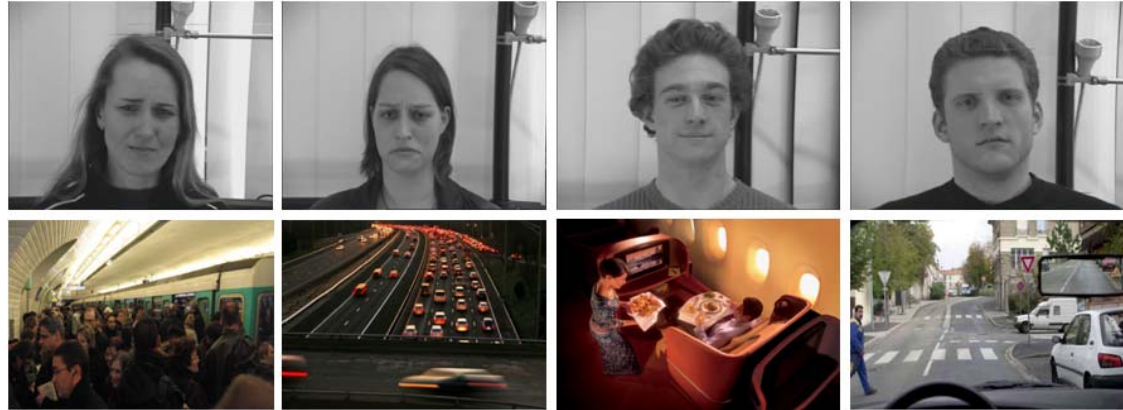


Dynamic facial expression recognition using a behavioural model



Thomas Robin

Michel Bierlaire

Javier Cruz

STRC 2009

10th september

The context

- Recent interest for emotion recognition in transportation

Driving assistance

- Safety
- Mobility

Well-being measuring of users

- Improve public transportation offers
- Improve car comfort

Abou-Zeid, M., Ben-Akiva, M. and Bierlaire, M. (2008). Happiness and travel behavior modification, *Proceedings of the European Transport Conference*, Leiden, The Netherlands.

The context

- Emotion: **mental** and **physiological** state associated with a wide variety of feelings, thoughts and behavior.
- Emotions signs easy to measure with non-intrusive techniques for transportation users:
 - Behavior
 - Face expression
 - Voice intonation

The context

Driving assistance



➔ Adapt car behavior
to a danger

Well being measuring



➔ Improve level
of service

Objectives

- Model the facial expression recognition made by a person looking at a face video sequence
- Model explicitly the **dynamic process**
- **No classification**
- Estimate the model on **behavioural** data (relax ground truth assumptions)

Outline

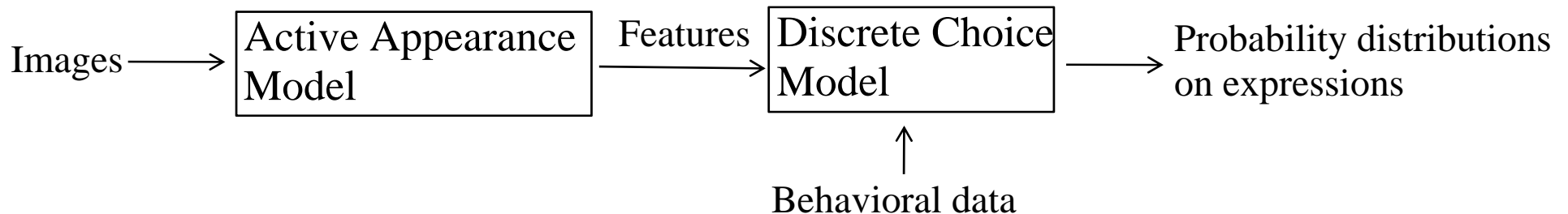
- **Introduction**
- **Data: video**
- **Features extraction**
- **Data: behavioral data**
- **Model framework**
- **Model estimation**
- **Model predictions**
- **Conclusion and Perspectives**

Introduction

- Static version of the work:

M.Sorci, M.Bierlaire, J-P.Thiran, J.Cruz, Th.Robin and G.Antonini (2008) Modeling human perception of static facial expressions, paper presented at *8th IEEE Int'l Conference on Automatic Face and Gesture Recognition*.

- ➔ - Images: Cohn-Kanade database
- Behavioral data: internet survey



- **Static** ➔ **Dynamic**

Introduction

- Dynamic framework inspired from dynamic model:

➔ Hidden Markov Model

- State transition process
- Measurement equation

➔ Choudhury, C. F. (2007) Model driving decisions with latent plans, Ph.D. Thesis, Massachusetts institute of technology.

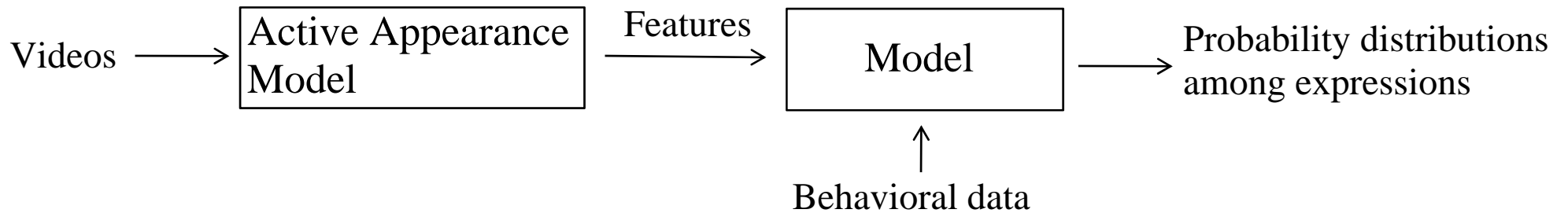
- Behavioral models
- Latent decisions

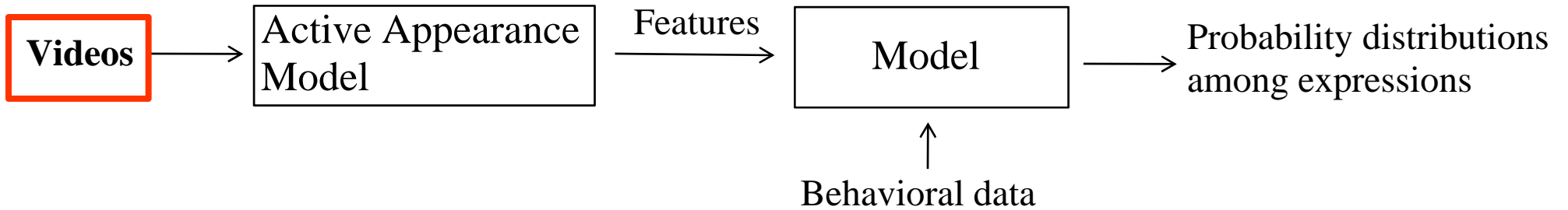
➔ J.L.Walker (2001) Extended discrete choice models: Integrated framework, flexible error structures, and latent variables, Ph.D. Thesis, Massachusetts Institute of Technology.

- Latent segmentation models

Introduction

- Model overview:

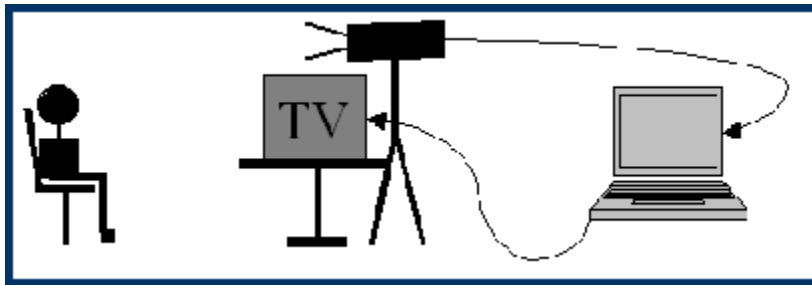




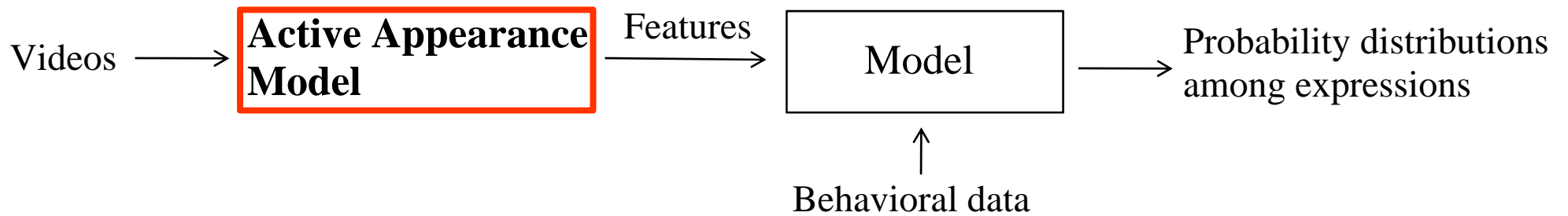
Data: video database

- The Technical University Munich database (TUM) Facial Expression and Emotion Database (FEED)

➔ Students faced to a video, natural expressions recorded



138 sequences, 18 subjects



Features extraction: Active Appearance Model

- Video = succession of images, called frames

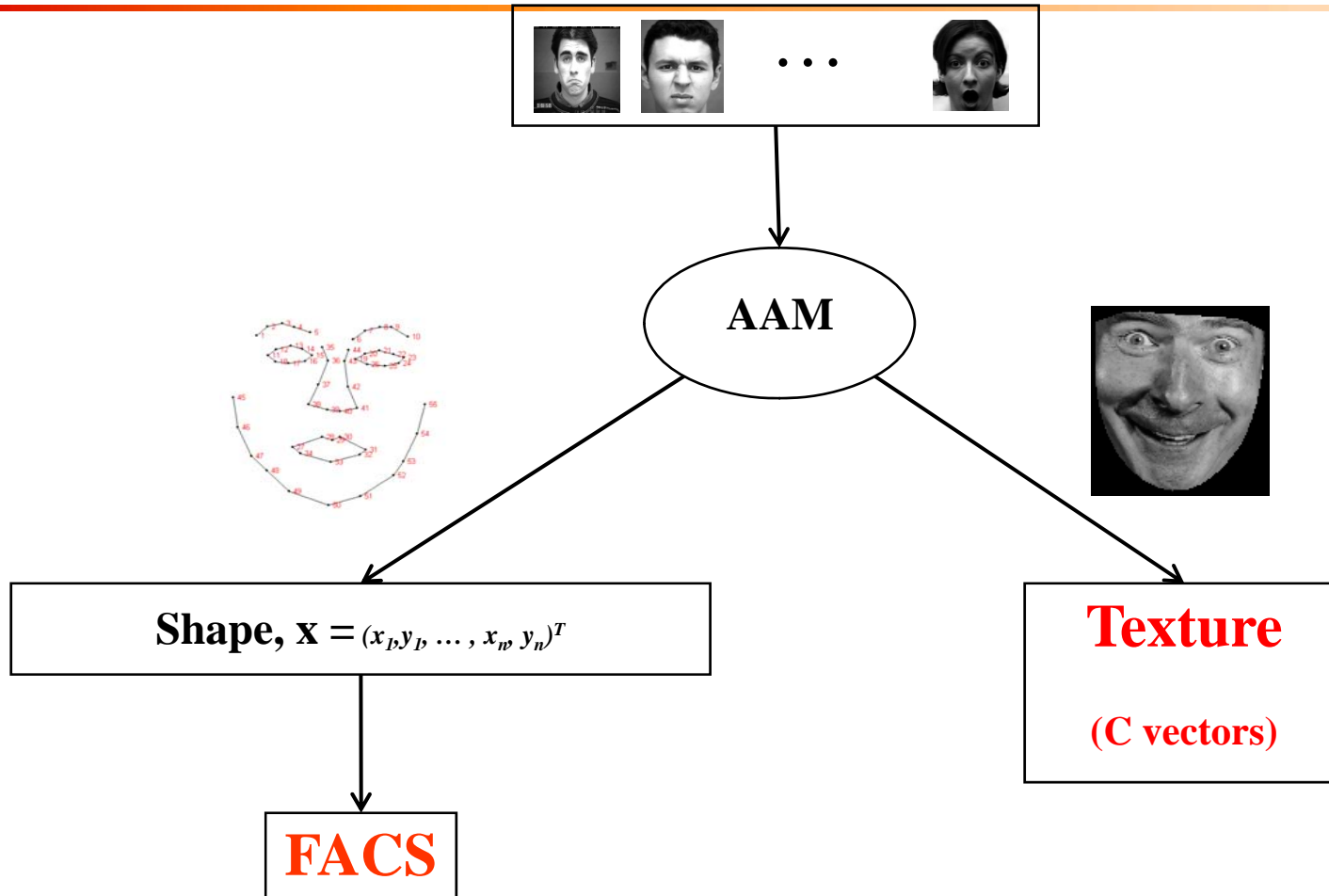
➡ information extracted on each frame

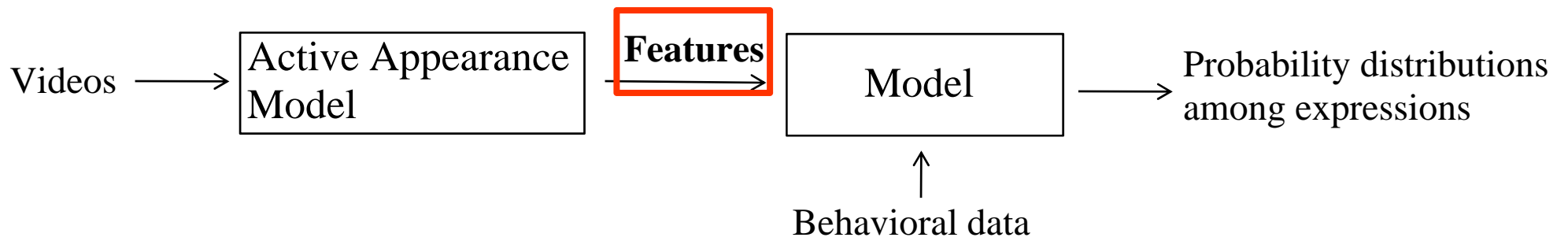
- Hypothesis: individual perception evolves at regular time step (1 s)

➡ |
- a video contains 25 frames per second
- first frame of each second retained



Features extraction: Active Appearance Model





















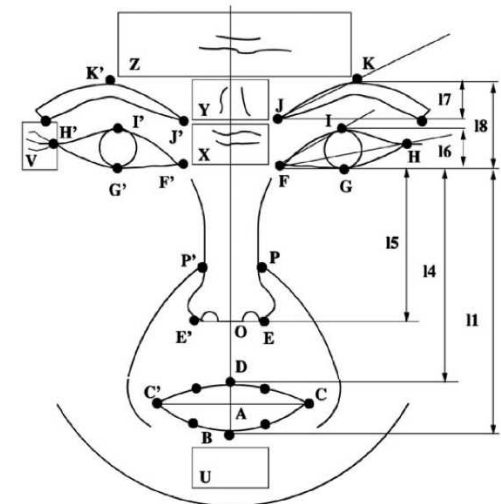


Features: Facial Action Coding System

- FACS developed by Ekman and Friesen (1978)
- Measurement units, called “Action Units” (AUs) associated to expressions

➔ leading standard for measuring facial expressions

AU1  Inner Brow Raiser	AU2  Outer Brow Raiser	AU4  Brow Lowerer	AU5  Upper Lid Raiser	AU6  Cheek Raiser	AU7  Lid Tightener
AU9  Nose Wrinkler	AU10  Upper Lip Raiser	AU12  Lip Corner Puller	AU15  Lip Corner Depressor	AU16  Lower Lip Depressor	AU17  Chin Raiser
AU20  Lip Stretcher	AU23  Lip Tightener	AU24  Lip Pressor	AU25  Lips part	AU26  Jaw Drop	AU27  Mouth Stretch



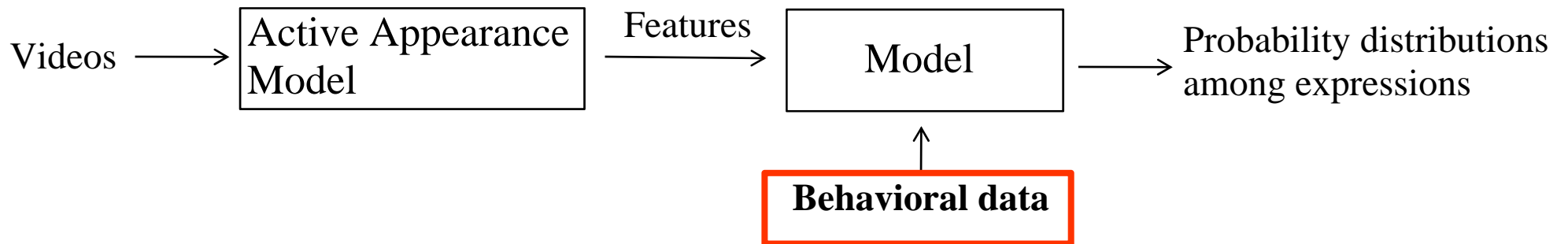
Features: C vectors

- Direct output of the Principal Component Analysis (PCA) conducted in the AAM

➔ It characterises both **face shape** and **face shadows**



→ C vector: 100 elements



Data: internet survey

- Survey conducted at the address below(English, French, Italian, Spanish):
<http://transp-or2.epfl.ch/videosurvey/>
- Respondents have to:
 - create an account
 - ➔ Socioeconomics attributes
 - label some video sequences with expressions
 - ➔ **observations**
- 1 database of video is used:
 - Facial Expression and Emotion Database (FEED)

Data: socio-economics

http://transp-or2.epfl.ch/videosurvey/index.php?include=createuser

Google t

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

LABORATOIRE TRANSPORT ET MOBILITE
FACIAL EXPRESSIONS EVALUATION SURVEY

INTER > TRANSP-OR > Facial Expressions Evaluation Survey

Status and links

Home

You are not connected

English

**Suggestions
(email us)**

E-mail address:
[text input]

Password:
[text input]

Login

Forgotten password

Why this socio-economics form?

The socio-economics fields are important in order to segment the labeler population based on those characteristics. The ethnic group is relevant to investigate the choice behavior of people when faced to videos of individuals belonging to the same or to another ethnic group.

Why a username?

An account is required so that you won't have to fill the form anytime you want to participate to the survey. If you find the survey too long you can stop whenever you want by logging off and restart from the first unlabeled image at your next login.

IMPORTANT: The e-mail address is only used to send you a new password, if you have forgotten it.

Create a new user

Birth Year: [0000]

Gender: Male Female

Language: [English]

Studies: [High School]

Ethnic group: [None]

Current location: [None]

Occupational category: [None]

E-mail address: [text input]

Password: [text input]

Password Confirmation: [text input]

OK

Data: labels

The screenshot shows a web browser window with the URL `http://transp-or2.epfl.ch/videosurvey/index.php?include=navigator`. The browser's address bar and search bar are visible. The page header includes the EPFL logo and the text "LABORATOIRE TRANSPORT ET MOBILITE" and "EVALUATION D'EXPRESSIONS FACIALES". Below the header, there is a navigation menu with "INTER > TRANSP-OR > Evaluation d'Expressions Faciales".

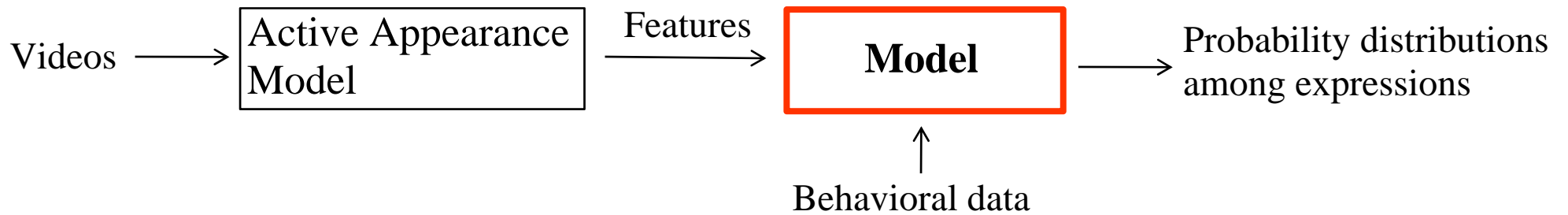
On the left side, there is a sidebar with the following content:

- Statut et liens**
- Accueil**
- Connecté en tant que : thomas.robin@epfl.ch
- [Se déconnecter](#)
- Suggestions (pour nous écrire)**

The main content area features a video player showing a woman's face. Below the video player, there is a questionnaire with the following options:

- Joie
- Surprise
- Peur
- Degout
- Tristesse
- Colere
- Neutre
- Autre
- Je ne sais pas

At the bottom of the questionnaire, there are navigation arrows and a button labeled "Valider le questionnaire".



Model: introduction

- Combination of 2 DCMs

Instantaneous expression perception sub-model

- Similar to static model
- Choice set: expressions
- Attributes: facial features

Video frames weighting sub-model

- Capture influence of each frame on final expression perception
- Choice set: frames (depend on the video)
- Attributes: dynamic features, such as facial features derivatives

Model: general framework

$$P_n(i/o) = \sum_{t=1}^{T_o} P_n(i/t, o) P_n(t/o)$$

Diagram illustrating the general framework for the model. The equation is annotated with red arrows indicating the components:

- expression** (arrow pointing down to $P_n(i/o)$)
- Duration of video o** (arrow pointing down to T_o)
- individual** (arrow pointing up to i)
- frame** (arrow pointing up to t)
- video** (arrow pointing up to o)

- $P_n(i/t, o)$: Instantaneous expression perception sub-model.
- $P_n(t/o)$: Video frames weighing sub-model.
- $P_n(t/o)$: Model.

Model: expression perception sub-model

$$P_n(i/o) = \sum_{t=1}^{T_o} P_n(i/t, o) P_n(t/o)$$

Diagram illustrating the expression perception sub-model equation:

- $P_n(i/o)$ is labeled "individual" (input from below) and "expression" (input from above).
- T_o is labeled "Duration of video o" (input from above).
- $P_n(i/t, o)$ is labeled "frame" (input from below) and is circled in red.
- $P_n(t/o)$ is labeled "video" (input from below).

- Choice set: 9 expressions (Happiness, Surprise, Fear, Disgust, Sadness, Anger, Neutral, Other, Don't know)
- Logit model
- Memory effect : captured in expression utilities

Model: expression perception sub-model

- Memory effect : captured in expression utilities

$$V_n(i/t, o) = V_{generic_n}(i/t, o) + a_{i,n} V_{generic_n}(i/t - 1, o)$$

↑
Estimated parameter

- Utility specification:
 - Alternative Specific Constants (**ASC**)
 - Measures corresponding to **AUs** (**FACS**)
 - Elements of **C vectors** (outputs of AAM)

Model: frames weighting sub-model

$$P_n(i/o) = \sum_{t=1}^{T_o} P_n(i/t, o) P_n(t/o)$$

Diagram illustrating the frames weighting sub-model equation:

- $P_n(i/o)$ is labeled "individual" (arrow pointing up).
- $P_n(i/t, o)$ is labeled "frame" (arrow pointing up).
- $P_n(t/o)$ is labeled "video" (arrow pointing up).
- T_o is labeled "Duration of video o" (arrow pointing down).
- $P_n(i/o)$ is labeled "expression" (arrow pointing down).

The term $P_n(t/o)$ is circled in red.

- Choice set: Frames of the videos (it depends on o)
- Logit model
- Utility specification: Derivatives of facial features

Model: frames weighting sub-model

High changes



**Expected chosen frame
to label the video**

Model: likelihood function

Number of observations relative to respondent n

Number of respondents Number of expressions Number of expressions

N O_n 9

$$l(\beta) = \prod_{n=1}^N \prod_{o=1}^{O_n} \prod_{i=1}^9 P_n(i/t, o, \beta)^{c_{i,o,n}}$$

Vector of parameters

Model estimation: general results

- Likelihood maximization
- Behavioral data: labels on the FEED videos (natural videos)
- Simultaneous sub-models estimation
- Estimation program based on the BIOGEME software
- General model fit:

Nb of observations:	294
Nb of parameters:	44
Null log-likelihood:	-645.98
Final log-likelihood:	-358.82
$\bar{\rho}^2$:	0.38

Model estimation: parameters values

Expression perception sub-model

Id	Parameter name	Value	Std-error	t-test 0
1	ASC_A	-7.83	5.31	-1.47
2	ASC_D	6.90	4.10	1.68
3	ASC_DK	-0.54	0.39	-1.40
4	ASC_F	-31.90	8.89	-3.59
5	ASC_H	24.23	5.31	4.56
6	ASC_O	5.31	3.14	1.69
7	ASC_SA	8.70	6.85	1.27
8	ASC_SU	-13.33	2.97	-4.48
9	b_broweye_l2_SA	570.49	134.49	4.24
10	b_broweye_l3_SU	70.73	19.54	3.62
11	b_broweye_r2_A_D_F_SA_SU	-99.24	26.90	-3.69
12	b_eye_angle_below_l_F	6.24	2.46	2.54
13	b_eye_angle_l_F_SA	17.33	5.01	3.46
14	b_eye_angle_r_F_SA	-10.17	3.08	-3.30
15	b_eye_brow_angle_l_SA	-16.58	3.95	-4.20
16	b_eye_mouth_dist_l2_D	-49.54	24.91	-1.99
17	b_eye_mouth_dist_l_H_O_SA	-97.20	36.70	-2.65
18	b_eye_nose_dist_l_A	248.02	36.42	6.81
19	b_eye_nose_dist_l_D_F_O_SA	101.16	22.25	4.55
20	b_eye_nose_dist_r_D_F_O_SA_A	-131.09	19.88	-6.59
21	b_leye_h_F	660.84	145.33	4.55
22	b_leye_h_SU	340.57	62.41	5.46
23	b_mouth_h_A_D_H_SA_F_SU	79.71	25.32	3.15
24	b_mouth_nose_dist2_A_SA	-283.30	56.02	-5.06
25	b_mouth_nose_dist_H	-324.71	52.45	-6.19
26	b_mouth_w_A_D_F_H_O	36.40	15.42	2.36
27	b_C_1_SU	90.35	20.63	4.38
28	b_C_1_F	153.47	28.59	5.37
29	b_C_1_D	115.28	19.57	5.89
30	b_C_1_A	170.99	27.17	6.29
31	b_C_2_H	23.23	10.47	2.22
32	b_C_2_SU	33.94	13.28	2.56
33	A_H	-0.70	0.13	-5.25
34	A_D	-0.15	0.10	-1.49
35	A_SA	-0.49	0.11	-4.28
36	A_A	-0.15	0.09	-1.58

ASCs

AUs

C vector

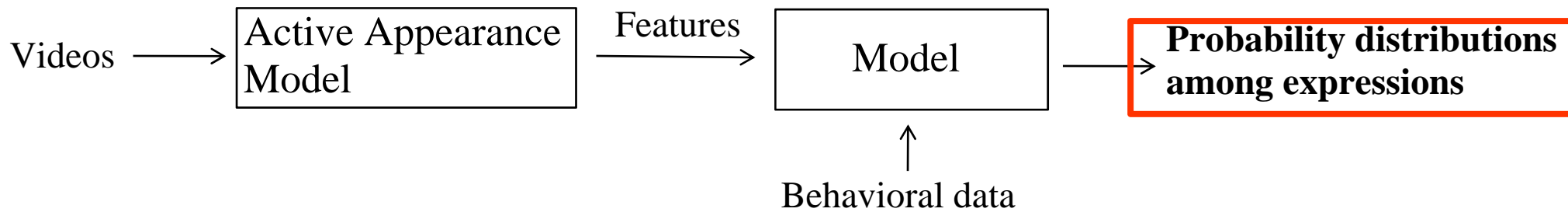
Memory effect

Frames weighting sub-model

Id	Parameter name	Value	Std-error	t-test 0
1	b_FRAME_C_1_deriv	-45.46	25.41	-1.79
2	b_FRAME_C_2_deriv	-224.99	72.18	-3.12
3	b_FRAME_C_3_deriv	240.01	79.08	3.04
4	b_FRAME_C_5_deriv	-73.34	27.28	-2.69
5	b_FRAME_eye_h_deriv	-805.69	226.21	-3.56
6	b_FRAME_eye_brow_angle_deriv	43.97	14.33	3.07
7	b_FRAME_mouth_h_deriv	1309.91	399.85	3.28
8	b_FRAME_mouth_w_deriv	-184.44	56.81	-3.25

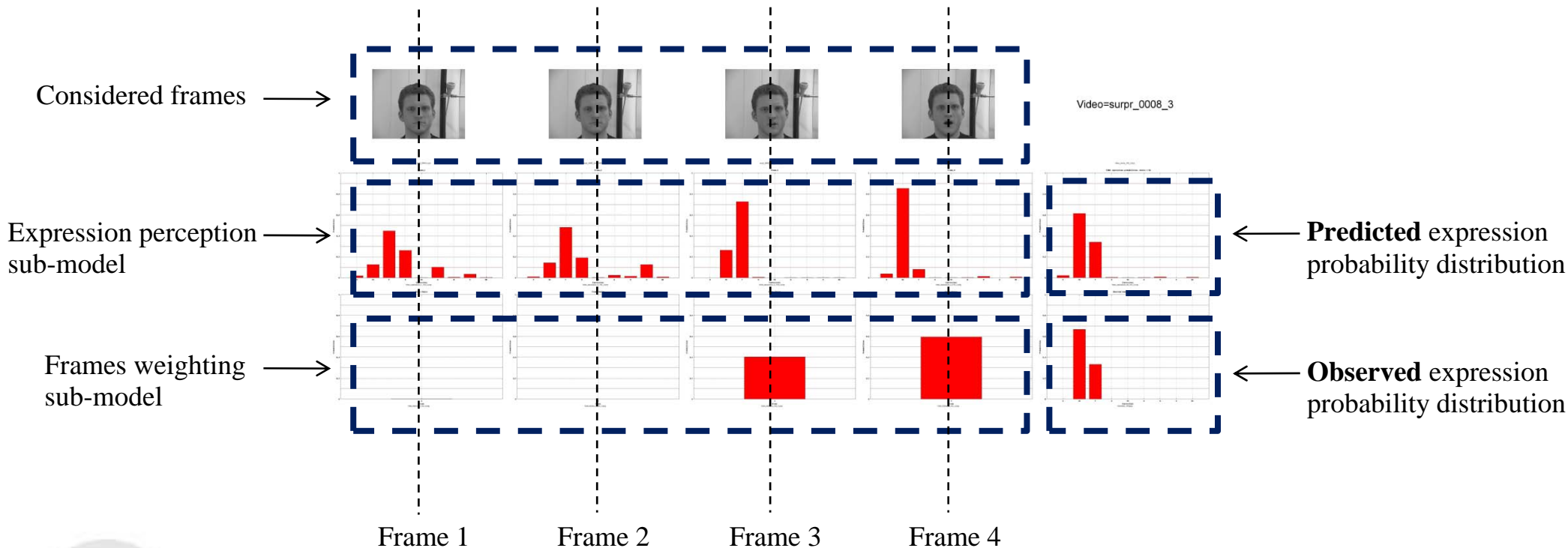
C vector deriv

AUs deriv



Model predictions: introduction

- Check model validity
- Prediction display example:

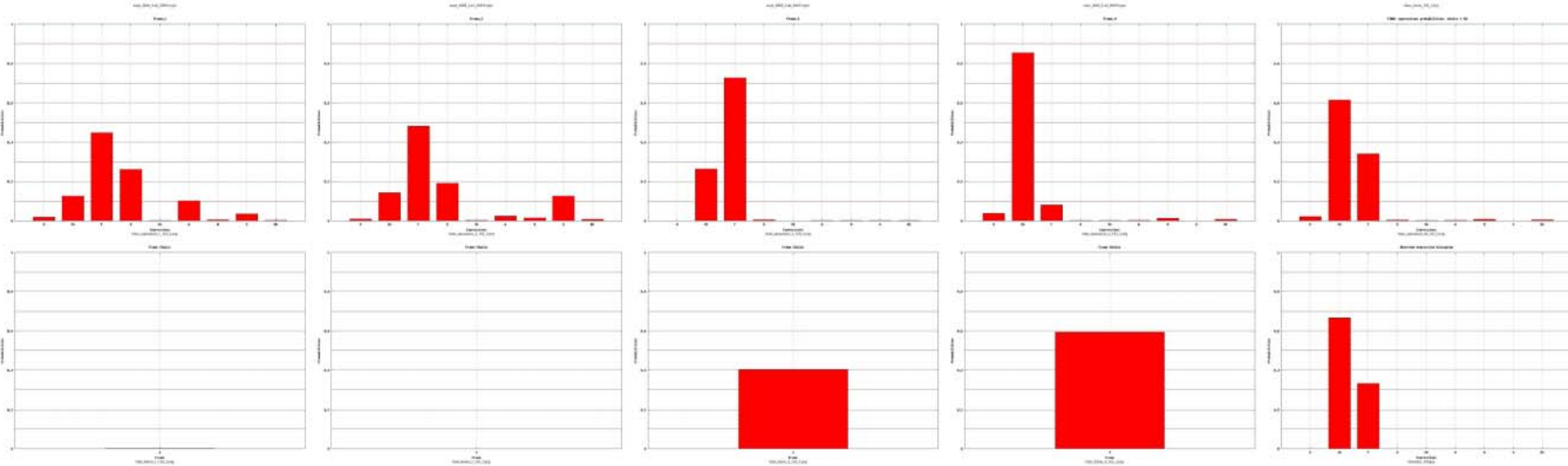


Model predictions: example 1

- Expressions order: **H, SU, F, D, SA, A, N, O, DK**



Video=surpr_0008_3

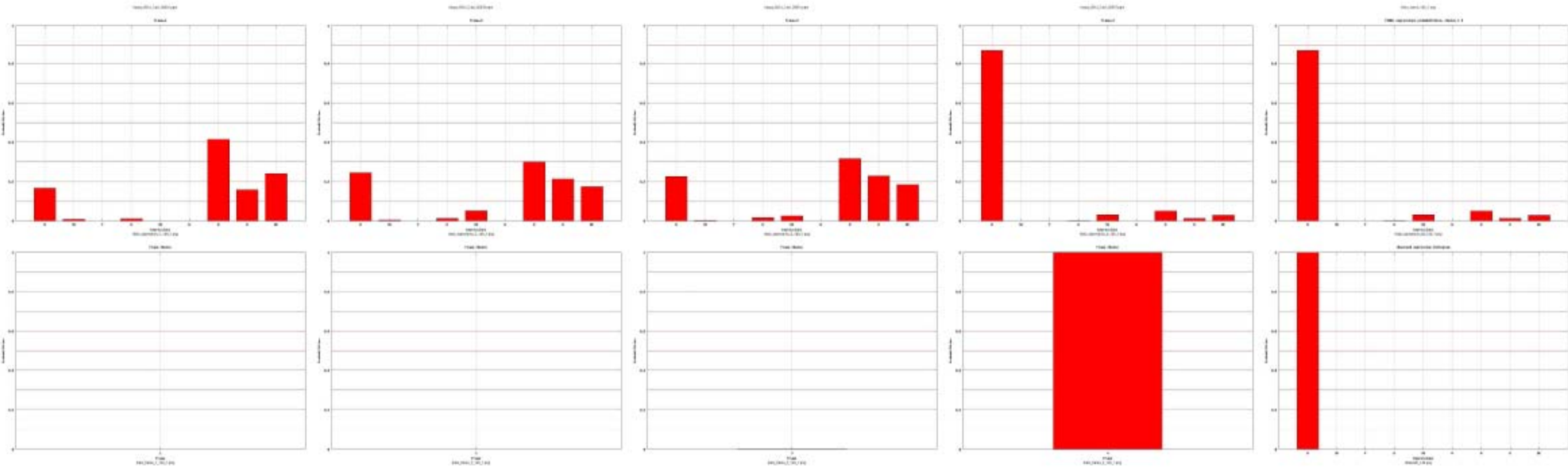


Model predictions: example 2

- Expressions order: **H, SU, F, D, SA, A, N, O, DK**



Video=happy_0014_2

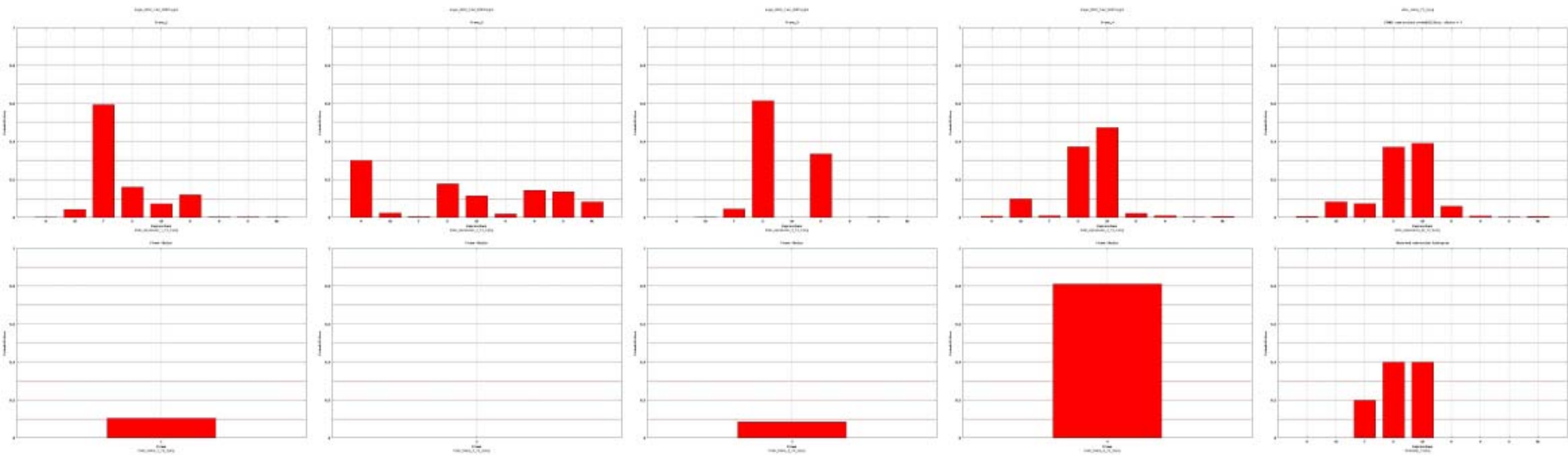


Model predictions: example 3

- Expressions order: **H, SU, F, D, SA, A, N, O, DK**



Video=anger_0002_2



Model pre-validation

- Comparison with simpler model: ASC model

➔ Only ASCs in expression perception sub-model

Proposed model		ASC model	
Nb of observations:	294	Nb of observations:	294
Nb of parameters:	44	Nb of parameters:	8
Null log-likelihood:	-645.98	Null log-likelihood:	-645.98
Final log-likelihood:	-358.82	Final log-likelihood:	-572.437
$\bar{\rho}^2$:	0.38	$\bar{\rho}^2$:	0.10

- Aggregated prediction results on estimation data : Outliers percentage

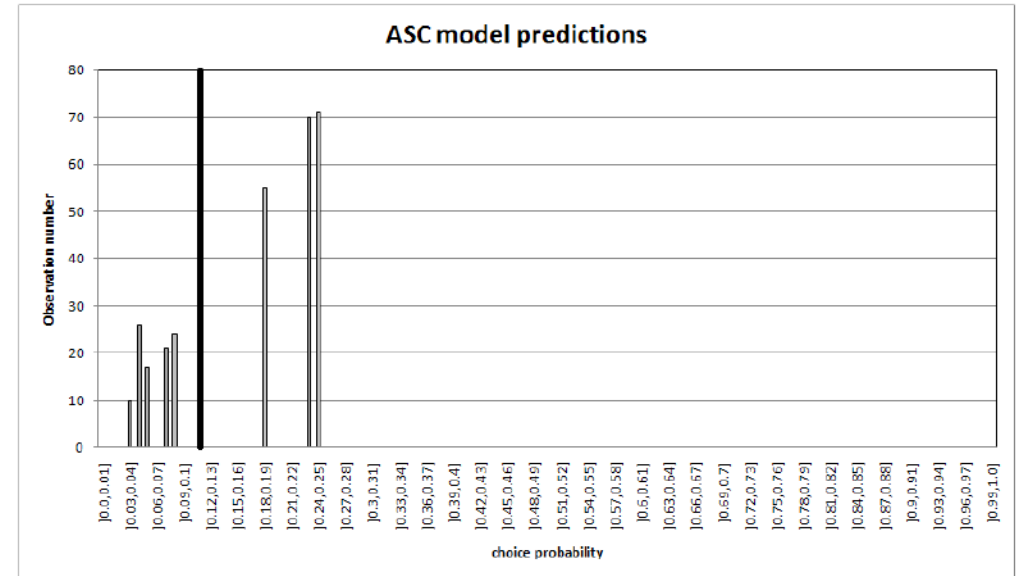
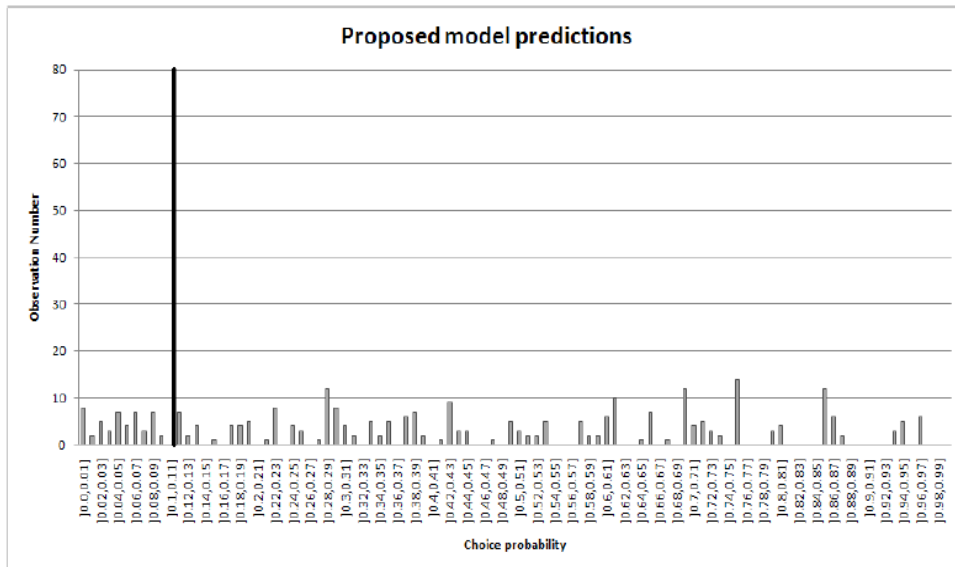
➔ Outlier: observation with choice probability less than $\frac{1}{9}$

Model pre-validation: outliers

- Outlier: observation with choice probability less than $\frac{1}{9}$

Proposed model	ASC model
16.33%	33.33%

- Choice probabilities histogram



Conclusions and Perspectives

- Conclusion:

- database of face video annotations
- new model framework
- model estimated using behavioral data
- pre-validated model

- Perspectives:

- implement the panel data effect
- estimate the model on more data (both videos and labels)
- use of another video database for validation

Thanks for your attention

<http://transp-or2.epfl.ch/videosurvey/>