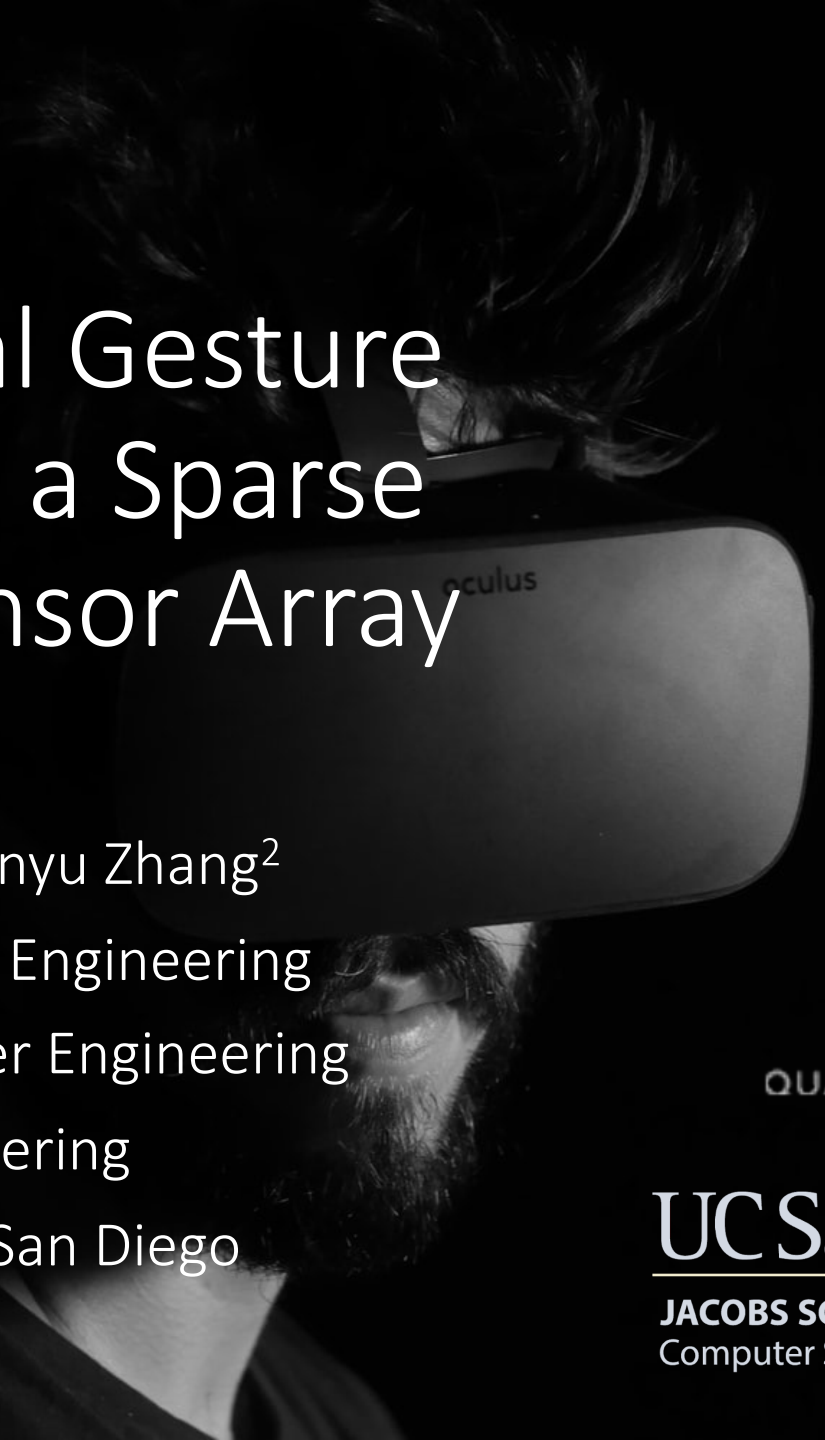
IPSN 2021
MAY 18-21, 2021
NASHVILLE, TENNESSEE, USA

ExGSense
Toward Facial Gesture Sensing with a Sparse Near-Eye Sensor Array



Chen Chen[1], Ke Sun[1], Xinyu Zhang[2]
[1]Computer Science and Engineering
[2]Electrical and Computer Engineering
Jacobs School of Engineering
University of California San Diego

QUALCOMM INSTITUTE
TRITONS
UC San Diego
JACOBS SCHOOL OF ENGINEERING
Computer Science and Engineering
UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

Collaborations Through VR

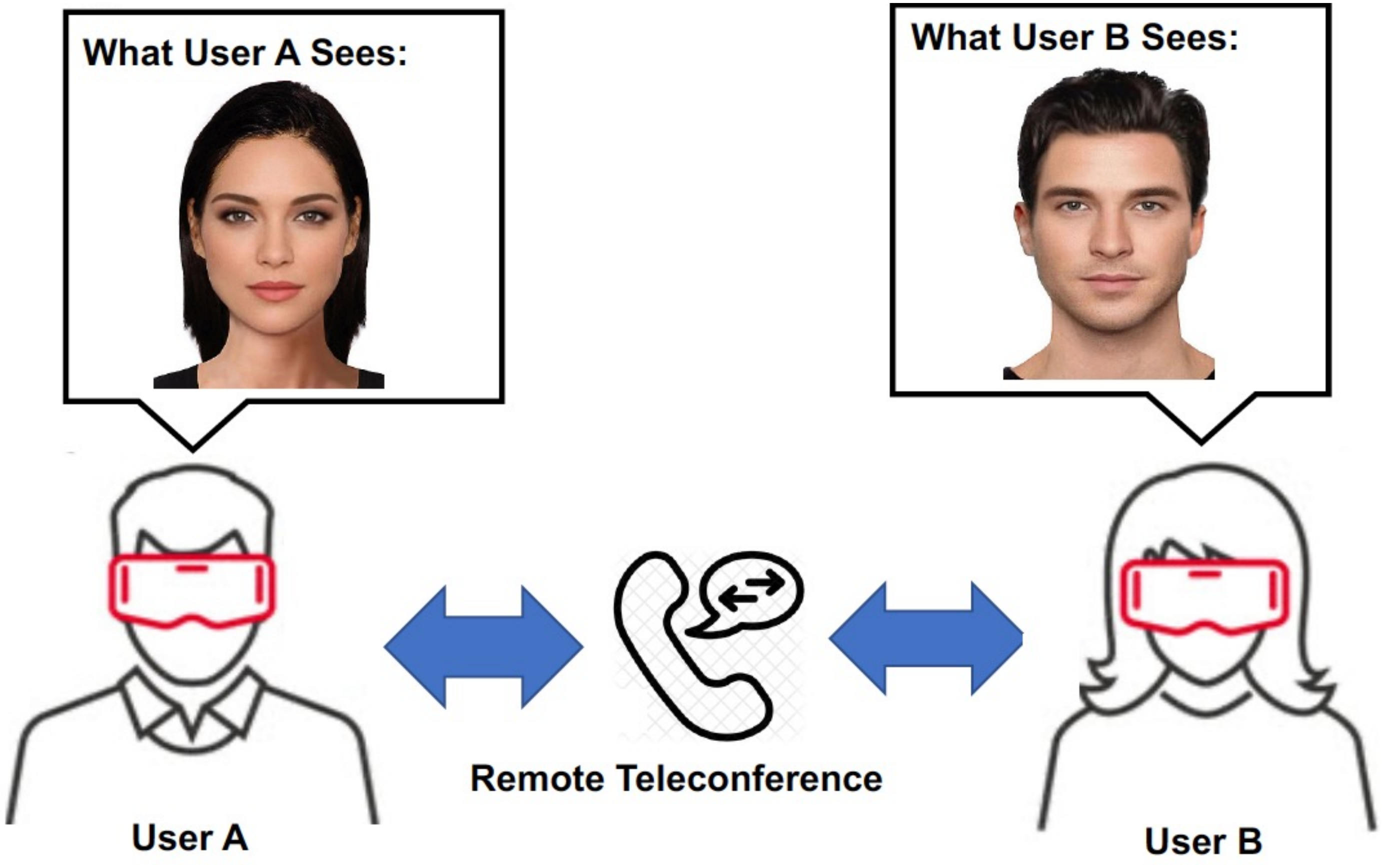
- Enables a new forms of telepresence applications;
- Empowers professionals to virtually connect with one to another;

Immersive Data Analysis,
(Rabaudo et. al., '20)

Training for Components Assembling
(Rumii software)

Sense of Co-Presence is the Key!

- Capturing full face of remote collaborators enables one to perceive the non-verbal cures of the remote collaborators, yet challenging!



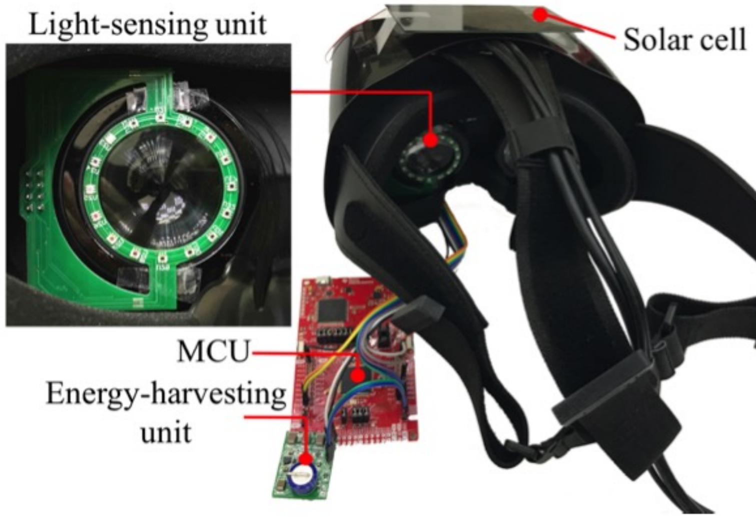
Existing Solutions

- Camera Based Solution (Inside or Outside of HMD);
  - Occlusions;
  - $$

Existing Solutions

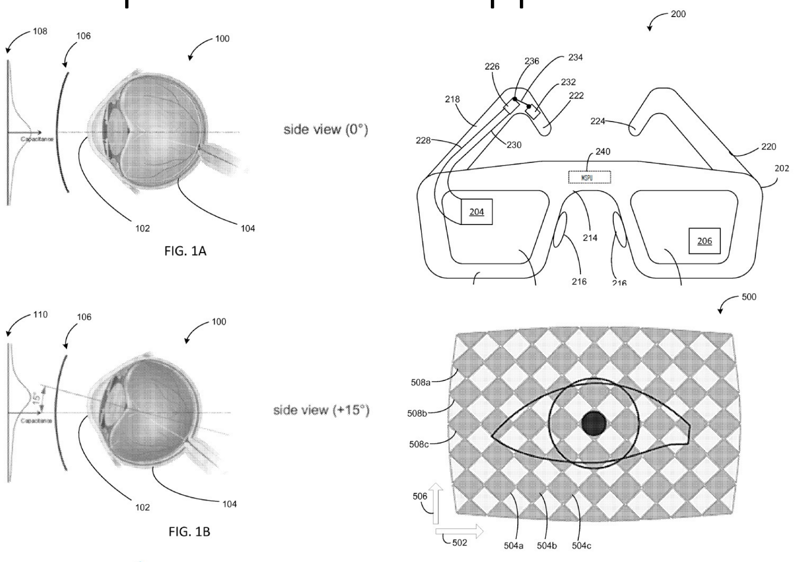
- Camera Based Solution (Inside or Outside of HMD);
  - Occlusions;
  - $$
- Proximity & Pressure Based Solution;
  - Complicated hardware design;
  - Can only detect partial facial gestures;

IR Based Approach



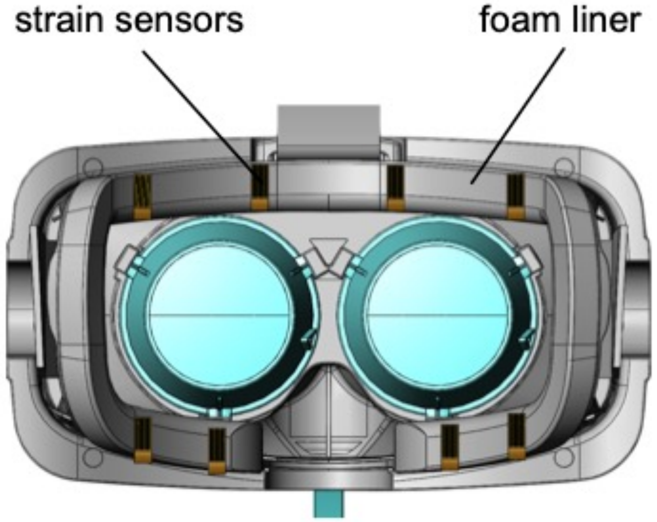
(LiGaze, Li *et. al.* ‘17)

Capacitive Based Approach



(Bergman *et. al.*, ‘15)

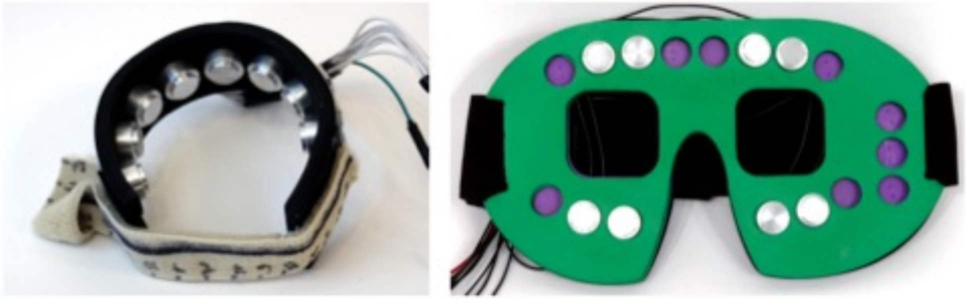
Strain Based Approach



(Li *et. al.*, ‘15)

Existing Solutions

- Camera Based Solution (Inside or Outside of HMD);
  - Occlusions;
  - $$
- Proximity & Pressure Based Solution;
  - Complicated hardware design;
  - Can only detect partial facial gestures;
- Interferometry and Tomography Based Solution;
  - Low sensing granularity;
  - Impacts from differences of underlying anatomical patterns across different users;



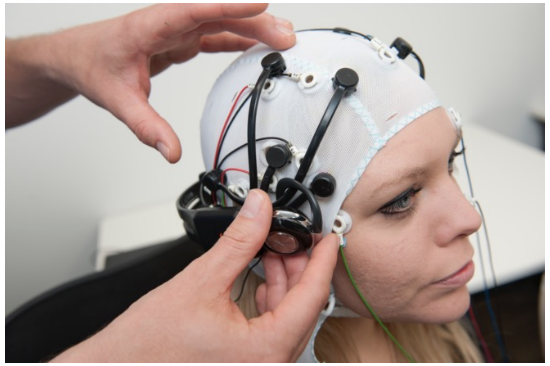
(Interferi, Iravantchi *et. al.*, CHI '19)

Existing Solutions

- Biosignal Based Solution;
  - Eye activities detections using EOG;
  - Facial muscle related gestures detections using EMG;
  - Coarse grained emotion detections, bulky as well as complicated setup using EEG;

(J!NE Related, Ishimaru *et. al.*, UbiComp ’14, Rostaminia *et. al.* ETRA ’19, IMWUT ‘19)

(AlterEgo, Kapur *et. al.*, IUI ‘18)



(Emotiv EPOC ‘18)

ExGSense

- **Sensing Modality:** We used the near-eye biopotential sensors to detect full face expressions. To achieve this, we explore the paradigm of *indirect* sensing where the lower facial gestures can be detected by the transducers resting on the upper face;

ExGSense

- Sensing Modality: We uses the near-eye biopotential sensors to detect full face expressions. To achieve this, we explore the paradigm of *indirect* sensing where the lower facial gestures can be detected by the transducers resting on the upper face;
- Detection Framework: We propose a dual-branch multiview representation learning pipeline, which can explicitly exploit the sensor diversities across time-frequency-spatial domains. We further propose a simple re-calibration approach for adapting the pretrained model for different users;

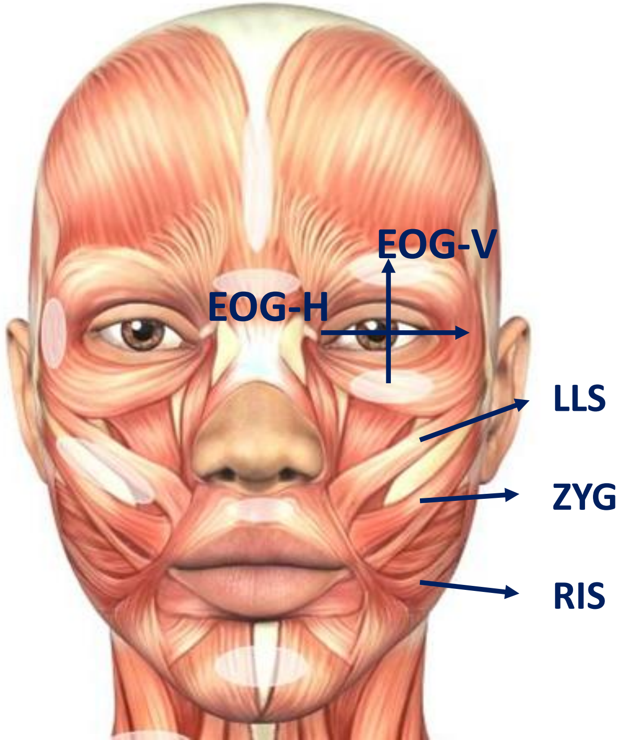
ExGSense

- Sensing Modality: We uses the near-eye biopotential sensors to detect full face expressions. To achieve this, we explore the paradigm of *indirect* sensing where the lower facial gestures can be detected by the transducers resting on the upper face;
- Detection Framework: We propose a dual-branch multiview representation learning pipeline, which can explicitly exploit the sensor diversities across time-frequency-spatial domains. We further propose a simple re-calibration approach for adapting the pretrained model for different users;
- Prototype & Evaluations: We build a proof-of-concept prototype of ExGSense and conduct the user study to verify its ability to concurrently track the fine-grained upper face eye and lower face mouth gestures by fully leveraging the underlying facial anatomy patterns;

Preliminary

ElectroOculoGraphy (EOG)

- ~μV;
- Induced by eye movements;
- Can be decomposed to EOG-H and EOG-V;



Preliminary

ElectroOculoGraphy (EOG)

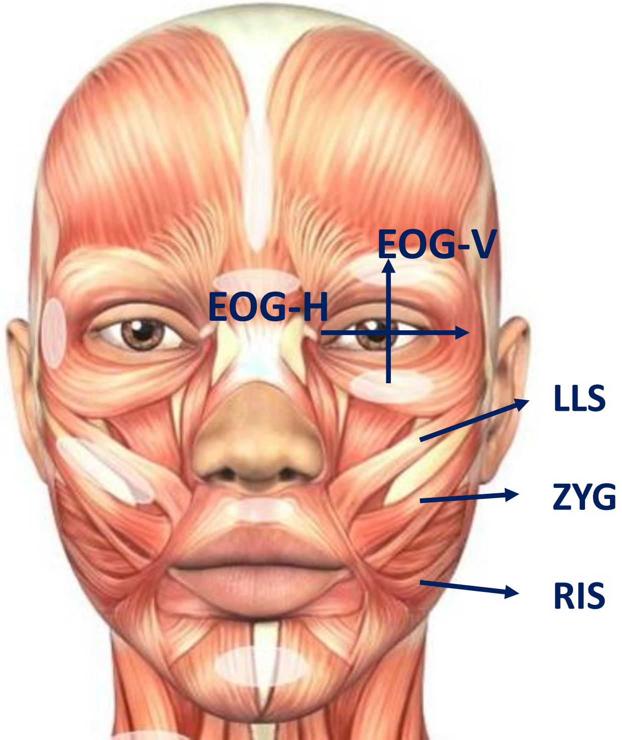
- ~μV;
- Induced by eye movements;
- Can be decomposed to EOG-H and EOG-V;

ElectroMyoGram (EMG)

- ~mV;
- Induced by muscle contract and relax;
- Indirect Sensing: using transducers resting on the upper face to detect the lower face mouth gestures;
- 3 Muscle Groups:
  - Levator Labii Superioris (LLS);
  - Zygomatics Majors (ZYG);
  - Risorius (RIS);

2 Unwanted Noise:

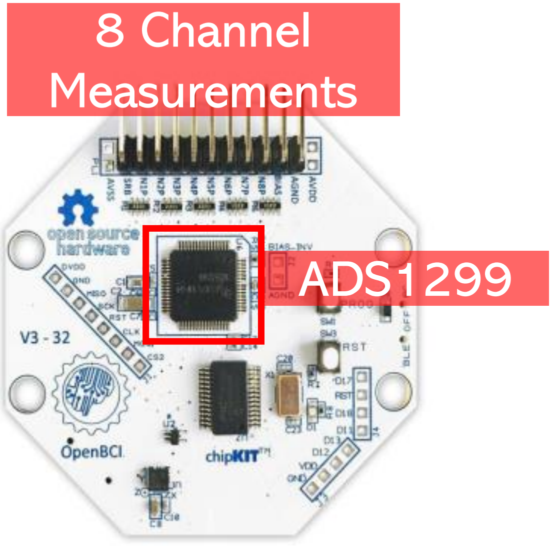
- Inter-Person Variations;
- Inter-Session Variations;



Implementations

Data Acquisitions:

- OpenBCI Cyton Board;
- 8 16-bit ADC (TI ADS1299);
- X24 Gain;
- Stream data to host PC through BLE 4.0;
- Sampling Frequency f_s= 250Hz;

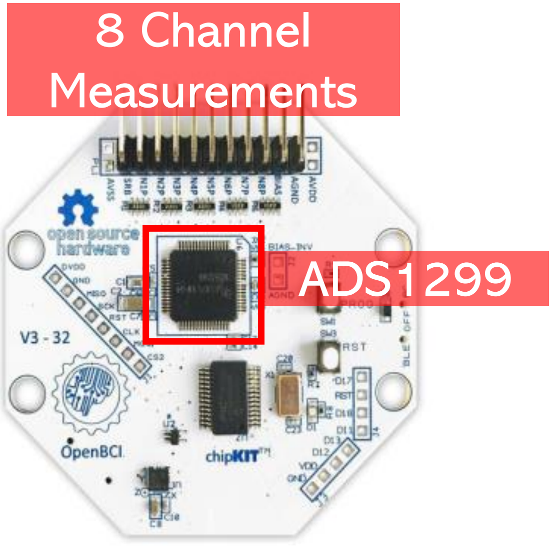


OpenBCI Cyton Bioprocessing Board

Implementations

Data Acquisitions:

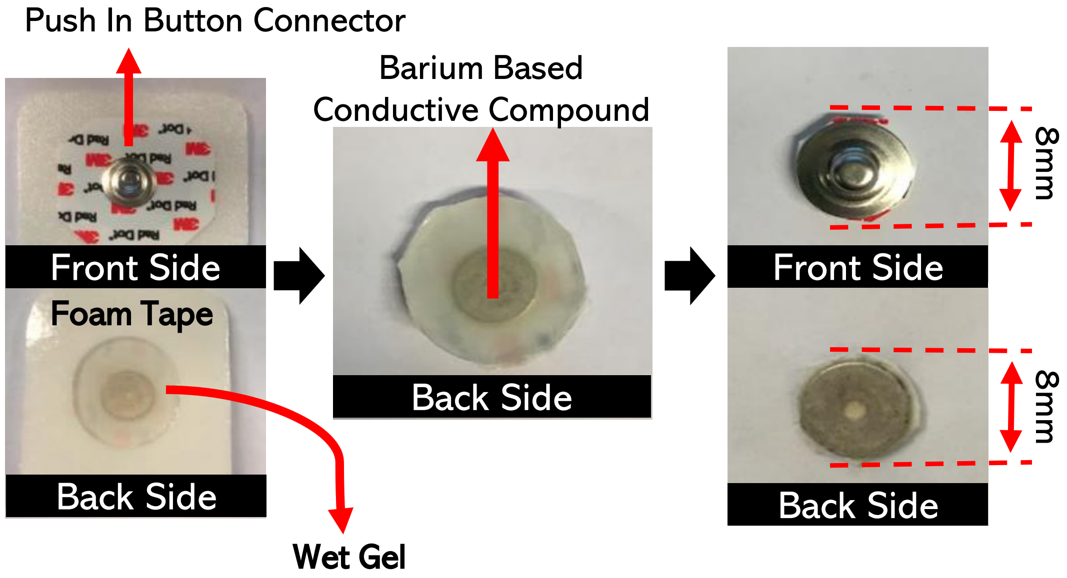
- OpenBCI Cyton Board;
- 8 16-bit ADC (TI ADS1299);
- X24 Gain;
- Stream data to host PC through BLE 4.0;
- Sampling Frequency f_s= 250Hz;



OpenBCI Cyton Bioprocessing Board

Transducer:

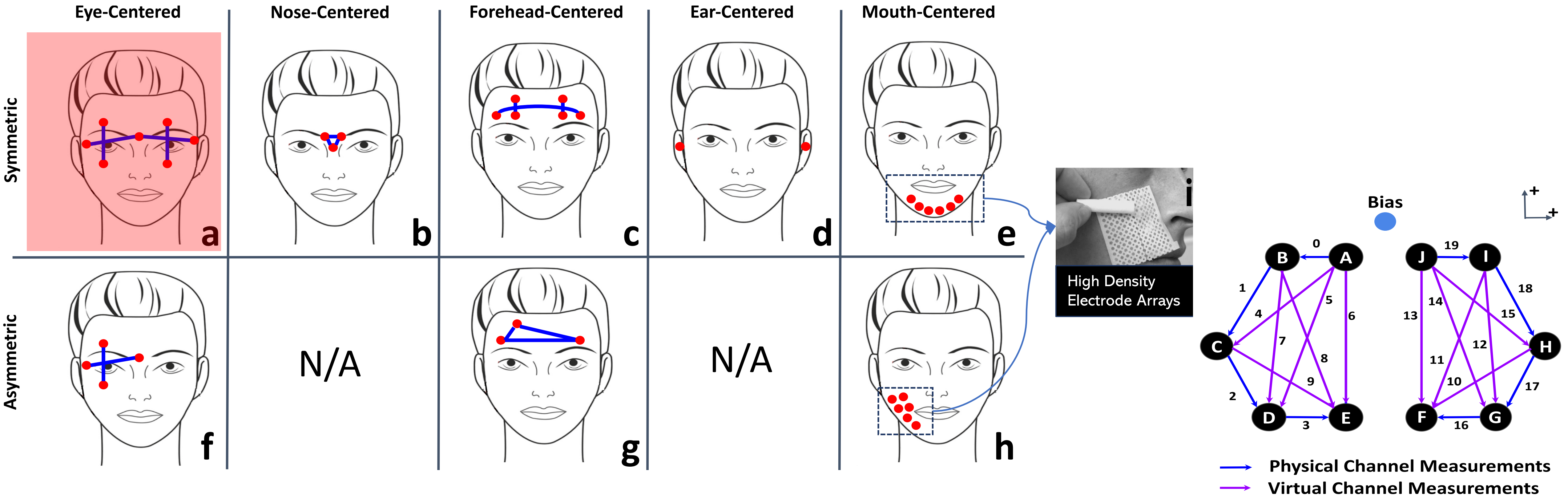
- Disposal wet electrodes can reduce noise, yet are less comfortable and impractical for mobile applications;
- Remove the wet gels to enable the comfortable reusable electrodes with reasonable signal quality degradations;



Transducer Arrangement

Indirect Sensing & Symmetric Design

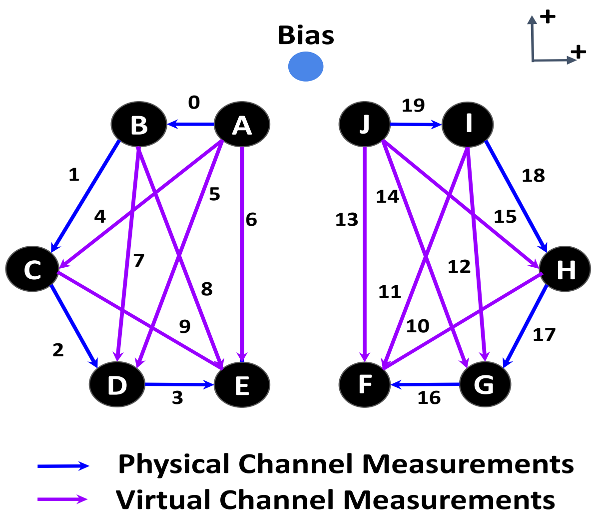
- Indirect Sensing: detect the lower facial gestures by only leveraging the transducers resting on the upper face;
- Symmetricity: introduce spatial redundancies and minimize imperfect sensing performance;



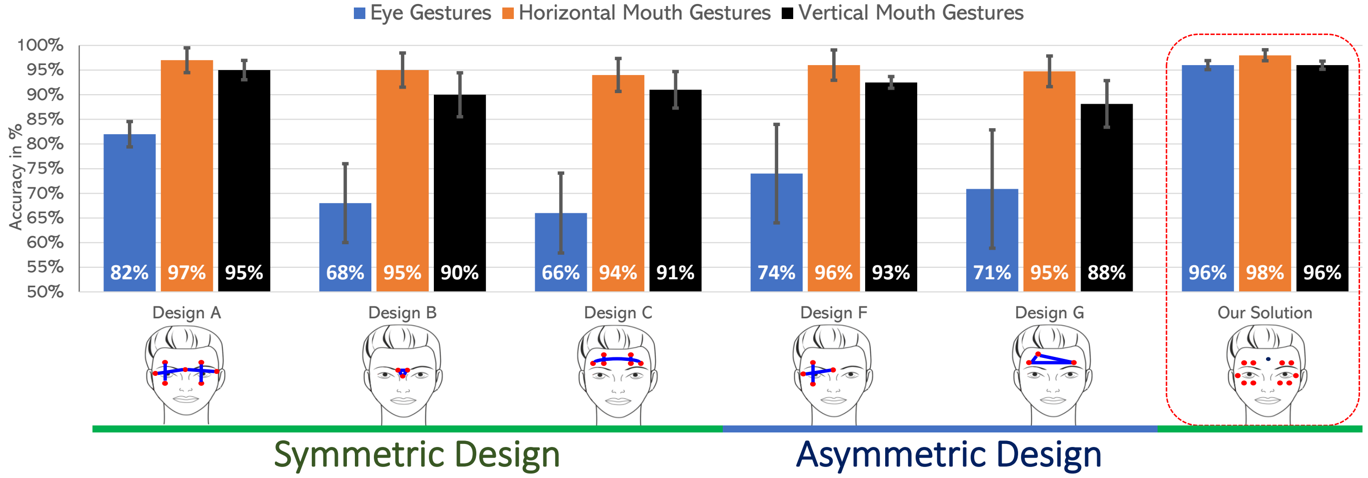
Transducer Arrangement

Virtual Channel Measurements

- Physical Channel Measurements (Blue): Direct collected by DAQ;
- Virtual Channel Measurements (Pink): Computed algorithmically;

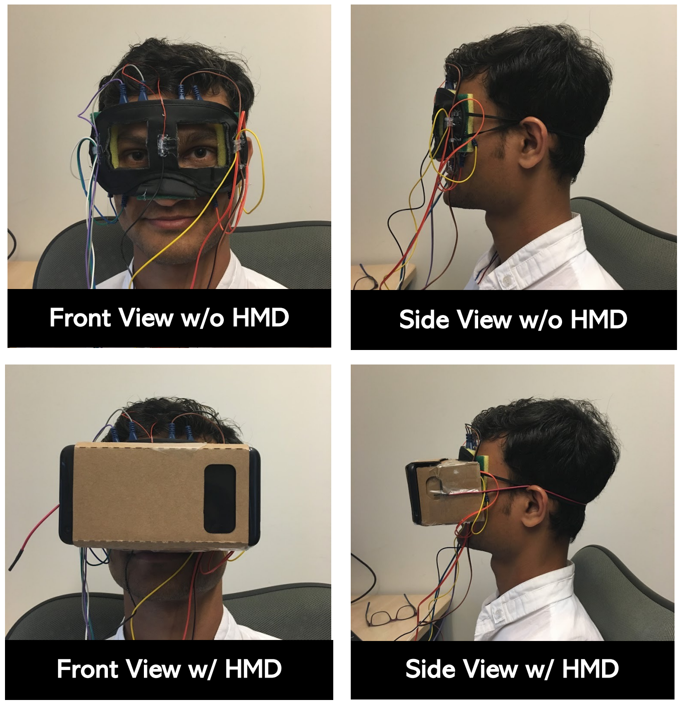


Micro Benchmark



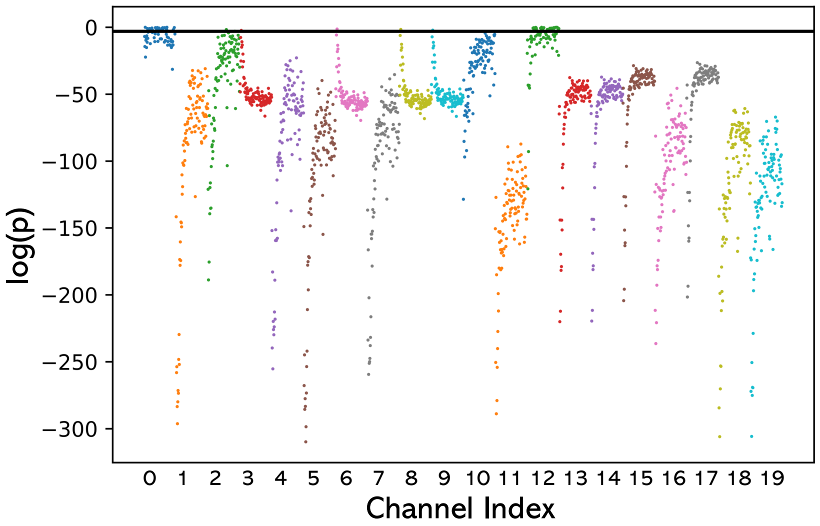
- 4 Participants;
- Eye Gestures: None, Blink, Gaze Up/Down/Left/Right;
- Horizontal Mouth Gestures: Small, Medium, Large;
- Vertical Mouth Gestures: Small Medium, Large;

Prototype

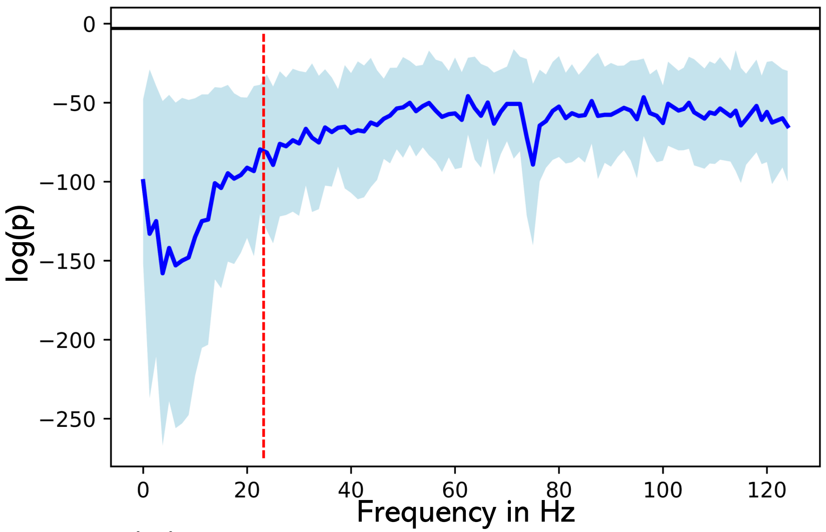


Algorithms – Finding Cut-Off Frequencies

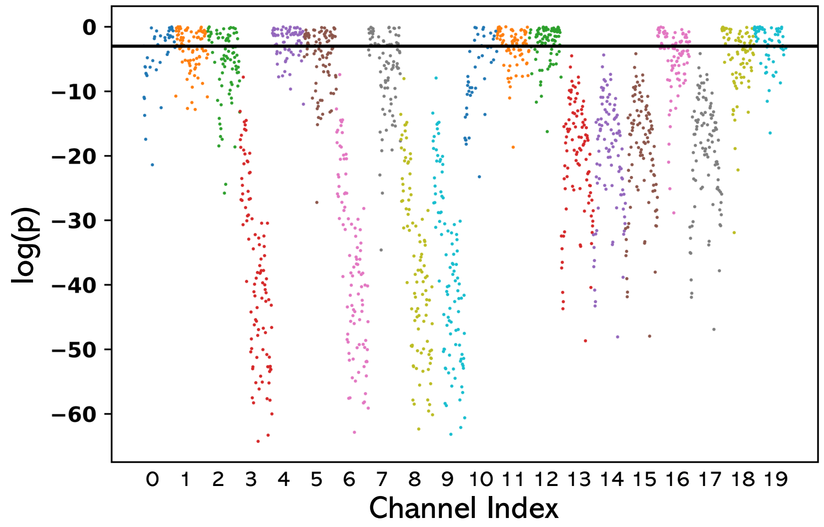
- Goal: Find optimal frequency that separates EMG & EOG;
- Only use FFT to perform the feature significance test;
- (a – b) p-value corresponding to each FFT bin at each channel;
- (c - d) p-value corresponding to each frequency, averaged across all channels;
- ~25Hz;



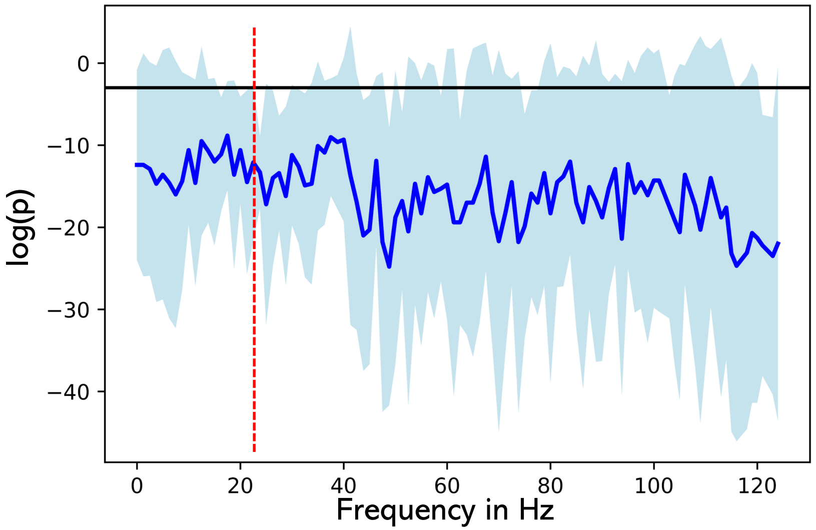
(a) Raw eye gesture results



(c) Avg. mouth gesture results



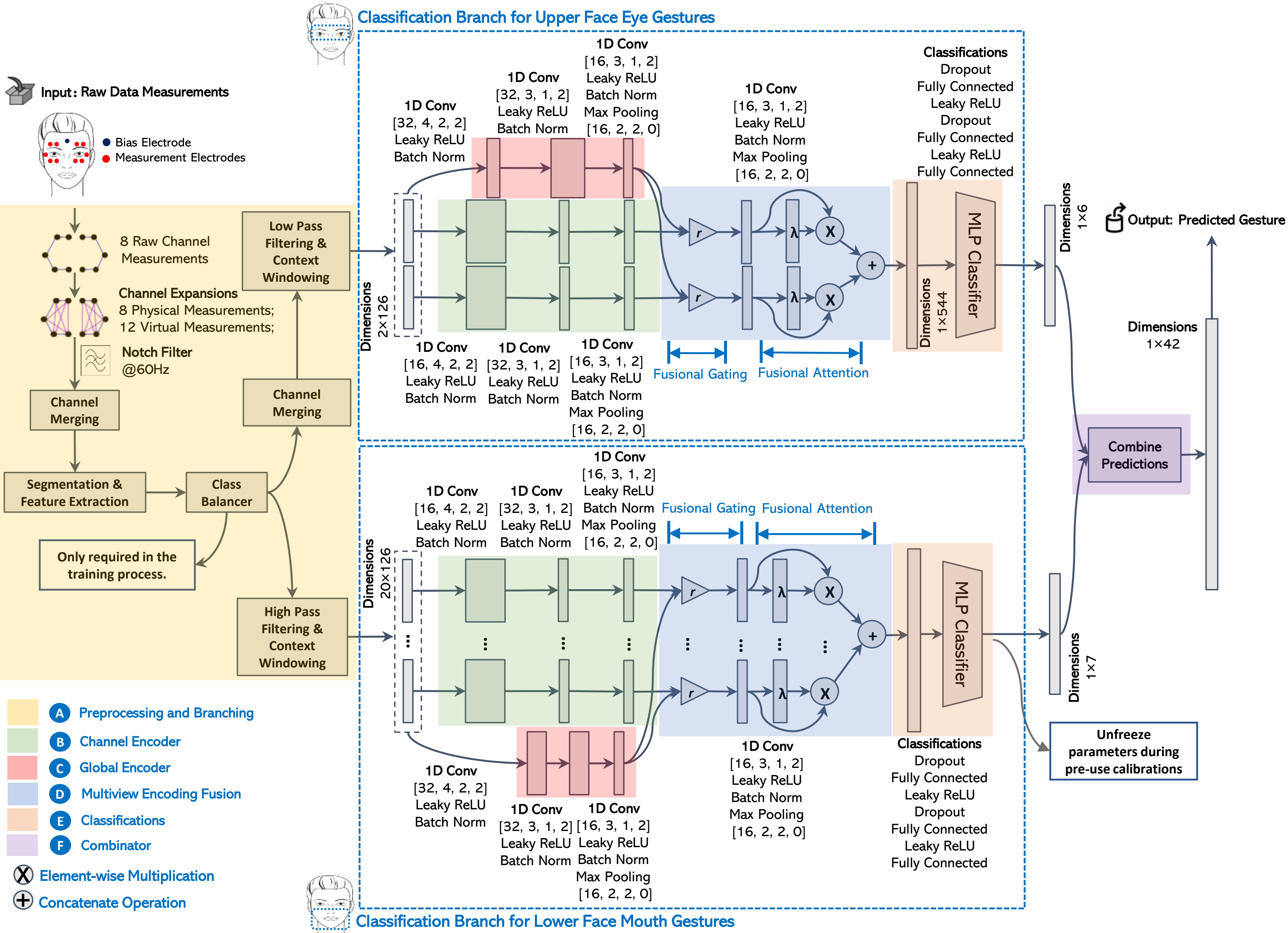
(b) Raw mouth gesture results



(d) Avg. mouth gesture results

Algorithms – Dual Branch Multi-View Classifications

- Channel & Global Encoders;
- Cross-Entropy Losses;
- Only the classification branch will be un-freezed for calibration purpose;
- PyTorch based implementation;



Sensing Events

- Emotion sensing and gesture detections are different;
- A Combinations of:

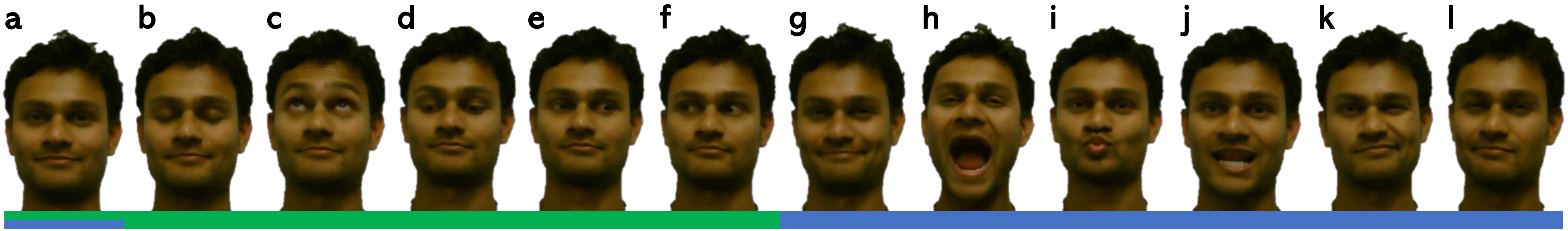
6 Eye Gestures [Green]:

- Neutral;
- Blinks;
- Gaze Looking Up;
- Gaze Looking Down;
- Gazing Looking Left;
- Gaze Looking Right;

7 Mouth Gestures [Blue]:

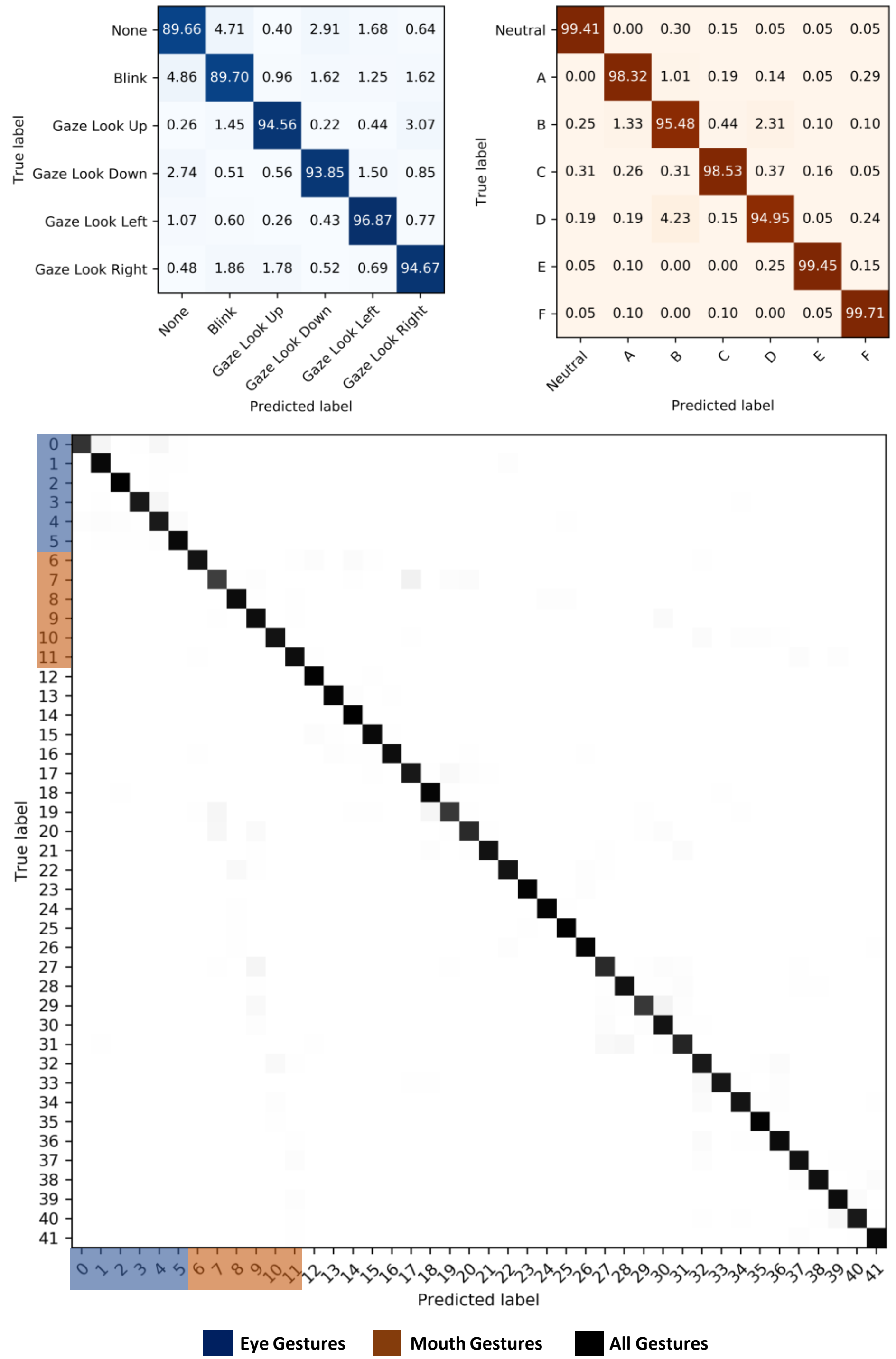
- Neutral;
- [A] Smile;
- [B] Mouth Open;
- [C]Kissy Mouth;
- [D] Tongue Touch Upper Teeth;
- [E] Raising Left Cheek;
- [F] Raising Right Cheek;

Index	Eye	Mouth	Index	Eye	Mouth
0	None	None	21	Gaze Right	B
1	Blink	None	22	Blink	C
2	Gaze Up	None	23	Gaze Up	C
3	Gaze Down	None	24	Gaze Down	C
4	Gaze Left	None	25	Gaze Left	C
5	Gaze Right	None	26	Gaze Right	C
6	None	A	27	Blink	D
7	None	B	28	Gaze Up	D
8	None	C	29	Gaze Down	D
9	None	D	30	Gaze Left	D
10	None	E	31	Gaze Right	D
11	None	F	32	Blink	E
12	Blink	A	33	Gaze Up	E
13	Gaze Up	A	34	Gaze Down	E
14	Gaze Down	A	35	Gaze Left	E
15	Gaze Left	A	36	Gaze Right	E
16	Gaze Right	A	37	Blink	F
17	Blink	B	38	Gaze Up	F
18	Gaze Up	B	39	Gaze Down	F
19	Gaze Down	B	40	Gaze Left	F
20	Gaze Left	B	41	Gaze Right	F



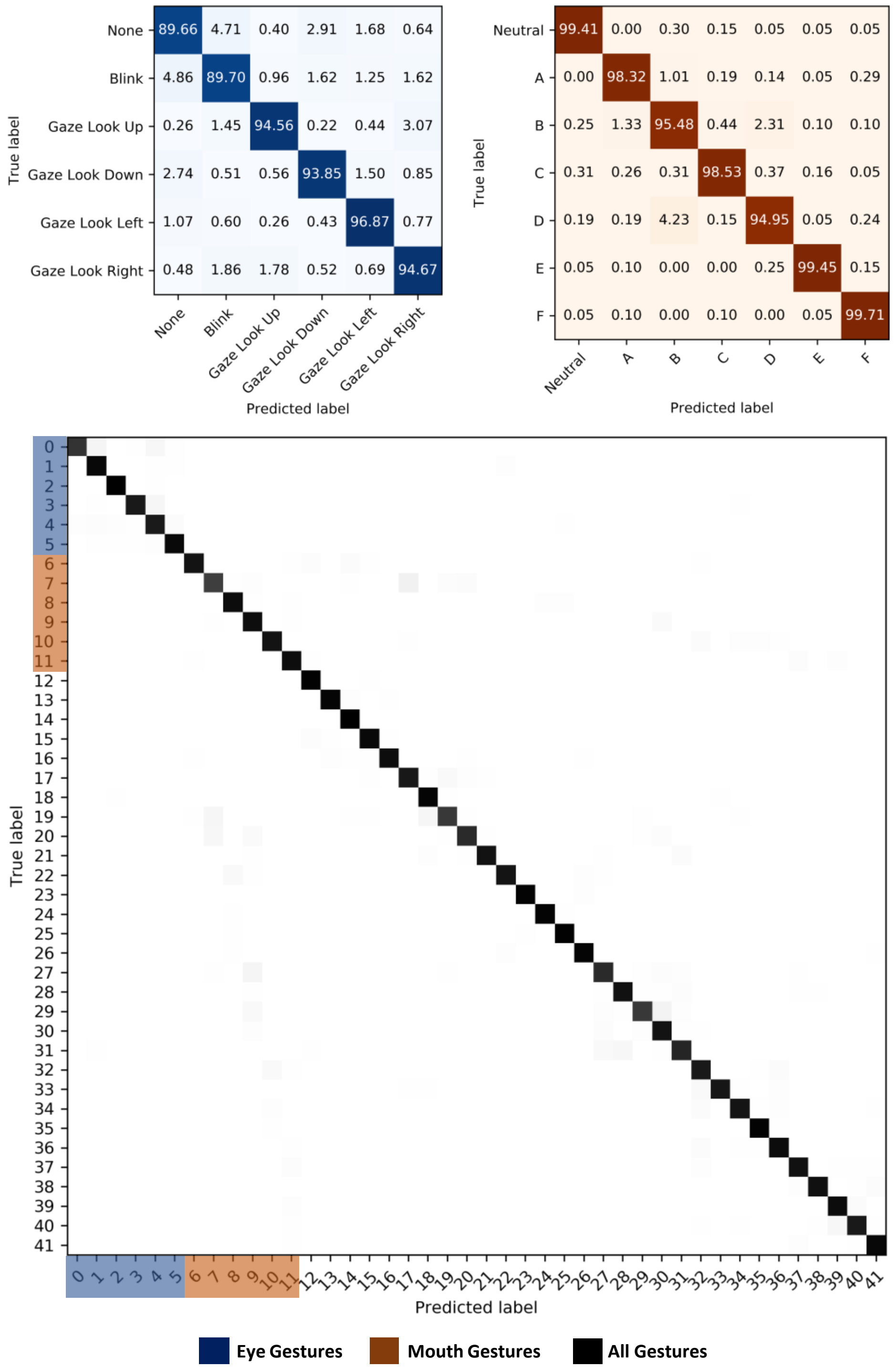
Single User Evaluations

- 10 participants;
- Train and test on the single user;
- The data of non-mixture gesture of each user is split into 70%, 10% and 20% for training, validation and testing;
- The model is train by 6 + 7 = 13 gestures;
- F1 score is used for balancing between precision and recall;



Single User Evaluations

- 10 participants;
- Train and test on the single user;
- The data of non-mixture gesture of each user is split into 70%, 10% and 20% for training, validation and testing;
- The model is train by 6 + 7 = 13 gestures;
- F1 score is used for balancing between precision and recall;
- Findings:
  - Overall ~93% accuracy;
  - Competitiveness of dual-branch classification pipeline (reduce # of training examples from 6 X 7 = 42 to 6 + 7 = 13);
  - Sensitivity of mouth gesture detections when mixed with eye gestures (~98%);



User Independent Evaluations without Partial Calibrations

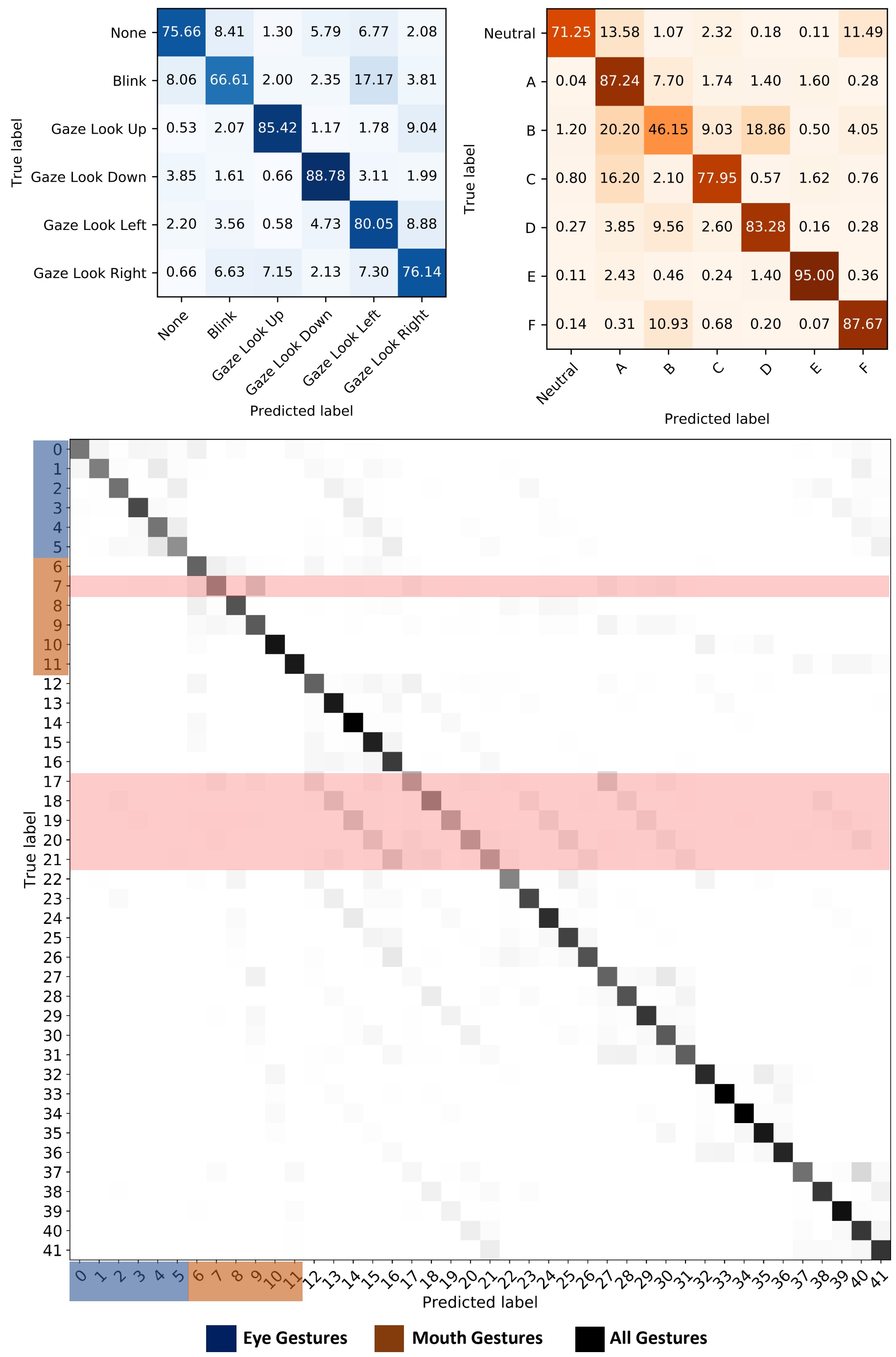
- Understand how ExGSense can be adapted to an "unseen" participants;

User Independent Evaluations without Partial Calibrations

- Understand how ExGSense can be adapted to an "unseen" participants;
- Leave one user out cross validation;

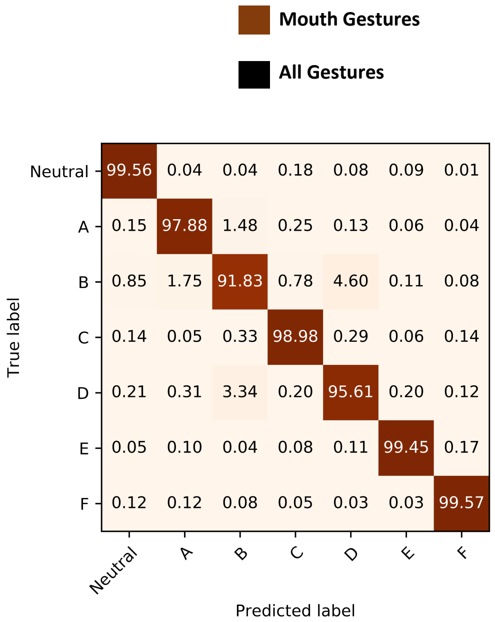
User Independent Evaluations <u>without</u> Partial Calibrations

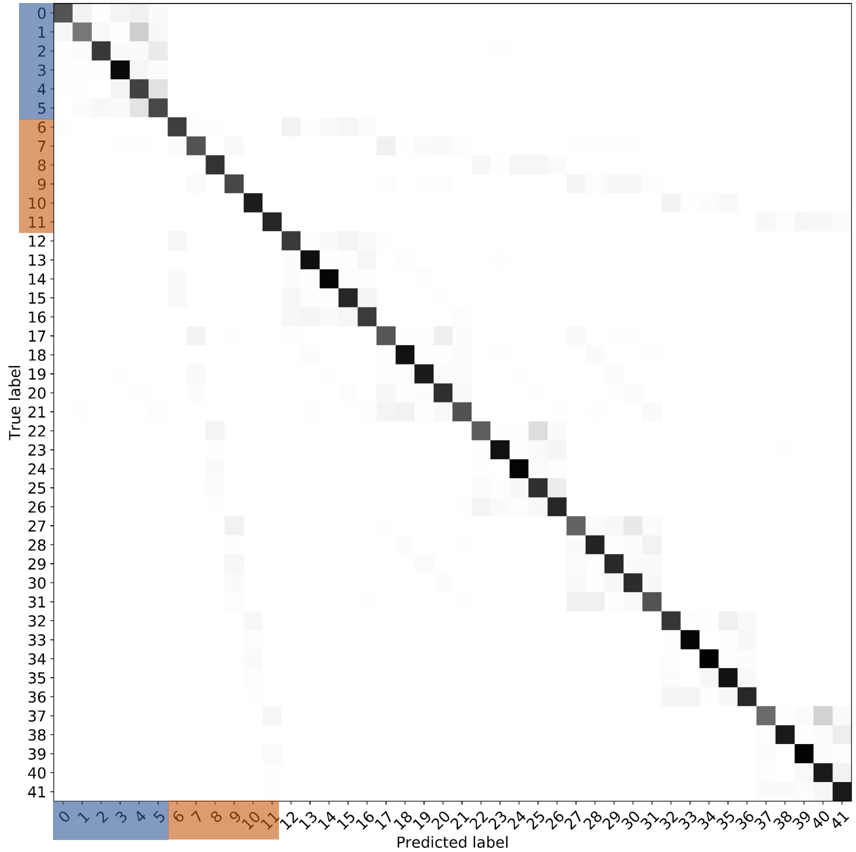
- Understand how ExGSense can be adapted to an “unseen” participants;
- Leave one user out cross validation;
- ~80% for eye and ~78% for mouth;
- Low performance for mouth gesture B (mouth open);

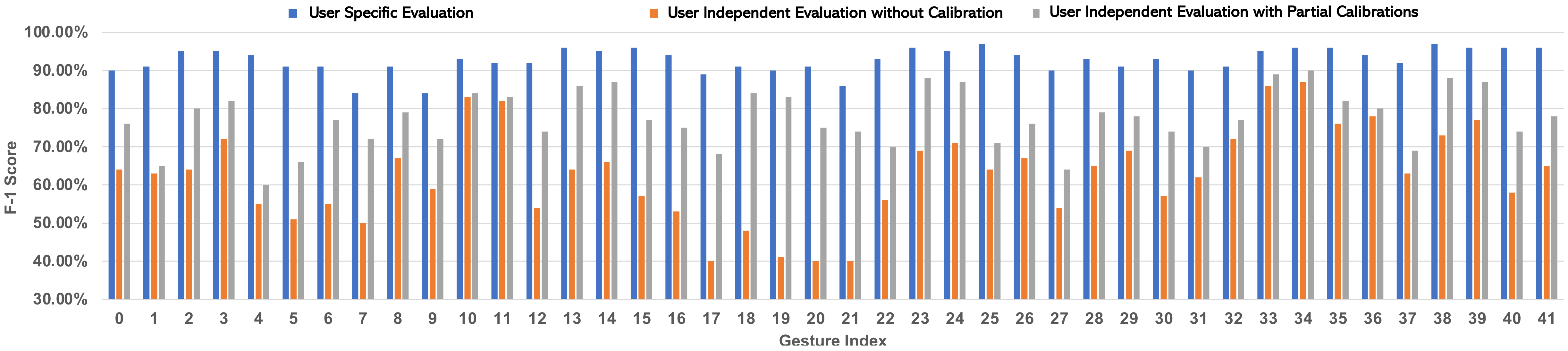


User Independent Evaluations with Partial Calibrations

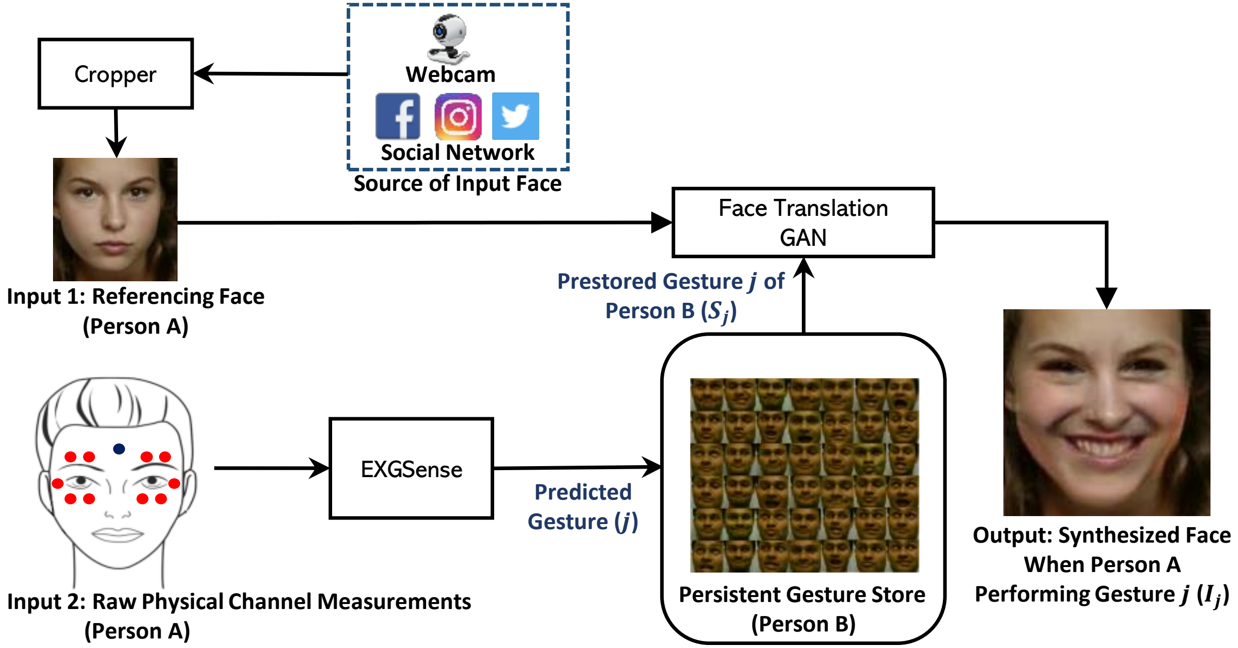
- A new user is asked to provide ~2min data for 6 non-mixture mouth gestures (exclude the none class);
- An overall accuracy of ~77%;

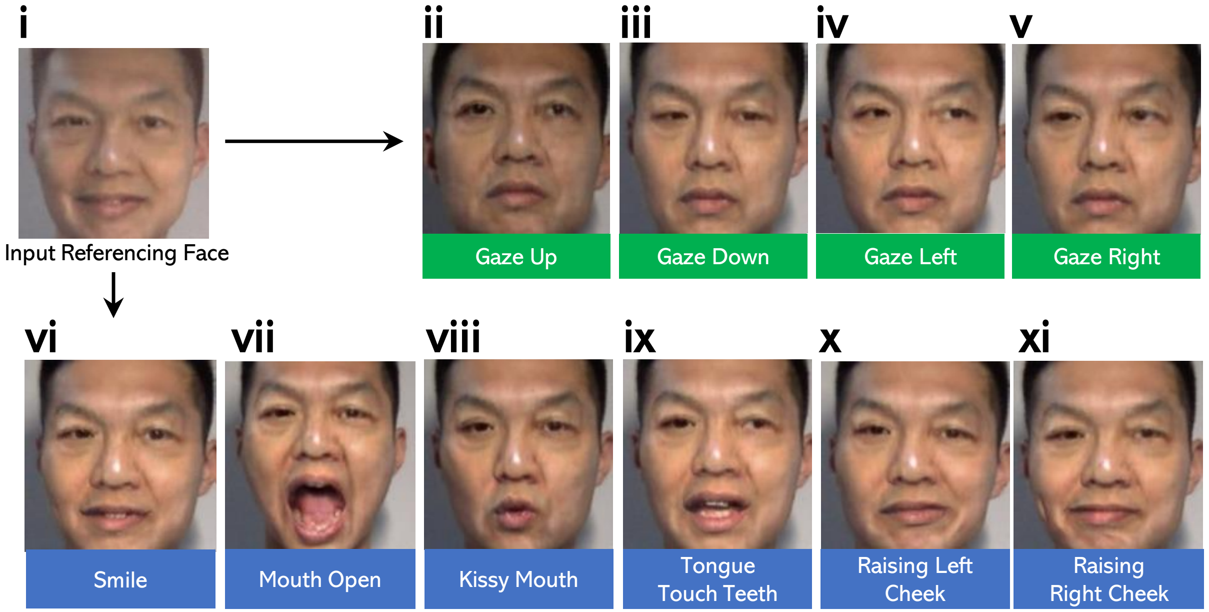






Examples


Cropper
Webcam
Social Network
Source of Input Face
Face Translation GAN
Input 1: Referencing Face (Person A)
Prestored Gesture j of Person B (S_j)
EXGSense
Predicted Gesture (j)
Persistent Gesture Store (Person B)
Output: Synthesized Face When Person A Performing Gesture j (I_j)
Input 2: Raw Physical Channel Measurements (Person A)


i
Input Referencing Face
ii
Gaze Up
iii
Gaze Down
iv
Gaze Left
v
Gaze Right
vi
Smile
vii
Mouth Open
viii
Kissy Mouth
ix
Tongue Touch Teeth
x
Raising Left Cheek
xi
Raising Right Cheek

Limitations & Future Work

Limitations & Future Work

- Real-time and continuous facial tracking and reconstructions;

Limitations & Future Work

- Real-time and continuous facial tracking and reconstructions;
- While transferring to new users, a small set of training samples are still required for model calibration purpose;

Limitations & Future Work

- Real-time and continuous facial tracking and reconstructions;
- While transferring to new users, a small set of training samples are still required for model calibration purpose;
- Continuous efforts on addressing multi-facet issues related to usability and incorporations into commercially available VR headset;

Limitations & Future Work

- Real-time and continuous facial tracking and reconstructions;
- While transferring to new users, a small set of training samples are still required for model calibration purpose;
- Continuous efforts on addressing multi-facet issues related to usability and incorporations into commercially available VR headset;
- A more diverse of participants;

Thank You!

Hope you enjoy our work!

For more detail, please refer to our manuscripts!

Contact: chenchen@ucsd.edu