SEWA

"Automatic Sentiment Analysis in the Wild"

Innovation Action

Horizon2020

Grant Agreement no. 645094

Deliverable D9.1
Annual Report 1

Deliverable Type:	R (Report)
Dissemination level:	CO (Consortium)
Month:	M12
Contractual delivery date:	January 31, 2016
Actual delivery date:	
Version:	0.1
Total number of pages:	14

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**	D9.1	**Title**	Annual Report 1
Work Package	**Number**	WP9	**Title**	Project coordination and management

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

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

Table of contents

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 199 successful data recording sessions using the aforementioned website. A total of 398 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.
- 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 will be further refined through semi-automatic correction.
- Identified a total of 540 representative segments (high/low arousal, high/low valence, and liking/disliking; in total 90 segments per culture group) from the SEWA corpus. These segments – titled "the core SEWA dataset" – will be 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.
- Released the SEWA database version 0.1 internally, as according to the data management plan, and prepared the web-portal and EULA for subsequent public database release.

1.2.2 WP2 – Low-level Feature Extraction

- Implementation and evaluation of a software tool openWord to generate Bag-of-Audio-Words (BoAW) representations from acoustic low-level descriptors (LLDs) for robust acoustic features.
- Feature enhancement by deep neural networks to improve acoustic features computed from noisy speech signals.
- Cross-corpus emotion analysis, i.e., testing models for emotion analysis on languages which are not included in the training data.
- Implementation of the incremental in-the-wild face alignment method for automatic facial landmark localisation.
- Generation of multi-lingual dictionaries for BoAW 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.

1.2.3 WP3 – Mid-level feature extraction

- Existing, state-of-the-art tracking algorithm has been used for extracting the features such as facial landmarks, 3D head pose, nods and tilts (WP3.1).
- Three methods for Facial Action Unit (AU) detection and intensity estimation have been developed (WP3.2).

- The methods for AU detection were trained and tested on two publicly available datasets of naturalistic facial behaviour coded in terms of AU intensity. On both datasets the proposed methods improved the state-of-the-art in automatic AU detection and intensity estimation.

1.2.4 WP4 – Continuous Affect and Sentiment Sensing in the Wild

- Work on WP4 will start in M15 (April 2016).

1.2.5 WP5 – Behaviour Similarity in the Wild – start M12 end M30

- Work on WP5 will start at the end of M12 (from 1st February 2016).

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

- Work on WP6 started just in M12 (January 2016).

1.2.7 WP7 – Integration, Applications and Evaluation

PlayGen have advanced the definition and design of the Chat Social Game.

- Refined and tested the concept underlying the Chat Social Game so that it is focused on a practical application with potential social and financial benefit.
- Clarified target user group and signed up 3 universities as partners to support user recruitment.
- Carried out 2 focus groups to define user needs.
- Developed initial game design concepts and mockups.
- Implemented initial prototype two-player chat-based game for debating called Sumobate.
- Progressed core technical functionality and advanced technical integration discussions.
- Planned evaluation approach.

RealEyes has focused on three major activities. First, they redefined targeted fields and the potential use of recommender engine with sentiment analysis support. Second, they worked on the computational framework that allows for testing ideas and methods for linking sentiment and emotion analysis with ad placement recommendation. Third, they built connections with different industrial players who could benefit from the SEWA results. In particular, they:

- Analysed the business interest in the use of sentiment analysis enhanced recommendations and found that the market of recommender systems is quite populated.
- Identified an already strong and increasing interest in media inventory optimization and online advertising.
- Initiated collaboration with potential future partners to help define target groups and clarify their needs.
- Obtained advert performance data from their partners which, in conjunction with social media performance and user rating, will be used for the work required to build and test the recommender engine. These profiles constitute an important part of the concept for using sentiment and emotion analysis for recommendation in an effective way.
- Implemented a first version for validating correspondences between emotion and (future) sentiment analysis and the quality of advertisements.
- Initial studies were conducted on modelling and clustering user and advert emotion profiles.

- Ran the first set of statistical analysis on signals derived from behavioural observations.
- Progressed core technical functionality and advanced technical integration discussions.
- Together with Imperial College London lead the organization of the Valorisation Board.

1.2.8 WP8 – Dissemination, Ethics, Communication and Exploitation

Workshops:

- The 300-VW 2015 (300 Videos in the Wild) – Facial Landmark Tracking in-the-wild Challenge & Workshop was organised as a satellite event of the IEEE International Conference on Computer Vision (ICCV 2015) in Santiago, Chile, in December 2015. SEWA sponsored the Prizes for the two winners of the challenge (USD 200 per winner; in total USD 400). (http://ibug.doc.ic.ac.uk/resources/300-VW/).
- The IBM-SEWA Cognitive Workshop 2015 was organised as a joint event between the IBM and the SEWA consortium with the aim to cross-fertilise the ideas on the state of affairs in cognitive computing and the future of it. The workshop was held in conjunction with the SEWA plenary meeting and the SEWA Valorisation Board meeting in October 2015, in London. (http://sewaproject.eu/ibmsewa15).
- The FERA 2015 (Facial Expression Recognition and Analysis Challenge) in facial action unit and facial expression detection in unconstrained images and videos was organized for the IEEE 11th International Conference on Automatic Face and Gesture Recognition in Ljubljana, Slovenia, in May 2015. SEWA sponsored 3 best paper prizes. (EUR 150 per winner, in the line with the prizes awarded at 2011 edition of FERA; in total EUR 450) (http://sspnet.eu/fera2015/).
- The WASA 2015 Workshop on Automatic Sentiment Analysis in the Wild was organized as a satellite event at the AAAC/IEEE 6th International Conference on Affective Computing and Intelligence Interaction (ACII 2015) in Xi'an, China, in September 2015. SEWA sponsored the Keynote Speaker (Jeffrey Cohn, USD 1000) and the Best Paper Award (USD 100). (http://sewaproject.eu/wasa15).
- The AV+EC 2015 (Audio/Visual + Emotion Challenge and Workshop) in the field of audio-visual behaviour understanding in the wild was organized for the ACM International Conference in Multimedia in Brisbane, Australia, in October 2015. This was the 5th AV+EC workshop so far and it includes physiological data for the first time. (http://sspnet.eu/avec2015/). SEWA sponsored the workshop by helping the data annotation.

Publications:

- “Sparkle: Adaptive Sample Based Scheduling for Cluster Computing”, *C. Hao, J. Shen, H. Zhang, X. Zhang, Y. Wu, M. Li.* In Proceedings of the 5th International Workshop on Cloud Data and Platforms (CloudDP), satellite event to EuroSys 2015.
- “Neural Conditional Ordinal Random Fields for Agreement Level Estimation”, *N. Rakicevic, O. Rudovic, S. Petridis and M. Pantic.* In Proceedings of the 1st International Workshop on Automatic Sentiment Analysis in the Wild (WASA), satellite event to ACII 2015.

- “Sentiment Apprehension in Human-Robot Interaction with NAO”, *J. Shen, O. Rudovic, S. Cheng and M. Pantic.* In Proceedings of the 1st International Workshop on Automatic Sentiment Analysis in the Wild (WASA), satellite event to ACII 2015.
- “Cross-Language Acoustic Emotion Recognition: An Overview and Some Tendencies”, *S. Feraru, D. Schuller, and B. Schuller.* In Proceedings of the 6th Bi-annual Conference on Affective Computing and Intelligent Interaction (ACII) 2015.
- “Detection of Negative Emotions in Speech Signals Using Bags-of-Audio-Words”, *F. Pokorny, F. Graf, F. Pernkopf and B. Schuller.* In Proceedings of the 1st International Workshop on Automatic Sentiment Analysis in the Wild (WASA), satellite event to ACII 2015.
- “Face Reading from Speech – Predicting Facial Action Units from Audio Cues”, *F. Ringeval, E. Marchi, M. Mehu, K. Scherer, and B. Schuller.* In Proceedings of INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association (ISCA), 2015.
- “Modelling User Affect and Sentiment in Intelligent User Interfaces”, *B. Schuller.* In Proceedings of the 20th ACM International Conference on Intelligent User Interfaces (IUI), 2015.
- “The First Affect Recognition Challenge Bridging Across Audio, Video, and Physiological Data”, *F. Ringeval, B. Schuller, M. Valstar, S. Jaiswal, E. Marchi, D. Lalanne, R. Cowie, and M. Pantic.* In Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge (AVEC), satellite event to ACM-MM 2015.
- “The 5th International Audio/Visual Emotion Challenge and Workshop”, *F. Ringeval, B. Schuller, M. Valstar, R. Cowie, and M. Pantic.* In Proceedings of the 23rd ACM International Conference on Multimedia (ACM-MM), 2015.
- “Speech Analysis in the Big Data Era”, *B. Schuller.* In Proceedings of the 18th International Conference on Text, Speech and Dialogue (TSD), satellite event of INTERSPEECH 2015, Lecture Notes in Artificial Intelligence (LNAI), Springer, 2015.
- “Variable-state Latent Conditional Random Fields for Facial Expression Recognition and Action Unit Detection”, *R. Walecki, O. Rudovic, V. Pavlovic, M. Pantic.* In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG), 2015.
- “Latent Trees for Estimating Intensity of Facial Action Units”, *S. Kaltwang, S. Todorovic, M. Pantic.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- “Second Facial Expression Recognition and Analysis Challenge”, *Michel F. Valstar, Timur Almaev, Jeffrey M. Girard, Gary McKeown, Marc Mehu, Lijun Yin, Maja Pantic and Jeffrey F. Cohn.* In Proceedings of the Automatic Face and Gesture Recognition (FG), 2015.
- “Fast and Exact Bi-Directional Fitting of Active Appearance Models”, *J. Kossaifi, G. Tzimiropoulos, M. Pantic.* In Proceedings of the Facial Expression Recognition and Analysis Challenge (FERA), satellite event to FG 2015.

- "Robust Statistical Face Frontalization", *C. Sagonas, Y. Panagakis, S. Zafeiriou, M. Pantic.* In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.
- "Multi-conditional Latent Variable Model for Joint Facial Action Unit Detection", *S. Eleftheriadis, O. Rudovic, M. Pantic.* In Proceedings of the International Conference on Computer Vision (ICCV), 2015.
- "Prediction based audio-visual fusion for classification of non-linguistic vocalisations", *S. Petridis and M. Pantic.* IEEE Transactions on Affective Computing, accepted for publication, 2015.
- "Probabilistic Slow Features for Behavior Analysis" *L. Zafeiriou, M. A. Nicolaou, S. Zafeiriou, S. Nikitidis, M. Pantic.* IEEE Transactions on Neural Networks and Learning Systems, accepted for publication, 2015.
- "Robust Correlated and Individual Component Analysis", *Y. Panagakis, M.A. Nicolaou, S. Zafeiriou, M. Pantic.* IEEE Transactions on Pattern Analysis & Machine Intelligence, accepted for publication, 2015.
- "Discrimination Between Native and Non-Native Speech Using Visual Features Only", *C. Georgakis, S. Petridis, M. Pantic.* IEEE Transactions on Man and Cybernetics, accepted for publication 2015.
- "Doubly Sparse Relevance Vector Machine for Continuous Facial Behavior Estimation", *S. Kaltwang, S. Todorovic, and M. Pantic.* IEEE Transactions on Pattern Analysis and Machine Intelligence, accepted for publication 2015.

Public Presentations:

- (Jan 2016) SEWA coordinator speaking at World Economic Forum
  Prof. Maja Pantic spoke at World Economic Forum in Davos on Emotional AI.
  https://webcasts.weforum.org/widget/1/davos2016?p=1&pi=1&hl=english&id=76123
- (Dec 2015) Dr Ognjen Rudovic and Dr Jie Shen (ICL) present SEWA technology at the Science Fair in Belgrade Serbia (a public event attended by 20,000+ children from elementary and high schools):
  http://festivalnauke.org/Program/Beogradski-sajam/Humaniji-nego-sto-mislite
- (Mar 2015) SEWA coordinator speaking at Royal Academy
  Prof. Maja Pantic spoke of SEWA and iBUG research at Royal Academy Event.
  https://www.royalacademy.org.uk/event/ra-schools-spring-symposium

Scientific Talks:

- (Nov' 2015) SEWA UP PI, Björn Schuller speaking at LIRIS Workshop on Emotion, Ecole Centrale de Lyon/Technicolor in an invited talk on the topic "Sound Affects".
- (Nov' 2015) SEWA coordinator speaking at ACPR 2015 as a Keynote speaker.
  http://acpr2015.org
- (Oct' 2015) SEWA UP PI, Björn Schuller speaking at "Orange Hour", GfK, Hamburg, in an invited talk on the topic "Emotions in the Voice: Making them Accessible for Consumer Research".

- (Oct' 2015) SEWA UP PI, Björn Schuller speaking at Cognitive Computing Workshop 2015 as a Keynote speaker.
  http://www3.imperial.ac.uk/newsandeventspggrp/imperialcollege/engineering/computing/eventssummary/event_16-9-2015-12-8-6
- (Sep' 2015) SEWA coordinator speaking at ACII 2015 as a Keynote speaker.
  http://www.acii2015.org
- (Sep' 2015) SEWA UP PI, Björn Schuller speaking at 18th International Conference on Text, Speech and Dialogue (TSD 2015) as a Keynote speaker.
  http://www.kiv.zcu.cz/tsd2015/index.php?page=speakers
- (Sep' 2015) SEWA coordinator speaking at ECMR 2015 as a Keynote speaker.
  https://lcas.lincoln.ac.uk/ecmr15/?q=node/1
- (Sep' 2015) SEWA UP PI, Björn Schuller speaking at Annual Meeting of the New Champions 2015 (AMNC, "Summer Davos"), BetaZone Session, World Economic Forum (WEF) as a Keynote speaker.
  https://agenda.weforum.org/news/world-economic-forum-honours-its-2015-young-scientists-community-at-annual-meeting-of-the-new-champions/
- (Sep' 2015) SEWA UP PI, Björn Schuller speaking at International Symposium on Companion-Technology (ISCT 2015, Expert Workshop of the SFB-TR 62) as a Keynote speaker.
  https://isct2015.informatik.uni-ulm.de/wp-content/uploads/2015/09/ISCT2015Program.pdf
- (Avg' 2015) SEWA UP PI, Björn Schuller speaking at SMART School on Computational Social and Behavioral Sciences as a Keynote speaker.
  http://www.smart-labex.fr/SMART_School_on_Computational_Social_and_Behavioral_Sciences.html
- (Jul' 2015) SEWA coordinator speaking at ICME 2015 as a Keynote speaker.
  http://www.icme2015.ieee-icme.org
- (Jun' 2015) SEWA UP PI, Björn Schuller speaking at 4th Machine Learning for Interactive Systems Workshop (MLIS 2015) held at the International Conference on Machine Learning (ICML'15).
  http://scuba.usc.edu/?q=workshops
- (Jun' 2015) SEWA UP PI, Björn Schuller speaking at UK SPEECH 2015 - 4th Meeting of the UK and Irish Speech Science and Technology Research Community as a Keynote speaker.
  http://www.ukspeech.org.uk/workshop/UKSpeech2015-programme.pdf
- (Jun' 2015) SEWA UP PI, Björn Schuller speaking at invited inaugural lecture, University of Passau.
  http://www.fim.uni-passau.de/fileadmin/files/dekanat/Veranstaltungshinweise/Antrittsvorlesung_Schuller_Wirth_Handschuh.pdf

- (May' 2015) SEWA UP PI, Björn Schuller speaking at invited tele-talk, Applied Signal Processing in Mental Health Workshop, University of Southern California as a Keynote speaker. http://scuba.usc.edu/?q=workshops
- (Apr' 2015) SEWA UP PI, Björn Schuller speaking at "Orange Hour" as a Keynote speaker.http://www.gfk-verein.org/en/events/orange-hour
- (Apr' 2015) SEWA UP PI, Björn Schuller speaking at 2015 International Symposium on Computational Psychophysiology as a Keynote speaker.http://www.taf.sdnu.edu.cn/enligsh.htm
- (Mar' 2015) SEWA UP PI, Björn Schuller speaking at NII Shonan Meeting: The Future of Human-Robot Spoken Dialogue: from Information Services to Virtual Assistants, Seminar 059 as a Keynote speaker. http://shonan.nii.ac.jp/shonan/blog/2013/12/10/the-future-of-human-robot-spoken-dialogue-from-information-services-to-virtual-assistants/

Press Coverage:

- (Jan 2016) SEWA coordinator interview for TV Serbia (RTS) on 4th Industrial Revolution http://www.rts.rs/page/tv/sr/story/20/RTS+1/2182120/Svetska+ekonomija+u+slobodnom+padu.html
- (Jan 2016) SEWA coordinator interview for France24 on Future AI http://www.france24.com/en/20160128-people-profit-davos-wef-global-economy-digital-revolution-robots-france-reform
- Wall Street Journal http://blogs.wsj.com/digits/2015/04/17/emotion-tracking-startup-gets-eu-funding-boost/
- Marketing http://www.marketingmagazine.co.uk/article/1343366/european-commission-issues-€36m-grant-tech-measures-content-likeability
- Campaign http://www.campaignlive.co.uk/youtube/article/1343366/european-commission-issues-36m-grant-tech-measures-content-likeability/
- Media Week http://www.mediaweek.co.uk/article/1343366/european-commission-issues-€36m-grant-tech-measures-content-likeability
- Brand Republic http://www.brandrepublic.com/article/1343366/european-commission-issues-€36m-grant-tech-measures-content-likeability
- The Drum http://www.thedrum.com/news/2015/04/17/company-tracks-human-emotions-webcams-awarded-26m-grant-eu-commission
- Research Live http://www.research-live.com/news/realeyes-eu-grant-for-measuring-ad-likeability/4013199.article
- DotRising http://www.dotrising.com/2015/04/20/european-commission-awards-2-6m-grant-to-find-tech-to-measure-content-likeability/
- Tech Investor News http://www.techinvestornews.com/Enterprise/Latest-Enterprise-News/realeyes-receives-grant-to-develop-likeability-tracking-webcam
- Computer Business Review http://www.cbronline.com/news/big-data/hardware/realeyes-receives-grant-to-develop-likeability-tracking-webcam--4557214

- (Sep' 2015) Computer lernt menschliches Verhaltenhttp://www.pnp.de/nachrichten/heute_in_ihrer_tageszeitung/bayern/1801562_Computer-lernt-menschliches-Verhalten.html
- (Jun' 2015) SEWA project to develop methods for automatic behavioural analysishttp://www.uni-passau.de/en/bereiche/press/press-releases/news/detail/sewa-project-to-develop-methods-for-automatic-behavioural-analysis/
- (Jun' 2015) Projekt SEWA entwickelt Methoden zur automatischen Verhaltensanalysehttp://www.uni-passau.de/bereiche/presse/pressemeldungen/meldung/detail/projekt-sewa-entwickelt-methoden-zur-automatischen-verhaltensanalyse/ (Feb' 2015) Björn Schuller: "Grosse Gefühle - Robotik und Emotionen",
  radio interview broadcast on radiobremen/Nordwestradio ,
  "Glauben und Wissen", Tina Würfel.

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.

Communication

- **Internet Presence:** The consortium has set up a web site: www.sewaproject.eu as a dissemination tool to be maintained during the project and beyond. It has been set up with general information about SEWA, members of the consortium, the objectives and results of the project, up-to-date news about all dissemination efforts, including the information about project presence in conferences, fairs, exhibitions, etc. The website facilitate subscription to project news and events, download of public deliverables, download of publications related to the project, download of released software tools and data, download of demonstration videos, and download of video-lectures recorded during the project-related events. It also facilitate a project-private space for filing deliverables and internal reports.
- **Press:** The full list of press coverage is listed in 1.2.8.
- **Industry Fairs:** ICL took part in The Festival Nauke in Belgrade 3 to 6 December 2015. http://festivalnauke.org/Program/Beogradski-sajam/Humaniji-nego-sto-mislite
  The aim was to promote the technology advances developed by the SEWA project to elementary and high-school children and to stimulate their interest in science, in general, and Artificial Intelligence and Human-Computer Interaction, in particular.

- **Workshops:** The full list of SEWA workshops is listed in 1.2.8.

Exploitation

- Realeyes' involvement in SEWA project has had a positive impact on its market position. It has strengthened its relationship with existing customers (especially those who took a role in the Valorisation Advisory Board, e.g., IPSOS). It also attracted attention of new customers. Of those, the most important is a significant commercial contract with one of the world's largest media agencies – Mediacom. This development had wide press coverage.
  http://www.mrweb.com/drno/news22038.htm
  http://digiday.com/agencies/facial-coding-saves-clients-millions-not-running-campaigns/
  http://realbusiness.co.uk/article/32747-marketing-firm-mediacom-to-track-consumer-emotions-with-uk-tech-startup-realeyes
- Realeyes has also partnered up with one of the Valorisation Board members, Xaxis. The goal of the partnership is to establish the links between emotional audience profiling and impact on the ad tech industry. This is a part of an ongoing collaboration and it is too early to disclose any results. But collaboration itself already paves the way for further commercial partnership between the two companies in the future.
- Through the course of exploration of possibilities for emotional profiling of ads and users, Realeyes has built a training and evaluation framework, which can be used for effective data analysis and predictive model training. A pilot study on their emotion-based predictive models suggests a direct link between emotions and effectiveness of an advertising video content. This is significant in the area of market research and, if these results are confirmed in subsequent studies, it would enable Realeyes to further strengthen its position in the market research industry. The pilot study in question was made possible in part thanks to the work of Realeyes on the SEWA ad recommendation engine.
- PlayGen's participation in the Valorisation board meeting and follow up conversations, lead to identifying a number of commercial opportunities, including jobs market, employability skills, dating and sales training.
- PlayGen organised a SEWA internal workshop and a set of meetings with project partners to explore potential business opportunities combining video chat, games and emotion detection technologies. It was decided to focus on exploration of potential market opportunities within employability skills for young people, in particular the potential role of emotion detection in enhancement of communication skills.

1.2.9 WP9 – Project co-ordination and management

- Creation of website: www.sewaproject.eu
- Creation of mailing list: sewa_formal@googlegroups.com
- Managed and submitted deliverables
- Project coordinator with ICL Research Contracts handled signing of the consortium agreement
- Project manager worked with consortium for pre-financing calculation
- Project manager highlighted policy and procedure changes from FP7 to H2020

- List of project meetings during this period:
  12 and 13 February 2015 – Kick off meeting – ICL, London, UK.
  6 and 20 March, 22 May, 19 June 2015 – Phone meetings.
  16 July 2015 – Plenary meeting – ICL, London, UK.
  4 September 2015 – Phone meeting.
  30 September and 1 October – Plenary and Valorisation Board meetings – ICL, London, UK.
  1 December 2015 and 11 January 2016 – Phone meetings.

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

- A total of 398 participants from 6 different cultural backgrounds (British, German, Hungarian, Serbian, Greek and Chinese) were recorded in the wild, 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. The data will be annotated 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. The SEWA database will be publicly released in a web-based searchable form. The SEWA database is the very first to contain in-the-wild recordings of people's reactive and interactive behaviours, to be demographically balanced (equal number of male and female subjects, equal number of subjects from each age group 20-30-40-50-60, in each age group there is at least one of the following interactive dyads male-male, male-female, female-female), having all subjects who are native speakers, to be annotated in terms of all visual, vocal, and verbal cues, affective dimensions, sentiment, and social signals such as rapport and mimicry.

1.3.2 WP2 – Low-level Feature Extraction

- Proven that Bag-of-Audio Words is able to predict emotions in terms of arousal and valence in a better way than all other known approaches and published results.
- 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) with different feature fusing schemes.
- 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

- Showed that by modelling AU segments in videos using a mixture of nominal and ordinal states improves the AU segmentation/detection over the state-of-the-art conditional random field models that employ either type of the states (i.e., nominal or ordinal).
- The proposed extension of the CORF (Conditional Ordinal Random Field) model, by defining its feature functions by means of Neural Networks, resulted in better estimation of AU intensities. The model was also applied to the task of agreement level estimation from the MAHNOB database – and it outperformed the existing methods applicable to the target task.

1.3.4 WP4 – Continuous Affect and Sentiment Sensing in the Wild

- Work on WP4 will start in M15 (April 2016).

1.3.5 WP5 – Behaviour Similarity in the Wild

- Work on WP5 will start at the end of M12 (from 1st February 2016).

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

- No findings so far as WP6 started just in M12 (January 2016).

1.3.7 WP7 – Integration, Applications and Evaluation

- Social Chat Game

Through a series of discussions with the Valorisaton Board and the project partners, it has been concluded that the application of the SEWA technology for Chat Social Game should represent a new approach to communication skills training utilising emotion detection technologies developed in SEWA, together with validation methodologies. The application is targeting young people aged 18+, who are either in educational institutions or have recently completed education, who are shortly embarking on a career, and who'd benefit from a light touch, fun and meaningful way of practicing negotiation and discussions that are part of the everyday working life (e.g. job interview, dealing effectively with customers, negotiating a reduction in rent with a landlord or being more effective at dealing with work colleagues). Feedback from employers and end-user focus groups has demonstrated that this has potential to develop students' negotiation and persuasion skills and increase the chances of their employability and their subsequent performance in what is usually their first job, in a cost-effective way. In addition to the gaps in influencing skills identified by employers and end-users, the focus on this aspect of communication appears to be particularly well suited for digital games, as there are clear objectives and possible ways of assessing the goals, as well as being a good fit with automatic emotion detection, as emotions play a significant role in interpersonal communication. Additionally as supported by the European Commission, this approach contributes to the promotion of recognition and validation of knowledge and skills acquisition through non-formal learning. Therefore this application aims to contribute to employability of young people, with obvious positive societal impacts.

In order to pre-empt integration, a simple two person chat social game was developed utilising on of the existing platforms being utilised in the platform. This provided both experience of the specific software as well as helped to identify future issues with respect to integration, delivery and evaluation.

- Advert Recommender System

The Advert Recommender System aims to integrate and validate a novel approach to ad placement optimization by analysing nonverbal behavioural cues of the users/customers. The great advantage of this method over traditional approaches is that ad placement can be made more personal, informative, effective and less annoying thus significantly improving the usefulness for both the advertisers and the potential customers. To this aim Realeyes did the following:

i) Built a framework to record audio, video, and questionnaire data online in response to video content. The framework has been used to collect user responses for over 150 advert videos (among which are the 4 advert videos used as stimuli material in the

SEWA data collection). Each of the 150 videos is associated with a performance measure (i.e., advert is successful/ unsuccessful), which was provided to us by our clients. This information and the behavioural responses by users constitute a unique dataset, which can be analysed for patterns of users' behaviour being predictive of the ultimate performance of an advert.

ii) A new framework has been designed and is being built that allows for quick statistical testing of "emotional profiles of ads" (which, in essence, determine which people would react how on a specific ad) and their use for predicting ads' performance.

iii) A pilot study has been carried out on the possible use of the collected data for building emotional profiles of adverts. Initial findings are positive and confirm our expectations, namely, that emotional reactions of users can be predictive of ads' performance and that they can be used for better targeting. Once we receive approval of our commercial partners to share the findings, we plan to publish these results in academic and market research papers.

1.3.8 WP8 – Dissemination, Ethics, Communication and Exploitation

(i) Dissemination & Communication

SEWA partners have increased the interest of general public in the field of automatic emotion recognition and corresponding applications. There are several evidences for this: a major article on SEWA in German national press (Passauer Neue Presse), a TV report on emotional agents partly filmed at the chair of PI Björn Schuller at University of Passau, invitation to the SEWA coordinator to speak on SEWA and related technologies at the World Economic Forum in Davos in January 2016, invitation to SEWA partners to present SEWA technology at the Science Fair in Belgrade (Serbia), and two TV interviews with the SEWA coordinator on emotional robots/technology and their role in the 4th Industrial Revolution (see 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 1st International Workshop on Automatic Sentiment Analysis in the Wild (WASA'15), which has has been organised by SEWA partners, 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 1.2.8).

(ii) Exploitation, Socio-Economic impact, and Social Implications of the project

While socio-economic impact of the project is clear, relating to the significant technological leverage that the industrial partners of the SEWA project benefit from, as explained in 1.3.7, social implications of the project are less clear, though possibly profound. Let us explain this in more detail.

According to the European Commission report in 2014, 14 million young people in the EU are not in employment, education and training, contributing over €162 billion in annual economic loss as well as additional long term personal and social costs. Since reliable training through education to secure employment no longer exists, governments in Europe including the UK government are investing in a range of schemes such as apprenticeships to help young people get a foothold on the job ladder. In the UK alone some 500,000 individual apprenticeships were offered in the year 2014/15 representing a 14% annual growth in number of apprenticeship placements. With each apprentice receiving support from the government including communication training by specialist organisations, companies offering apprenticeship placements consistently cite lack of soft skills such as ability to communicate well as a critical barrier for young people to succeed.

In addition to half a million young people seeking apprenticeship, in the UK, over 2 million people graduate from universities and higher education establishment per year, the figure in Europe for tertiary education is over 20 million individuals, the great majority of whom will then be embarking on a career which requires amongst other skills, to have the communication skills necessary to succeed in job interviews as well as in the majority of instances to then fit into working environment through effective communication skills.

With a potential European market size of some 30 million individual young people annually, a large proportion of who appear to benefit from better communication skills, the strategies for exploitation of the outcome of SEWA will explore a number of avenues. In the first instance the focus will be in the UK market, since PlayGen is in the UK, it will exploit the good links it has with local and national government agencies including the department of education as well as a large number of UK universities, to develop a commercial offering centred around improving student employability, increasing apprenticeship success rates, highlighting the benefits in reduction of training cost and long term success of individuals. From a given potential addressable market of some 2.5 million in the UK, and armed with scientific evidence that SEWA delivers better communication skills through non-formal setting it may be estimated that even a modest uptake of some 5% could potentially generate revenues of €1.2m annually. Of course this only represents the UK, extending the solution to the English speaking nations and further leveraging the languages supported by SEWA, it is feasible to imagine far larger revenue in due course.

Whilst the current focus is on employability skills, the core technologies and game at a meta-level can be extended for a wide range of applications and communication skills, from sales training to customer support training to dating applications. We will continue to monitor and explore potential markets as we develop, evaluate and validate the approach from both technology and commercial aspects, for instance exploring freemium and tiered models aimed both at B2C and B2B markets in addition to the strategies highlighted above.

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
D8.1	Overall Dissemination Plan	WP8	UP	R	CO	M3 (Apr 15)	6/5/2015	N/A	Submitted
D8.2	Data Management Plan	WP8	ICL	R	CO	M6 (Jul 15)	23/7/2015	N/A	Submitted
D2.1	Improved acoustic feature extractor	WP2	UP	DEM	CO/PU[1]	M9 (Oct 15)	29/10/15	N/A	Submitted
D7.1	Report on user requirements for SEWA applications	WP7	RealEyes	R	PU	M12 (Jan 16)	05/02/16	N/A	Submitted
D9.1	Annual Report 1	WP9	ICL	R	PU	M12 (Jan 16)	10/02/16	N/A	Submitted

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
MS1	SEWA DB + SEWA applications requirements	WP1, WP7	ICL	1/2/2016	Data collected and annotated, workshop of Valorisation Advisory Board held and SEWA applications' requirements defined	YES	N/A	N/A

4. Critical implementation risks and mitigation actions

4.1 Foreseen Risks

Risk Number	Description of Risk	Work Packages Concerned	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 WP2
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 WP8
R13	Loss of key technology / application partner	WP8, WP9	See description in WP8

4.2 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**	**Task 3.1.** The existing databases of spontaneous facial behaviour recorded in laboratory settings (UNSBC Shoulder PAIN (Lucey et al 2011) & DISFA (Mavadatti et al 2013) datasets) have been used for training the initial versions of the FAU detection tool in WP3. These initial versions of the tool are then used in a semi-automatic manner to speed up the annotation of the SEWA database.
R2	**M1-9**	**YES**	**NO**	The feature extraction performance is near real-time.
R3	**M1-13**	**YES**	**Potentially**	**Task 2.1** Our efforts to create noise robust features, or rather feature representations for real-life emotion recognition from speech, succeeded. ASR has problems with strong dialect, noise and cross-talk. Influence on emotion recognition must be studied later on (see also R5).
R4	**M3-15**	**YES**	**In progress**	**Task 3.2** A spatio-temporal representation of facial changes cannot be optimised for real-time performance (20-25 fps). The work in T3.2 is currently based on facial landmark location and their trajectories in time. Alternative static appearance-based features are being investigated.
R5	**M15-27**	**NO**	**NO**	Not applicable yet.
R6	**M15-27**	**YES**	**NO**	The core SEWA dataset (540 representitive short segments) will be annotated first. The progress towards this end is already well advanced. This data-set will be used to develop the methods in WP4.
R7	**M12-30**	**NO**	**NO**	Not applicable yet
R8	**M12-42**	**NO**	**NO**	Not applicable yet
R9	**M12-42**	**NO**	**NO**	Not applicable 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**	Not applicable yet
R12	**M1-42**	**NO**		No competing technology has emerged from another academic or industrial institution
R13	**M1-42**	**NO**		No loss of key technology / application partner has occured.

4.3 Innovation

From PlayGen

Activities developed within the project	**Number**	**Explanation (not for inputting to final report)**
Prototypes	3	1 prototype for each version
Testing activities	4	1 set of initial focus groups 3 round of user testing (1 for each version)
Clinical trials	NA	NA

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

New product (good or service)	**[YES]**
New process	**[YES]**
New method	**[YES]**

From RealEyes

Activities developed within the project	**Number**	**Explanation (not for inputting to final report)**
Prototypes	3	1 prototype for each version
Testing activities	3	3 round of user testing
Clinical trials	NA	NA

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

New product (good or service)	**[YES]**
New process	**[YES]**
New method	**[YES]**

5. Gender of R&D participants involved in the project

Beneficiaries	Number F including third parties (if appropriate)	Number M including third parties (if appropriate)	Total Including third parties (if appropriate)
ICL	3	9	12
UP	1	3	4
PlayGen	2	3	5
RealEyes	0	4	4

6. Periodic Technical Report

6.1 Project as a whole: Progress on project objectives

(i) SEWA technology uses **robust features** for audio-visual human behaviour analysis in the wild. It combines existing technologies for robust audio feature extraction, including state-of-the-art techniques for cancelling non-stationary noise, leading to an improved accuracy in audio-based recognition of positive and negative valence/ arousal. Further, it uses existing techniques for accurate detection of a very dense set of facial landmarks, which perform robustly for unseen subjects and independently of variations in pose, expression, illumination, background, occlusion, and image quality. This work relates to WP2.

This objective has been achieved in the first year of the project. Section 6.3 of this report details the approach to robust feature extraction used by SEWA.

(ii) SEWA technology performs **robust and accurate detection and intensity estimation of facial actions** (i.e. Facial Action Units (FAU)) in the wild. It combines previously proposed methodologies for both spatiotemporal representation of facial appearance changes and context-adaptive dynamic estimation of FAU intensity levels, train/adapt those to novel data recorded in the wild (see also objective (vii)), and achieves an F-1 score of 50% calculated on average for all 45 FAUs and their intensity scored on 5-level Likert scale. This work relates to WP3.

The work on this objective is still ongoing. Section 6.4 of this report details the progress so far on robust joint detection and estimation of FAUs and their intensities used by SEWA.

(iii) SEWA technology produces **continuously-valued estimates of sentiment and affect dimension**s in the wild. In analogy to the state of the art approaches to inferring continuously valued predictions of affective dimensions (e.g. valence and arousal) from audio-visual data, and by using appropriate segmentation of the input data streams into meaningful episodes (e.g., utterances between speech pauses, facial actions between expressionless facial displays), SEWA technology realises fully automatic continuously-valued sentiment and affect dimensions prediction from audio-visual data recorded in the wild with a Mean Squared Error of 0.1 or less on average. This work relates to WP4.

The work on this objective did not start yet (WP4 starts in M15 of the project). The description of WP4 in the original DoW details our intended approach to audio-visual continuous sentiment and affect dimensions analysis in the wild.

(iv) SEWA technology measures **spatio-temporal similarity of two behavioural episodes** observed in the wild; that is, it answers the question “are these two behaviours similar?” instead of “what is the displayed behaviour?”. The aim is to achieve 80% accuracy in labelling two behaviours as being similar. This work relates to WP5.

The work on this objective did not start yet (WP5 starts in M12 of the project). The description of WP5 in the original DoW details our intended approach to measuring spatio-temporal behaviour similarity used by SEWA.

(v) SEWA technology **recognises audio-visual behaviour matching and inter-personal sentimen**t shown by two people interacting in the wild by means of an FF-HCI application. It detects audio-visual behaviour matching such as mimicry, rapport, and empathy (mutually alike sentiment), with an average F-1 score of 70% by using either prediction-based approach (predict the behaviour of one person based on the behaviour of the other person, and vice versa) or behaviour-similarity-based approach described in objective (iv) and detailed in WP5. It assesses interpersonal sentiment with an average of F-1 score of 50% for 5 levels of Likert scale of how positive/ negative was the sentiment during the dyad, based on the presence, frequency and duration of mirroring, rapport episodes, and using existing regression and/or multi-class classification methodologies. This work relates to WP6.

The work on this objective did not start yet (WP6 starts in M12 of the project). The description of WP6 in the original DoW details our intended approach used by SEWA.

(vi) SEWA technology **adapts to the contextual situation and target individuals** or observation. It is based on incremental learning, autonomous self-learning, and context-driven methods, and it achieves 5% more accurate recognition results when using person- and context-sensitive models than when using person- and context-independent models. Statistical significance of the results will be ensured. WP2-WP4 explicitly include comparisons of person/ context-driven and person/ context-independent models. These validation studies also include studies on gender-based models.

This objective has been achieved in terms of using person-adaptive tracking algorithms. See section 6.3.2 for details.

The work on this objective is still ongoing when it comes to behaviour interpretation algorithms. Section 6.4 of this report details the progress so far on context-sensitive models for estimation of FAUs and their intensities used by SEWA. The work on context-sensitive estimation of higher-level behaviours like sentiment and affective dimensions did not commence yet (the work on relevant WPs – WP4-WP6 – will start from M12).

(vii) SEWA releases a **large volume of audio-visual data of human behaviour recorded in the wild** together with expert annotations in the form of SEWA DB.

The work on this objective is well ongoing. See section 6.2 for details.

(viii) The **applications of SEWA technology (Advert Recommender System and Social Chat Game) are user-centric and sentiment-driven**, enhancing the users' experience (i.e., their satisfaction with the provided recommendation, or their enjoyment of the social game). This work relates to WP7.

The work on this objective is still ongoing. Section 6.8 of this report details the progress so far on applications of SEWA technology.

(ix) SEWA ensures a **large spill over of the knowledge acquired to European industries** by means of the SEWA Valorisation Advisory Board comprising industrial representatives of different branches.

The work on this objective is well ongoing. Section 6.9 of this report details the progress so far.

(x) SEWA **improves the competitive position of RealEyes, PlayGen, and other European industries** (directly and indirectly) related to the SEWA project through provision of highly innovative high-value technology for robust and accurate automatic audio-visual human behaviour analysis, including affect and sentiment analysis.

The work on this objective is well ongoing. Section 1.2.8 (on exploitation) details the benefits that the SEWA project brought to RealEyes and PlayGen already.

(xi) SEWA ensures a **wide public engagement** by organising a number of high-profile public demonstrations of the SEWA technology.

The work on this objective is well ongoing. Section 1.3.8 details a number of high-profile public demonstrations of the SEWA technology.

6.2 Progress on WP1 – SEWA DB collection, annotation and release

6.2.1 Task 1.1: Ethical Approval

This has been done by amending the existing ethical approval for studies on "Multimodal Analysis of Human Nonverbal Behaviour in Real-World Settings", issued by the Imperial College Research Ethics Committee (ICREC ref no: ICREC_8_2_3), as to include UP team member (Prof Bjoern Schuller) and RealEyes team member (Antonios Oikonomopoulos) and extend the validity of the ethical approval beyond the end of the SEWA project (i.e., 16-02-2020). In addition, an ethical approval was obtained by the Ethical Advisory Board, as explained in 1.2.8 above, and forwarded to the Commission as part of deliverable D8.2, submitted on 24 July 2015.

The members of the Ethical Advisory Board: 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.

6.2.2 Task 1.2: SEWA data acquisition

An important aspect of the SEWA project lies in collecting suitable datasets of enough labelled examples to facilitate the development of robust tools for automatic machine understanding of human behaviours. To create such dataset, a data collection experiment has been conducted, resulting in a large in-the-wild audio-visual corpus containing a wide variety of spontaneous expressions of emotions and sentiment.

In this experiment, participants were divided into pairs based on their cultural background, age and gender. During initial sign-up, participants were asked to complete a questionnaire of demographic measures including gender, age, cultural background, education, personality traits, and familiarity with the other person in the pair. To promote natural interactions, participants within each pair were required to know each other personally in advance of the experiment. Each pair of the participants then took part in two parts of the experiment, resulting in two sets of recordings.

Experimental Setup Part 1: Each participant was asked to watch 4 adverts, each one being around 60 seconds long. These adverts had been chosen to elicit mental states including amusement, liking and boredom. After watching the advert, the participant was also asked to fill-in a questionnaire to self-report his/her emotional state and sentiment toward the advert.

Experimental Setup Part 2: After watching the 4th advert, the two participants were asked to discuss the advert they had just watched by means of the video-chat function provided by the SEWA data collection website. On average, the conversation was 3 minutes long. The discussion was intended to elicit further reactions and opinions about the advert and the advertised product, such as whether the advertised is to be purchased, whether it is to be recommended to others, what are the best parts of the advert, whether the advert is appropriate, how it can be enhanced, etc. After the discussion, each participant was asked to fill-in a questionnaire to self-report his/her emotional state and sentiment toward the discussion.

The SEWA data collection experiment was conducted using a website specifically built for this task. The website (http://videochat.sewaproject.eu) utilises WebRTC/OpenTok to facilitate the playing of adverts, video-chat, and synchronized audio/video recording using the microphone and webcam on the participants' own computer. This setup allowed the participants to be recorded in truly unconstrained "in-the-wild" environments with various lighting conditions, poses, background noise levels, and sensor qualities.

During the SEWA experiment, 199 recording sessions have been successful, with a total of 398 subjects being recorded. The subjects were coming from 6 different cultural backgrounds: British, German, Hungarian, Serbian, Greek, and Chinese. 201 of the participants are male, 197 are female, resulting in a gender ratio (male / female) of 1.020. The participants cover 4 age groups: 18~29, 30~39, 40~49, and 50+, with the 18~29 group being most numerous. The detailed participant demographics are shown in Table below.

Table: SEWA Participant Demographics

Cultural Background		Age Group		Length of acquaintance between dyads (in years)		Self-Reported Familiarity Rating	
British	66	**18~29**	203	**<1**	80	**Not Familiar**	9
				1	30		
German	64			**2**	39	**Slightly Familiar**	13
		30~39	94	**3**	40		
Hungarian	70			**4**	37	**Somewhat Familiar**	35
				5~9	55		
Serbian	72	**40~49**	25	**10~14**	20	**Moderately Familiar**	114
				15~19	22		
Greek	56			**20+**	75	**Extremely Familiar**	227
		50+	76				
Chinese	70						

A total of 1990 audio-visual recording clips (5 clips per subject: 4 recorded during the advert-watching part and 1 recorded during the video-chat part) were collected during the experiment. The clip duration ranges from 30 seconds to 3 minutes. The total length of the entire SEWA corpus is 2661 minutes (44.33 hours). The spatial resolution of the video recordings is 480x360 pixels for the advert-watching part or (effectively) 320x240 pixels for the video-chat part, respectively. Due to the wide spread of the participants' computers' hardware capacity, the video clips' frame rate varies between 10 FPS and 50 FPS, and the audio recordings' sample rate is either 44.1 kHz or 48 kHz. Audio and video signals were synchronized by means of time-stamping using the recording computer's reference clock, achieving a theoretical synchronisation accuracy of less than 16 ms.

6.2.3 Task 1.3: SEWA data annotation

Two versions (ComParE and GeMAPSv01a) of low-level acoustic feature descriptors and the 49 facial landmarks have been automatically extracted for all SEWA recordings. The methods used to extract these low-level features are described in section 6.3. Some facial landmarks localization results on the SEWA recordings are shown in Figure below. As illustrated by the figure, the facial landmark tracking results are reasonably accurate in most cases when the face is in a near-frontal pose and is mostly unclouded, even under challenging illumination conditions. The tracking results become worse when the face is at a profile angle and/or there is significant occlusion in the facial region (or when part of the face is not inside the camera's view). These frames will be corrected in a semi-automatic manner (Chrysos et al. 2015) to obtain the final annotation.

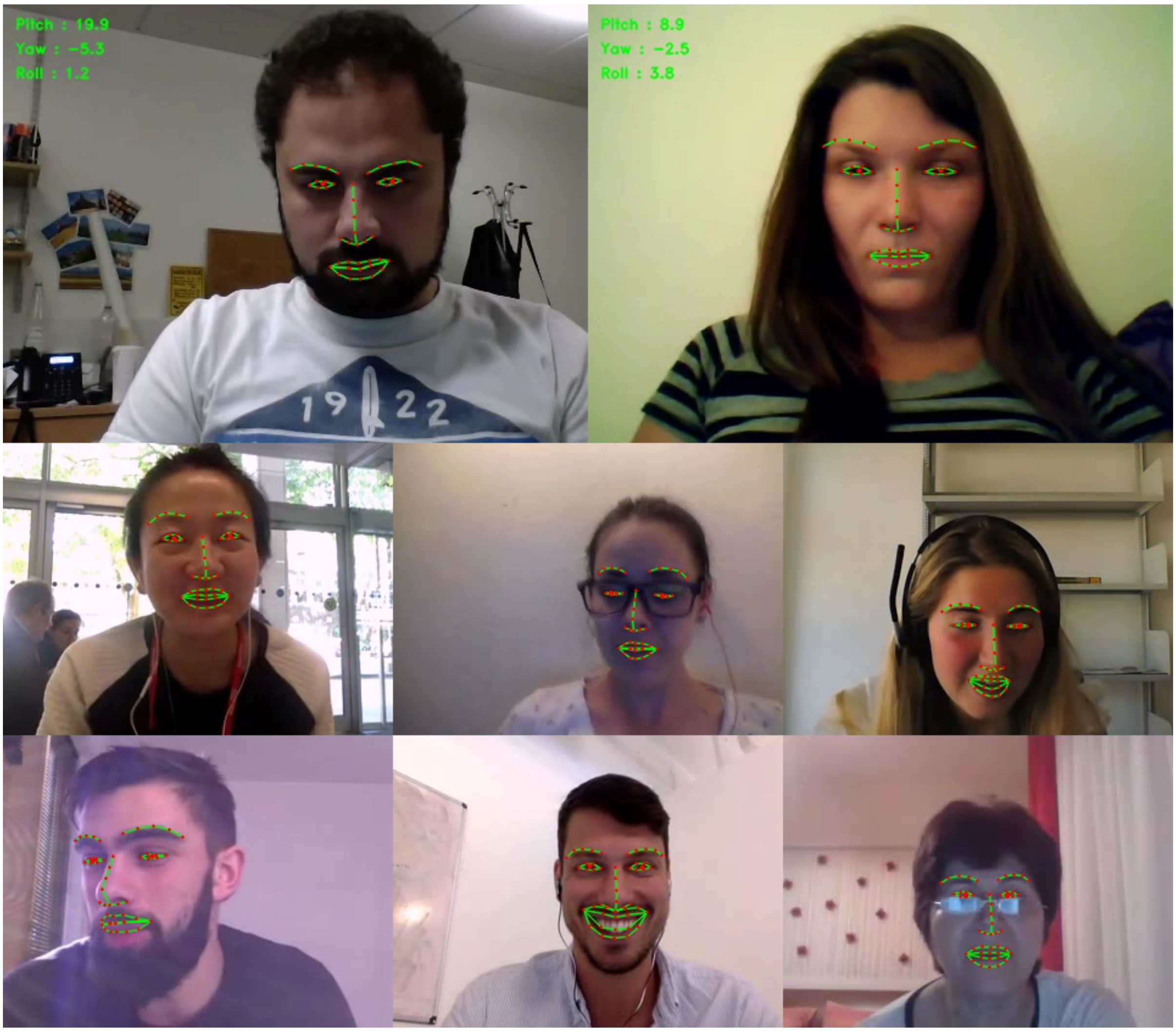


Figure: The automatic facial landmark localisation results obtained using the Chehra tracker (see section 6.2 for details). The top row shows the results on the advert-watching recordings and the two bottom rows show the results on the video-chat recordings.

We have also selected a list of 540 representative (in terms of arousal, valence, and liking/ disliking) segments from the SEWA corpus to be 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 -- “the core SEWA dataset”. Specifically, 90 segments were selected from the recordings of each culture group by the consensus of at least two expert annotators from the same cultural background (thus to be able to recognise the subtle display of emotions and/or sentiment specific to that culture). Within these 90 segments, 15 are of high arousal, 15 low arousal, 15 high valence, 15 low valence, 15 liking and 15 disliking. The duration of the segments ranges from 15 seconds to 1 minute. These 540 segments will form the core SEWA dataset, on which all SEWA technology will be demonstrated. A sample selection of these segments can be found at https://www.youtube.com/playlist?list=PL7d_oG0-e2PiofbLyHfIJegFBT3YEI01i.

For these selected segments, the annotations of valance, arousal and liking/ disliking were performed by annotators using joysticks and our purpose-built software tool for annotation. 5 annotators were recruited from each culture group to annotate the segments of the same culture. Moreover, to study the effect of different modalities, each segment has been annotated three times, first with audio data only, then video only and finally with both audio and video data. The annotation of FAU was conducted semi-autonomously using the tool based on variable-state latent conditional random fields (Walecki et al. 2015).

6.2.4 Task 1.4: SEWA database design and release

The SEWA database is organised into a nested list of folders, grouped by the participants’ cultural background. Each folder stores the data collected from a single recording session. The folder is further divided into two subfolders, each storing the data (including demographic profile, audio-visual data recorded during the experiment, and annotations) of one participant in the pair. Both the session folders and the participant folders are name sequentially, reflecting the order of the data being archived. The layout of the participant folder's content is as follows.

1. The participant’s demographic data is saved in a Java-script object notation (JSON) file. The participant’s identity (i.e., name and e-mail address) is removed from the record to anonymise the data.
2. Other data are organised into 5 sequentially named subfolders. The first 4 subfolders (“Advert0001” to “Advert0004”) contain the data recorded during the 4 advert-watching sessions and the 5th subfolder (“VideoChat0001”) contains the recordings of the online video-chat between the two participants. Specifically, each of these folders includes the following files and subfolders.

   a) An AVI file storing the video recording of the participant's reaction to the advert/discussion. The videos has a frame rate of 10~50 FPS (depending on the hardware capacity of the participant’s computer) and a spatial resolution of either 480x360 pixels (for the advert-watching recordings) or 320x240 pixels (for the video-chat recording).

   b) A WAV or AAC file storing the audio recordings of the participant's reaction to the advert/discussion. The audio data has a sample rate of either 44.1 kHz or 48 kHz.

c) A time-stamp file in JSON format, indicating start/end time of the recording.

d) A JSON file with the self-report provided by the participant at the end of the session.

e) A folder containing all available annotations. Each type of labels is saved in its own subfolder, of which the exact folder structure and/or file format may vary. "ReadMe" files are included in the subfolders to explain the specific data organisation method.

Along with the data, a comprehensive help document is also provided to give detailed explanation on the semantics of all JSON files being used in the database. For completeness, the database also contains the localised stimulus (adverts) used by participants from different cultural backgrounds.

Last but not least, the 540 representative segments and their annotations are also included in the SEWA database. They are stored in a separate folder adopting a similar internal structure to that of the main SEWA audio-visual corpus.

A web-portal for the SEWA database is currently under development. The portal will allow easy access and search of the available recordings according to various evidences (i.e. annotations of key cues like facial actions, mirroring, rapport, and liking/disliking) and according to various metadata (gender, age, cultural background, occlusions, etc.). 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 will be made available to researchers for academic-use only. To comply with clauses stated in the Informed Consent signed by the recorded participants (see the description of Ethical Issues above and the relevant appendices), all non-academic/commercial use of the data is prohibited. To enforce this restriction, an end-user license agreement (EULA) is prepared (see Appendix). 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, data will be protected by SSL (Secure Sockets Layer) with an encryption key. If at any point, the administrators of the SEWA database and/or SEWA researchers have a reasonable doubt that an authorised user does not act in accordance to the signed EULA, he/she will be declined the access to the database.

References:

- G. Chrysos, et al. (2015), "Offline deformable face tracking in arbitrary videos." ICCV-W, First Facial Landmark Tracking in-the-Wild Challenge and Workshop, 2015.
- R. Walecki, et al. (2015), "Variable-state latent conditional random fields for facial expression recognition and action unit detection." Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG'15), 2015.

6.3 Progress on WP 2 – Low-level Feature Extraction

6.3.1 Task 2.1: Environmentally robust acoustic features

The goal of Task 2.1 is to develop an improved acoustic feature extraction module, able to realise recognition of sentiment and emotion ’in the wild’, and in particular, being robust to additive and convolutional noise. Additive noises are sounds that superimpose the speech signal, e.g. traffic noise, babbling or technical noise introduced by the recording device. Convolutional noise usually presents itself in reverberation, depending on the recording environment.

For this purpose, we pursued different approaches. First of all, we investigated which acoustic features from **standard feature sets** are distorted less when introducing artificial noise by computing their linear correlation (Pearson’s correlation coefficient), i.e. the similarity between noisy and clean features. Figure below shows this similarity based on one recording from the RECOLA database (Ringeval et al., 2013) and three different noise types. The illustrated results are exemplary for many speech samples. We could figure out that energy, pitch, spectral and cepstral features are most robust against noise. Delta coefficients should be used with caution as they are heavily distorted by reverberation. Voice quality features (jitter and shimmer), zero-crossing rate, HNR and voicing probability should not be used in SEWA as they are severely distorted by noise. Therefore, we will focus on the LLDs energy, pitch and most spectral and cepstral features.

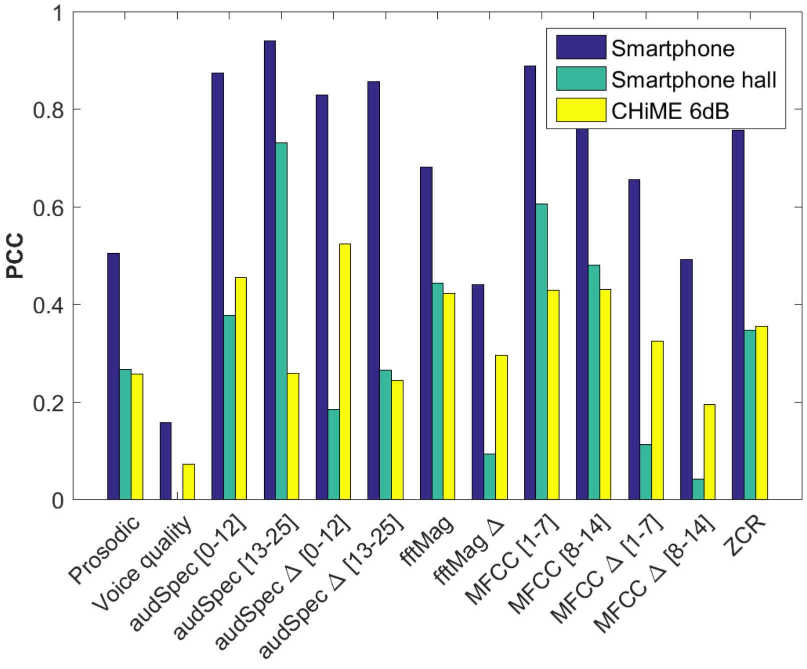


Figure: Pearson’s correlation coefficient between acoustic features computed from a clean speech signal and noisy versions of the same signal (RECOLA database)

Then, we employed the **Bag-of-Audio-Words (BoAW)** approach which is inspired by the Bag-of-Words in text retrieval and known to be robust against noise as it implies a vector quantization step. In this method, acoustic features (low-level descriptors, LLDs) are assigned to a limited number of templates (audio words), based on their similarity to them, and then the number of assignments is counted for each audio word. This results in a histogram (bag) of audio words which is then input to a classifier, such as support vector machine (SVM) or neural network (NN). Figure below exemplifies this approach.

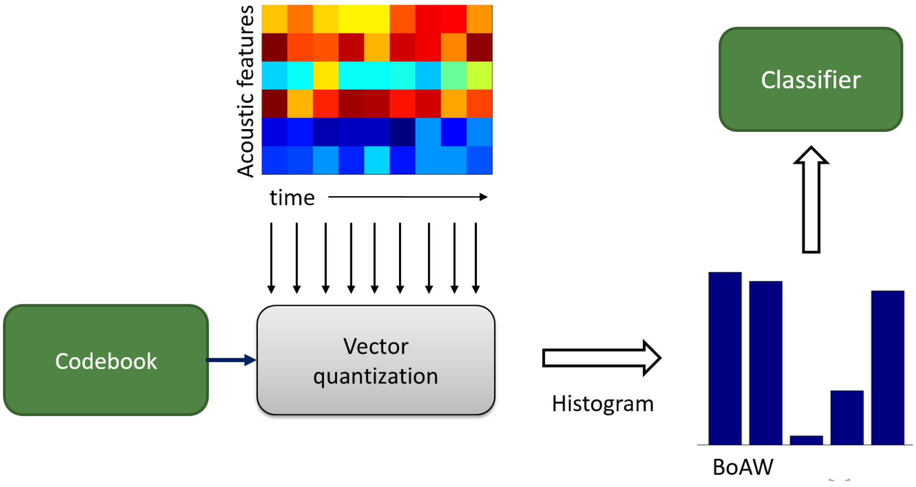


Figure: Basic principle of the Bag-of-Audio-Words (BoAW) method

The tool ***openWord*** has been implemented to investigate the BoAW method and will be used in the further progress of the SEWA project.

We can show that the BoAW method leads to a better recognition performance for several tasks, including emotion recognition in terms of arousal and valence, than standard feature sets, even if only MFCC features and energy are used. This finding is illustrated in the following table, where the results of prediction of arousal/valence with the baseline (i.e. state-of-the-art acoustic features without BoAW) are put in contrast to the prediction with the BoAW method. The accuracy is given in terms of concordance correlation coefficient (CCC), which is more reliable than the linear correlation (PCC). The results for development set (on which the method is optimised) and test set (previously unseen data) of RECOLA are shown.

Labels – data set	**Baseline [CCC]**	**BoAW [CCC]**
Arousal-dev	.287	.647
Arousal-test	.228	.587
Valence-dev	.069	.401
Valence-test	.068	.327

These results show that BoAW has an outstanding performance with the task at hand. Investigations on further emotion-related databases also prove that BoAW outperforms the standard methods in audio classification [Pokorny et al., 2015].

Moreover, we were able to prove that the performance is quite competitive in case of noisy audio signals, if **multi-condition training (MCT)** is applied. MCT means that we show the emotion recognition system both clean and noisy examples of speech with different emotions during training. For example, CCC for the prediction of arousal on the development set is still 0.584 if noise typical for smartphone recordings and reverberation is added. Thus, also on an artificially distorted version of RECOLA provides excellent results in terms of prediction accuracy, even better than the baseline on clean signals. Overall, this shows us that BoAW will lead to a high robustness in the wild. In future experiments, we will also employ the SEWA database for further investigations on the topic.

Besides BoAW by vector quantization, we have shown that the learning of BoAW-like representations using **Deep Semi-NMF** (non-negative matrix factorization) is also an option for the creation of new types of acoustic features. NMF basically reduces the dimensionality of the acoustic features over time as it tries to find several templates, i.e. generalizations, of feature vectors and then indicates which templates are present most at certain instants of time. However, as this method seemed not as profitable as vector quantization, we are focusing on standard BoAW in SEWA.

Furthermore, we investigated the possibility of **feature enhancement** by deep-learning of long short-term memory recurrent neural networks (LSTM), a common type of neural networks in machine learning and artificial intelligence. Feature enhancement means that once we have features extracted from noisy speech we want to remove the disturbing effect of noise from these features directly. We have shown that feature enhancement using **BLSTM-RNN** (bidirectional long-short term memory recurrent neural networks) can improve the prediction in case of additive noise. The performance for the prediction of arousal in terms of CCC without and with feature enhancement is illustrated in the following table. Noise at different levels (SNR) has been added to the RECOLA database. ‘0dB’ stands for the ‘worst case’, where the speech signal has the same level as the noise signal.

CCC	**clean**	**12 dB**	**9 dB**	**6 dB**	**3 dB**	**0 dB**
State-of-the-art	.661	.556	.526	.472	.420	.329
Feature enhancement	.467	.648	.631	.612	.521	.368

It must be investigated, if this approach is capable to enhance BoAW representations from noisy signals in the same way.

Recognizing emotion across-language is a key requirement within SEWA. We were able to show that training a recognition system in one or several languages and then applying it to previously unknown languages, works in principle [Feraru et al., 2015], although the performance is not outstanding. These investigations were using standard acoustic feature sets as the baseline features from INTERSPEECH 2013 Computational Paralinguistics Challenge (ComParE) [Schuller et al., 2013].

Concerning language-independence, we have shown that training a speech emotion recognition system on certain languages and using it on an unknown language, work in principle with the state-of-the-art, although the performance depends on the respective characteristics of language and culture.

As a **conclusion**, our efforts to create noise robust features or rather feature representations for real-life emotion recognition from speech succeeded. We tested our algorithms on realistic data, as the RECOLA database and evaluated also that both BoAW and feature enhancement works on a noisy version of SEWA database. We have shown that the results with the proposed methods are better than the state-of-the-art and we will repeat our investigations on the SEWA data, once the annotation procedure is finished.

In general, the BoAW method with openWord and multi-condition training and feature enhancement using BLSTM will be our main approaches to getting noise robust acoustic features in the scope of SEWA.

6.3.2 Task 2.2: Environmentally robust visual features

This task addressed the problem of joint detection of faces and facial landmarks in input videos. The main requirement here is the robustness to large and steady changes in head pose, illumination, occlusion, and facial expression. Several algorithms have been tested to this end (i.e., Tzimiropoulos & Pantic, 2014; Asthana et al., 2014, 2015) and the Chehra facial landmark tracker [Asthana et al., 2015] had the best performance on the SEWA data and was selected for further use in the project.

The Chehra facial landmark tracker [Asthana et al., 2015] is based on incremental training of discriminative models that use a cascade of linear regressors to learn the mapping from facial texture to the shape. For this, we exploit the fact that the cascade of regressors is trained using the Monte-Carlo sampling methodologies and present a very efficient methodology which can incrementally update all linear regressors in cascade in parallel. The advantage of our approach includes: (1) It is capable of adding new training samples and updating the model, without re-training from scratch, thereby, constantly increasing robustness of the generic model; (2) The tracker can automatically tailor themselves to the subject being tracked and the imaging conditions using image sequences, and hence, become person-specific over time.

The robustness of the Chehra model is currently enhanced further by training it on large amount of data including SEWA data annotated in a semi-automatic fashion in terms of facial landmarks (see Task 1.3).

6.3.3 Task 2.3: Cross-lingual language-related features

In Task 2.1, the *openWord* toolkit to generate bag-of-audio-words representations of utterances as acoustic features for emotion classification or regression has been implemented.

In this task, this toolkit was used to generate multi-lingual BoAW features for multi-language sentiment analysis. Furthermore, to make BoAW representation closer to real words, we generate it from fully automatic syllabification of speech data, or rather generate BoAW with variable length and investigate the performance of this amendment.

Moreover, we investigated the state-of-the-art of linguistic features, which are used by other researchers. With tools such as LIWC, the number or percentage of positive/negative words/phrases in the whole utterance are computed and listed separately. Linguistic features such as sentiment scores are provided to combine the information in both positive and negative sides to give a final decision. Besides, a novel word vector representation method using Google toolkit *word2vec* is also of our interest for affective analysis assignment.

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.

References:

- Schuller, B., et al. (2013) “The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism.” In: INTERSPEECH 2013: 14th Annual Conference of the International Speech Communication Association, Lyon, France, 25-29 August 2013.
- Ringeval, F., et al. (2013) "Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions." Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013.
- Feraru, S., et al. (2015): "Cross-Language Acoustic Emotion Recognition: An Overview and Some Tendencies", In Proceedings of 6th biannual Conference on Affective Computing and Intelligent Interaction (ACII 2015), (Xi’an, P. R. China), AAAC, IEEE
- Pokorny F., et al. (2015): "Detection of Negative Emotions in Speech Signals Using Bags-of-Audio-Words", 1st International Workshop on Automatic Sentiment Analysis in the Wild (WASA’15), satellite event to ACII 2015.
- Tzimiropoulos G. & Pantic M. (2014): “Gauss-Newton Constrained Local Models”, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR’14), pp. 1851-1858.
- Asthana, A., et al. (2014): “Incremental face alignment in the wild.” In IEEE Conference on Computer Vision and Pattern Recognition (CVPR’14), pp. 1859-1866.
- Asthana, A., et al. (2015): “From pixels to response maps: Discriminative image filtering for face alignment in the wild”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 6, pp. 1312-1320.

6.4 Progress on WP3 – Mid-level feature extraction

The tasks in this WP are split so as to obtain the necessary mid-level features needed for high-level interpretation of human behaviour including sentiment.

6.4.1 Task 3.1: Extraction of head and face-touching gestures

The extraction of the mid-level features was done in terms of head-pose and facial landmarks using the Chehra facial landmark tracker [Asthana et al., 2015]. The extracted head-gestures in terms of head-pose will be used in T3.2 and T3.3 to perform fusion of facial landmarks and these gestures to achieve more robust estimation of AU intensity levels. The automated extraction of touching –the-face gestures have been attempted in two ways by analysis of dynamic hand movement and static face touching. To retrain existing tools for automated detection of these gestures, the SEWA video clips are currently being manually annotated. This will be followed by a validation round.

6.4.2 Task 3.2: The work done so far

T3.2 has mainly been concerned with improving over the existing methods for intensity estimation of AUs as well as AU segmentation from image sequences. Improving the quality of the extracted mid-level features (intensity estimation and detection of AUs) was necessary in order to facilitate the high-level interpretation of human behaviour like sentiment. The developed approaches are described below.

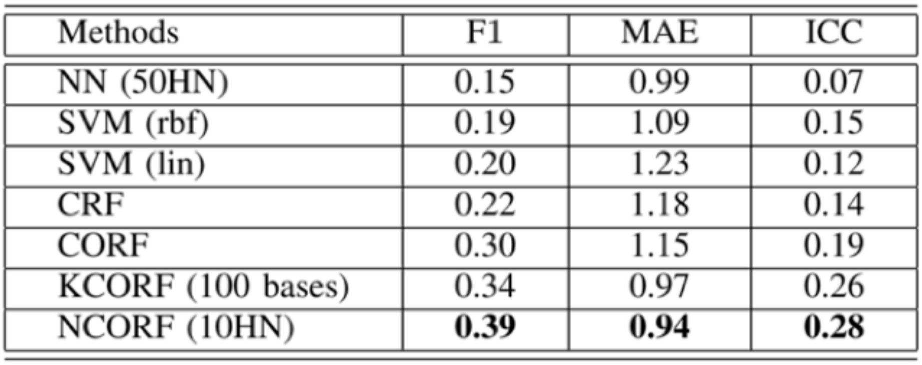
a) Frame-level-intensity estimation

The main contribution of the work done here is a novel methodology for dynamic intensity estimation of facial actions (the target mid-level features). It offers an effective nonlinear modelling of facial features for the target task by combining the Neural Networks (NN) for nonlinear feature transformation and Conditional Ordinal Random Fields' (CORF) dynamical ordinal modelling capabilities [Rudovic et al., 2012].

- A Neural Conditional Ordinal Random Field (NCORF) model has been proposed where the usual CORF's linear projection has been extended to a nonlinear one provided through a Neural Network.
- Since the annotation of the suitable SEWA data is pending, we tested the proposed approach on the task of estimating inter-person agreement intensity levels from images of spontaneously displayed facial expressions [Rakicevic et al. 2015]. We addressed this problem using existing MAHNOB dataset. Once it is ready, the (partially) annotated SEWA dataset will be used to re-train/evaluate the proposed method and perform AU intensity estimation (and estimation of the intensity of liking/disliking, valence, and arousal in WP4). The MAHNOB-Mimicry database has been used and manually annotated per frame, specifically in terms of ordinal agreement levels according to the Likert scale (Strong disagreement, Disagreement, Neutral, Agreement, Strong Agreement).
- The input features used are the normalised coordinates of the 49 tracked facial points.
- In total, five 15 min videos containing 5 different subjects discussing various topics have been employed and segmented.

- In total, 329 sequences have been used, on average 80 frames long.
- 5 fold subject-independent cross validation of the compared models.
- The performance of the proposed model has been compared with static baselines such as: Neural Networks and Support Vector Machines (various kernels); as well as dynamic state of the art models such as: Conditional Random Fields (CRF), linear CORFs and Kernel CORFs (different kernel sizes).

Table: Performance of different methods

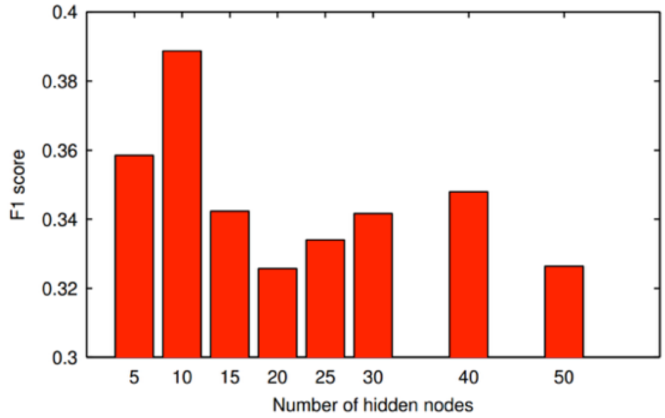


Methods	F1	MAE	ICC
NN (50HN)	0.15	0.99	0.07
SVM (rbf)	0.19	1.09	0.15
SVM (lin)	0.20	1.23	0.12
CRF	0.22	1.18	0.14
CORF	0.30	1.15	0.19
KCORF (100 bases)	0.34	0.97	0.26
NCORF (10HN)	**0.39**	**0.94**	**0.28**

In Table above, we show that the proposed method (NCORF) outperforms the state-of-the-art methods in the task for agreement level estimation from facial images. This is mainly attributed to the newly introduced NN layer in the CORF model that allowed us to find better low-level feature representations from the input facial landmarks. As evidenced by the results, these are needed to obtain better estimation of the target mid-level features (in this case, agreement level). We showed this using three commonly employed evaluation measures for intensity estimation – Intra-class correlation (ICC), mean absolute error (MAE) and the standard F-1 score.

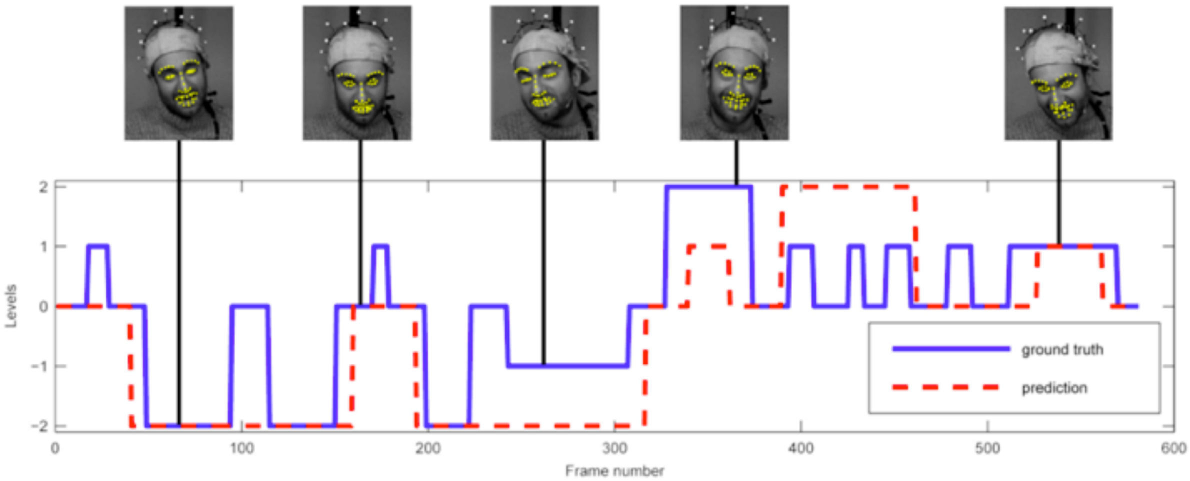
To find optimal structure of the NN, and thus improve estimation of the target intensity levels, various architectures for the NN part of the NCORF model have been evaluated and the best performing was the one layer with 10 hidden nodes (see Figure below).

Figure: F1 score vs. number of hidden layers in NCORF



Finally, from Figure below we can see that the NCORF model is able to discriminate well between more subtle intensity levels e.g. strong (dis)agreement vs. (dis)agreement. This is also very important for the task of AU intensity estimation in order to accurately distinguish among the AU intensity levels.

Figure: Intensity estimation of agreement in an example sequence



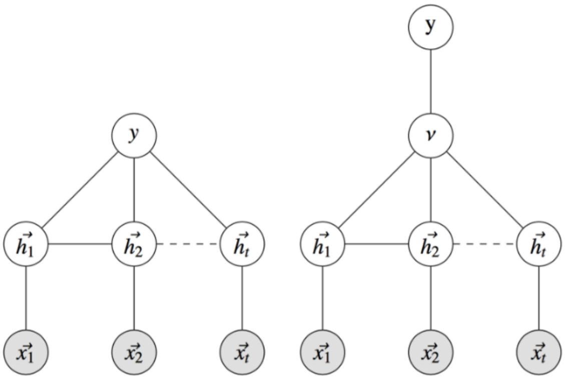
This work was a necessary step in order to develop a more complex model for feature extraction from facial appearance using the latest advances in deep neural networks (DNNs). DNNs have recently shown a great performance in various related tasks (from object detection to face analysis). The ongoing work consists of extensions of the NCORF model to include feature extraction based on DNN, and also perform fusion of mid-level features from T3.1 (head-pose, tilts, nodes), facial landmarks and audio-features. When applied to the SEWA dataset, these are expected to give significant improvements in extraction of target mi-level features (AUs and their intensity) over the existing methods. As some AU cannot be accurately or at all detected from facial landmarks solely, we are currently performing evaluation of the NCORF model with appearance based features (Local Binary patterns extracted around the locations of facial landmarks). We use these sparse features instead of the proposed dynamic appearance features (see T3.2) in order to reduce the computational load of the used models and achieve time AU estimation performance.

b) Facial action unit segmentation from image sequences

In order to perform extraction of mid-level features in T3.2, the first step is localization of image sequence segments where the target AU is active, i.e., have intensity level higher than 0. Although one may try identifying such segments directly by applying the intensity level detectors, this results in a low performance of the extractors due to the highly imbalanced nature of the active vs. non-active segments, leading to an overall high false-positive rates in the detection task. To address this, we proposed a novel method for detection of active segments of AUs from images sequences. The critical task here is to separate modelling of active (intensity higher than 0) and non-active segments of AUs.

We addressed this by introducing a novel extension of the state-of-the-art Latent Conditional Random Fields (L-CRF) framework [Wang et al. 2006] and its extensions Conditional Ordinal Random Fields [Rudovic et al 2012]. These frameworks allow us to efficiently encode dynamics through the latent states accounting for the temporal consistency in emotion expression. However, its latent states are typically assumed to be either unordered (nominal), or ordered (ordinal). Yet, such approach is often too restrictive. For instance, in the case of AU detection, the goal is to discriminate between the segments of an image sequence in which this AU is active or inactive. While the sequence segments containing activation of the target AU may better be described using ordinal latent states (corresponding to the AU intensity levels), the inactive segments (i.e., where this AU does not occur) may better be described using unordered (nominal) latent states, as no assumption can be made about their underlying structure (since they can contain either neutral faces or activations of non-target AUs). To this end, we introduced a novel Variable-state L-CRF (VSL-CRF) model [Walecki et al. 2015] for classification of image sequences that, in contrast to existing L-CRF models, has flexibility to use either nominal or ordinal latent states for modelling the underlying dynamics of target sequences. The proposed model selects automatically the optimal latent states for each target sequence. The outline of the proposed VSL-CRF is given in Figure below.

Figure: The graph structure of the (left) traditional Latent CRF models H-CRF/H-CORF, and (right) proposed VSL-CRF model. In H- CRF/H-CORF, the latent states h, relating the observation sequence x = {x1, . . . , xT } to the target label y (e.g., emotion or AU activation), are allowed to be either nominal or ordinal, while in VSL-CRF the la- tent variable ν = {nominal, ordinal} performs automatic selection of the optimal latent states for each sequence.

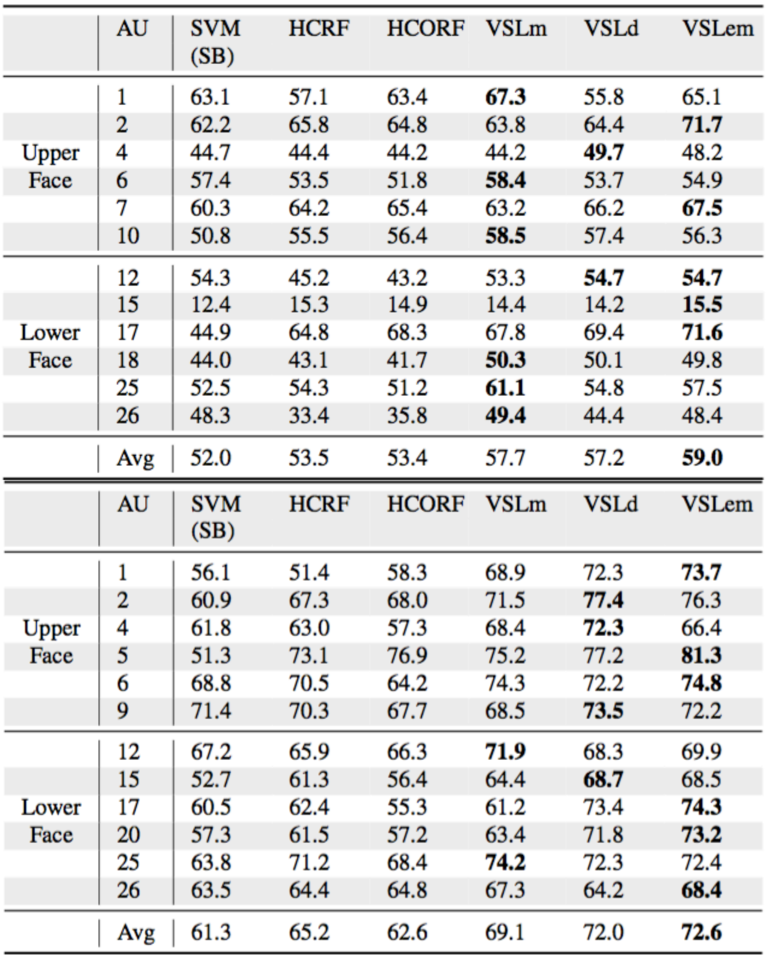


Experiments and evaluation:

- We created a training datasets that consists of active and not-active subsequences of upper and lower face AU
- We used the locations of 49 facial points, extracted from target images sequences using Chehra facial landmark tracker [Asthana et al., 2015].
- The pre-processing of the features was performed by first applying Procrustes analysis to align the facial points to the mean faces of the datasets.
- We then applied PCA to reduce the feature size, retaining 97% of energy.

- We applied three learning algorithms based on max-pooling, the Expectation Maximization-like learning of the latent states and the direct optimization by marginalizing out the latent states.
- We used the graph-Laplacian regularization of the model parameters, for efficient training of the proposed VSL-CRF model. This result in a model that is less prone to overfitting of target data compared to when traditional maximum-likelihood learning (ML) approach is used, as in L-CRF models such as H-CRF and H-CORF.
- We show on three publicly available datasets, the CK+, the GEMEP-FERA and the DISFA database that, the VSL-CRF model can better learn the underlying dynamics of target facial expressions and outperforms the standard HCRF methods and 11 recently published related methods for frame and sequence classification.

Table: F1-sequence-based results on the DISFA dataset (up) and the GEMEP-FERA dataset (down)



	AU	SVM (SB)	HCRF	HCORF	VSLm	VSLd	VSLem
Upper Face	1	63.1	57.1	63.4	**67.3**	55.8	65.1
	2	62.2	65.8	64.8	63.8	64.4	**71.7**
	4	44.7	44.4	44.2	44.2	**49.7**	48.2
	6	57.4	53.5	51.8	**58.4**	53.7	54.9
	7	60.3	64.2	65.4	63.2	66.2	**67.5**
	10	50.8	55.5	56.4	**58.5**	57.4	56.3
Lower Face	12	54.3	45.2	43.2	53.3	**54.7**	**54.7**
	15	12.4	15.3	14.9	14.4	14.2	**15.5**
	17	44.9	64.8	68.3	67.8	69.4	**71.6**
	18	44.0	43.1	41.7	**50.3**	50.1	49.8
	25	52.5	54.3	51.2	**61.1**	54.8	57.5
	26	48.3	33.4	35.8	**49.4**	44.4	48.4
	Avg	52.0	53.5	53.4	57.7	57.2	**59.0**

	AU	SVM (SB)	HCRF	HCORF	VSLm	VSLd	VSLem
Upper Face	1	56.1	51.4	58.3	68.9	72.3	**73.7**
	2	60.9	67.3	68.0	71.5	**77.4**	76.3
	4	61.8	63.0	57.3	68.4	**72.3**	66.4
	5	51.3	73.1	76.9	75.2	77.2	**81.3**
	6	68.8	70.5	64.2	74.3	72.2	**74.8**
	9	71.4	70.3	67.7	68.5	**73.5**	72.2
Lower Face	12	67.2	65.9	66.3	**71.9**	68.3	69.9
	15	52.7	61.3	56.4	64.4	**68.7**	68.5
	17	60.5	62.4	55.3	61.2	73.4	**74.3**
	20	57.3	61.5	57.2	63.4	71.8	**73.2**
	25	63.8	71.2	68.4	**74.2**	72.3	72.4
	26	63.5	64.4	64.8	67.3	64.2	**68.4**
	Avg	61.3	65.2	62.6	69.1	72.0	**72.6**

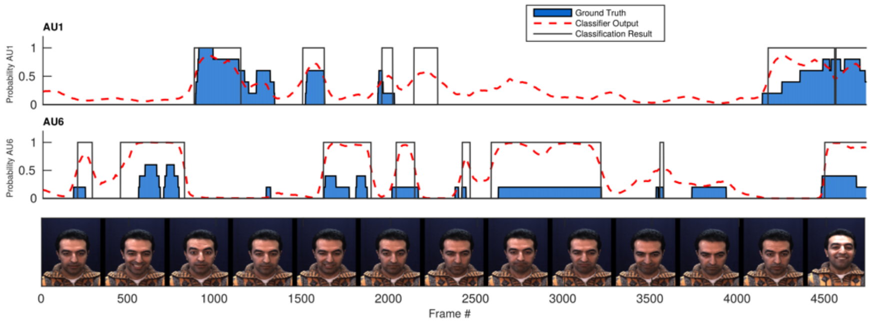


Figure: The segmentation of target AUs obtained using the VSL-CRF model with 3 hidden states and a window size of 8 frames

In Table above, we show that the three versions of the proposed method outperform the state-of-the-art approaches for the target task. Furthermore, we show that the VSLem approach (that perform EM learning of the model parameters) achieves the best results on average. This is because the proposed learning strategy is less prone to overfitting compared to the other two optimization approaches proposed. Finally, Figure above depicts an image sequence from DISFA dataset where the goal is to identify active segments of AU 1 and 6 of the target subject. As can be seen, in most cases, the proposed method accurately segments the target sequences. As with NCORF model, this model is trained with input features based on facial landmarks. Currently, we ate investigating the use of appearance-based features to tackle the AUs that cannot be detected solely from facial landmarks. Furthermore, we extracted these features from two publicly available datasets of spontaneously displayed AUs (DISFA and PAIN). These are used to train the VSLem methods for AU segmentation that will be used for automated segmentation of the SEWA videos.

c) Joint Estimation of AUs using a novel feature-fusion approach

We also developed a novel methodology for AU detection from static images. The main aim of this approach is to be able to efficiently exploit both facial points and appearance –based features (LBPs extracted around the target landmarks) and perform simultaneous detection of multiple AUs from static images. The proposed Multi-conditional Latent Variable Model (MC-LVM) [Eleftheriadis et al., 2015; see Figure below] has been trained and evaluated on three publicly available datasets: CK+ of AUs from posed data, and DISFA and PAIN datasets of AUs from spontaneously displayed facial expressions. The proposed method is based on Gaussian Processes (GP) for non-linear function learning and outperforms the state-of-the-art methods for the target task. The outline of the method is depicted below.

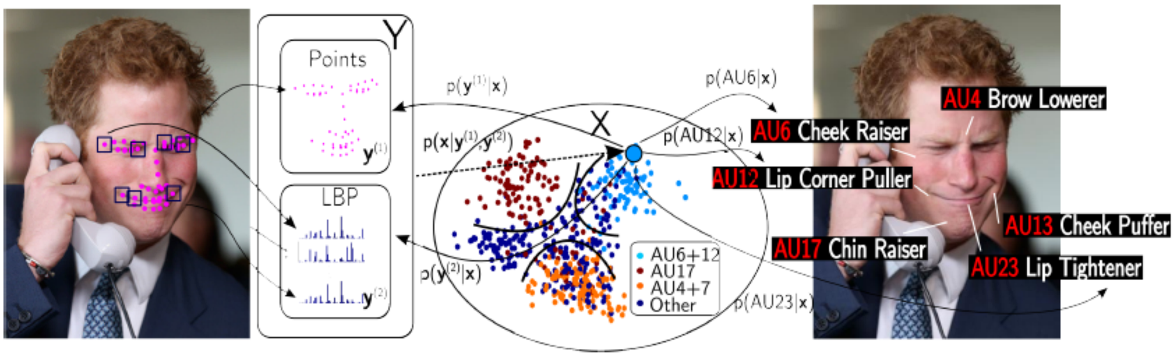


Figure: The proposed MC-LVM. The geometrical and appearance input features, y^(1) and y^(2), are first projected onto the shared manifold X. The fusion is attained via GP conditionals, p(y^ (1) |x) and p(y^(2)| x), that generate the inputs. Classification is performed on the manifold via simultaneously learned logistic functions p(z^(c)| x) for multiple AU detection. The subspace is regularized using constraints imposed on both latent positions and output classifiers, encoding local and global dependencies among the AUs.

In Figure below we show that the proposed approach effectively leverages the information from the two facial representation by achieving improved estimation of AUs (in terms of AUs). We show this on three datasets and on AUs (1, 2, 4, 6, 7, 12, 15, 17) for CK+, AUs (1, 2, 4, 6, 12, 15, 17) for DISFA and AUs (4, 6, 7, 9, 10, 43).

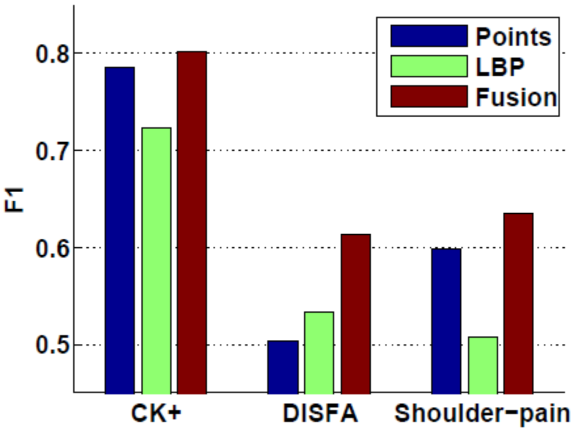
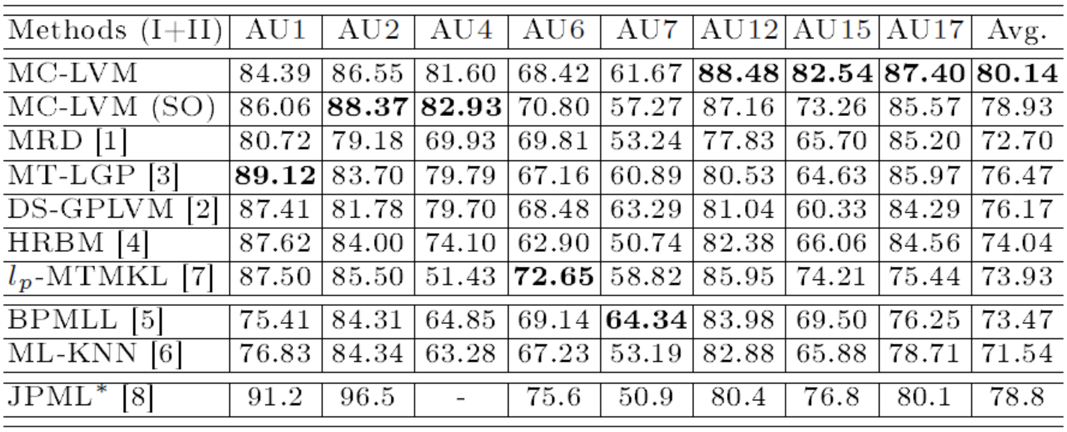


Figure: Average F-1 for AU estimation using the proposed MC-LVM approach

The main results showing that the proposed MC-LVM outperforms a number of state-of-the-art approaches for the target task are summarized in Table below. Furthermore, we showed the advantages of joint AU modelling (MC-LVM) over the same model with independent modelling of AUs (MC-LVM (SO)). For description of compared methods, see [Eleftheriadis et al 2015].

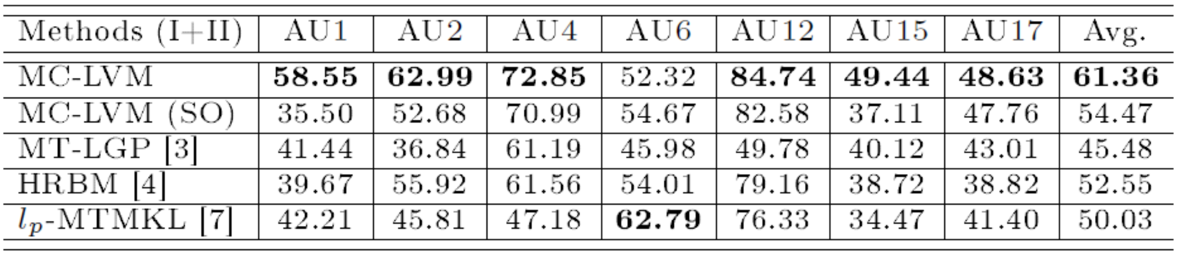
Table: F1 score for joint AU detection

CK+



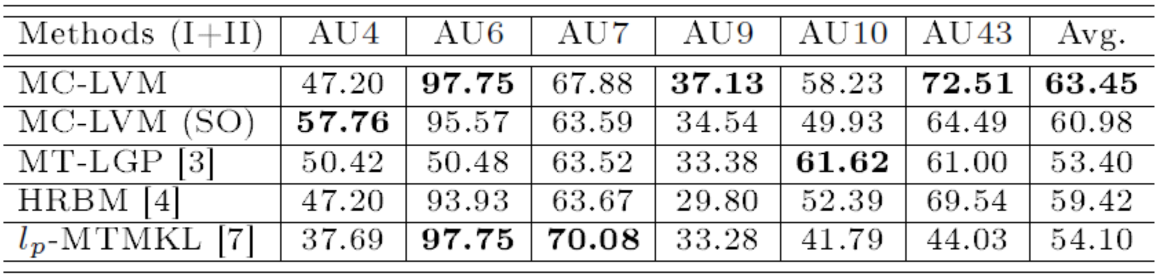
Methods (I+II)	AU1	AU2	AU4	AU6	AU7	AU12	AU15	AU17	Avg.
MC-LVM	84.39	86.55	81.60	68.42	61.67	**88.48**	**82.54**	**87.40**	**80.14**
MC-LVM (SO)	86.06	**88.37**	**82.93**	70.80	57.27	87.16	73.26	85.57	78.93
MRD [1]	80.72	79.18	69.93	69.81	53.24	77.83	65.70	85.20	72.70
MT-LGP [3]	**89.12**	83.70	79.79	67.16	60.89	80.53	64.63	85.97	76.47
DS-GPLVM [2]	87.41	81.78	79.70	68.48	63.29	81.04	60.33	84.29	76.17
HRBM [4]	87.62	84.00	74.10	62.90	50.74	82.38	66.06	84.56	74.04
l_p-MTMKL [7]	87.50	85.50	51.43	**72.65**	58.82	85.95	74.21	75.44	73.93
BPMLL [5]	75.41	84.31	64.85	69.14	**64.34**	83.98	69.50	76.25	73.47
ML-KNN [6]	76.83	84.34	63.28	67.23	53.19	82.88	65.88	78.71	71.54
JPML* [8]	91.2	96.5	-	75.6	50.9	80.4	76.8	80.1	78.8

DISFA



Methods (I+II)	AU1	AU2	AU4	AU6	AU12	AU15	AU17	Avg.
MC-LVM	**58.55**	**62.99**	**72.85**	52.32	**84.74**	**49.44**	**48.63**	**61.36**
MC-LVM (SO)	35.50	52.68	70.99	54.67	82.58	37.11	47.76	54.47
MT-LGP [3]	41.44	36.84	61.19	45.98	49.78	40.12	43.01	45.48
HRBM [4]	39.67	55.92	61.56	54.01	79.16	38.72	38.82	52.55
l_p-MTMKL [7]	42.21	45.81	47.18	**62.79**	76.33	34.47	41.40	50.03

PAIN



Methods (I+II)	AU4	AU6	AU7	AU9	AU10	AU43	Avg.
MC-LVM	47.20	**97.75**	67.88	**37.13**	58.23	**72.51**	**63.45**
MC-LVM (SO)	**57.76**	95.57	63.59	34.54	49.93	64.49	60.98
MT-LGP [3]	50.42	50.48	63.52	33.38	**61.62**	61.00	53.40
HRBM [4]	47.20	93.93	63.67	29.80	52.39	69.54	59.42
l_p-MTMKL [7]	37.69	**97.75**	**70.08**	33.28	41.79	44.03	54.10

To summarize, for task T3.2 we proposed three novel methods for AU detection, intensity estimation and active segment detection form static face images and image sequences. These methods account for different aspects of the context in which the target facial AUs occur: while all the methods focus on modeling of the context questions “when“ and “how“, the methods in (a) & (b) account for the time and single-feature modeling, respectively, while the method described in (c) focuses on the context questions “when“ by modeling co-occurences of different AUs, and “how“ through the feature fusion. We also performed preliminary studies on modeling of the context question “who“ using the context-sensitive CORF model (cs-CORF, Rudovic et al., 2015) by applying it to the spontaneous data of AUs from DISFA&UNBC-PAIN datasets. We are currently investigating the performance of this method when applied to the task of the AU intensity estimation from the SEWA dataset, as well as its extensions that combine the joint modeling of AUs and sequence-segmentation approaches described in (a)-(c).

References:

- Rakicevic et al. (2015): “Neural Conditional Ordinal Random Fields for Agreement Level Estimation”, In Proceedings of 1st Int’l Workshop on Automatic Sentiment Analysis in the Wild (WASA), satellite event to ACII 2015.
- Eleftheriadis et al. (2015): “Multi-conditional Latent Variable Model for Joint Facial Action Unit Detection”, In Proc. of the Int’l Conf. on Computer Vision (ICCV’15).
- Walecki et al. (2015): “Variable-state Latent Conditional Random Fields for Facial Expression Recognition and Action Unit Detection”, In Proc. of the IEEE Int’l Conf. on Automatic Face and Gesture Recognition (FG’15).
- Rudovic et al. (2015): “Context-sensitive Dynamic Ordinal Regression for Intensity Estimation of Facial Action Units”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 37, pp. 944-958, 2015.
- Rudovic et al. (2012): “Multi-output Laplacian Dynamic Ordinal Regression for Facial Expression Recognition and Intensity Estimation”, In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR’12).
- Wang et al. (2006): “Hidden conditional random fields for gesture recognition”, In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR’06).

6.5 Progress on WP4 – Continuous Affect and Sentiment Sensing in the Wild

Work on WP4 will start in M15 (April 2016).

6.6 Progress on WP5 – Behaviour Similarity in the Wild

Work on WP5 will start at the end of M12 (from 1st February 2016).

6.7 Progress on WP6 – Temporal Behaviour-Patterning and Interpersonal Sentiment in the Wild

Work on WP6 started just in M12 (January 2016).

6.8 Progress on WP7 – Integration, Applications and Evaluation

6.8.1 Chat Social Game by PlayGen

a) Objectives

The overall objective of this stream of the WP is to develop a Chat Social Game application where tools from WP2-6 will be incorporated. This WP started in M6. The overall objectives during this reporting period (M1-12) were:

- To complete the definition phase of the game development cycle: specify target users, behaviours, needs and corresponding functionalities for Chat Social Game v1
- To develop initial designs for the game
- To plan how the tools from WP2-6 will be technically integrated into the game.

b) Tasks

Task 7.2: SEWA Chat Social Game – development of Chat Social Game version 1

Task 7.3: User Groups and user studies – definition of user groups, discussion with users for Chat Social Game version 1

Task 7.5: Evaluation – plan evaluation approach

c) Results

The objectives for development of Chat Social Game during M1-12 were achieved and development of the Chat Social Game is well under way. Specific results included:

- A comprehensive assessment of potential exploitation of SEWA automatic voice and audio analysis was carried out. This was used to adapt and develop the plans for the Chat Social Game so as to deliver the most impact from the application. The focus of the Chat Social game was refined as a result of this analysis to focus on feedback that could improve communication skills. Full details of the criteria and assessment are available in the document 'Chat Social Game workplan and exploitation v2.0' available on the SEWA project website (internal use, accessible by the Commission).

- The target user group for the Chat Social Game was specified – university students aged 18+. This group was selected for two reasons: 1) the potential benefit they would get from a game to develop negotiation and persuasion skills using videochat, and 2) the data collection process showed that videochat is a more natural mode of communication for younger age groups, who would therefore be more suitable for testing a 'proof of concept' application. PlayGen received endorsement for this rationale from three academic institutions and partners in the Valorisation Board. Three universities (Imperial, Queen Mary's University of London and City University) have been signed up as partners to support recruitment of users for focus groups, testing and publicising each release of the Chat Social Game.

- Two user focus groups have been conducted in order to identify user needs, test the broad concept of the game and identify technical specifications. The focus groups explored the advantages and disadvantages of existing tools to develop persuasion and negotiation skills, students' views of an ideal chat social game, their ideas about game mechanics that could support engagement and feedback on the initial Chat Social Game concept. Full results of the focus groups and a corresponding mapping onto technical specifications for the game are available in D7.1 Report on user requirements.

- An initial game design concept was developed prior to user focus groups. This was refined following the user groups and a first version mockup of design and visuals that meet user needs has been developed.

- The core technical functionality to select a chat partner from a social network and carry out a videochat has been developed. Critically, this has been developed in a way that reflects the storage and technical integration requirements of the SEWA automatic face and audio analysis that will be incorporated in V2.

- The evaluation approach for the Chat Social Game has been planned. This has included setting a target number of 100 users to test each version and developing a strategy to meet these targets. The core strategy will be to ensure that the game itself is useful. Prompts will also be built in to re-engage students with new versions – Users will be issued new challenges (e.g. new topics to debate, new goals for number of debates or scores to achieve in a fixed period, comparisons with peer scores) and invited to join specific sessions when a new version of the game is released. Social game mechanics such as leaderboards, social proof and peer feedback will be used to motivate users to put their skills to the test again. Incentives for structured testing sessions will be used as a fallback if user numbers for any version fall short of the target.

  A strategy has also been developed for including a small cohort of users that test each version of the game successively. We propose recruiting a small cohort of users (25-30) to test each version of the game in succession at the end of the project once all 3 versions are available. This will be more likely to identify benefits and challenges related purely to the experience of the different game versions rather than contextual differences for the students. However, recruitment of users for V1 of the game will target first years so that some of these users are still available to test later versions of the game, even assuming a high dropout rate. Full details of the strategy to ensure sufficient user engagement for evaluation purposes are available in the document 'User data collection strategy' on the SEWA website (internal use, accessible by the Commission).

  The initial evaluation plan has also identified the key metrics that will need to be captured during gameplay in order to assess the effectiveness of the SEWA automatic face and audio analysis compared with self-reported feedback alone. This has been incorporated into the technical specifications of the Chat Social Game.

6.8.2 Advert Recommender Engine by RealEyes

a) Objectives

The main goal of this subtask within WP7 is to develop an Advert Recommender Engine tool that will integrate novel methods to be developed in WP2-WP6. The objectives within the reporting period (Y1, starting from M6) were the following:

- To refine the original concept of enhancing recommendation systems by incorporating user feedback via sentiment analysis on behavioural cues.
- To develop initial designs for the assessment of the performance of the enhanced methods.
- To plan how the tools developed by the consortium partners will be integrated.

b) Tasks

Task 7.1: Sentiment-driven multimedia-content recommender version 1

Task 7.3: User groups and user studies – definition of user groups, discussions with commercial partners, collection of data for evaluation of recommender engine

Task 7.5: Evaluation – plan evaluation approach

c) Results

The objectives for development of Advert Recommender Engine during M1 - 12 were achieved. In particular, the results include the following:

- Conducting market analysis with help of RealEyes' clients and partners to assess the business opportunities created by a video recommender engine. This involved in depth discussions with RealEyes sales and product development team, analysis of the type of services currently purchased from RealEyes, feedback from company clients, advisors and industry experts. Based on the analysis a product plan has been developed for a video recommender engine. We then updated the definition of the recommendation engine in line with the findings of the market research and product development plans. The change enhanced and focused recommendation engine on programmatic advertising industry where it could be offered in the form of Software as a Service.

- As a consequence of the changes in the exploitation plan, the target user groups have been redefined, changing the target user groups from movie fans and movie rental companies to online programmatic video advertisement market. Online video advertisement market consists of two main user groups: consumers of online video advertisements and video advertisers. The latter can be further split into ad exchange platforms, DMPs, DSPs and SSPs.

- The evaluation approach for the Advert Recommender Engine has been planned. This has included both of the user groups. To measure the impact of advert recommendation engine on video advertisement consumers we will test 200 participants in each of the control and exposed groups. Control group participants will be subject to random ad video presentation, while exposed group will be subject to ad videos recommended to them by the engine. The difference in audio/visual emotional and survey response will be used to evaluate the recommendation engine's performance. For this part of evaluation we have built a framework to collect audio, video and questionnaire data in response to video content for control and expose groups.

- The impact of advert recommendation engine on online AdTech companies will be measured through lift of audio/visual and survey response of participants as well as lift in social media performance or sales performance of the tested advertisements. Although this part of evaluation is for the later stage of the project, we have already identified key partners in collaboration with whom the recommendation engine will be developed and evaluated. These customers include big FMCG company, who has already provided us with valuable sales lift data, Maker Studios social media performance data, Xaxis as one of the partners in the Valorisation board will be working with us, we are also working on establishing partnerships with AOL One Video and Annalect (Neustar).

- For the purposes of mass data collection needed to train and test a recommender system, we have built a framework to record audio, video, and questionnaire data online in response to video content. The framework has been used to collect user responses for over 150 advert videos (among which are the 4 advert videos used as stimuli material in the SEWA data collection). Each of the 150 videos is associated with a performance measure (i.e., advert is successful/ unsuccessful), which was provided to us by our clients. This information and the behavioural responses by users constitute a unique dataset, which can be analysed for patterns of users' behaviour being predictive of the ultimate performance of an advert. In addition, a new framework has been designed and is being built that allows for quick statistical testing of "emotional profiles of ads" (which, in essence, determine which people would react how on a specific ad) and their use for predicting ads' performance. Finally, a pilot study has been carried out on the possible use of the collected data for building emotional profiles of adverts. Initial findings are positive and confirm our expectations, namely, that emotional reactions of users can be predictive of ads' performance and that they can be used for better targeting. Once we receive approval of our commercial partners to share the findings, we plan to publish these results in academic and market research papers.

6.9 Progress on WP8 – Dissemination, Ethics, Communication and Exploitation

6.9.1 Task 8.1: SEWA website and e-services

The consortium has set up a web site: www.sewaproject.eu as a dissemination tool to be maintained during the project and beyond. It has been set up with general information about SEWA, members of the consortium, the objectives and results of the project, up-to-date news about all dissemination efforts, including the information about project presence in conferences, fairs, exhibitions, etc. The website facilitate subscription to project news and events, download of public deliverables, download of publications related to the project, download of released software tools and data, download of demonstration videos, and download of video-lectures recorded during the project-related events. It also facilitate a project-private space for filing deliverables and internal reports.

Statistics related to the SEWA website/ facebook are as follows:

- Total number of visitors: 3323
- Average number of pages seen per visit: 1,1
- Mostly used browser languages: English and German
- Facebook (https://www.facebook.com/sewaprojecteu/): 48 likes

6.9.2 Task 8.2: Valorisation Advisory Board

To get further advice on the design and architecture of SEWA applications and to investigate commercialization potential of each aspect of SEWA technology, it was agreed at the start of the project to form a Valorisation Advisory Board open to interested external parties that have a business broadly related to the HCI and FF-HCI. The list of partners in the Valorization Board is as follows:

- Market research industry – IPSOS
- Online telecommunication and interviewing – Jobatar
- Online advertising industry – VisualDNA and Xaxis
- News industry – Financial Times
- Automotive industry – Daimler
- Basic technology providers (data collection platforms and signal processing) -- Anaii d.o.o, and audEERING.

Organisation of the initial Valorisation Board Meeting was the first important milestone of the project (M1). The meeting was held on the 1st October 2015, and was attended by the SEWA partners and the representatives of Jobatar, Visual DNA, Anaii d.o.o, Financial Times, IPSOS Connect, and audEERING.

The meeting was organised in in a workshop style meeting. Members of the board were updated on the goals of SEWA project, the planned industrial applications and consulted on the needs of their respective industries. One of the main goals of the meeting was to collect board members feedback on the usefulness of the developed technologies and applications, as well as hear recommendations for further commercialization potentials of the developed technologies. Agenda of the board meeting consisted of short presentations by each partner of the key aspects of their respective organizations and industries, presentation by SEWA consortium partners of motivations and objectives of SEWA project and an open panel discussion of relevance of SEWA project in various industries, potential commercialization avenues, unexplored possibilities, impact plan, showcases for SEWA applications.

During the meeting CEO of Neuro and Behavior Science at IPSOS, Mrs Elissa Moses was elected a chairman of the Valorisation Advisory Board. Under her supervision a collective feedback of all partners was assembled into a SEWA Valorisation Board Notes document and distributed with all partners of SEWA project. It was agreed to send Valorisation Board members regular updates (every 6 months) on the progress of SEWA project. The target period for the next Valorisation Board meeting was set for May 2016.

The report of the Valorisation Board is available on the SEWA website (internal use, accessible by the Commission).

6.9.3 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, workshop organisation, scientific talks at conferences and workshops, presentations to general public, and press coverage is given in section 1.2.8.

6.9.4 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 is to achieve vision-based detection of face and facial landmarks in unconstrained indoor environments. The first versions oft he 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), could be 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.

6.9.5 Task 8.5: Engagement with the public

SEWA partners have increased the interest of general public in the field of automatic emotion recognition and corresponding applications. There are several evidences for this:

- a major article on SEWA in German national press (Passauer Neue Presse) – UP PI Björn Schuller gave an interview on our current research in SEWA and other related projects. An article, including pictures from the SEWA database recordings (the subjects gave their explicit consent to appear in the newspaper) was released thereupon.
- a German TV (ARD) documentary about social agents was partly filmed at the chair of UP PI Björn Schuller -- the researchers of the group were filmed while interacting with an emotional agent and discussing the technology behind.
- invitation to the SEWA coordinator to speak on SEWA and related technologies at the World Economic Forum in Davos in January 2016.
- invitation to SEWA partners to present SEWA technology at the Science Fair in Belgrade (Serbia) – the fair was visited by 20,000 children and their parents and the SEWA stand was continuously crowded for all four days of the fair.
- two TV interviews with the SEWA coordinator on emotional robots/technology and their role in the 4th Industrial Revolution (see 1.2.8 for details).
- PlayGen highlighted the project at the annual applied games and gamification event in London called Gaminomics on 11th June 2015, see http://gaminomics.com
- More than a dozen other popular-press coverage oft he project

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

In M6, the data dissemination plan for the SEWA project was delivered in D8.2 and it specifies the data dissemination strategy in detail.

According to the plan, the SEWA dataset will be shared with the scientific community within ethics guidelines and regulations. We have built a web-portal for the SEWA dataset and prepared the end-user license agreement (EULA). The exact progress on this effort is detailed in section 6.2.

6.9.7 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 international, 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.

6.9.8 Task 8.8: Organisation of challenges and benchmarking

Once SEWA database has been released, scientific challenges, i.e. competitions between research groups, will be organised to work on the data and to improve and compare the performance of methods for automatic prediction of sentiment, arousal, and valence. The consortium has profound experience in organising such challenges in the framework of academic conferences.

Examples for previous efforts are the series of INTERSPEECH ComParE (Computational Paralinguistics Challenge) and AVEC (Audio/Visual Emotion Challenge and Workshop).

References:

- Schuller, B., et al. (2013) “The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism.” In: INTERSPEECH 2013: 14th Annual Conference of the International Speech Communication Association, Lyon, France, 25-29 August 2013.

- Ringeval, F., et al. (2015) "AVEC 2015: The 5th International Audio/Visual Emotion Challenge and Workshop." Proceedings of the 23rd Annual ACM Conference on Multimedia Conference. ACM, 2015.

6.10 Progress on WP9 – Project co-ordination and management

6.10.1 Task 9.1: Coordination of the consortium's activities

The project coordinator and project manager organized three plenary meetings, seven phone meetings, Valorisation Advisory Board meeting, and IBM Cognitive Workshop (see section 1.2.9 for details).

The project manager implemented mailing lists for consortium members and Valorisation Board members.

The project manager assisted with organization of sub-team meetings.

6.10.2 Task 9.2: Quality control and work plan monitoring

Agreed action list in written form is composed by the project coordinator following each plenary and/or phone meeting. The project manager follows the plan and issues reminders.
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 M1 to M6 financial transactions and usage of man months.

6.10.3 Task 9.3: Reporting to the European Commission

The coordinator has submitted five deliverables:
D8.1 Overall Dissemination Plan
D8.2 Data Management Plan with Annex for Ethical Approval from the Ethical Advisory Board
D2.1 Improved Acoustic Feature Extractor
D7.1 User Requirements for SEWA applications
D9.1 Annual Report for Year 1

6.10.4 Task 9.4: Legal and Contractual management

ICL's dedicated EU team handled the consortium agreement negotiation and pre-financing distribution.

7. Impact

SEWA contributes to the expected impacts of the ICT-22-2014 Call – Multimodal and Natural Computer Interaction – in several ways. Foremost, SEWA is expected to have a profound impact on the advancement of the state of the art in natural, multimodal human-computer interfaces, by advancing a number of underpinning technologies, including:

- **Automatic human behaviour understanding:** SEWA works on technology for sensing, detecting and interpreting facial, vocal and verbal behavioural cues observed in the wild, i.e., 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. Thus, SEWA enables interactions that are unscripted, based on spontaneous communication patterns of users, observed in unconstrained everyday life environments of the user, and hence more natural for the user.

- **Context sensing and adaptation:** SEWA works on technology that learns from interactions observed in the wild to model its user and the mapping from multimodal input to user sentiment, affect and intentions (e.g. rapport or empathy) by means of online and context-sensitive learning frameworks. Thus, SEWA enables automatic personalisation and context adaptation of the interaction and, in turn, more efficient, intuitive and seamless interaction with the user.

- **Machine learning:** SEWA works on technology for audio-visual behaviour similarity measurement that is 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?”. If attained, this 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.

- **Computer-mediated face-to-face interaction:** SEWA works on technology for automatic analysis of dyadic interactions in the wild and predictive modelling and detection of temporal patterning of the facial, vocal and verbal signals in interpersonal sentiment/ liking, mirroring, rapport and empathy shown by the two interactants. Thus, SEWA will enhance our understanding of interaction patterns shown in computer-mediated dyadic interactions and will facilitate seamless and efficient FF-HCI by offering prompts on smoothness of the interaction (i.e. on shown rapport and empathy).

SEWA contributes to all **expected impacts defined specifically for Innovation Actions** within ICT-22-2014 Call – Multimodal and Natural Computer Interaction.

- SEWA **improves the competitive position of RealEyes, PlayGen, and other European industries** (directly and indirectly) through provision of highly innovative *solutions* to robust and accurate automatic audio-visual human behaviour analysis, including affect and sentiment analysis. To wit, SEWA technology *directly addresses the specific needs of RealEyes* that specialises in automatic measurement and analysis of respondent's behavioural reactions to a variety of online stimuli and providing crucial consumer insight. By introducing novel, more reliable and quantitative methods in market research, SEWA plays a very important role for market-research industry overall, and it does so for RealEyes in particular; to wit, we expect a significant shift of the competitive position of RealEyes that the innovative technology and sentiment-based recommendation engine proposed by SEWA will bring about. It also *addresses the specific needs of PlayGen* that specialises in creating playful solutions that engage the audience, measure competencies and attitudes, and influence behaviour. SEWA technology is all about engaging users, enhancing their experience, measuring their attitudes (sentiment, rapport, empathy) and facilitating seamless and efficient user-centric interface design, and is expected to improve the quality of PlayGen solutions significantly. Furthermore, SEWA technology could be used in multitude of applications including entertainment (e.g. enjoyment and excitement measurement in fun rides), novel healthcare technologies (e.g. drug effectiveness assessment based on automatic measurement of increase in positive affect), serious games (e.g. training in negotiation skills by monitoring use of rapport and empathy), to mention but a few examples.

- SEWA ensures a **large spill over of the knowledge acquired to European industries** by means of the SEWA Valorisation Advisory Board comprising industrial representatives of different branches.

8. Update of the plan for exploitation and dissemination of result

Exploatation plan

PlayGen has carried out the first phase of analysis of the exploitable market for SEWA technology and SEWA applications. This analysis assessed existing computer-mediated face-to-face interactions as well as offline face-to-face interactions that could be migrated to videochat using SEWA automatic analysis. After applying criteria related to feasibility, acceptability and impact we concluded that communication skills training and job interview skills training were domains of highest exploitation potential for PlayGen. Full detail of this analysis is available in the document 'Chat Social Game workplan and exploitation v2.0' available on the SEWA project website (internal use, accessible by the Commission).

The plan for exploitation will therefore be focussed on these domains, building on lessons from the proof of concept application to develop student's persuasion and negotiation skills. As the concept is tested, discussion with the Valorisation Board will explore potential to applications to develop communication and interviewing skills with other audiences. PlayGen will also use every opportunity to advance discussions with new clients about how communications skills training modules could further enhance games and simulations.

Based on its own market analysis and feedback from existing business partners RealEyes has concluded that there is a substantially larger exploitation potential in the online ad placement optimization compared to multimedia recommendation systems with static inventory. On one hand, trend analysis demonstrated that online advertising becomes dominant, and on the other hand, this domain is much more versatile and the service we could provide with the enhanced support with SEWA technologies could be useful in more than one ways. The rationale behind this change in the exploitation strategy is documented in available on the SEWA project website. In turn the exploitation plan will focus on the particular use of our technology integrated into existing and soon to be launched services aiming to provide more powerful personalized advertising on online media platforms. Once the fundamental concept is tested, further discussions with the Valorisation Board will help us shape the development plan and get prepared for tests under real life conditions.

Dissemination plan

In M3, the dissemination plan of SEWA project was delivered in D8.1 to specify the dissemination strategy in detail. The activities spread the project achievements and the R&D knowledge gained during the project. Three different target audiences were identified: the general public, the scientific community, and the industry. For each of these groups, dissemination follows different strategies and uses different media.

For the general public, SEWA website and mass media the most important ways of informing the public. Scientific dissemination is accomplished via the well-established channels such as journals, conferences and workshops. Furthermore, once the SEWA database has been released, there will be scientific challenges on these data, organised by the consortium. A list of our recent publications acknowledging SEWA is part of this report.

For industrial dissemination, a meeting with the Valorisation Advisory Board has taken place in October 2015 and the board is seriously interested in several future meetings. Besides, the industrial partners of the consortium visit industrial fairs and create multimedia material to demonstrate the applications of SEWA technology.

9. Update of the data management plan

Deliverable D8.2 Data Management Plan was submitted at M6. There is no further update to this plan.

10. Deviation from Annex 1

10.1 Tasks

The detail of Task 7.2 deviates from the description set out in the DoW as more comprehensive market analysis identified an application and target user group with higher exploitation potential.

The core functionality of the Chat Social Game application will be very similar to the DoA. Some adaptations have been incorporated to make it more useful for users, increase user testing opportunities and alignment with likely exploitation application.

Area	Description in DoW	Proposed development
Objective of social game	Enjoyment	Develop communication skills
User test group	Users of _connect me	Students 18+ (_connect me platform has been discontinued because people didn't use the functionality to communicate with people in their social network that they don't know well)
User interface	Enables visualisation of social network	Enables visualisation of social network Incorporates useful social game mechanics to drive frequency of engagement with the game and increase learning e.g. reputation, feedback
Automatic estimation of sentiment, rapport and empathy	Used to develop clusters of liked people	Used to suggest more appropriate partners for discussion
Feedback	Users receive simple binary feedback on sentiment: another user is either 'liked' or not on the basis of observed behaviours	Users receive more detailed feedback that is immediately useful. This includes stated feedback from the other person as well as feedback on observed levels of sentiment and agreement during the conversation.

However the fundamental objective and timing of Task 7.2. - to develop 3 versions of a Chat Social Game that demonstrates how SEWA automatic analysis can be applied– remains the same. Following the retargetting of this task in M1-6, it started on schedule in M6 and development of V1 is well on track.

An updated version of WP7 is available on the SEWA website. As explained in Section 2, the target application area and the potential users differ to some extent from those described in the original concept.

The application retargeting resulted in the following modifications:

Area	**Description**	**Proposed changes**
Objective	Provide more sensitive and user specific information during the process of direct media recommendation in online movie renting services	Provide more sensitive and user specific information to improve online ad placement (implicit recommendation)
Users	2 groups of users: 1, customers who want to make an educated decision based on the recommendations given by the service 2, service provider who wants to increase user satisfaction	2 groups of users: 1, in market customers who visit a website and wants to be targeted by the most relevant ads 2, platform providers and ad owners who want to maximize the impact. (direct customer response, liking, visit frequency, etc.)

The central principle and the main functionality, however, remains the same: provide improved content recommendation to users based on sentiment and emotion analysis beyond direct verbal feedback. In turn, the challenges to be addressed as well as the development tasks all remain the same, thus there is no need to modify the timeline or reschedule some of the subtasks.

An updated version of WP7 is available on the SEWA website as the document ‘Chat Social Game workplan and exploitation v2.0’ (internal use, accessible by the Commission).