shortly transfer the employment of their Budapest based staff to this local entity. The new entity is called Realeyes Kft and it is 100% owned by Realeyes OÜ. From a practical purpose the same individuals will continue working on the project but the employing entity will change for those that are Budapest based.

Realeyes OU and Realeyes KFT, and their LEARS are now fully validated, and we are in the final stages of the process of submitting a formal amendment. The Project Officer was informed of this development from the start and his approval was sought.

# APPENDIX 1 – Letters of Understanding

Monday, February 6, 2017

## Letter of understanding

**Re: Commercial exploitation of results of project SEWA**

This letter of understanding is between RealEyes and Chair of Complex & Intelligent System of the University of Passau. The parties are engaged in a European Commission funded H2020 Innovation Action project called SEWA as joint industry and academic R&D.

This letter of understanding is based on the SEWA project's Consortium Agreement signed by the parties.

In order to ensure graceful execution of business and commercialisation plans of RealEyes arising from the development of integrated applications in the SEWA project, Chair of Complex & Intelligent System of the University of Passau commit to provision of all software and foreground IP generated in the SEWA project for the purpose of sentiment analysis.

The software and foreground IP includes: source code, algorithms, compiled binaries, docker instance image(s), and training data as developed during the SEWA project lifetime and used to train the algorithms in question. Chair of Complex & Intelligent System of the University of Passau provides a royalty-free, non- exclusive, non-transferable, perpetual licence to PlayGen to use, integrate, and distribute commercial products based on the aforementioned software developed during the SEWA project.

Whereas RealEyes may refactor or optimize the of code or algorithms of the aforementioned software, RealEyes grants Complex & Intelligent System of the University of Passau licence to such iterations on provision of protection of their business interest in so far as the use by Complex & Intelligent System of the University of Passau remains for the purpose of scientific research and direct results based on RealEyes's efforts are not shared with or licensed to other commercial entities.

SEWA PI for Passau: Bjoern Schuller

SEWA PI for RealEyes: Elnar Hajiyev

Monday, February 6, 2017

## Letter of understanding

**Re: Commercial exploitation of results of project SEWA**

This letter of understanding is between PlayGen and Chair of Complex & Intelligent System of the University of Passau. The parties are engaged in a European Commission funded H2020 Innovation Action project called SEWA as joint industry and academic R&D.

This letter of understanding is based on the SEWA project's Consortium Agreement signed by the parties.

In order to ensure graceful execution of business and commercialisation plans of PlayGen arising from the development of integrated applications in the SEWA project, Chair of Complex & Intelligent System of the University of Passau commit to provision of all software and foreground IP generated in the SEWA project for the purpose of sentiment analysis.

The software and foreground IP includes: source code, algorithms, compiled binaries, docker instance image(s), and training data as developed during the SEWA project lifetime and used to train the algorithms in question. Chair of Complex & Intelligent System of the University of Passau provides a royalty-free, non- exclusive, non-transferable, perpetual licence to PlayGen to use, integrate, and distribute commercial products based on the aforementioned software developed during the SEWA project.

Whereas PlayGen may refactor or optimize the of code or algorithms of the aforementioned software, PlayGen grants Complex & Intelligent System of the University of Passau licence to such iterations on provision of protection of their business interest in so far as the use by Complex & Intelligent System of the University of Passau remains for the purpose of scientific research and direct results based on PlayGen's efforts are not shared with or licensed to other commercial entities.

SEWA PI for Passau: Bjoern Schuller

SEWA PI for PlayGen: Kam Star

Monday, February 06, 2017

## Letter of understanding

**Re: Commercial exploitation of results of project SEWA**

This letter of understanding is between PlayGen and iBUG group of Imperial College. The parties are engaged in a European Commission funded H2020 Innovation Action project called SEWA as joint industry and academic R&D. This letter of understanding is based on the SEWA project's Consortium Agreement signed by the parties.

In order to ensure graceful execution of business and commercialisation plans of PlayGen arising from the development of integrated applications in the SEWA project, iBUG group of Imperial College commit to provision of all software and foreground IP generated in the SEWA project for the purpose of sentiment analysis.

The software and foreground IP includes: algorithms, compiled binaries, docker instance image(s), and training data as developed during the SEWA project lifetime and used to train the algorithms in question. iBUG group of Imperial College provides a royalty-free, non-exclusive, non-transferable, perpetual licence to PlayGen to use, integrate, and distribute commercial products based on the aforementioned software developed during the SEWA project.

Whereas PlayGen may refactor or optimize the of code or algorithms of the aforementioned software, PlayGen grants non-transferable licence to such iterations on provision of protection of their business interest in so far as the use by iBUG group of Imperial College remains for the purpose of scientific research and direct results based on PlayGen's efforts are not shared with or licensed to other commercial entities.

SEWA PI for the iBUG Group: Maja Pantic

SEWA PI for PlayGen: Kam Star

Monday, February 06, 2017

## Letter of understanding

**Re: Commercial exploitation of results of project SEWA**

This letter of understanding is between RealEyes and iBUG group of Imperial College. The parties are engaged in a European Commission funded H2020 Innovation Action project called SEWA as joint industry and academic R&D. This letter of understanding is based on the SEWA project's Consortium Agreement signed by the parties.

In order to ensure graceful execution of business and commercialisation plans of RealEyes arising from the development of integrated applications in the SEWA project, iBUG group of Imperial College commit to provision of all software and foreground IP generated in the SEWA project for the purpose of sentiment analysis.

The software and foreground IP includes: algorithms, compiled binaries, docker instance image(s), and training data as developed during the SEWA project lifetime and used to train the algorithms in question. iBUG group of Imperial College provides a royalty-free, non-exclusive, non-transferable, perpetual licence to RealEyes to use, integrate, and distribute commercial products based on the aforementioned software developed during the SEWA project.

Whereas RealEyes may refactor or optimize the of code or algorithms of the aforementioned software, RealEyes grants non-transferable licence to such iterations on provision of protection of their business interest in so far as the use by iBUG group of Imperial College remains for the purpose of scientific research and direct results based on RealEyes's efforts are not shared with or licensed to other commercial entities.

SEWA PI for the iBUG Group: Maja Pantic

SEWA PI for RealEyes: Elnar Hajiyev