

***BAH* Dataset for Ambivalence/Hesitancy Recognition in Videos for Behavioural Change**

Manuela González-González², Soufiane Belharbi¹, Muhammad Osama Zeeshan¹, Masoumeh Sharafi¹, Muhammad Haseeb Aslam¹, Marco Pedersoli¹, Alessandro Lameiras Koerich², Simon L Bacon², Eric Granger¹

¹ LIVIA, ETS Montreal, Canada

² Dept. of Health, Kinesiology & Applied Physiology, Concordia University, Montreal, Canada



CHAIRE DOUBLE FRSQ
EN IA ET EN SANTÉ
NUMÉRIQUE



FRSQ DOUBLE CHAIR
IN AI AND DIGITAL
HEALTH



**LABORATOIRE
D'IMAGERIE, DE VISION
ET D'INTELLIGENCE
ARTIFICIELLE**

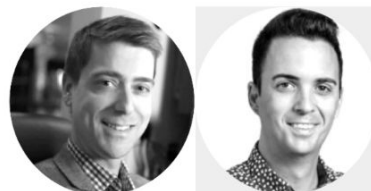
Accredited lab at ETS since 25+ years: 14 regular and 10+ associate

Ben Ayed, Ismail



Cardinal, Patrick

Desrosiers, Christian



Dolz, Jose

Duchaine, Vincent



Granger, Eric

Koerich, Alessandro



Pedersoli, Marco

Sabourin, Robert



Toews, Matthew

Wong, Tony



Menelau-Cruz, Rafael

Forouzanfar, Mohamad



Shateri, Mohammadhadi



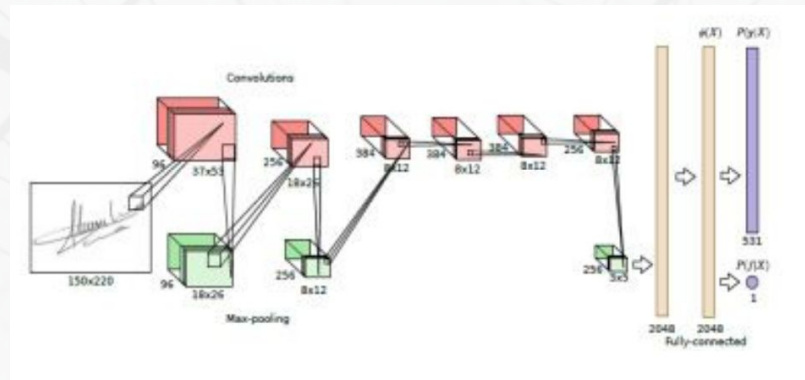
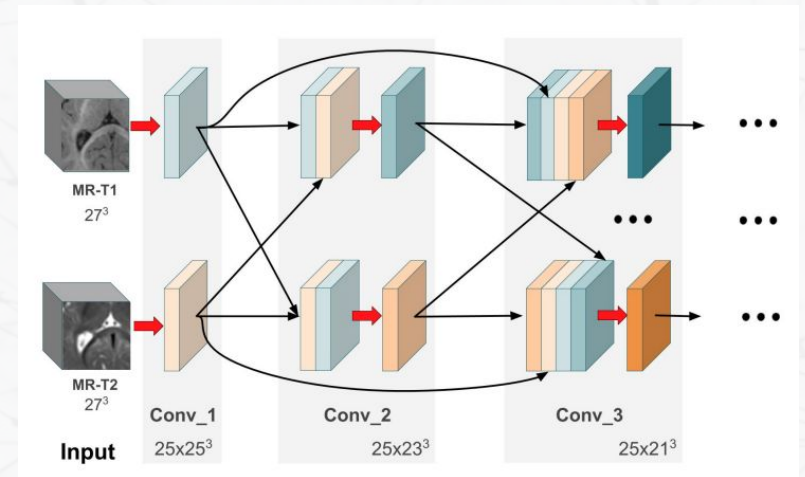
<https://www.etsmtl.ca/en/labs/livia>



**LABORATOIRE
D'IMAGERIE, DE VISION
ET D'INTELLIGENCE
ARTIFICIELLE**

Research axes:

- Machine learning
- Computer vision: perception in 2D and 3D scenes
- Pattern recognition in static and dynamically-changing environments
- Information fusion
- Optimization of complex systems

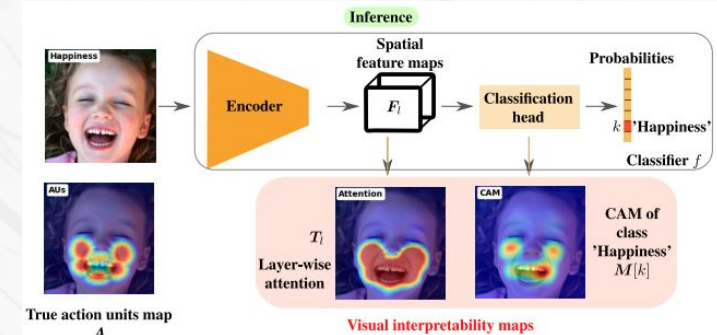
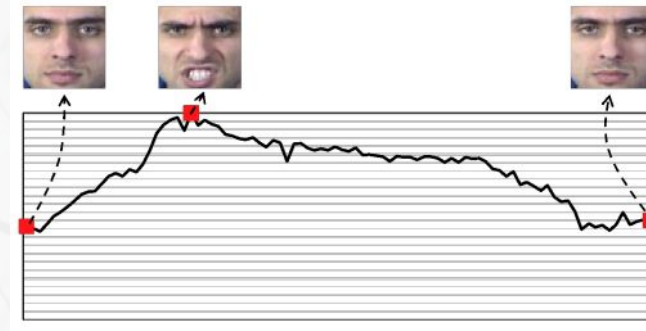
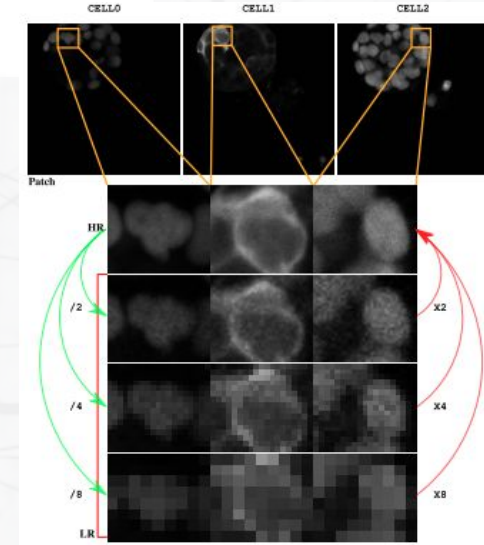
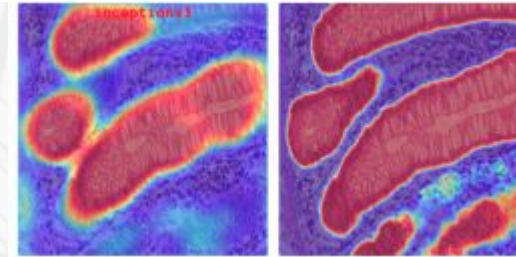
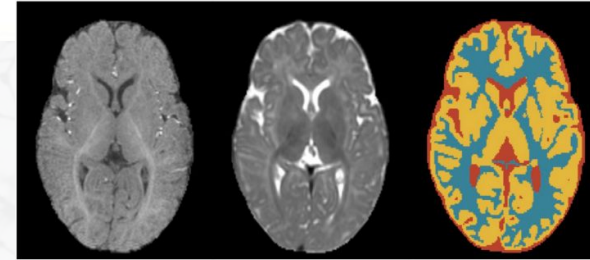




LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTIFICIELLE

Applications axes:

- Analysis of medical, aerial images
- Video analytics and surveillance
- Biometrics (face, voice and signature)
- Document analysis
- Affective computing



Context: Behavioral change and online interventions

Health-related behaviour change focuses on strategies to support individuals in adopting and maintaining healthy behaviours to:

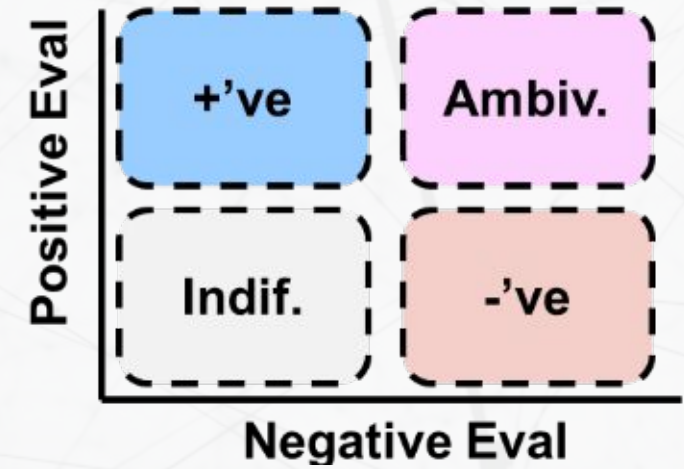
- Prevent or manage chronic diseases
- Reduce early mortality
- Improve mental health and well-being

In **in-person interventions**, it is important for therapists/clinicians to determine **ambivalence and hesitancy** to help overcome it.



<https://www.bozemancounseling.org/blog/2024/3/24/exploring-therapy-a-guide-to-different-types-of-mental-health-therapists>

Context: Behavioral change and online interventions



Ambivalence and hesitancy:

- The **simultaneous** presence of **competing positive and negative** feelings, ideas, thoughts or emotions towards one same object or goal.
- **Simultaneous** feeling of being **resistant and willing** to do something.
- Conflicting/subtle affect

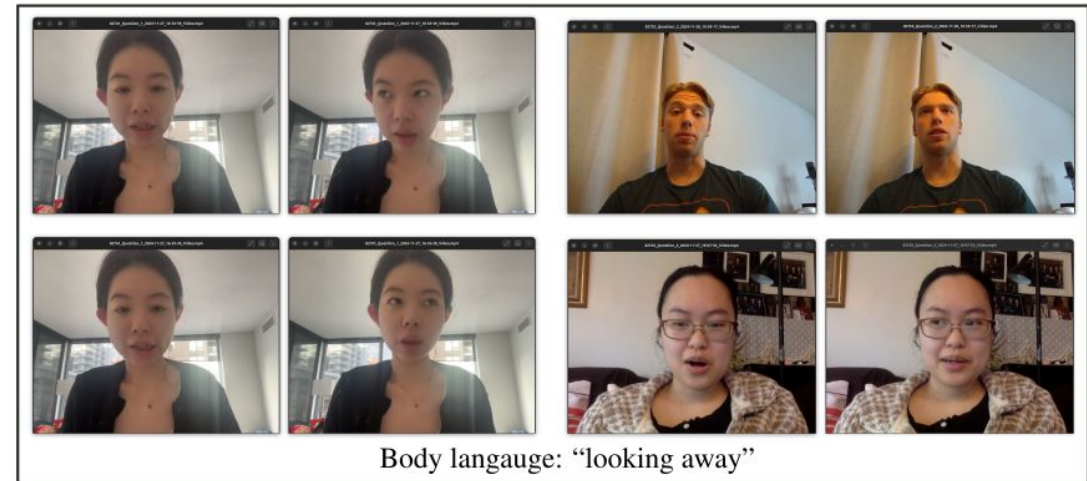
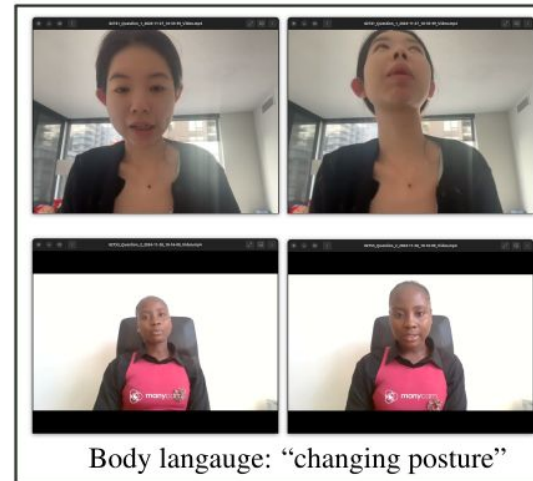
Context: Behavioral change and online interventions

Online interventions:

- Personalized digital health (eHealth) interventions
- Easily scalable / cheap
- Fully automatic
- Requires an automatic system to recognize ambivalence/hesitancy moments to act accordingly

No available A/H datasets.

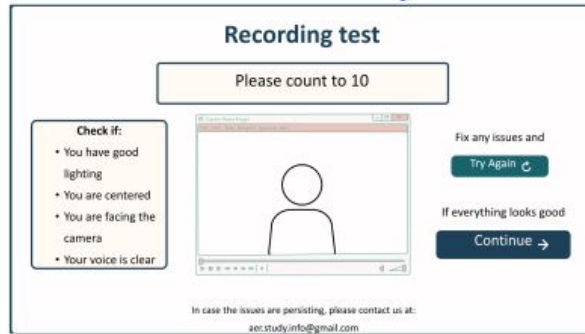
BAH Dataset



- **BAH: Behavioural Ambivalence/Hesitancy** (A/H) dataset
- Task: A/H recognition in videos
- 300 participants in Canada
- Online videos: answers to 7 predefined questions
- 1,427 videos (~10.6 hours where ~1.8 hours contain A/H)
- 916,618 frames
- Annotation: video/frame level, cues, inconsistencies

BAH Dataset: Collection

- 1 Our online platform for data collection
www.aerstudy.ca



Participant



Task: answer the designed questions while recording themselves via webcam

- 2 Data transfer
by Administrator

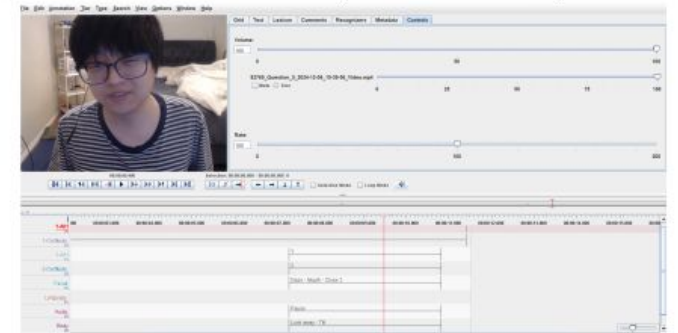


Our local secured
storage server for videos

Secured access



- 3 Video annotation for A/H at multiple levels:
video-, frame-level, cues



Our behavioural expert annotator

BAH Dataset: Collection

1 Our online platform for data collection
www.aerstudy.ca

Recording test

Please count to 10

Check if:

- You have good lighting
- You are centered
- You are facing the camera
- Your voice is clear



Fix any issues and
Try Again ↺

If everything looks good
Continue →

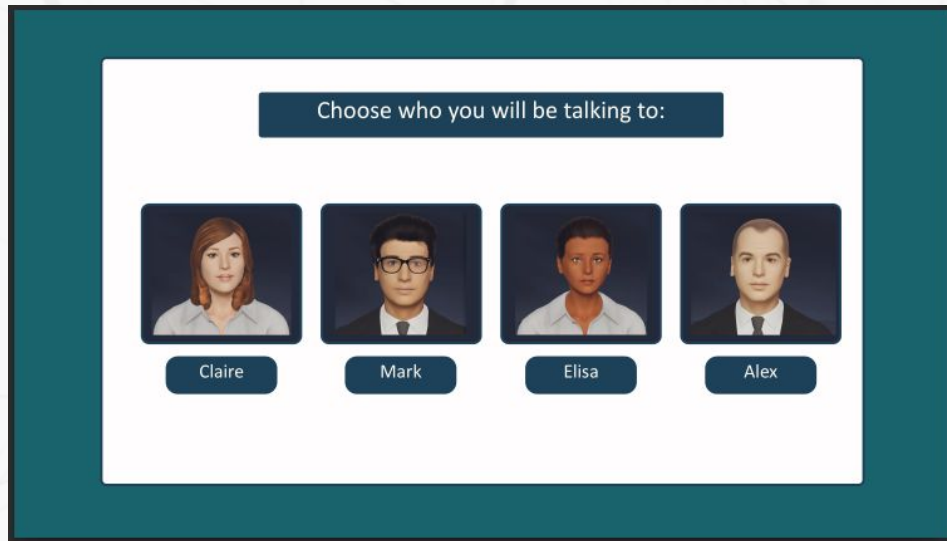
In case the issues are persisting, please contact us at:
aerstudy.info@gmail.com

Participant

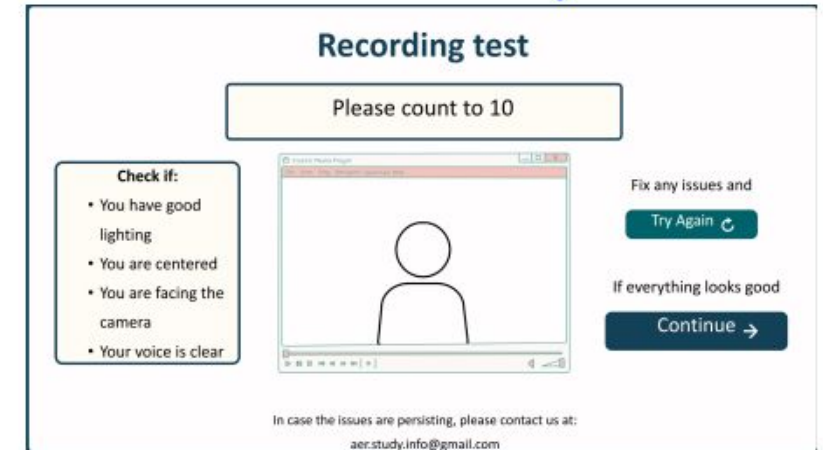


Task: answer the designed questions while recording themselves via webcam

BAH Dataset: Collection



1 Our online platform for data collection
www.aerstudy.ca

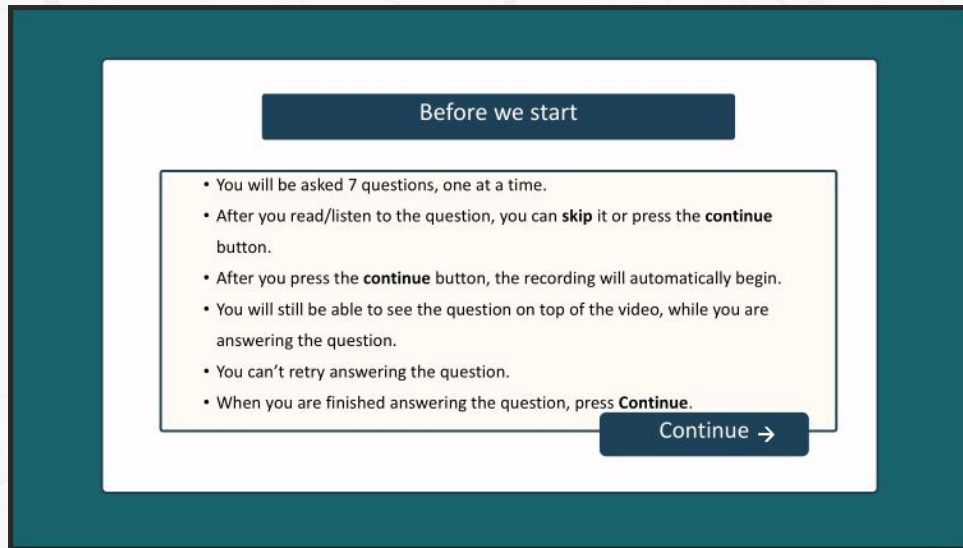


Participant

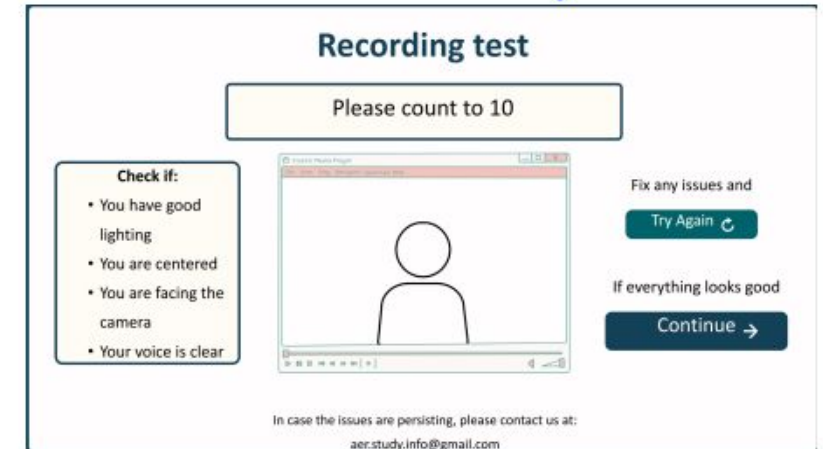


Task: answer the designed questions while recording themselves via webcam

BAH Dataset: Collection



1 Our online platform for data collection www.aerstudy.ca

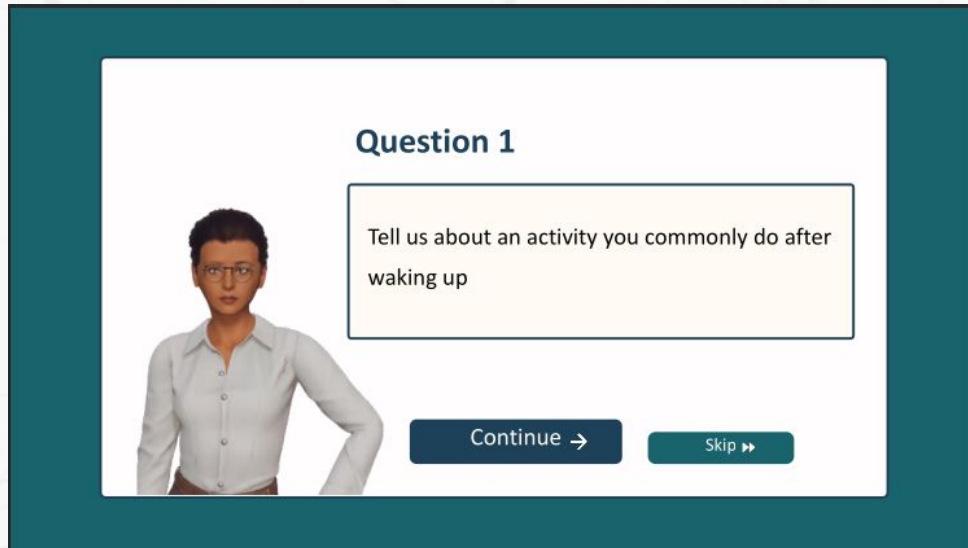


Participant

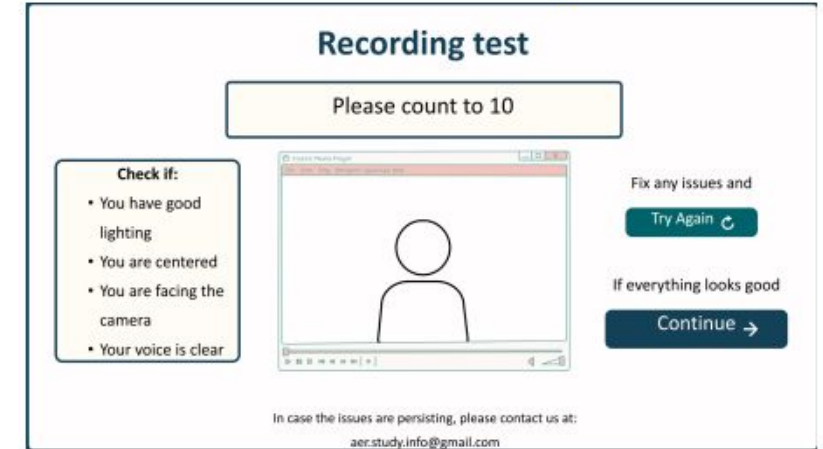


Task: answer the designed questions while recording themselves via webcam

BAH Dataset: Collection



1 Our online platform for data collection
www.aerstudy.ca



Participant



Task: answer the designed questions while recording themselves via webcam

BAH Dataset: Collection

Question no.	Response	Prompt
1	Neutral	Tell us about an activity you commonly do after waking up.
2	Positive	Talk about an activity that brings you joy, for example, a hobby. Tell us why.
3	Negative	Talk about an activity you dislike doing, for example, a chore or something you find boring or annoying. Tell us why.
4	Ambivalent	Tell us about something you enjoy doing but wish you stopped doing (like a guilty pleasure) or something you don't do but wish you did.
5	Willing	Tell us about an activity you are almost always willing to do, for example with friends, at work, at home.
6	Resistant	Tell us about something people around you do, but that you would not be willing to do, for example, with friends, at work, at home.
7	Hesitant	Tell us about something you could have done already but haven't done yet, for example, something you are procrastinating or haven't made up your mind about.

1 Our online platform for data collection
www.aerstudy.ca

Recording test

Please count to 10

Check if:

- You have good lighting
- You are centered
- You are facing the camera
- Your voice is clear



Fix any issues and
Try Again ↺

If everything looks good
Continue →

In case the issues are persisting, please contact us at:
aer.study.info@gmail.com

Participant



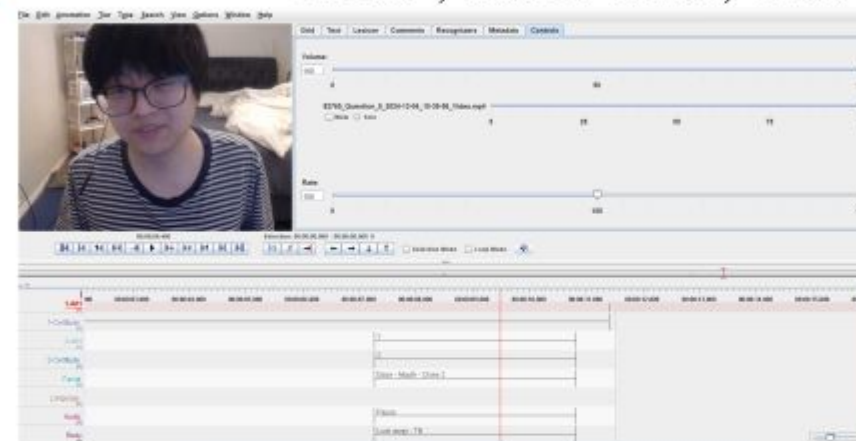
Task: answer the designed questions while recording themselves via webcam

BAH Dataset: Annotation

- 3 behavioral expert annotators
- Use of codebooks: annotation, cues (face, body, audio, language), cross-modality inconsistencies
- Annotation: indicate the presence/absence of A/H
- Levels: videos/frames, timestamps: start/end A/H
- Additional information: annotation cues, inconsistencies

3

Video annotation for A/H at multiple levels:
video-, frame-level, cues



Our behavioural expert annotator

BAH Dataset: Annotation

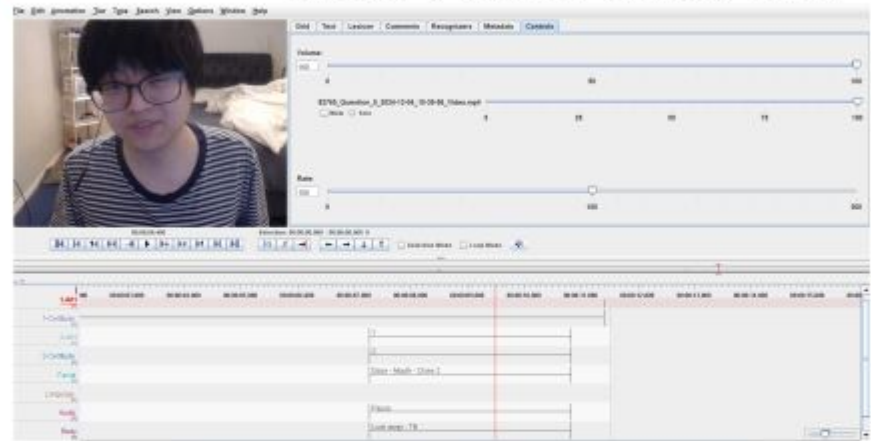


	00:00:05.000	00:00:06.000	00:00:07.000	00:00:08.000	00:00:09.000
AH [2]	Y				
Certitude [2]	2				
Facial [2]	gaze-close 2-eyebrow-eye roll				
Language [1]					
Audio [1]	stutter-breath				
Body [1]	posture-sigh-tilt				
Inconsistency [1]					

3

SS

Video annotation for A/H at multiple levels:
video-, frame-level, cues



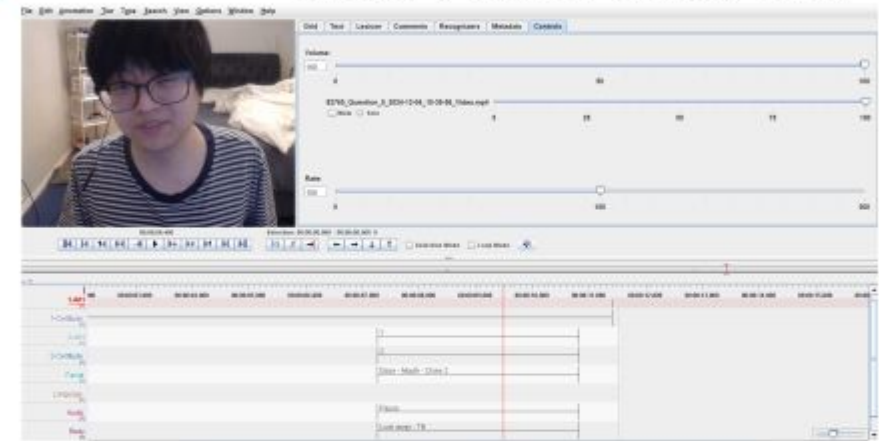
Our behavioural expert annotator

BAH Dataset: Annotation

Annotation levels	Description	Annotated variables				
		Presence of AH	Level of certitude	Time stamps	Modality used	Cues
Level 1	Global annotation	Yes	Yes	No	No	No
Level 2	Frame level	Yes	Yes	Yes	No	No
Level 3	Modality focused	Yes	Yes	Yes	Yes	No
Level 4	Cue focused	Yes	Yes	Yes	Yes	Yes

3

Video annotation for A/H at multiple levels: video-, frame-level, cues



Our behavioural expert annotator

BAH Dataset: Annotation

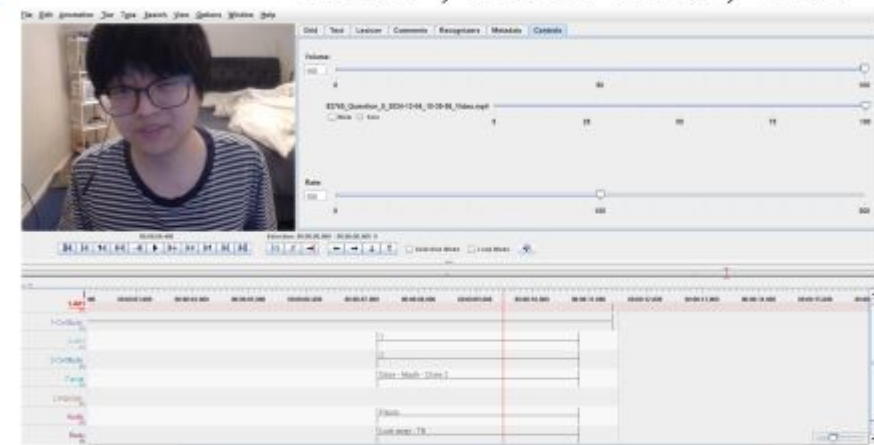
Term	Definition
Ambivalence/Hesitancy	The simultaneous presence of competing positive and negative feelings, ideas, thoughts, or emotions towards one same object or goal. A state in which a person has not entirely made up their mind about doing something; when they aren't fully decided on how to act (towards a behaviour or object; not necessarily the goal behaviour; excluding towards language or answering questions)
Facial Cues	Different motions of the muscles in the face. Facial expressions commonly occur around the mouth and eyes, including changes in a person's gaze. They can be used to assess a person's emotional state.
Language Cues	Includes verbal/speech-based expressions of ambivalence or hesitancy. Some common verbal expressions can include the use of 'I want to... but...', 'mmmm', among others.
Audio Cues	Changes in a person's non-verbal language, such as changes in tone, speed and pitch.
Body Cues	Non-verbal signals that include gestures, body posture and movements. Some of the cues that can be annotated as body language are hand movements, head tilts, shoulders shrugging and sighs (chest movement).
Cross-modal inconsistency Cues	Simultaneous incompatibility between two or more modalities or different types of cues. For example, this could be represented by someone saying 'yes' while shaking their head side to side.

Table 8 BAH dataset annotation codebook: definitions.

3

Video annotation for A/H at multiple levels:
video-, frame-level, cues

SS



Our behavioural expert annotator

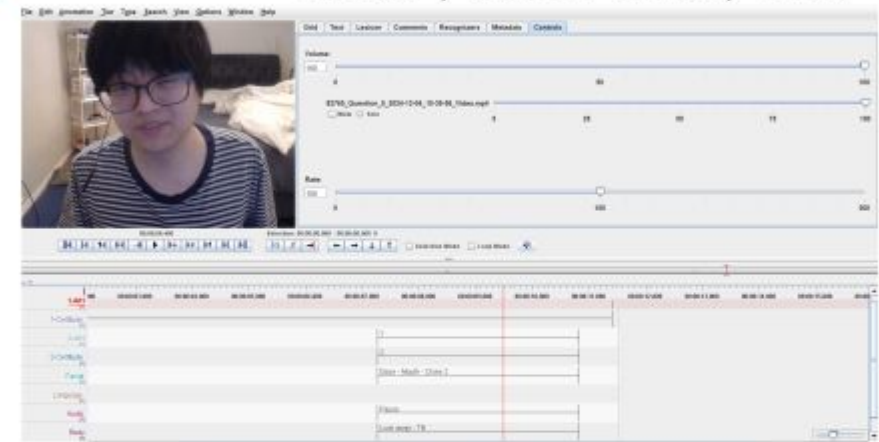
BAH Dataset: Annotation

Body cues	Definition
Look away	Moving the orientation of the head away from the baseline position such that eyes or the gaze will look away. Includes the head facing down, head facing up, looking down, looking up, looking from side to side, lowering head, raising head.
Shake	Turning the head from side to side, it can be done with repetitive head movements or with a slight turn of the head to one or both sides. Includes shaking head "no". Rotation is on the horizontal plane
Tilt	Angling the head to the side without focusing on something else, and holding the position. Changing the position of the head so it is in a sloping position. It can be accompanied by changes in the gaze but not necessarily. Includes head tilting up and down, tilting head to the side, tilted head. Includes bobbling head.
Throw	Throwing the head in a rapid movement in a particular direction.
Sigh	Movements of the chest, shoulder or head that accompany a sigh or a deep breath. It includes long sigh, deep breath, sigh, big sigh. Noticeable bringing the chest or diaphragm muscles up and down. Change determined in comparison to the person's own baseline.
Nod	Moving the head up and down. Lowering and raising the head, it can be done by slight or clearly marked movements. Includes movements such as back and forward or a single small nod.
Shrug	Raising of the shoulders, it can be a momentary or slight rise or a longer movement where one or both shoulders is raised. It includes shrugging shoulders, shrugs
Hands	Movements or placement of the hands that differs from baseline
Posture	Movements in the overall positioning of the spine, body or arms (independent from the head). The changes are determined by each person's baseline. It includes movements like readjusting in the seat, sloughing, turning to the sides. Needs to involve more than just the head. Excludes shrugging.
Scratch	Movements in the hands and arms to scratch or caress another part of the body or face. It includes scratching head, scratching neck, scratching eyes, scratching chin
Restless	Rhythmic and repeated movements. Can be swaying, shaking, being jittery.

Table 12 BAH dataset annotation codebook: body cues.

3

Video annotation for A/H at multiple levels:
video-, frame-level, cues



Our behavioural expert annotator

BAH Dataset: Annotation

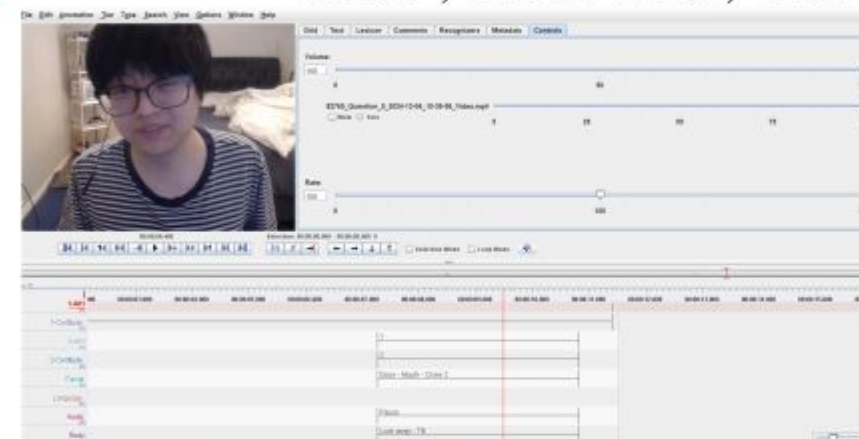
Cross-modal inconsistency	Definition
FL	Face and language/speech do not match. E.g., looking uncomfortable while saying yes, looking annoyed or uncomfortable while saying they are happy, smiling while saying they are worried.
FA	Face and audio do not match. E.g., speaks in a sad, energetic tone while smiling.
FB	Face and body do not match. E.g., Nodding while looking afraid or concerned, showing disgust but leaning forward
LA	Language/speech and audio do not match. E.g., speaks in a sad, energetic tone while saying they are happy.
LB	Language/speech and body do not match. E.g., seems like they are about to say something but do not, nod is discrepant with verbal speech, shaking head while saying yes
AB	Body and language/speech do not match. E.g., unengaged tone while nodding (in agreement)

Table 13 BAH dataset annotation codebook: cross-modal inconsistency cues - occurring simultaneously.

3

Video annotation for A/H at multiple levels:
video-, frame-level, cues

SS



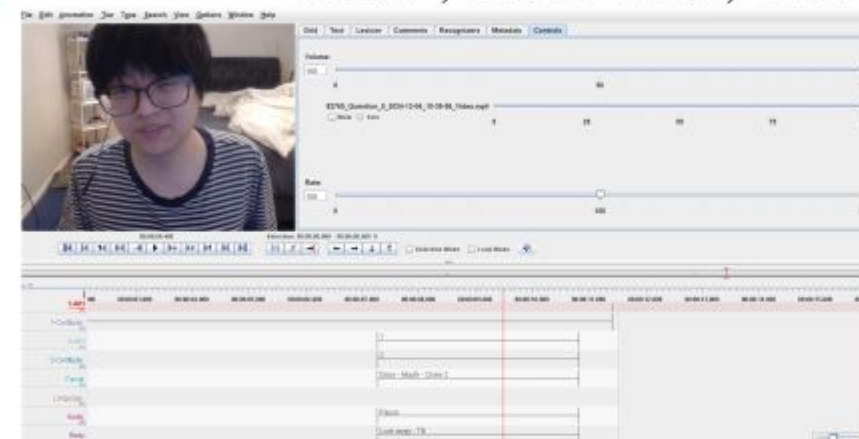
Our behavioural expert annotator

BAH Dataset: Annotation

```
92991 Videos/82569/Visite_1/82569_Question_2_2024-10-02_16-02-30_Video.mp4:
92992 all_cues:
92993 - audio: null
92994   body: look away
92995   facial: Eyebrows-hanging-gaze
92996   inconsistencies: null
92997   language: filler sound-hedging
92998 - audio: breath
92999   body: sigh-nod-restless
93000   facial: gaze-close 2
93001   inconsistencies: LB - FL
93002   language: positive-opp
93003 - audio: slow
93004   body: null
93005   facial: gaze
93006   inconsistencies: null
93007   language: positive-filler sound-repetition
93008 annotator_id: MGG
93009 certainty_ah:
93010 - 1
93011 - 2
93012 - 1
93013 fr_detailed_ah:
93014 - - 55
93015 - - 134
93016 - - 311
93017 - - 389
93018 - - 755
93019 - - 868
93020 frame_annotation:
93021 - - Videos/82569/Visite_1/82569_Question_2_2024-10-02_16-02-30_Video.mp4/frame-0.jpg
93022 - 0
93023 - - Videos/82569/Visite_1/82569_Question_2_2024-10-02_16-02-30_Video.mp4/frame-1.jpg
93024 - 0
93025 - - Videos/82569/Visite_1/82569_Question_2_2024-10-02_16-02-30_Video.mp4/frame-2.jpg
93026 - 0
93027 ...
93028 global_ah: 1
93029 time_detailed_ah:
93030 - - '00:00:02.360'
93031 - - '00:00:05.630'
93032 - - '00:00:13.040'
93033 - - '00:00:16.260'
93034 - - '00:00:31.540'
93035 - - '00:00:36.230'
```

3

Video annotation for A/H at multiple levels:
video-, frame-level, cues



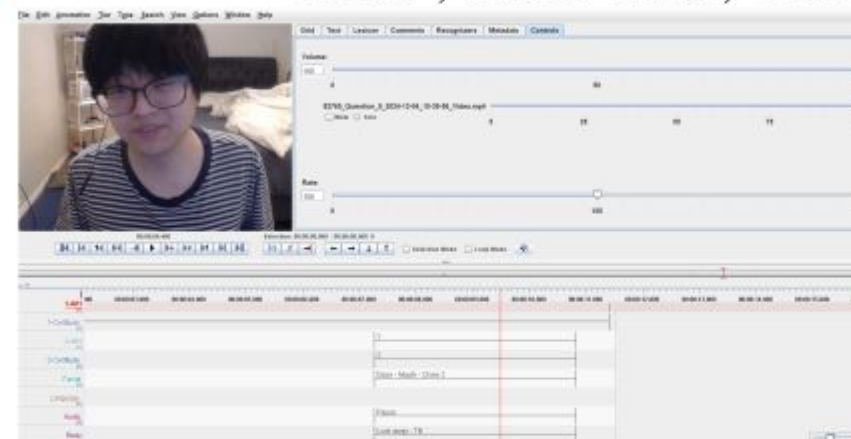
Our behavioural expert annotator

BAH Dataset: Annotation

```
transcript:
chunks:
- language: english
  text: 'The activity that brings me joy, like a hobby, I guess it's cycling.'
  timestamp: !!python/tuple
    - 0.8
    - 5.3
- language: english
  text: 'I do like to go, especially in the summertime, take a long bicycle ride.'
  timestamp: !!python/tuple
    - 5.54
    - 9.42
- language: english
  text: 'I have an e-bike, so it's not super strenuous, but it's more of an
    enjoyable activity where'
  timestamp: !!python/tuple
    - 9.56
    - 14.54
- language: english
  text: 'I could relax and just take in the scenery.'
  timestamp: !!python/tuple
    - 14.54
    - 17.34
- language: english
  text: 'And it's de-stressing going out in nature, and again, for not really
    doing something'
  timestamp: !!python/tuple
    - 18.4
    - 23.52
- language: english
  text: 'super strenuous, but for example, having the wind to your face and the
    sunshine coming'
  timestamp: !!python/tuple
    - 23.52
    - 29.36
- language: english
  text: 'down on a nice beautiful day and going on a nice route. It is enjoyable
    and relaxing for me.'
  timestamp: !!python/tuple
    - 0.0
    - 8.4
text: 'The activity that brings me joy, like a hobby, I guess it's cycling.
I do like to go, especially in the summertime, take a long bicycle ride. I have
an e-bike, so it's not super strenuous, but it's more of an enjoyable activity
where I could relax and just take in the scenery. And it's de-stressing going
out in nature, and again, for not really doing something super strenuous, but
for example, having the wind to your face and the sunshine coming down on a
nice beautiful day and going on a nice route. It is enjoyable and relaxing for
me.'
```

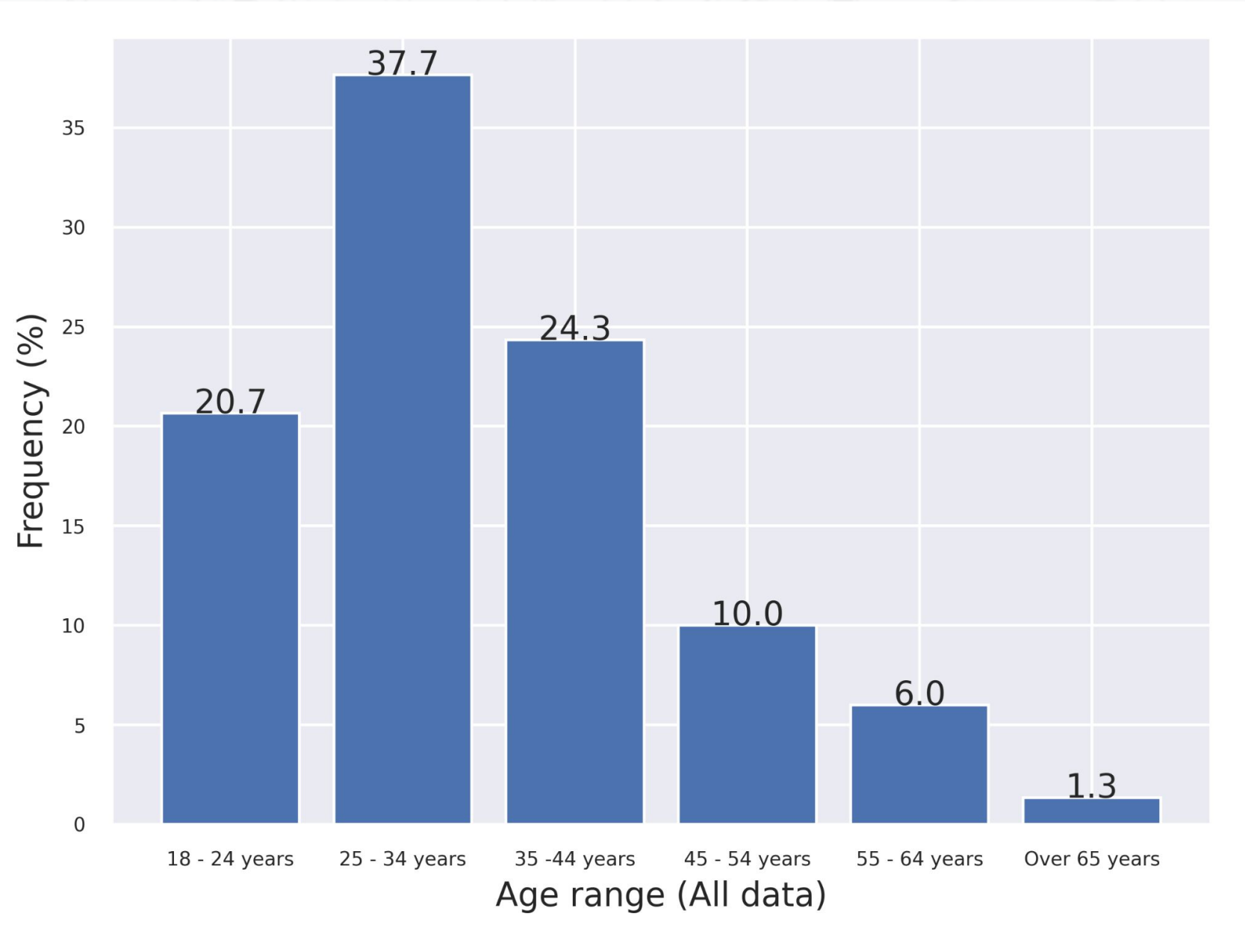
3

Video annotation for A/H at multiple levels:
video-, frame-level, cues

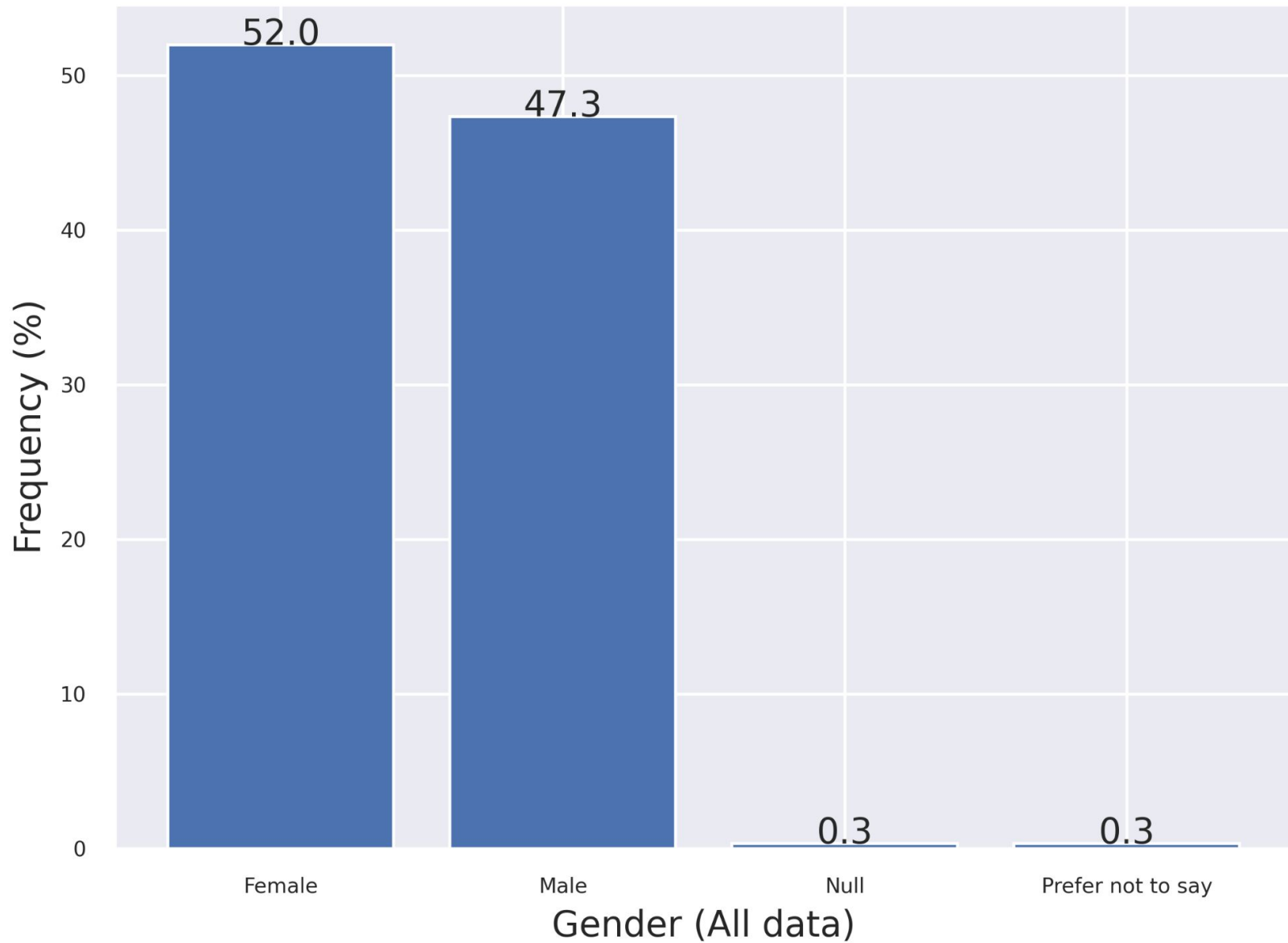


Our behavioural expert annotator

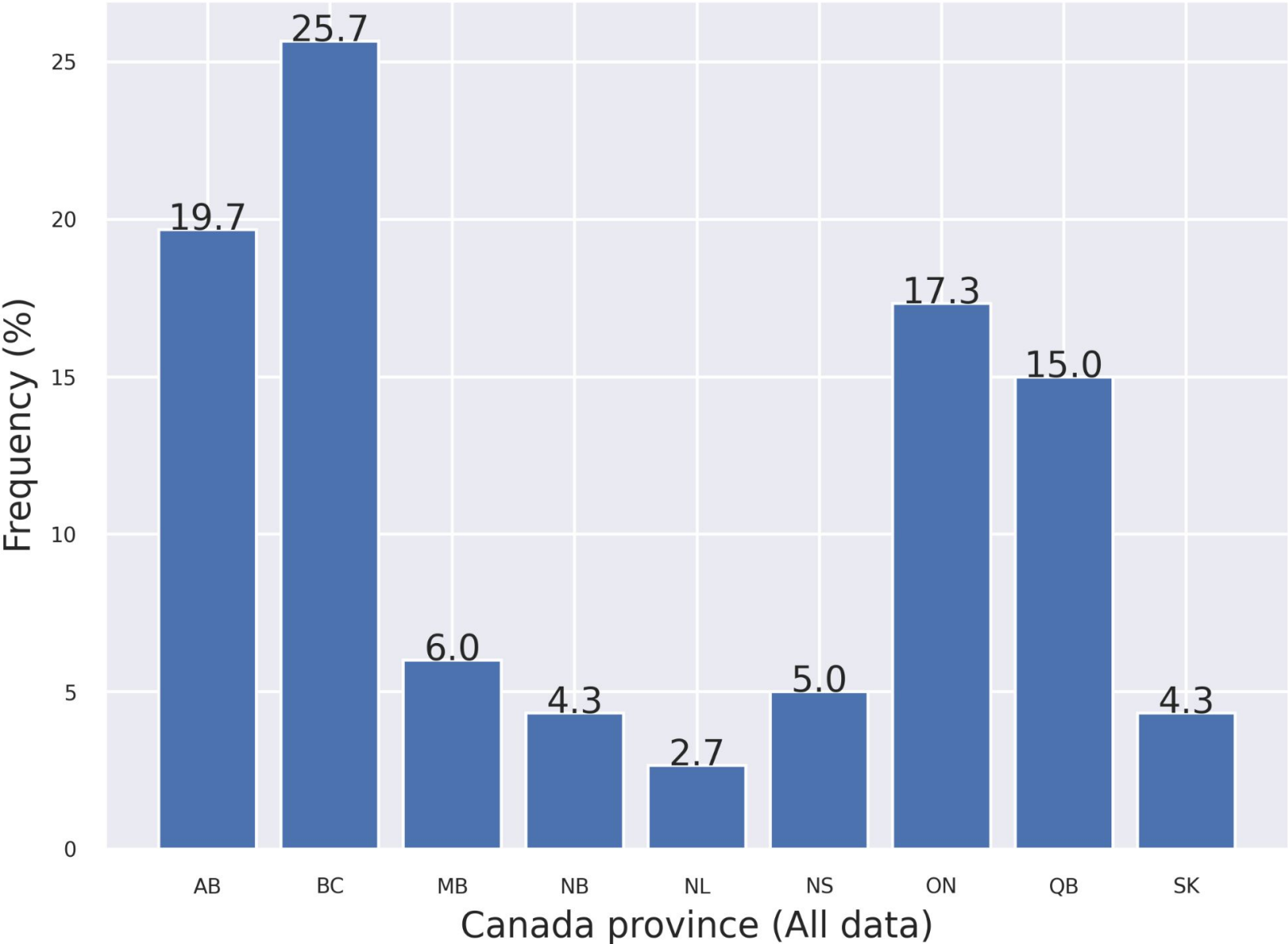
BAH Dataset: Statistics – Age



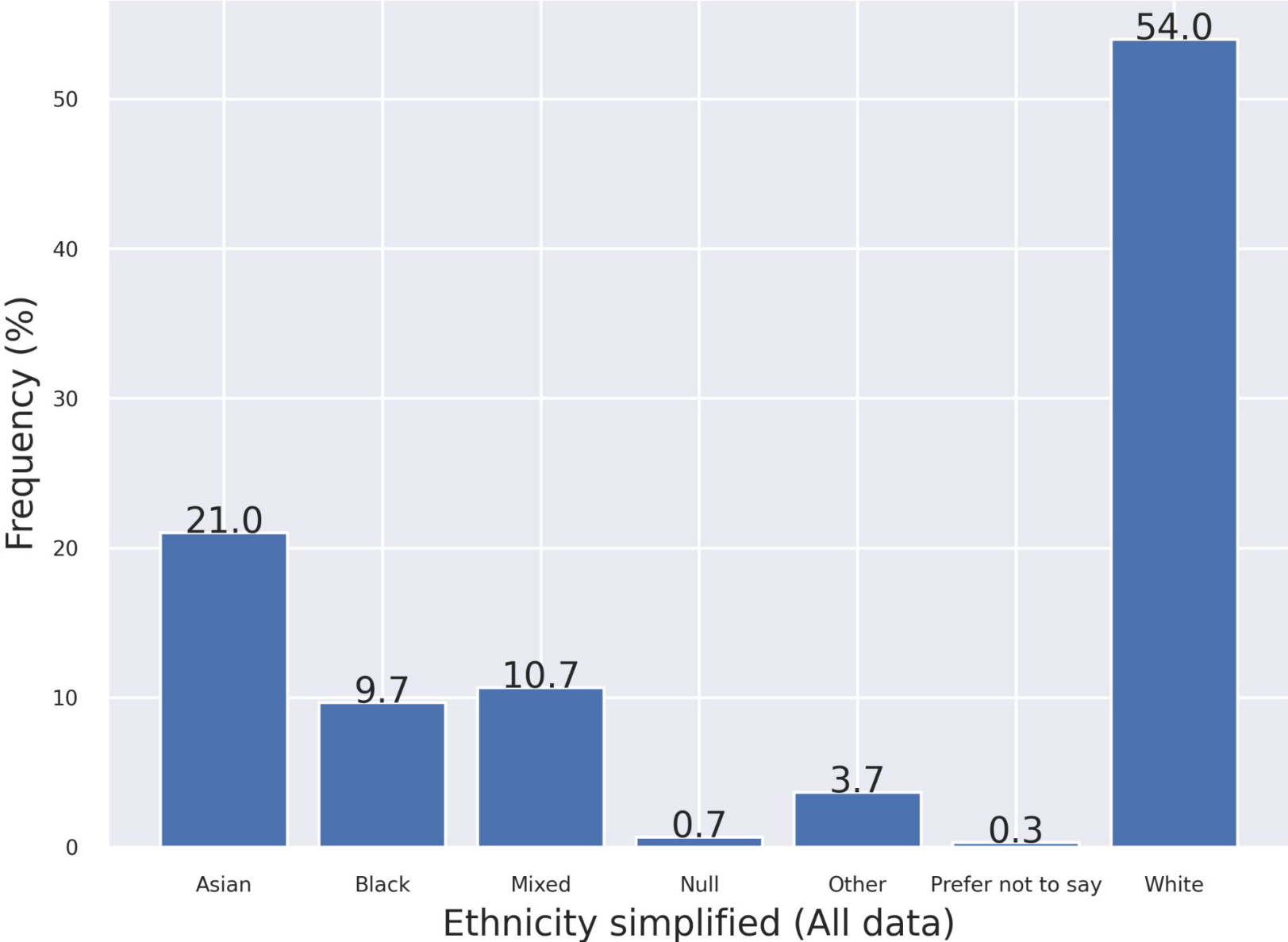
BAH Dataset: Statistics – Gender



BAH Dataset: Statistics – Province



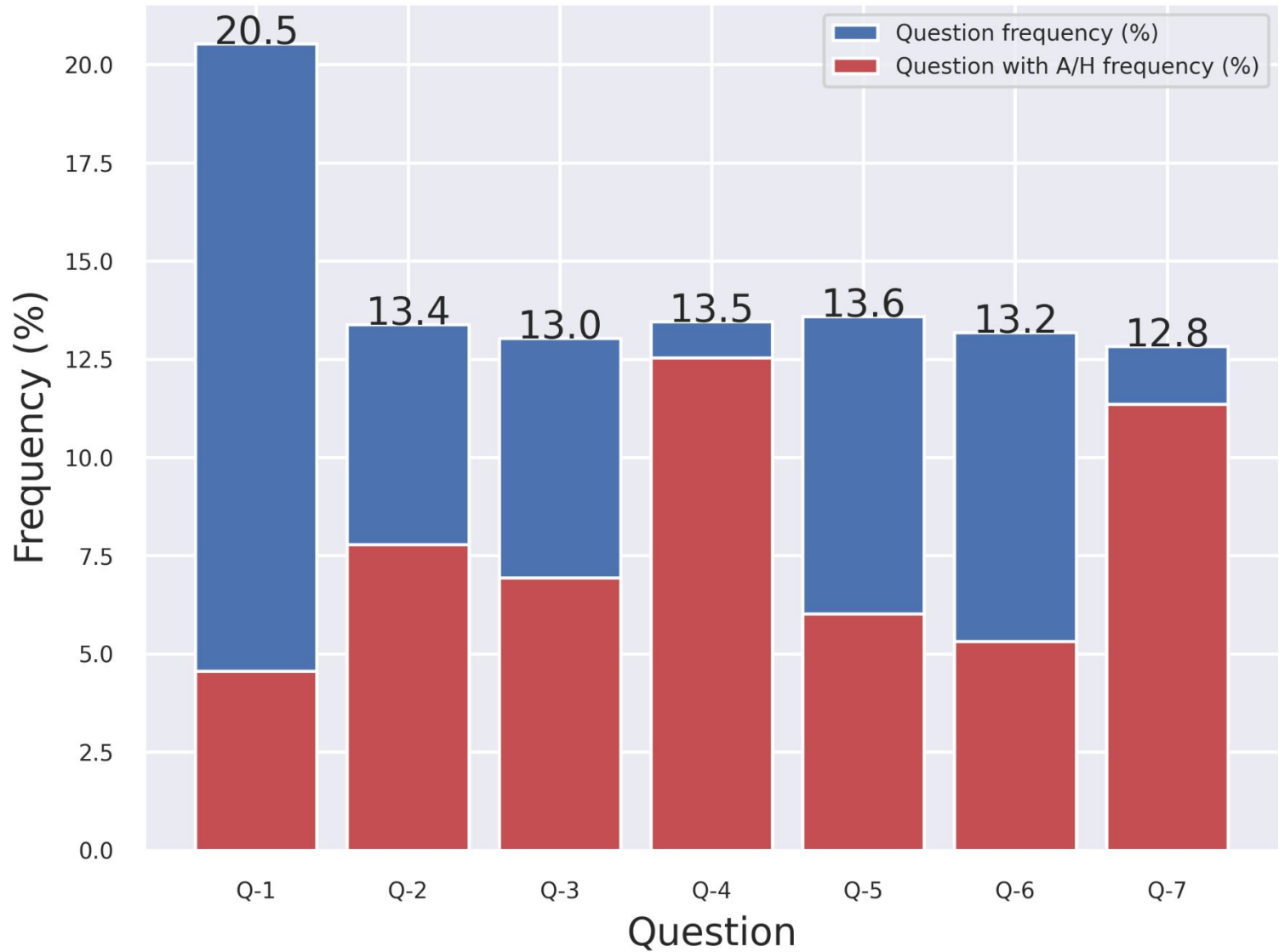
BAH Dataset: Statistics – Ethnicity



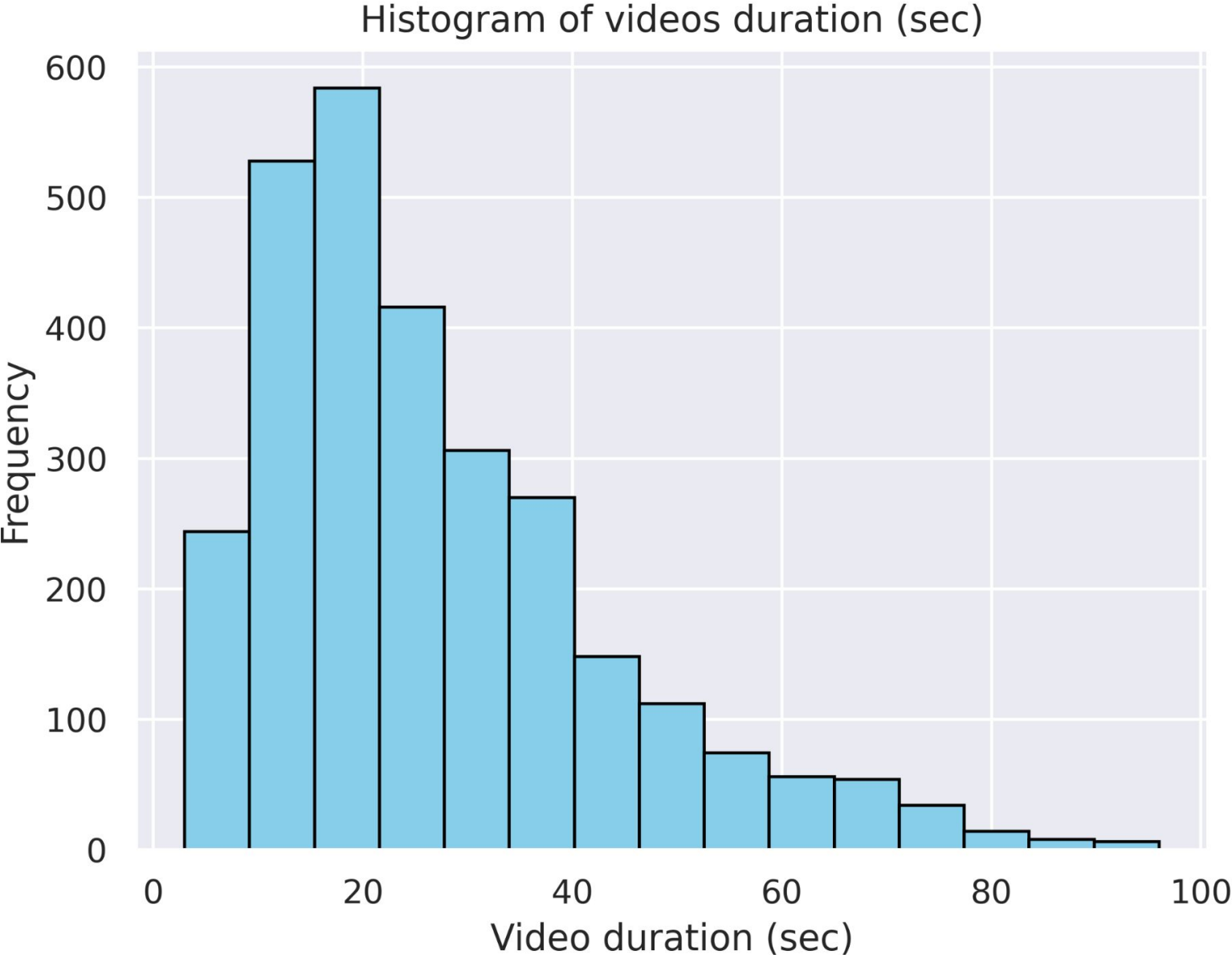
BAH Dataset: Statistics – Birth Country

- Algeria: 1
- Australia: 2
- Bangladesh: 2
- Belize: 1
- Bulgaria: 1
- **Canada: 132**
- **China: 9**
- Colombia: 1
- France: 1
- Germany: 3
- Ghana: 1
- India: 3
- Japan: 1
- Kenya: 2
- Macedonia: 1
- New Zealand: 1
- **Nigeria: 10**
- Null: 1
- Peru: 1
- Philippines: 4
- Russian Federation: 1
- Saint Lucia: 2
- Sri Lanka: 2
- Taiwan: 1
- Trinidad and Tobago: 1
- Tunisia: 1
- Turkey: 2
- Ukraine: 1
- United Arab Emirates: 1
- United Kingdom: 5
- United States: 4
- Vietnam: 1

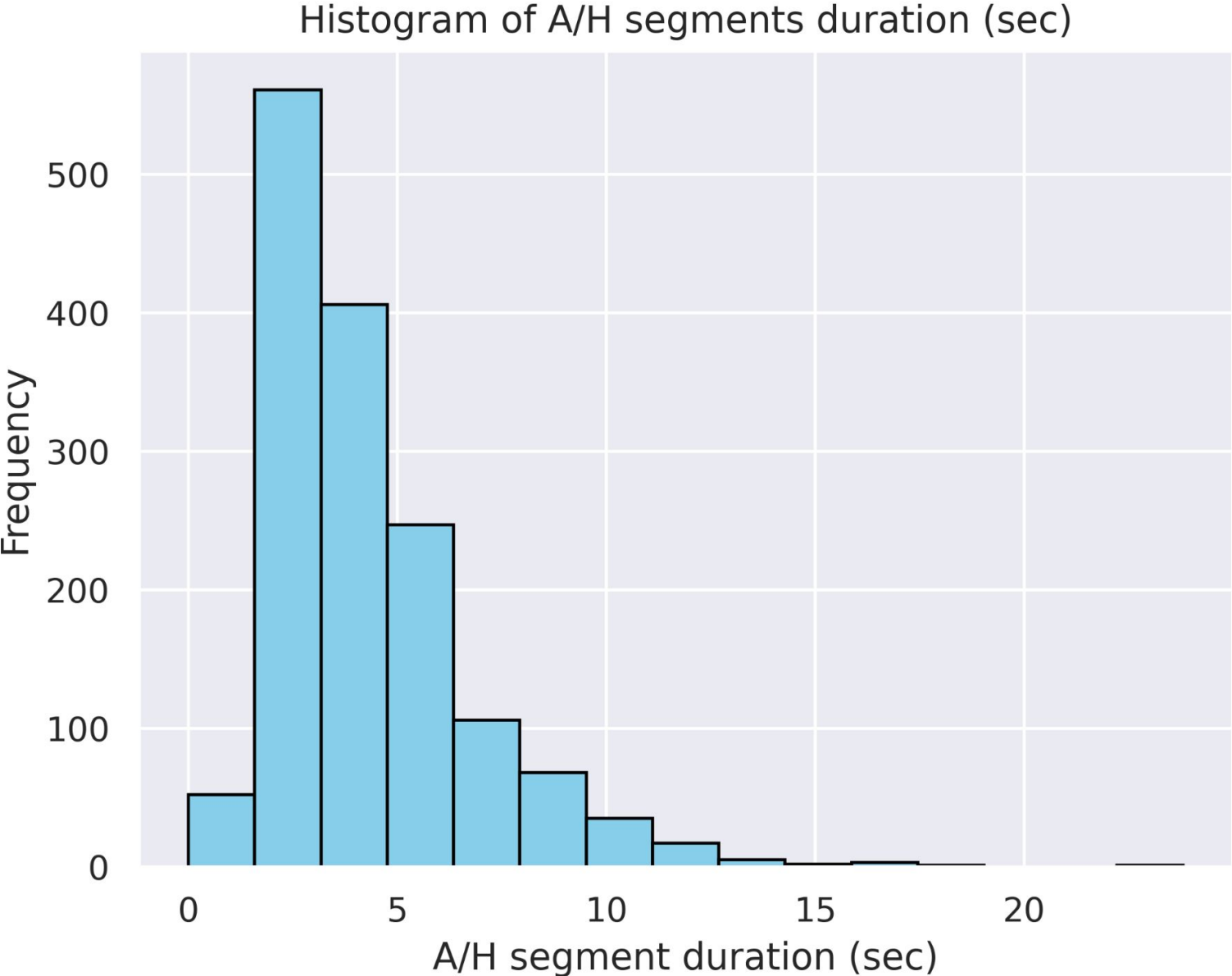
BAH Dataset: Statistics – Questions / A/H



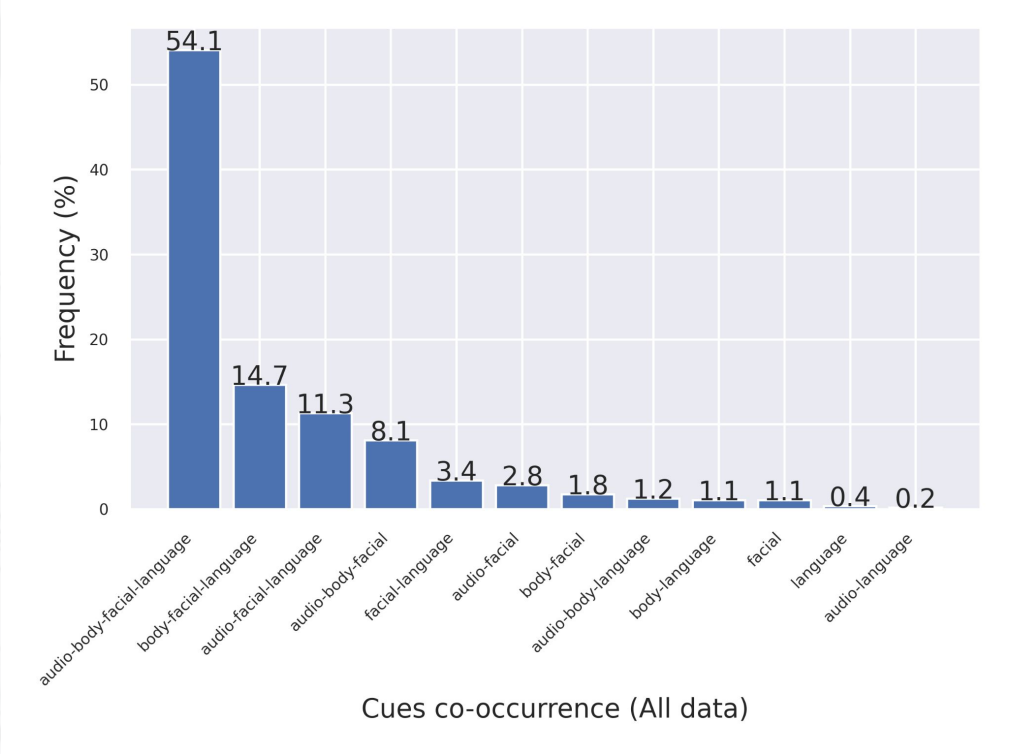
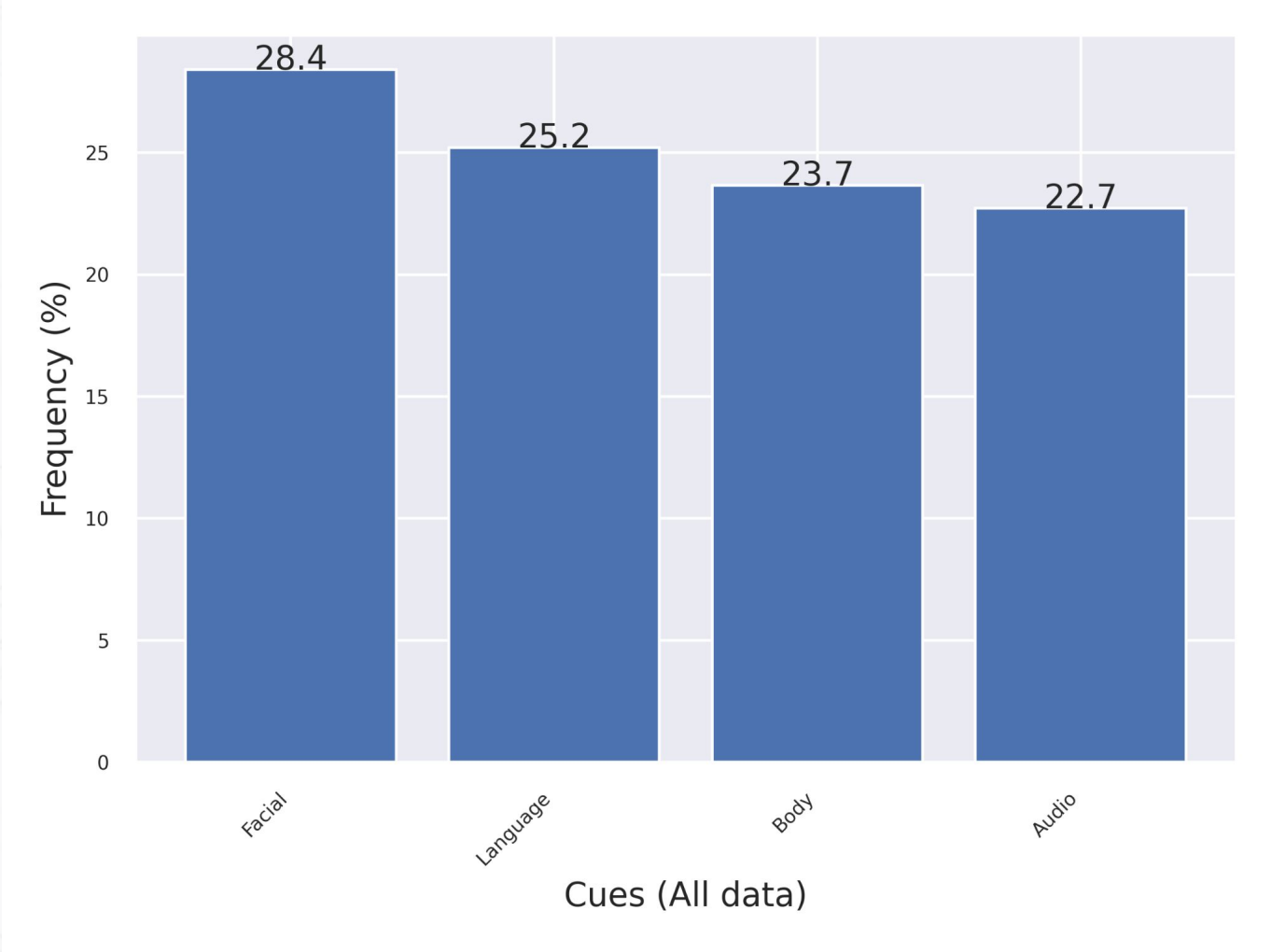
BAH Dataset: Statistics – Videos duration



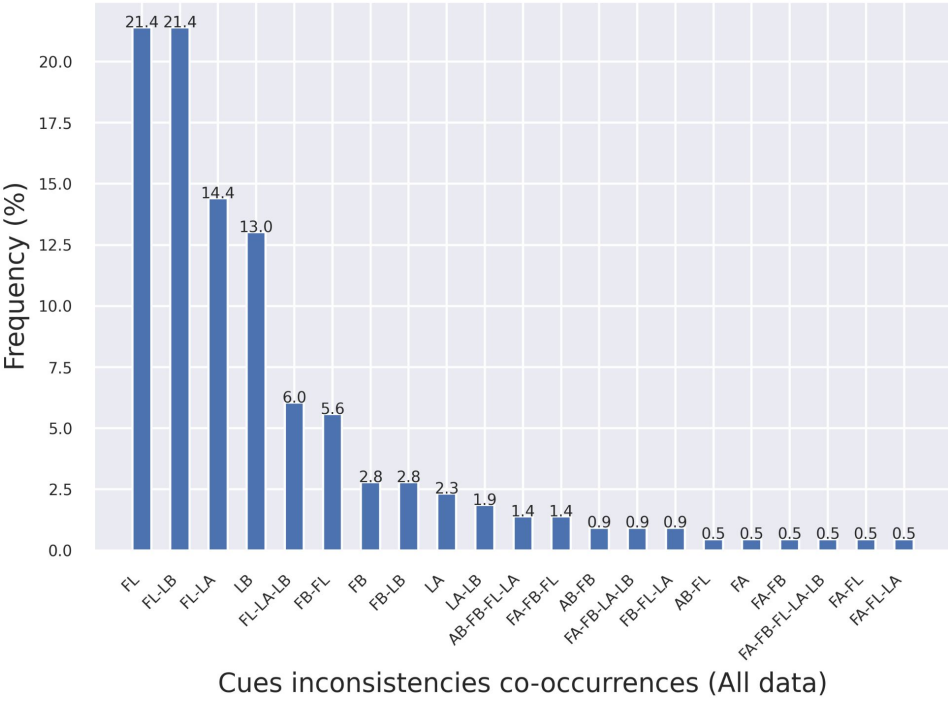
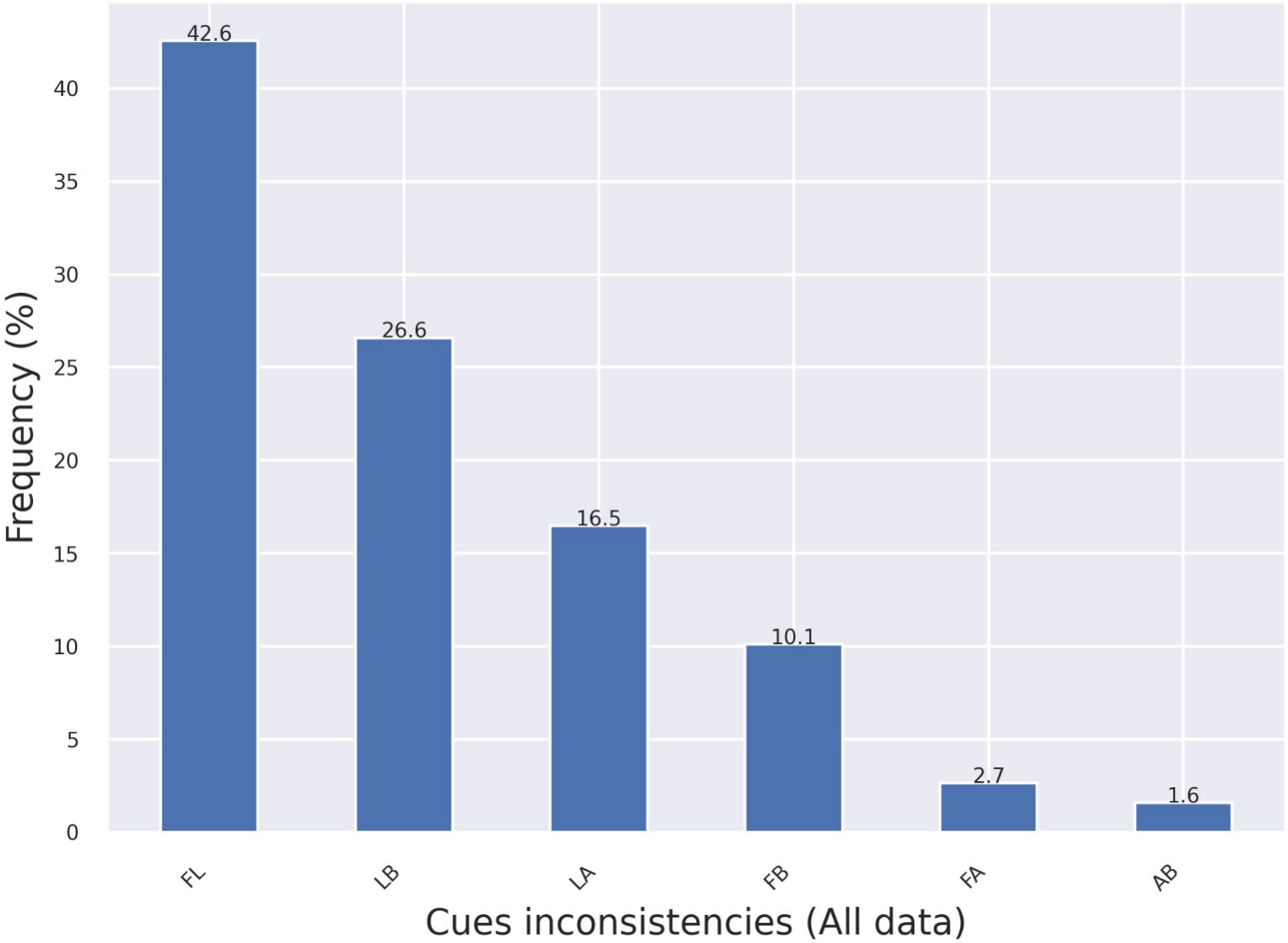
BAH Dataset: Statistics – A/H segments duration



BAH Dataset: Statistics – Annotation Cues



BAH Dataset: Statistics – Annotation Cues



BAH Dataset: License

- *BAH* is open-access
- Research, non-commercial and not-for-profit use



Behavioural Ambivalence and Hesitancy (BAH) Dataset

End User Licence Agreement – EULA

Please read carefully the following terms and conditions and any accompanying documentation before you download and/or use the Behavioural Ambivalence and Hesitancy (BAH) Dataset.

The BAH dataset is managed through the Montréal Behavioural Medicine Centre (MBMC) and the Imaging, Vision and Artificial Intelligence Laboratory (LIVIA).

If you have questions about the dataset, contact manuela.gonzalez@mail.concordia.ca or soufiane.belharbi@livia.etsmtl.ca

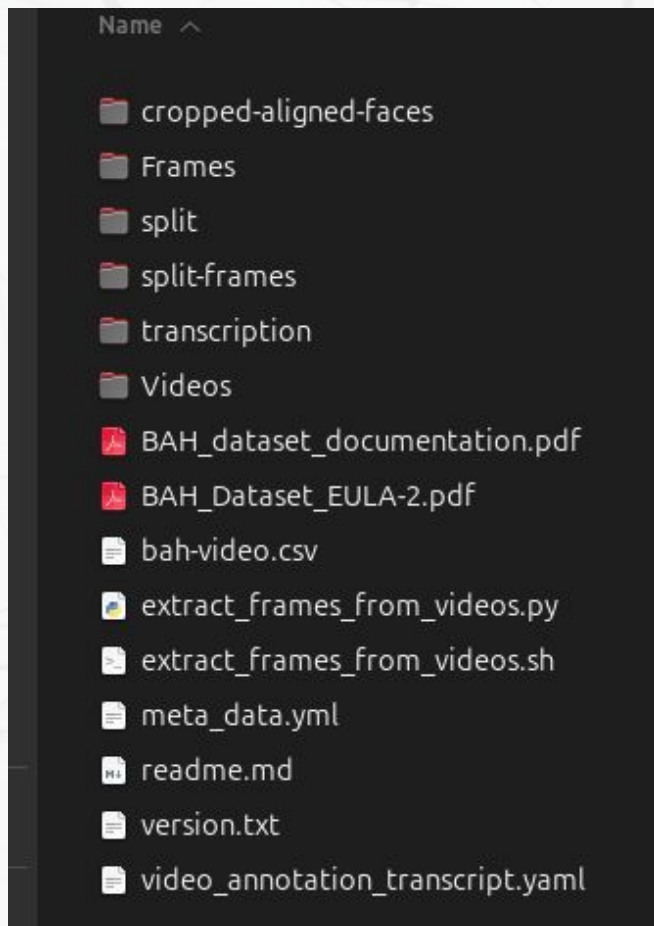
By signing this document the user, i.e., the person and their team who will make use of the dataset, agrees to the following terms.

The dataset includes both the actual data (raw/pre-processed) as well as the annotations.

Definitions:

Licensor: This is the owner of the BAH dataset (see below)

BAH Dataset: Access



BAH dataset: Download

To download BAH dataset, please fill in the following form which includes signing and uploading the End-User License Agreement (EULA). You will receive a link to download *BAH* dataset.

- PLEASE FILL IN THE DATASET REQUEST FORM CAREFULLY TO AVOID ERRORS/DELAYS.
- PLEASE FOLLOW THE NEXT INSTRUCTIONS.

- BAH DATASET REQUEST FORM: <https://www.crhscm.ca/redcap/surveys/?s=LDMDDJR3AT9P37JY>



Request *BAH* dataset at: <https://github.com/sbelharbi/bah-dataset>

BAH Dataset: Benchmarks / Challenges

arXiv paper: <https://arxiv.org/abs/2505.19328>

“BAH Dataset for Ambivalence/Hesitancy Recognition in Videos for Behavioural Change”, 2025



Experiments on A/H recognition:

- Supervised
- Zero-shot
- Domain adaptation / personalization

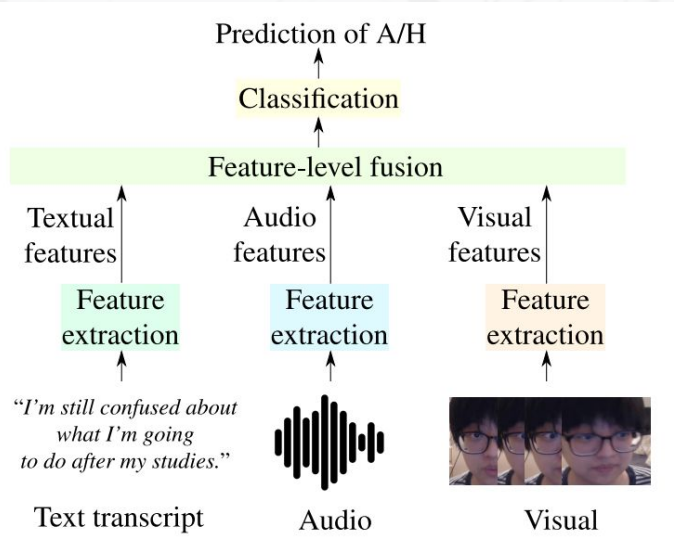
BAH Dataset: Benchmarks / Challenges



arXiv paper: <https://arxiv.org/abs/2505.19328>
“BAH Dataset for Ambivalence/Hesitancy Recognition in Videos for Behavioural Change”, 2025

Supervised A/H recognition:

- Importance of temporal/context modeling
- Importance of multi-modal learning
- Importance of fusion



Backbone	Without context		With context (TCN)	
	AVGF1	AP	AVGF1	AP
APViT (Xue et al., 2022)	0.5289	0.2221	0.5227	0.2217
ResNet18 (He et al., 2016)	0.4513	0.2064	0.5066	0.1842
ResNet34 (He et al., 2016)	0.4922	0.2107	0.4989	0.1683
ResNet50 (He et al., 2016)	0.4525	0.2124	0.5183	0.2068
ResNet101 (He et al., 2016)	0.5301	0.2280	0.5370	0.2191
ResNet152 (He et al., 2016)	0.5076	0.2096	0.5454	0.2674

Fusion type	AVGF1	AP
LFAN (Zhang et al., 2023) (cvprw,2023)	0.5557	0.2416
CAN (Zhang et al., 2023) (cvprw,2023)	0.5708	0.2559
MT (Waligora et al., 2024) (cvprw,2024)	0.4834	0.2108
JMT (Waligora et al., 2024) (cvprw,2024)	0.5350	0.2335

Modalities	AVGF1	AP
Visual	0.5454	0.2674
Audio	0.5372	0.2520
Text	0.5309	0.2047
Visual + Audio	0.5636	0.2818
Visual + Text	0.5756	0.2809
Audio + Text	0.5644	0.2459
Visual + Audio + Text	0.5557	0.2416

***BAH* Dataset: Benchmarks / Challenges**

Supervised A/H recognition:

- Still a very difficult task
- Code, models are available!

Code: <https://github.com/sbelharbi/bah-dataset>



BAH Dataset: Events

- ABAW8 - CVPR2025
- Future challenges (ABAW10?)

AH Recognition Challenge

If you want to participate *in this Challenge* you should follow the below procedure for registration.

i) fill in and sign [this EULA](#);

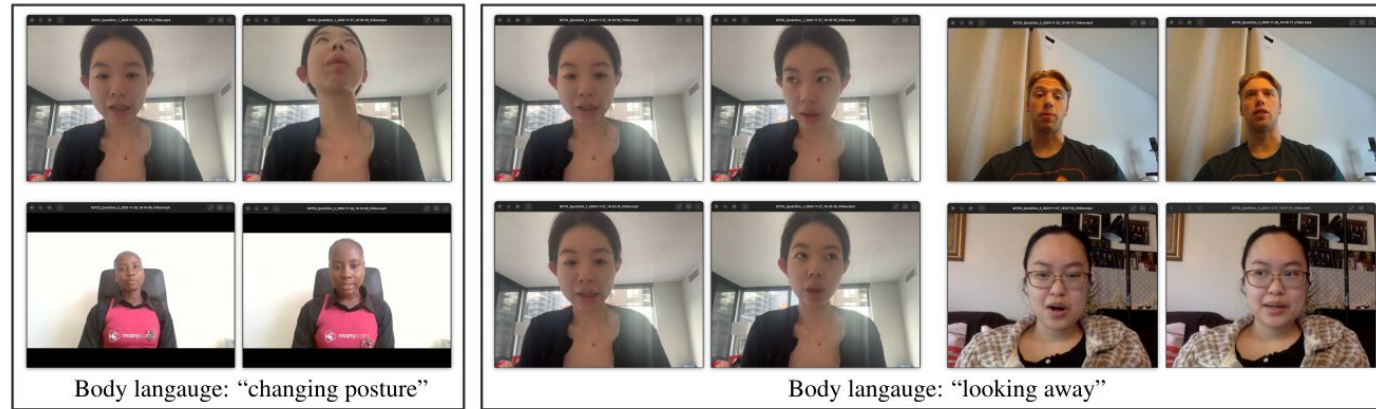
ii) fill in [this form](#), and submit it. Please use this form to upload your signed EULA;

iii) both documents must be signed by a person with a full-time faculty position at a university, higher education institution, or equivalent organization. The signee cannot be a student (UG/PG/Ph.D.).

Once both documents are received, we will contact you to provide data access including the raw BAH dataset, cropped-aligned faces at each frame, transcripts with timestamps, frame and video annotation, and documentation.

<https://affective-behavior-analysis-in-the-wild.github.io/8th/>

Questions?



Code: <https://github.com/sbelharbi/bah-dataset>

arXiv paper: <https://arxiv.org/abs/2505.19328>

“BAH Dataset for Ambivalence/Hesitancy Recognition in Videos for Behavioural Change”, 2025

For more discussions, please visit us at the **poster session**

