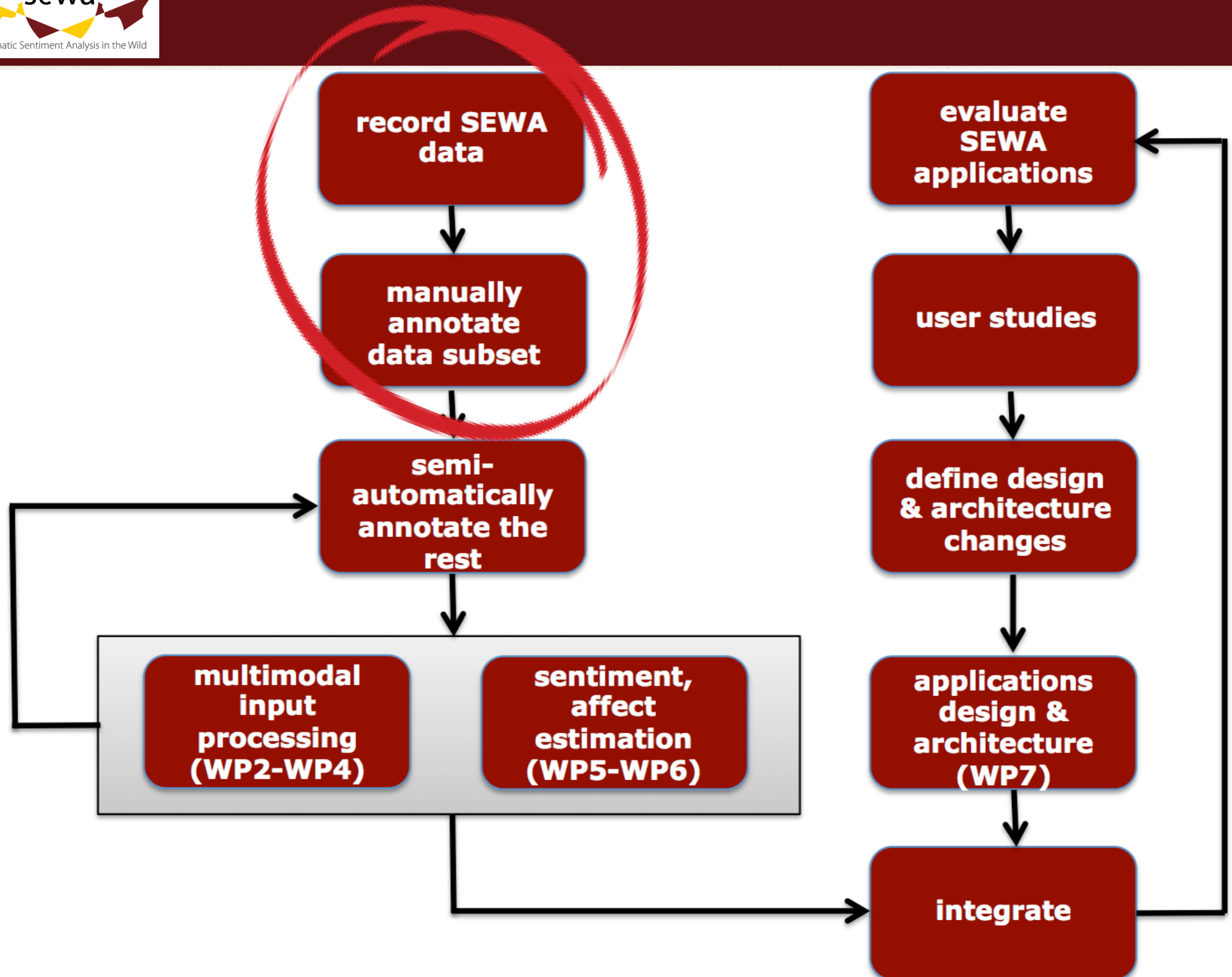


WP1: SEWA Database: Data Collection, Annotation and Release

Jie Shen



Automatic Sentiment Analysis in the Wild



data release



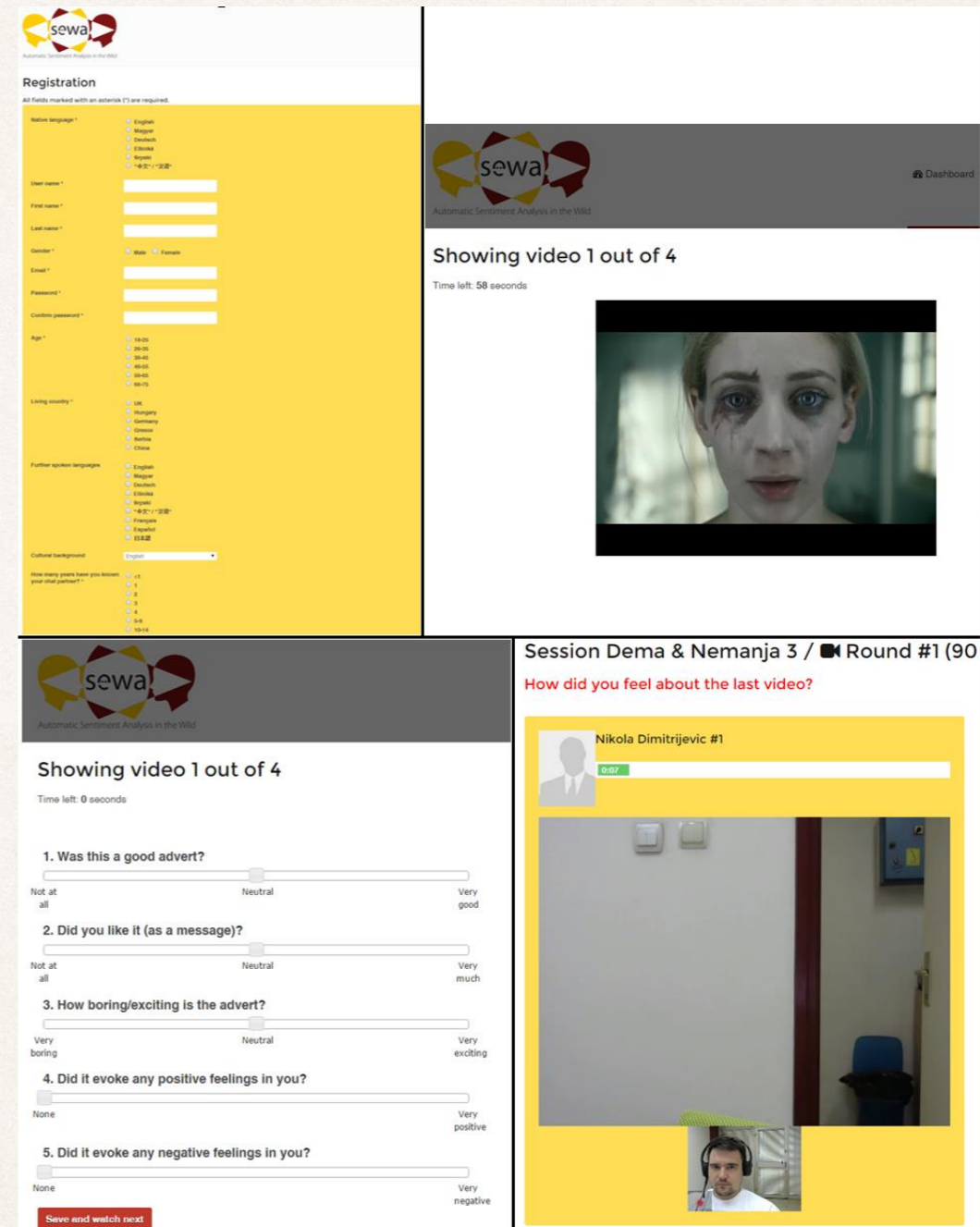
Milestones						M1						M2						M3						M4
Month	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39	42			
WP1		Data acquisition and annotation					SEWA DB design and release																	
WP2	Development of robust and cross-language audio-visual features																							
WP3		Development of behavioural feature extraction (body language, FAU, vocalisations, etc.)																						
WP4							Development of continuous-valued audio-visual sentiment models																	
WP5						Development of behaviour similarity measures																		
WP6						Development of mimicry, rapport, recognition																		
WP7		Iterative requirements engineering and application development																						
WP8	Dissemination and communication activities; ethical review																							
WP9	Coordination and management																							

Objectives

- ❖ Online SEWA DB containing annotated audio-visual behaviour in the wild (D1.1)

Data Collection

- ❖ Different *cultural background*, *gender* and *age*.
- ❖ Fill-in basic *demographic information*.
- ❖ Watch **4 adverts**, we record reactions.
- ❖ Discuss the 4th advert through *video-chat*.
- ❖ All in their *native language*.



The screenshots illustrate the user experience on the sewa platform. The top-left screenshot shows a registration form with fields for native language, user name, first name, last name, gender, email, password, confirm password, age, living country, further spoken languages, and cultural background. The top-right screenshot shows a video player interface with the text 'Showing video 1 out of 4' and a 'Time left: 58 seconds' indicator. The bottom-left screenshot shows a feedback survey with five questions, each with a horizontal slider for response: '1. Was this a good advert?', '2. Did you like it (as a message)?', '3. How boring/exciting is the advert?', '4. Did it evoke any positive feelings in you?', and '5. Did it evoke any negative feelings in you?'. The bottom-right screenshot shows a video chat session with a participant named 'Nikola Dimitrijevic #1' and a 'Save and watch next' button.

Subject Demographics

- ❖ 408 subjects (204 pairs), 6 different cultures (British, German, Hungarian, Serbian, Greek and Chinese), 202 Female, 206 Male

Cultural Background		Age Group		Years Known the Other Participant		Self-Reported Familiarity Rating	
British	66	18~29	203	<1	80	Not	9
				1	30	Familiar	
German	64	30~39	94	2	39	Slightly	13
				3	40	Familiar	
				4	37	Somewhat	35
Hungarian	70	40~49	25	5~9	55	Familiar	
				10~14	20	Moderately	114
Serbian	72	50~59	46	15~19	22	Familiar	
				20+	75	Extremely	227
Greek	56	60+	30			Familiar	
Chinese	70						

Data Composition

- ❖ **1525 minutes** of people's *reaction to adverts*.
- ❖ **568 minutes** of *video-chat recordings*
- ❖ Video resolution: **320x240~640x360 pixels**, frame rate: **20~30 fps**
- ❖ Audio sample rate: **44.1~48.0 kHz**
- ❖ We created the fully annotated **basic SEWA dataset**: 538 short (10~30s) video-chat segments of low / high valence, low / high arousal, and liking / disliking.

SEWA Recording: An Example



Annotations

- ❖ Low-Level Features: 1) Facial Landmarks, and 2) Audio Low-Level Descriptors
- ❖ Mid-Level Features: 1) Hand Gestures, 2) Head Gestures, 3) Facial Action Units, and 4) Audio Transcript
- ❖ Others: 1) Valence, Arousal and Liking / Disliking, 2) Template Behaviours, 3) Episodes of Agreement / Disagreement, and 4) Episodes of Mimicry

Facial Landmarks

- ❖ Annotated semi-autonomously for the basic SEWA dataset:
 - ❖ Applied a generic tracker.
 - ❖ Manually corrected 1/8 of the frames.
 - ❖ Trained person-specific models to be applied to all frames.
 - ❖ Finally, the results were further verified and, if necessary, corrected by hand.



Audio Low-Level Descriptors

- ❖ Audio low-level descriptor (LLD) features (e.g., loudness, MFCC, spectrum characteristics, etc.) are provided for the whole SEWA database.
- ❖ LLDs were extracted in 10ms steps.
- ❖ We provide two sets of LLD features:
 - ❖ ComPareELLD: 65 LLDs
 - ❖ GeMAPSv01aLLD: 18 LLDs

Hand Gesture

- ❖ Annotated for all video-chat recordings in 5-frame steps:
 - ❖ Hand not visible (89.08%)
 - ❖ Hand touching head (3.32%)
 - ❖ Hand in a static position (0.63%)
 - ❖ Display of hand gestures (2.39%)
 - ❖ Other hand movements (3.68%)



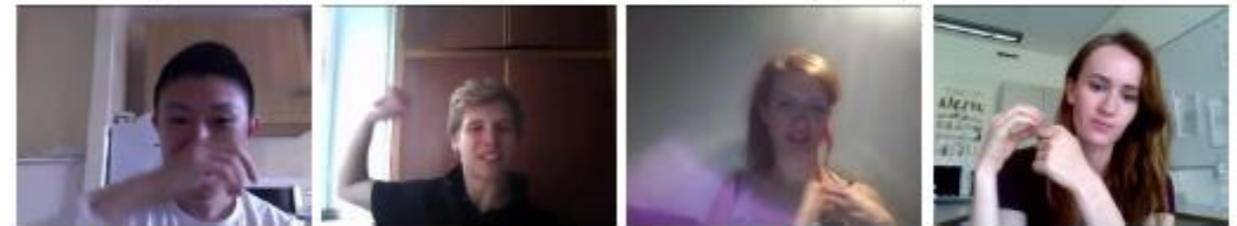
Hands are not visible in 89.08% (565535) of the frames in 99.50% (396) of the videos.



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



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



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

Head Gesture

- ❖ Head nod / shake annotated for the basic SEWA dataset.
- ❖ 282 head nods.
- ❖ 122 head shakes.
- ❖ Display of head gesture is culture-dependent.



[Nod]



[Shake]

Facial Action Units

- ❖ Annotated semi-autonomously for the basic SEWA dataset:
 - ❖ FAU1: Inner eyebrow raiser (109 examples)
 - ❖ FAU2: Outer eyebrow raiser (79 examples)
 - ❖ FAU4: Eyebrow lowerer (94 examples)
 - ❖ FAU12: Lip corner puller (104 examples)
 - ❖ FAU17: Chin raiser (61 examples)
- ❖ The FAU annotation is not exhaustive.

Facial Action Units: Examples



[FAU1]



[FAU2]



[FAU4]



[FAU12]



[FAU17]

Transcript

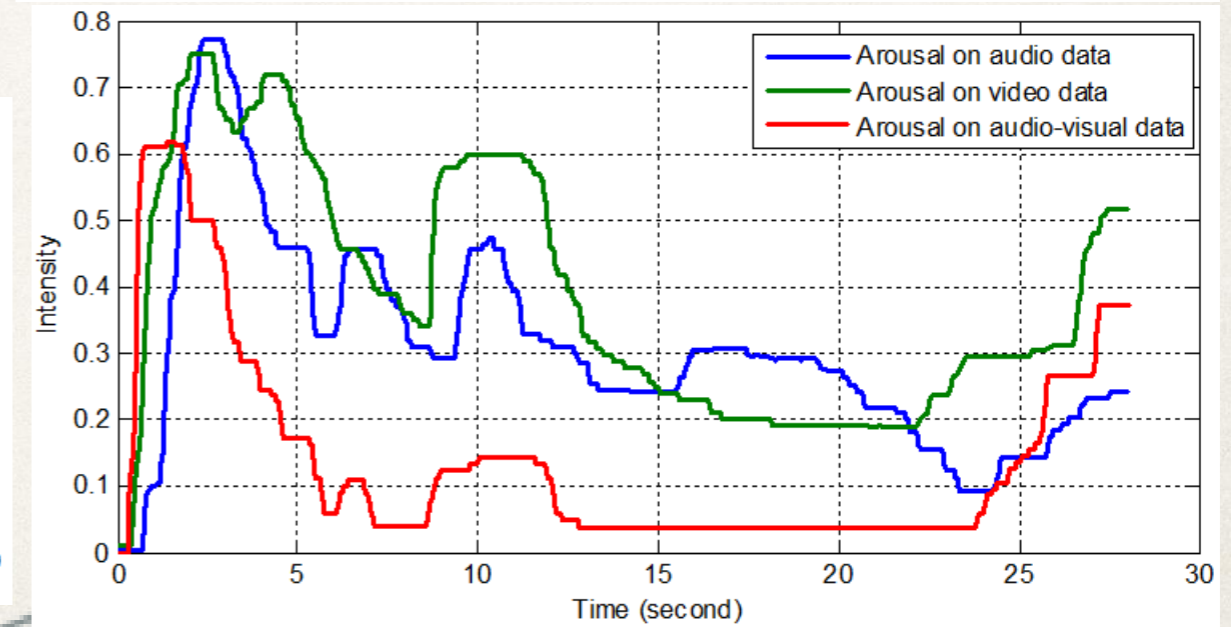
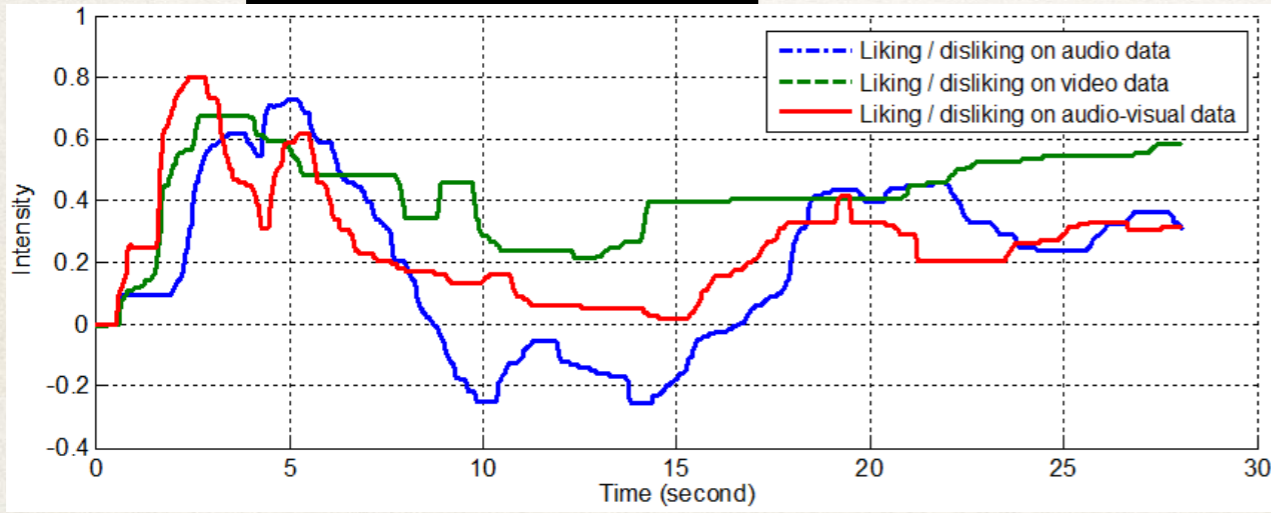
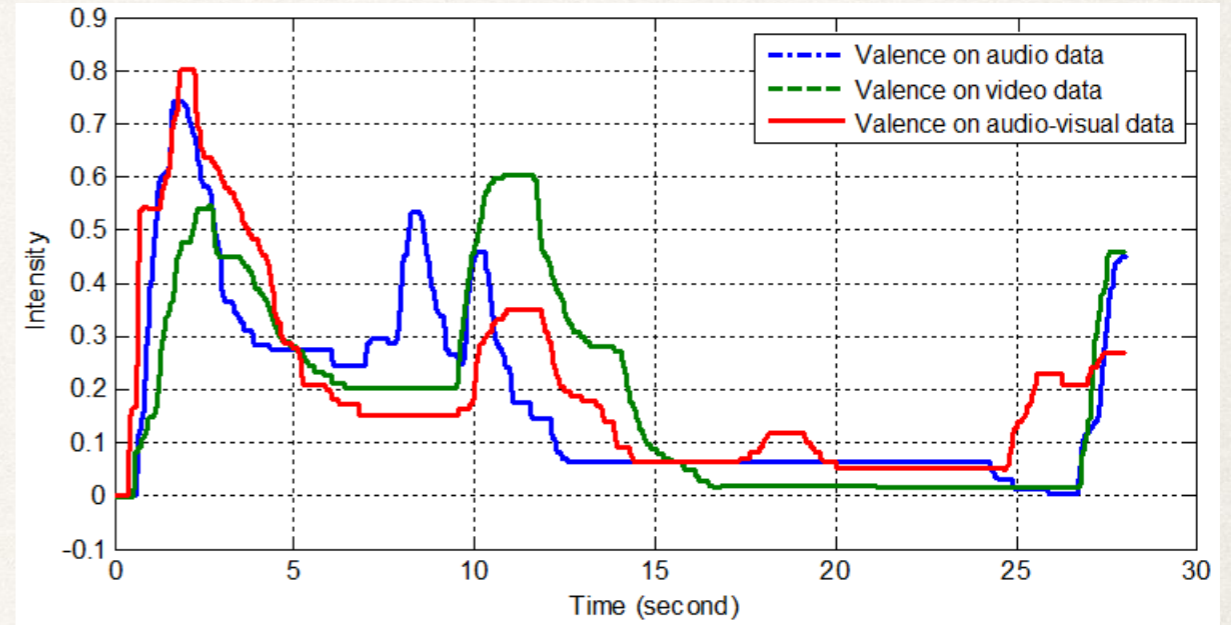
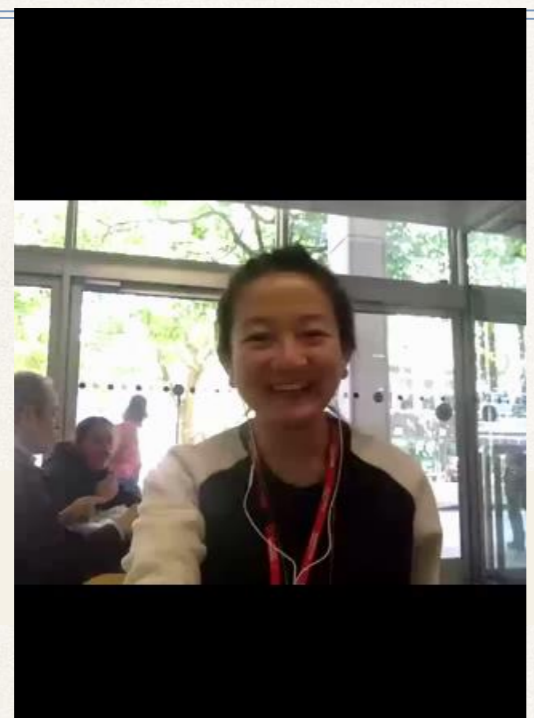
- ❖ Transcribed verbal content and some non-verbal cues in all video-chat recordings

```
1 start_time,end_time,subject,text
2 6.208465,6.818082,66,"Szevasz "
3 11.582718,12.416930,65,"Na hogy tetszettek?"
4 12.850079,14.037227,66,"Én most nem hallak"
5 16.620077,18.256416,66,"Várjál most hallak, felhangosítalak"
6 17.069268,17.502417,65,"Miért nem hallasz?"
7 22.283095,24.448839,66,"Szóval mit gondolsz a legutolsó videóról?"
8 27.464837,27.577135,65,"Hát..."
9 27.817773,29.277645,65,"A legutolsóóóóó..."
10 29.422028,31.122538,65,"???"
11 31.459431,31.956750,66,"Micsoda?"
12 32.502196,36.994109,65,"???"
13 34.651897,35.630493,65,"?Egyenlő szárú?"
14 37.010152,37.764151,65,"Nekem az tetszett"
15 37.892492,40.122405,65,"???"
16 40.122405,42.400447,66,"Amúgy a csap az nagyon jó ötlet"
17 44.293467,44.806828,66,"Mondjuk sze..."
18 45.400403,50.774655,66,"Kicsit elhúzták, mert mondjuk az első kettőnél le lehetett vágni, hogy miről van szó."
19 51.913676,54.159632,66,"Tehát felesleges volt"
20 54.416313,54.945717,65,"???"
21 55.699717,56.598099,65,"Meg a végére tették"
```


Valence, Arousal and (Dis)Liking

- ❖ Annotated for the basic SEWA dataset.
- ❖ 5 annotators for each culture.
- ❖ Each annotation task was repeated 3 times:
 - ❖ On audio data
 - ❖ On video data
 - ❖ On audio-visual data

Valence, Arousal and (Dis)Liking



Template Behaviours

❖ Templates selected from the basic SEWA dataset:

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

Template Behaviours: Examples



[Low Valance]
[Hungarian]



[High Arousal]
[Greek]



[Liking]
[German]



[Disliking]
[Chinese]

(Dis)Agreement Episodes

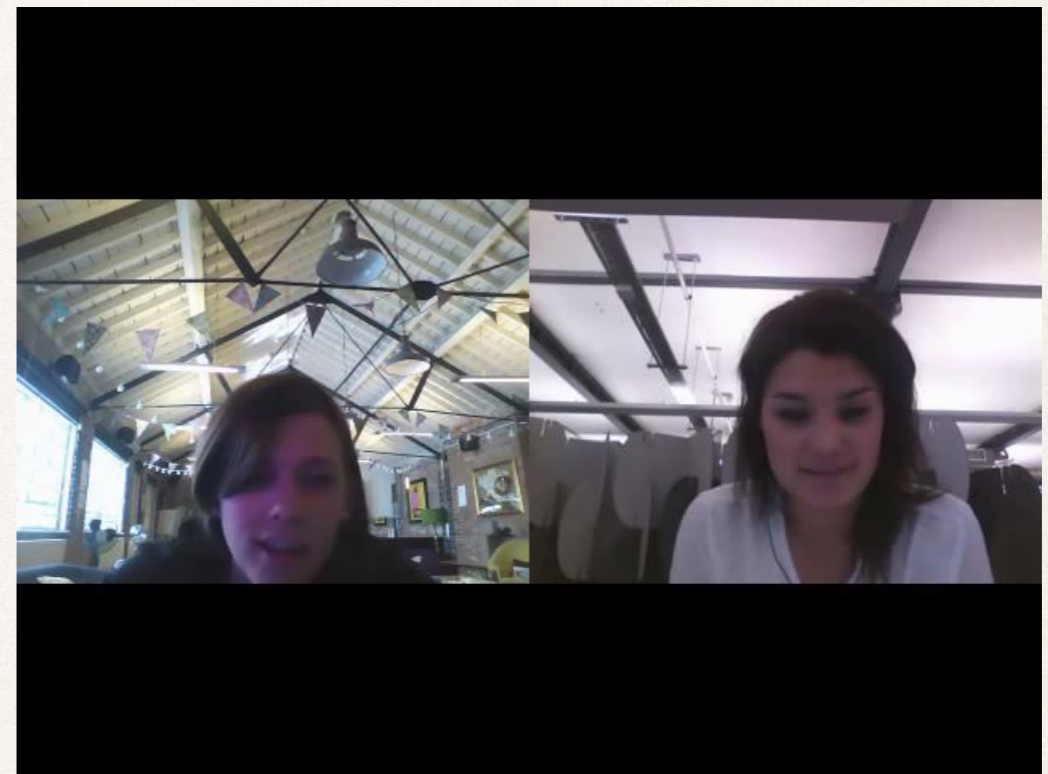
- ❖ Agreement / disagreement episodes selected from the video-chat recordings:

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

(Dis)Agreement Episodes: Examples



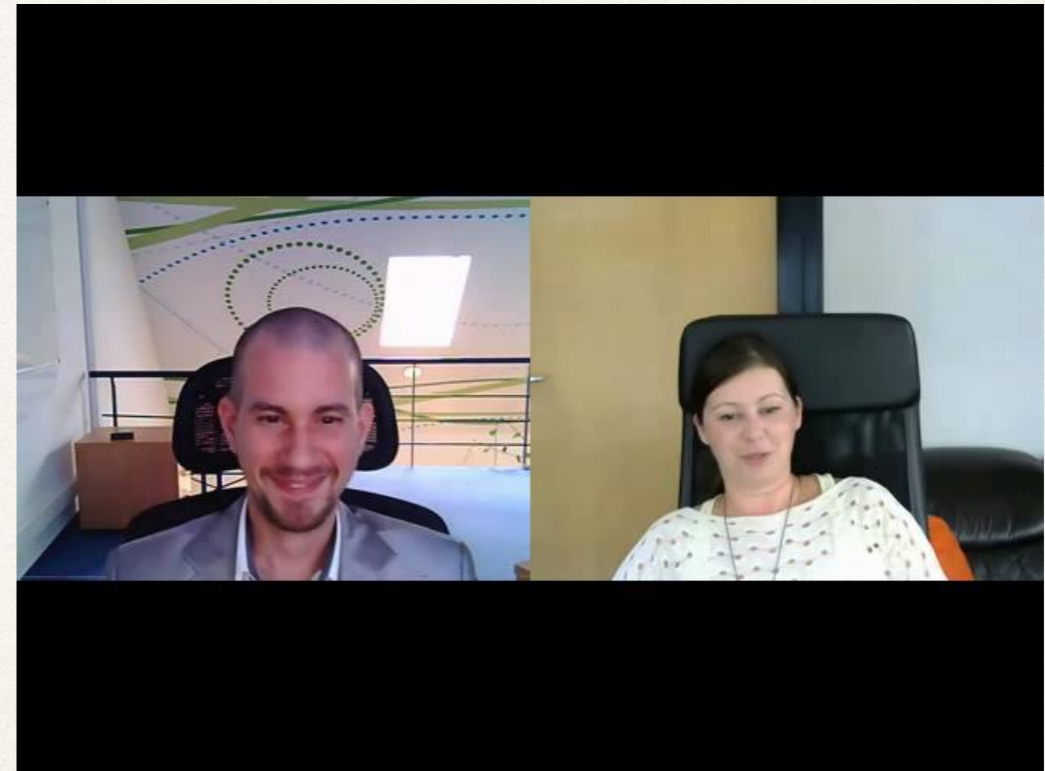
[Strong Agreement, British]



[Strong Disagreement, British]

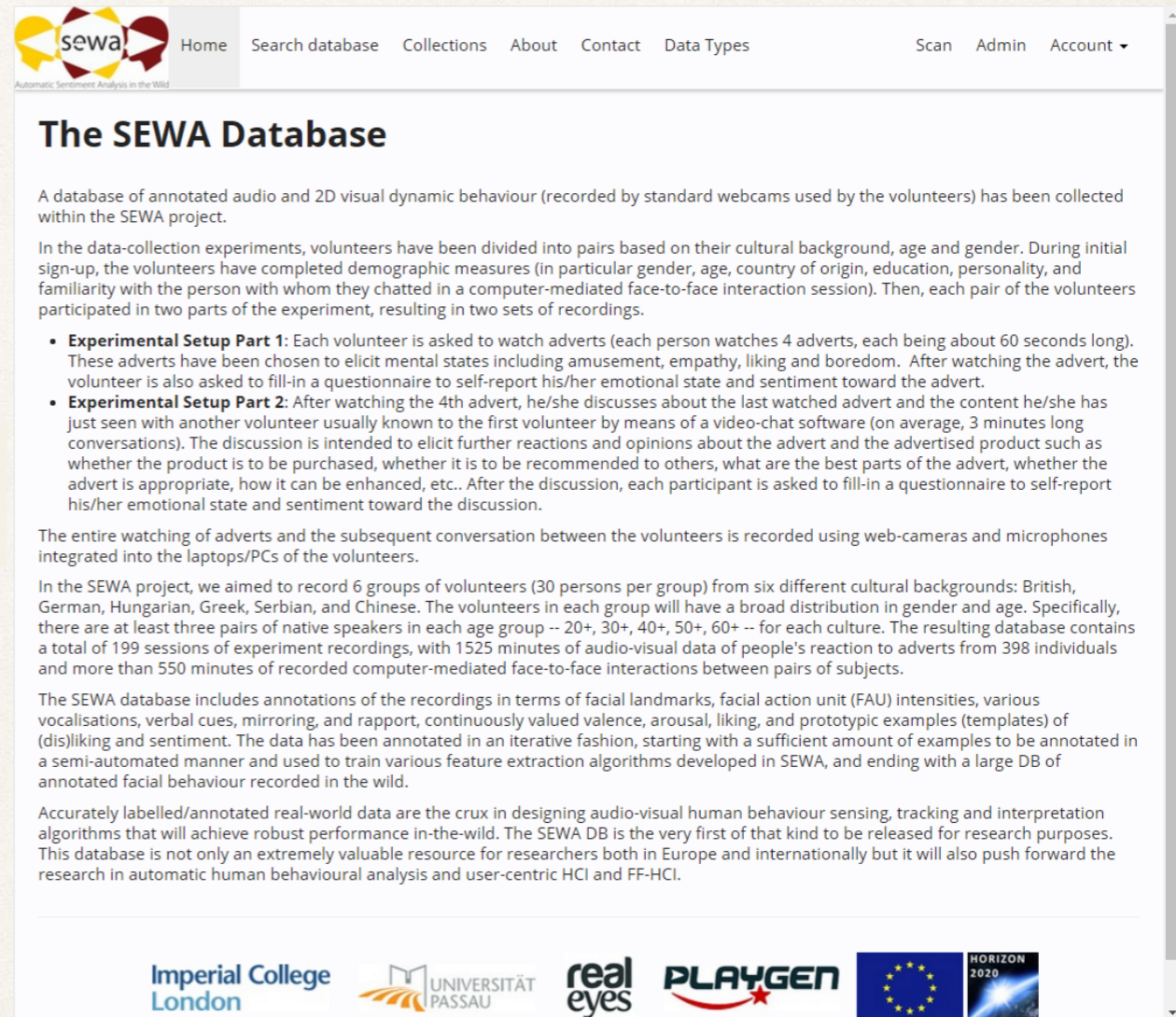
Mimicry Episodes

- ❖ 197 mimicry episodes (48 British, 31 German, 39 Hungarian, 20 Serbian, 41 Greek and 17 Chinese) extracted from the video-chat recordings.



SEWA Database Release

- ❖ Released online:
<http://db.sewaproject.eu/>
- ❖ Access only granted to academics who signed the EULA.
- ❖ Provides comprehensive search filters for ease of access.



The SEWA Database

A database of annotated audio and 2D visual dynamic behaviour (recorded by standard webcams used by the volunteers) has been collected within the SEWA project.

In the data-collection experiments, volunteers have been divided into pairs based on their cultural background, age and gender. During initial sign-up, the volunteers have completed demographic measures (in particular gender, age, country of origin, education, personality, and familiarity with the person with whom they chatted in a computer-mediated face-to-face interaction session). Then, each pair of the volunteers participated in two parts of the experiment, resulting in two sets of recordings.

- **Experimental Setup Part 1:** Each volunteer is asked to watch adverts (each person watches 4 adverts, each being about 60 seconds long). These adverts have been chosen to elicit mental states including amusement, empathy, liking and boredom. After watching the advert, the volunteer is 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, he/she discusses about the last watched advert and the content he/she has just seen with another volunteer usually known to the first volunteer by means of a video-chat software (on average, 3 minutes long conversations). The discussion is intended to elicit further reactions and opinions about the advert and the advertised product such as whether the product 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 is asked to fill-in a questionnaire to self-report his/her emotional state and sentiment toward the discussion.

The entire watching of adverts and the subsequent conversation between the volunteers is recorded using web-cameras and microphones integrated into the laptops/PCs of the volunteers.

In the SEWA project, we aimed to record 6 groups of volunteers (30 persons per group) from six different cultural backgrounds: British, German, Hungarian, Greek, Serbian, and Chinese. The volunteers in each group will have a broad distribution in gender and age. Specifically, there are at least three pairs of native speakers in each age group -- 20+, 30+, 40+, 50+, 60+ -- for each culture. The resulting database contains a total of 199 sessions of experiment recordings, with 1525 minutes of audio-visual data of people's reaction to adverts from 398 individuals and more than 550 minutes of recorded computer-mediated face-to-face interactions between pairs of subjects.

The SEWA database includes annotations of the recordings in terms of facial landmarks, facial action unit (FAU) intensities, various vocalisations, verbal cues, mirroring, and rapport, continuously valued valence, arousal, liking, and prototypic examples (templates) of (dis)liking and sentiment. The data has been annotated in an iterative fashion, starting with a sufficient amount of examples to be annotated in a semi-automated manner and used to train various feature extraction algorithms developed in SEWA, and ending with a large DB of annotated facial behaviour recorded in the wild.

Accurately labelled/annotated real-world data are the crux in designing audio-visual human behaviour sensing, tracking and interpretation algorithms that will achieve robust performance in-the-wild. The SEWA DB is the very first of that kind to be released for research purposes. This database is not only an extremely valuable resource for researchers both in Europe and internationally but it will also push forward the research in automatic human behavioural analysis and user-centric HCI and FF-HCI.

Imperial College London UNIVERSITÄT PASSAU real eyes PLAYGEN HORIZON 2020

Objectives

- ❖ Online SEWA DB containing annotated audio-visual behaviour in the wild (D1.1)

WP1: SEWA Database: Data Collection, Annotation and Release

Jie Shen



Automatic Sentiment Analysis in the Wild