AVCAffe: A Large Scale Audio-Visual Dataset of Cognitive Load and Affect for Remote Work

Pritam Sarkar[1, 2] Aaron Posen[1] Ali Etemad[1]

[1] Queen's University, Canada [2] Vector Institute

https://pritamsarkar.com/AVCAffe

Introduction

Motivation

Due to the recent COVID-19 pandemic, *remote work* is the new reality of work for millions across the world. While this new paradigm has a number of advantages such as enabling social distancing and flexible hours, it brings about a number of challenges that were less common in in-person work environments. For instance, studies have shown that *remote work settings could contribute to increased cognitive load and fatigues in individuals* due to the reasons including but not limited to:

- back-to-back work-related meetings with minimal physical mobility in-between,
- the inability to effectively perceive and transmit non-verbal expressive cues,
- the need to apply intense focus on the screen with minimal variation.

In order to better understand mental health and manage the impact of remote work meetings on individuals, it is necessary to design and develop tools capable of quantifying factors such as cognitive load and affect in relevant settings.

Problem Statement

The lack of a large-scale audio-visual dataset consists of cognitive load and affective ground truths to study human behavior.

Our Contribution

We introduce AVCAffe, the first Audio-Visual dataset consisting of Cognitive load in addition to Affect attributes. Currently, AVCAffe is the largest collected affective dataset (in the English language) consisting of 108 hrs. of audio-visual recordings of 106 participants from 18 different countries.

AVCAffe

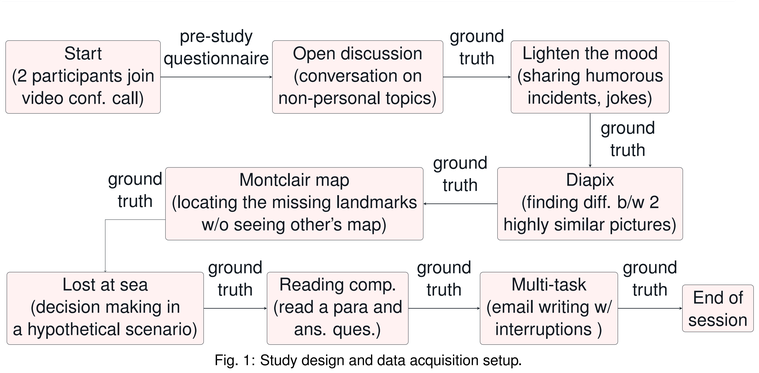

Fig. 1: Study design and data acquisition setup.

# Subjects. 106	**Duration.** 108 hrs.	**# Clips.** 58,112
Video. Resolution: 456×256; Rate: 25fps;		**Audio.** Freq.: 44.1KHz.
Gender. Male: 52, Female: 53, Non-Binary: 1.	**Age:** 18 to 20 : 8; 21 to 30 : 75; 31 to 40 : 17; 41 to 50 : 2; 51 to 60 : 4.	
Countries of origin: Bangladesh(1), Brazil(2), Canada(67), China(3), Ecuador(1), Egypt(1), Germany(1), Hong Kong(1), India(11), Iran(4), Ireland(1), Jordan(1), Mexico(4), Nigeria(2), Pakistan(2), Sweden(1), USA(2), Vietnam(1)		
Ground truths. Arousal, Valence, Mental Demand, Temporal Demand, Effort, Physical Demand, Performance, and Frustration		

Tab. 1: Dataset statistics.

	Participant A			Participant B		
Arousal	Wide-awake	Excited	Wide-awake	Wide-awake	Excited	Wide-awake
Valence	Pleasant	Unsatisfied	Pleased	Pleasant	Pleased	Pleasant
Effort	8	17	16	2	16	12
Mental demand	3	12	13	4	12	15
Temporal demand	0	20	11	3	16	6

Fig. 2: Sample clips along with self-reported affect and cognitive load scores during different tasks.

AVCAffe (Contd.)

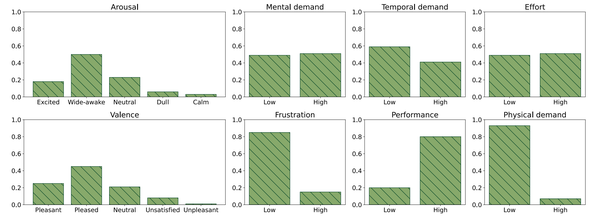

Fig. 3: Class distribution of AVCAffe.

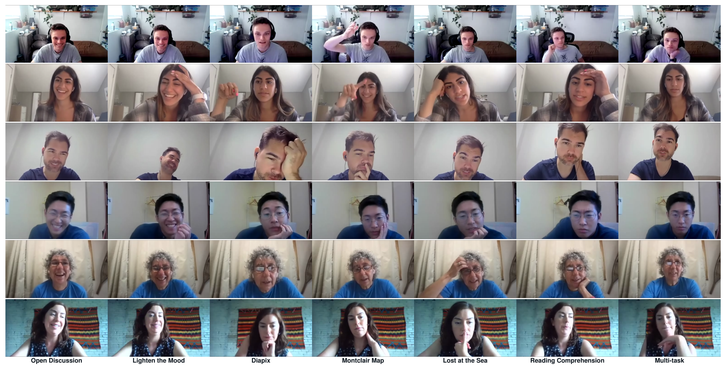

Fig. 4: Sample representative frames of AVCAffe during different tasks.

Analysis

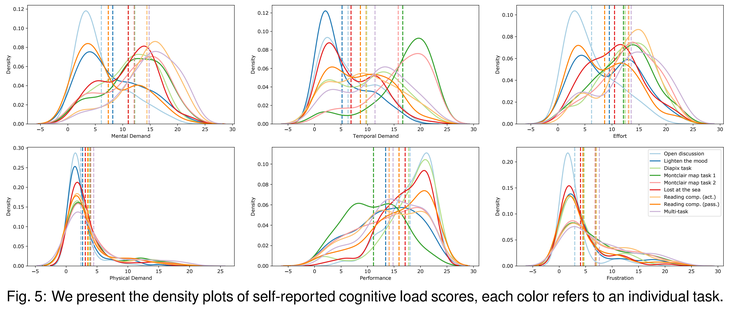

Fig. 5: We present the density plots of self-reported cognitive load scores, each color refers to an individual task. Left to right in increasing order of cognitive load.

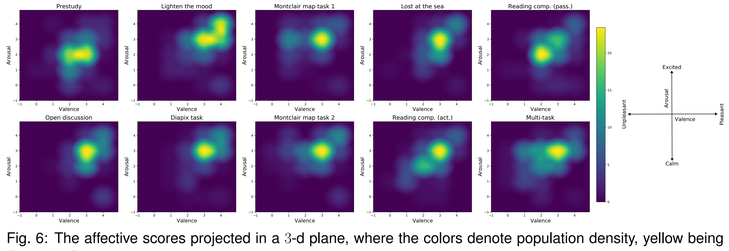

Fig. 6: The affective scores projected in a 3-d plane, where the colors denote population density, yellow being the most dense.

Analysis (Contd.)

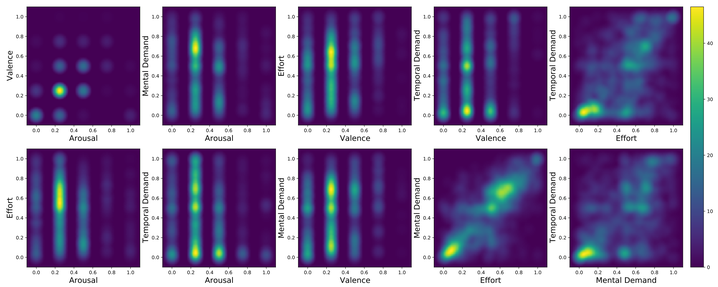

Fig. 7: We project the self-reported arousal and valence scores on a 3-d space. The x and *y* axes represent valence and arousal respectively, and color denotes population density with yellow being the most dense.

Baselines

Audio	Visual	#Params	Mental D. Short	Mental D. Long	Effort Short	Effort Long	Temporal D. Short	Temporal D. Long	Arousal Short	Arousal Long	Valence Short	Valence Long
Random Classifier		-	51.3	45.6	48.0	43.6	35.1	33.9	32.9	26.2	34.0	30.3
VGG-16	✗	14.7M	58.8	61.2	58.8	62.1	57.9	56.7	38.3	36.1	40.3	39.1
ResNet-18	✗	11.2M	58.2	60.7	57.0	60.8	58.2	54.4	38.1	30.4	39.3	36.3
✗	MC3-18	11.7M	60.4	61.0	61.4	63.8	60.0	59.4	41.4	34.0	42.0	38.8
✗	ResNet3D-18	33.4M	59.3	59.0	61.0	62.7	58.5	57.9	37.8	30.9	41.9	39.5
✗	R(2+1)D-18	31.5M	60.5	59.6	**65.5**	67.7	59.6	54.9	39.7	33.3	38.7	34.9
VGG-16	MC3-18	47.4M	59.4	60.2	59.7	66.2	60.8	61.4	41.3	38.9	40.3	**41.7**
VGG-16	ResNet3D-18	69.1M	**65.0**	**65.6**	59.7	60.5	59.2	60.3	40.7	37.3	**43.9**	39.4
VGG-16	R(2+1)D-18	67.2M	60.1	64.7	59.7	**69.4**	60.4	**66.7**	42.1	**40.5**	41.1	39.5
ResNet-18	MC3-18	43.9M	61.3	60.2	59.4	62.1	58.8	57.7	42.4	36.0	41.4	39.2
ResNet-18	ResNet3D-18	65.6M	58.8	61.2	60.7	64.4	**61.2**	61.7	42.6	35.1	39.8	39.1
ResNet-18	R(2+1D)-18	63.7M	60.4	62.7	60.8	61.1	58.6	59.0	**44.0**	39.5	40.9	37.7

Tab. 2: Baselines on AVCAffe are presented. The best F1-scores in each subcategory (audio-only/visual-only/audio-visual) are underlined and best scores of each label are highlighted in bold. Here, Random Classifier refers to a randomly initialized classifier with no training which serves as a reference point to understand the performance of different models.

Summary

- We present a novel audio-visual database of cognitive load and affect collected in a setup resembling 'remote work meetings'. To the best of our knowledge, AVCAffe is the first audio-visual dataset comprised of cognitive load annotations. Moreover, AVCAffe is the largest affective dataset in the English language.
- AVCAffe enables the researchers to study the impact of remote work meetings on cognitive load and affect, and broadly mental health and human behavior.
- We believe AVCAffe would be a challenging benchmark for the research community given the inherent difficulty of classifying affect and cognitive load in particular.
- In addition to understanding cognitive load and affect, AVCAffe can be further used to conduct research in several other areas like audio-visual speech recognition, lip reading from visual input, and long video summarization, among others.
- The authors do not foresee any major negative societal impact, however, all ethical considerations that apply to other audio-visual affective datasets may equally apply here.

Acknowledgement

We are grateful to the Bank of Montreal and Mitacs for funding this research. We are also thankful to SciNet HPC Consortium for helping with the computation resources.