Behavioural modeling of dynamic facial expression recognition

Thomas Robin, Michel Bierlaire, Javier Cruz

28th august 2008

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

The context

Face expression decision

human **Face's video sequence**

Applications:

- **Driver's attention state;**
- **Smart meeting rooms;**
- **Human-Machine interfaces.**

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Objectives

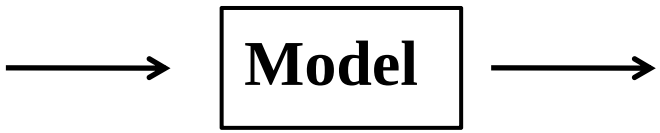
- Model the facial expression recognition made by a person looking at a face video sequence
- Model explicitely the **dynamic process**
- Estimate the model on **behavioural** data (not classification)

Outline

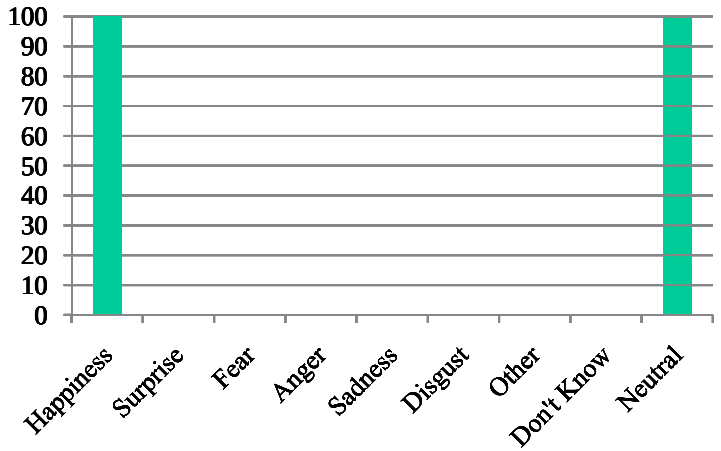
- **Introduction**
- **Features extraction**
- **Data:**
  - Video data bases
  - Internet survey
- **Model:**
  - State transition process
  - Measurement equation
  - Likelihood function
- **Conclusion and Perspectives**

Introduction

Input:



Output:



TRANSP-OR

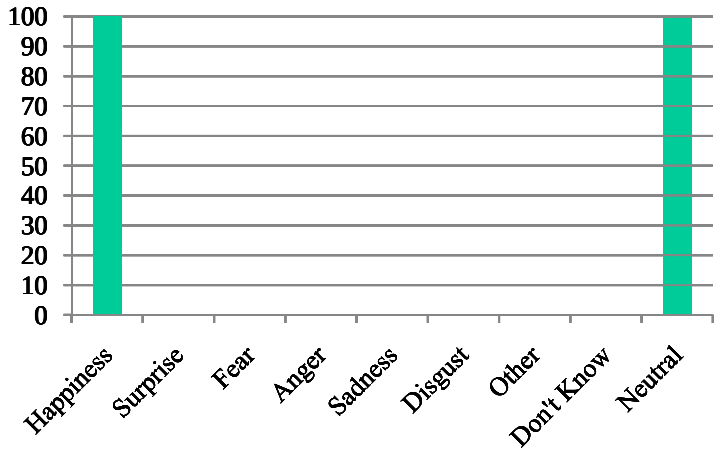
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Introduction

Input:

Output:

Video → **Model** →

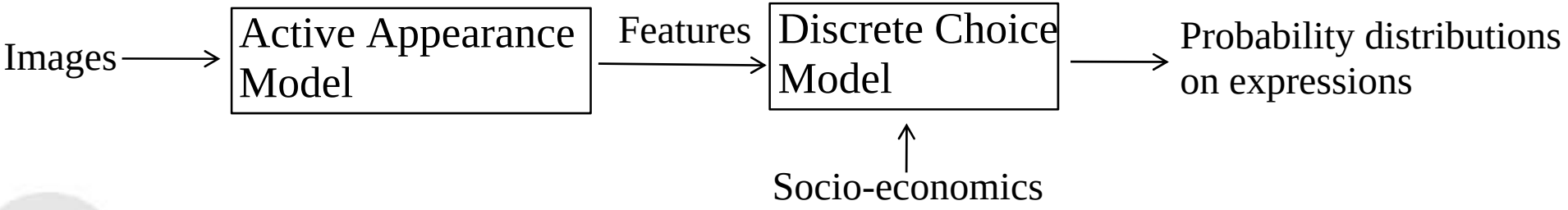


Introduction

- 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, *8th IEEE Int'l Conference on Automatic Face and Gesture Recognition.*

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



Introduction

- Inspired from dynamic model:

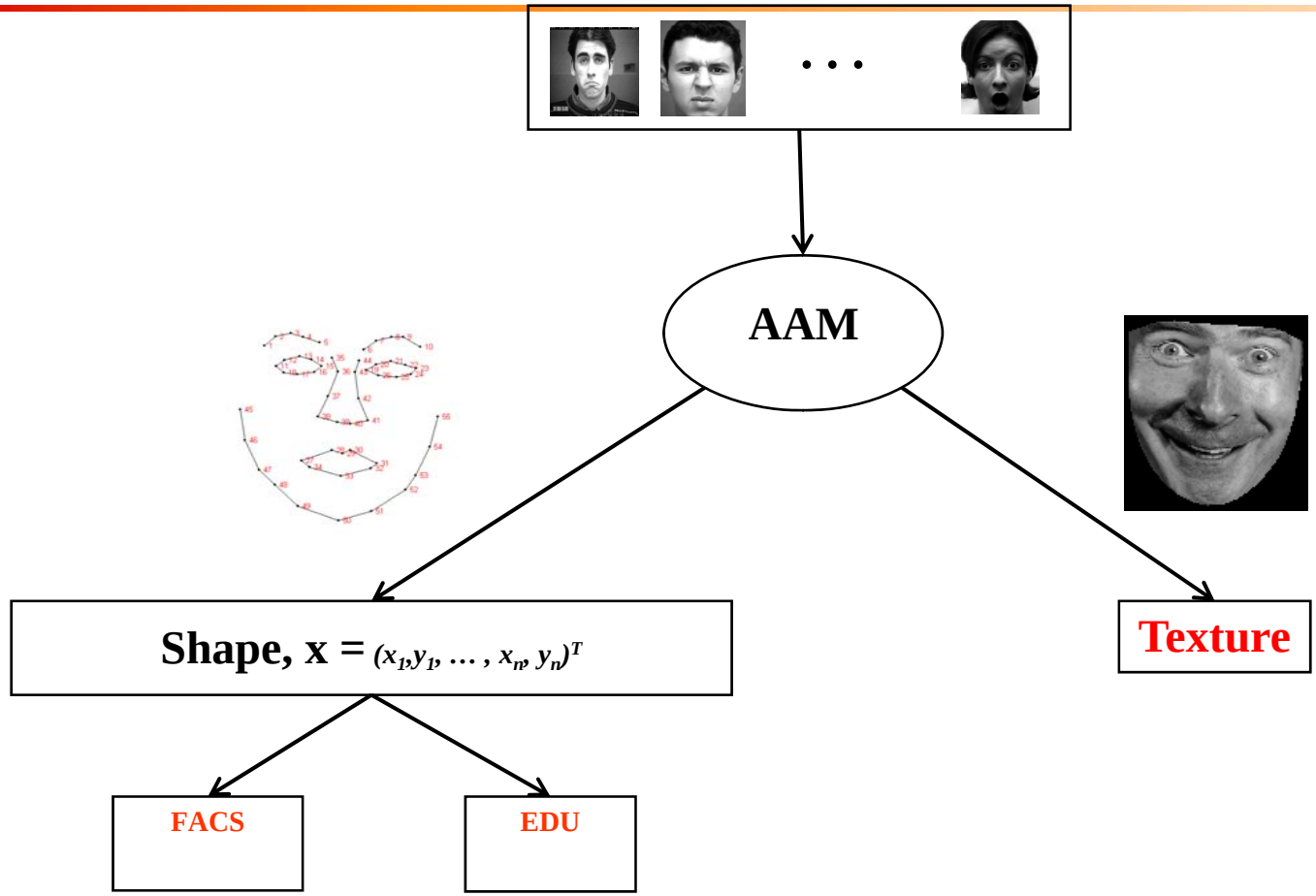
→ Hidden Markov Model

| - State transition process
| - Measurement equation

→ Choudhury, C. F. (2007). *Model Driving Decisions with Latent Plans*, PhD thesis, Massachusetts institue of technology.

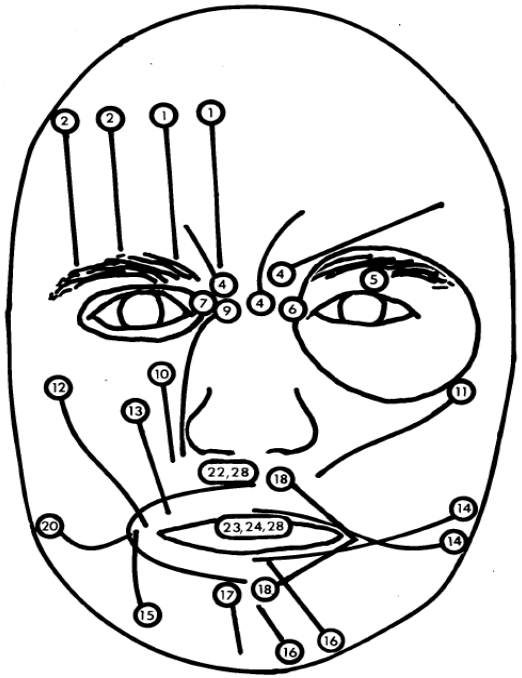
| - Latent decisions
| - Estimation by likelihood maximization

Features extraction: Active Appearance Model



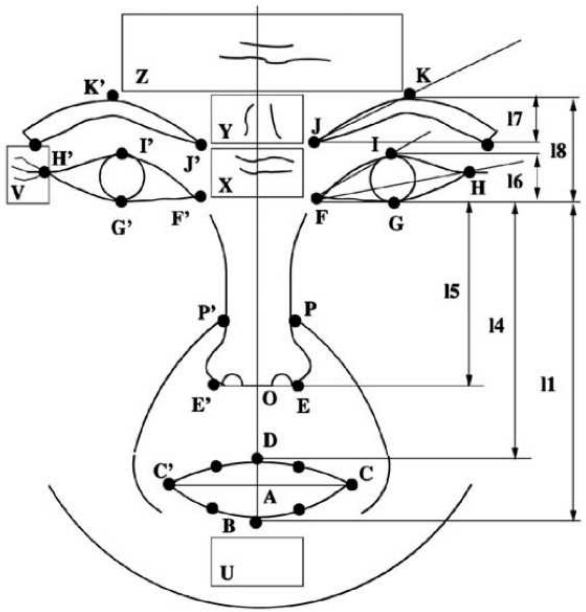
Features extraction: Active Appearance Model FACS

- In 1978 Ekman and Friesen developed the Facial Action Coding System
- Mesurement units: “Action Units” (Aus)
  - AUs are contractions or relaxations of one or more muscles
    - 46 AUs account for changes in facial expression
    - 12 AUs describe changes in gaze direction and head orientation



The FACS has become the leading standard for measuring facial expressions

Features extraction: Active Appearance Model FACS



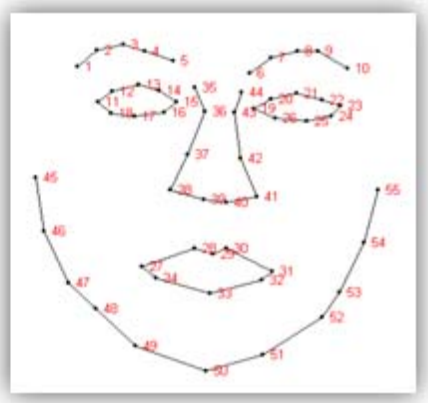
AU1  Inner Brow Raiser	AU2  Outer Brow Raiser	AU4  Brow Lowerer	AU5  Upper Lid Raiser	AU6  Cheek Raiser	AU7  Lid Tightener
AU9  Nose Wrinkler	AU10  Upper Lip Raiser	AU12  Lip Corner Puller	AU15  Lip Corner Depressor	AU16  Lower Lip Depressor	AU17  Chin Raiser
AU20  Lip Stretcher	AU23  Lip Tightener	AU24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

TRANSP-OR

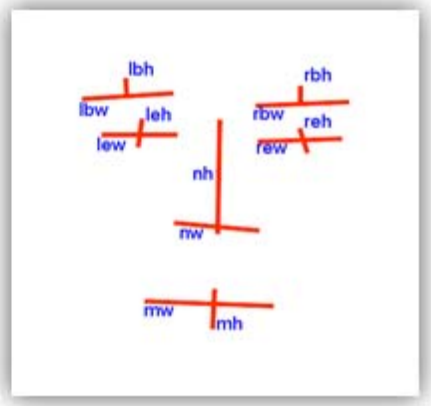
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Features extraction: Active Appearance Model EDU

- Expression Descriptive Units by Antonini, Sorci, Bierlaire and Thiran in « Discrete Choice Models for Static Facial Expression Recognition »



EDU1	$\frac{lew+rew}{leh+reh}$	**EDU8**	$\frac{leh+reh}{lbh+rbh}$
EDU2	$\frac{lbw}{lbh}$	**EDU9**	$\frac{lew}{nw}$
EDU3	$\frac{rbw}{rbh}$	**EDU10**	$\frac{nw}{mw}$
EDU4	$\frac{mw}{mh}$	**EDU11**	EDU2 / EDU4
EDU5	$\frac{nh}{nw}$	**EDU12**	EDU3 / EDU4
EDU6	$\frac{lew}{mw}$	**EDU13**	EDU2 / EDU10
EDU7	$\frac{leh}{mh}$	**EDU14**	EDU3 / EDU10



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Features extraction: Active Appearance Model Texture

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

- 2 databases of video are used:
  - Cohn-Kanade
  - Technical University Munich (TUM)

Data: video database

- The Cohn-Kanade database

  ➡ Actors **playing** expressions, according to the Facial Action Coding System (**FACS**)

55 sequences, 11 subjects

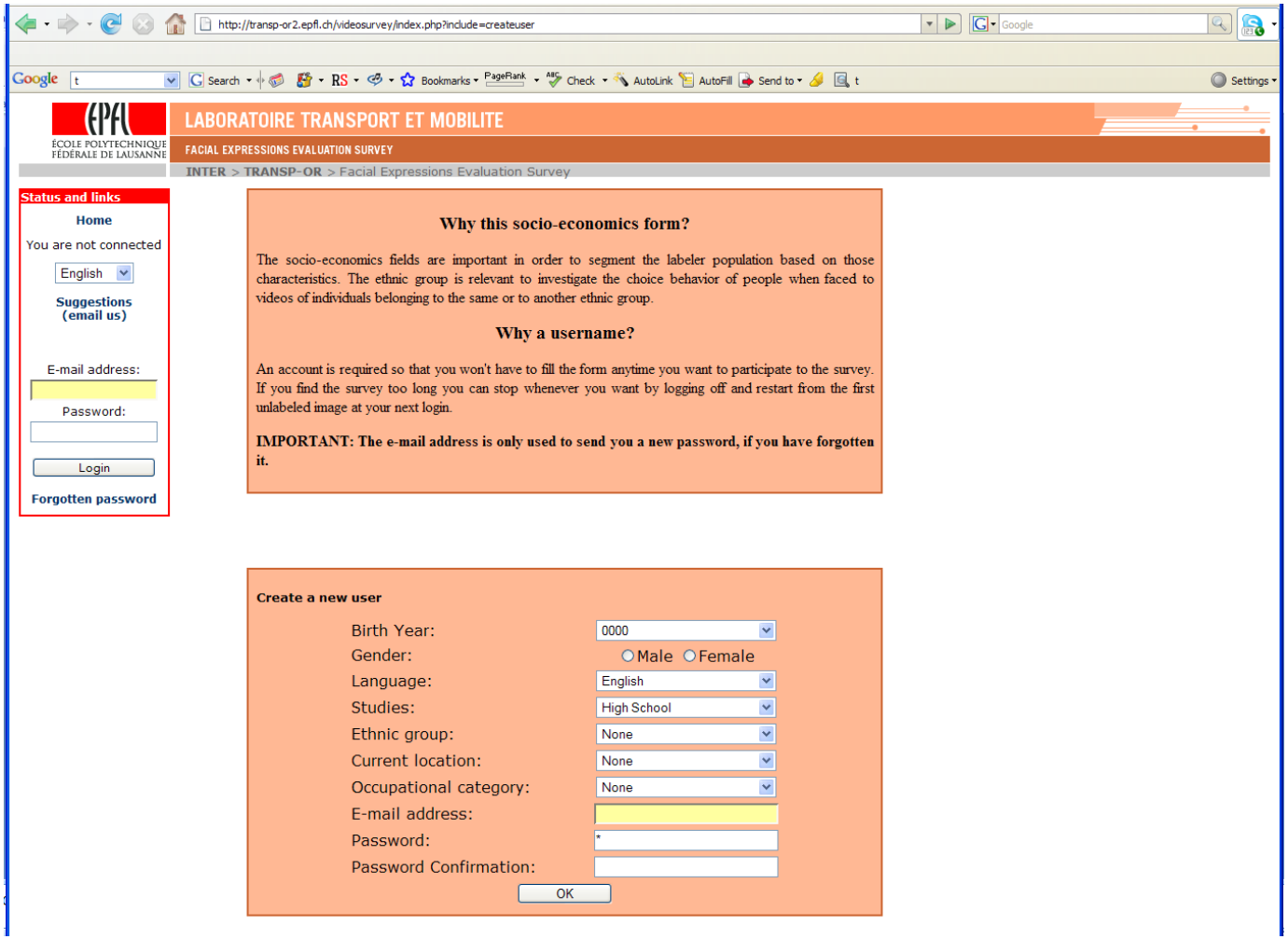
Data: video database

- The Technical University Munich database (TUM)

  ➡ Students faced to a video, natural expressions recorded

399 sequences, 18 subjects

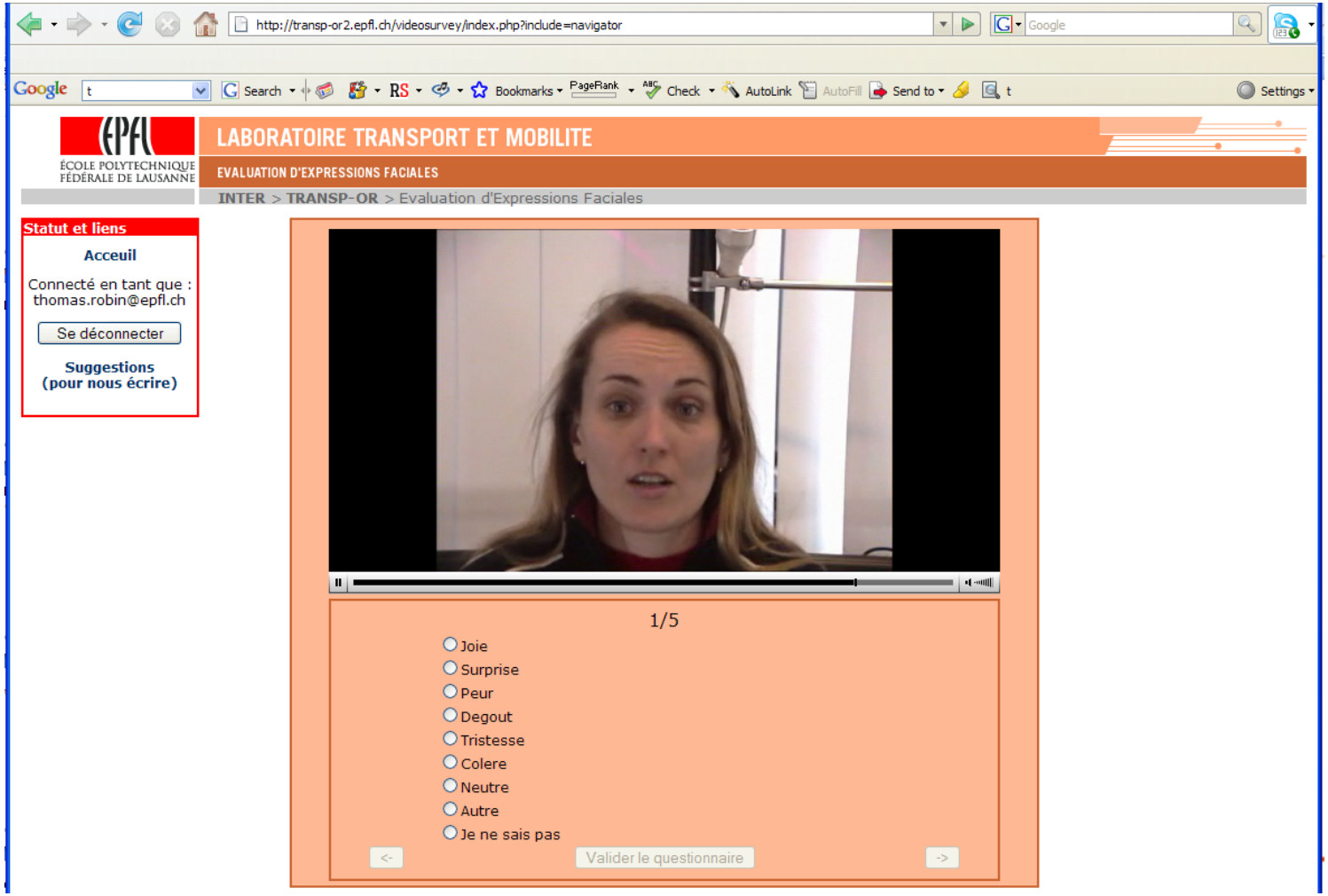
Data: socio-economics



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data: labels



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model: introduction

- Dynamic evaluation of the video sequence → Video sequence labeled with one expression

(1) (2)

- (1) : modeling the dynamic evaluation
  → **The state transition process**

- (1) → (2) : Link between dynamic evaluation and label
  → **The measurement equation**

Model: state transition process

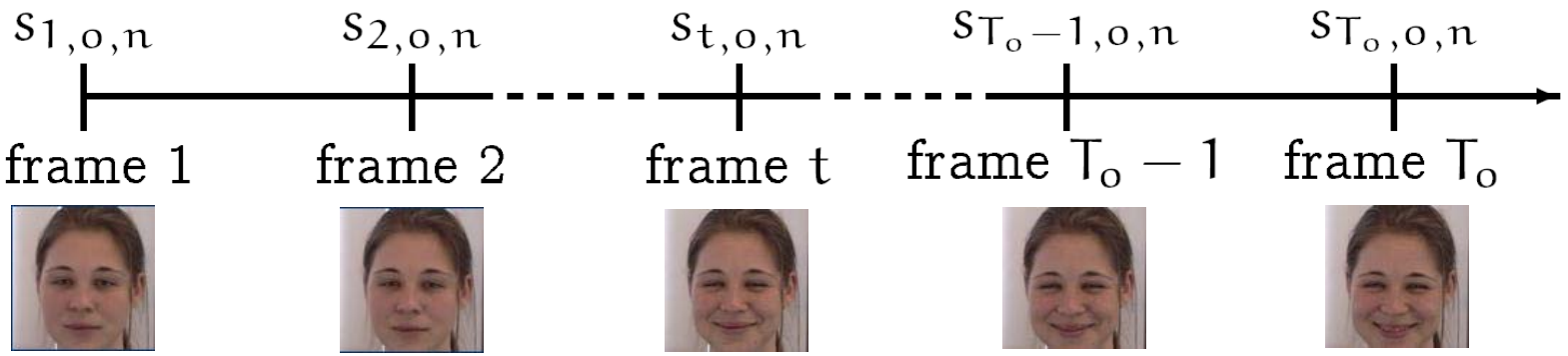
- Modeling of the dynamic evaluation of a video sequence

  - i: expression
  - n: respondent
  - N: total number of respondents
  - O_n: number of video sequences labelled by the respondent n
  - t: frame of a video sequence
  - o: video sequence
  - T_o: total number of frames in the video sequence o

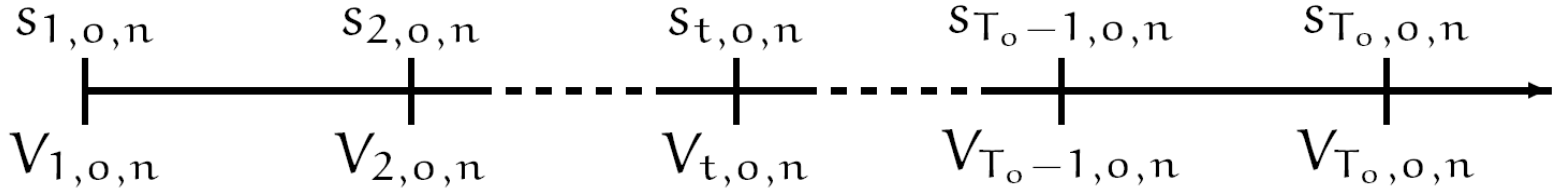
Model: state transition process

- The video sequence **o** watched by the respondent **n**:

  a state $s_{t,o,n}$ associated to each frame ***t***



- A vector of utility functions $V_{t,o,n}$ associated to the state $s_{t,o,n}$



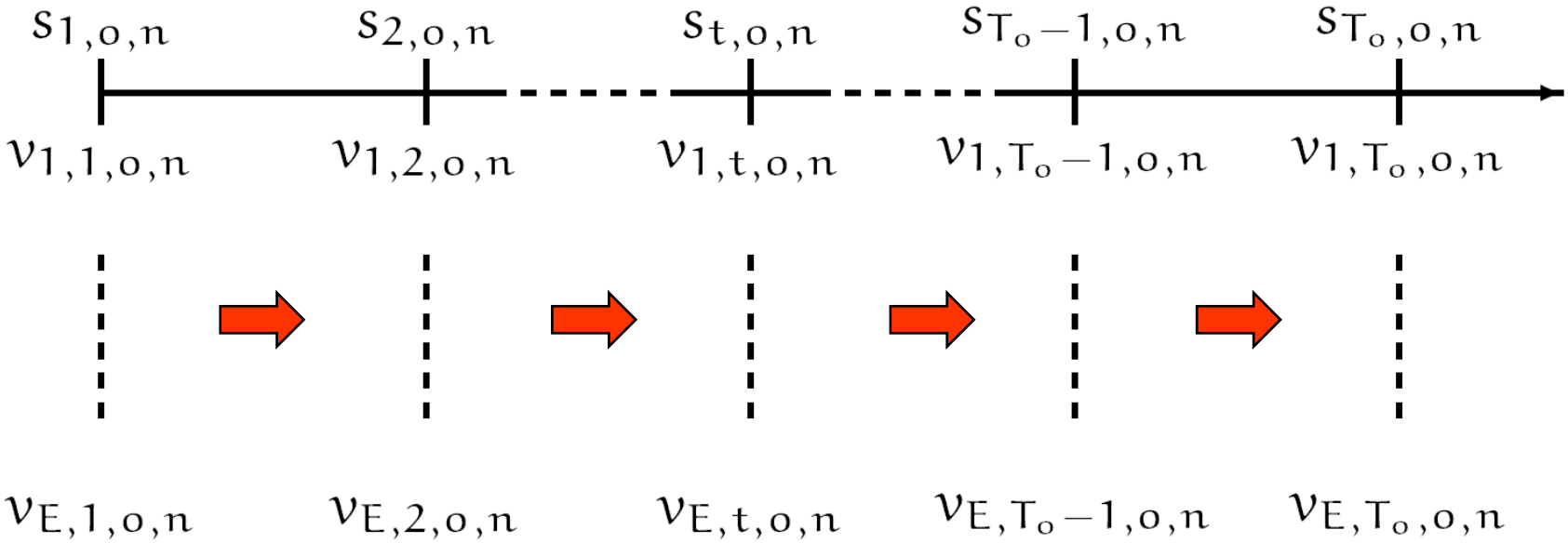
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model: state transition process

- Model the transition between the states $\{s_{t,o,n}\}_{t \leq T_o}$

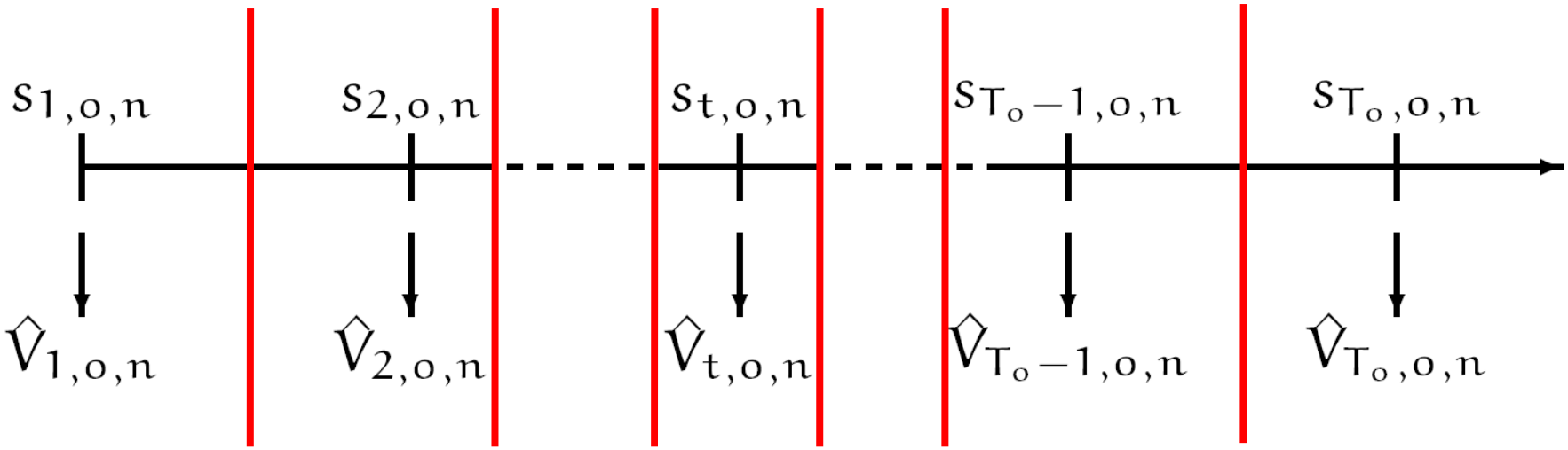
$$V_{t,o,n} = \{v_{1,t,o,n}, v_{2,t,o,n}, \ldots, v_{E,t,o,n}\}$$



Model: state transition process

- $\hat{V}_{t,o,n}$: specific vector of "**static**" utility functions capturing the respondent perception of the frame ***t***

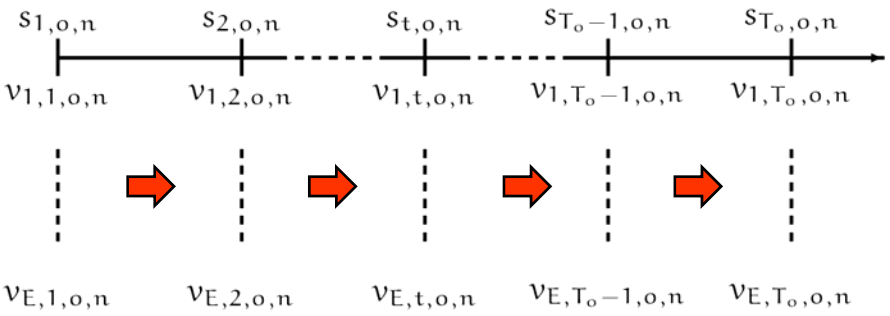
$$\hat{V}_{t,o,n} = \{\hat{v}_{1,t,o,n}, \hat{v}_{2,t,o,n}, \ldots, \hat{v}_{E,t,o,n}\}$$



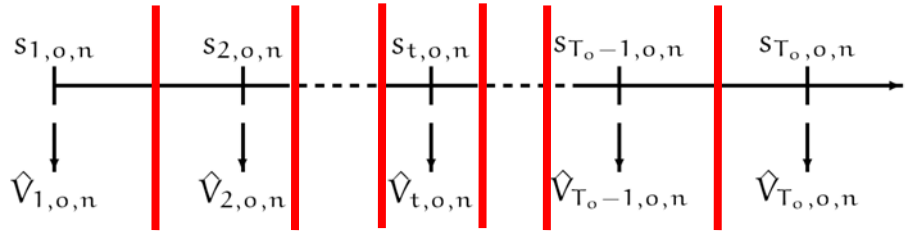
Model: state transition process

- Link between $V_{t,o,n}$ and $\hat{V}_{t,o,n}$

dynamic



static



$$V_{t,o,n} = \sum_{a=1}^{t} A^{t-a} \hat{V}_{a,o,n} + \xi_n$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model: state transition process

$$V_{t,o,n} = \sum_{a=1}^{t} A^{t-a} \hat{V}_{a,o,n} + \xi_n$$

Dynamic utilities vector

Transition matrix with attenuation memory effect

Static utilities of the frame ***a***

Respondent ***n*** unobserved influences

- Remarks:
  - A: in $R^{E \times E}$ can be set diagonal and universal to ease the model identification
  - ξ_n: depends only on the respondent, we supposed it $N(0, \sigma)$ distributed

Model: measurement equation

- Association of a random utility $u_{i,t,o,n}$ for each frame ***t*** of the video sequence ***o*** watched by the respondent ***n*** and for each expression ***i***

$$u_{i,t,o,n} = v_{i,t,o,n} + \epsilon_{i,t,o,n}$$

Random utilities vector

Dynamic utilities vector (**state transition process**)

Random error term distributed $EV(0, \mu)$

Discrete choice models framework

Model: measurement equation

- $P_{o,n}(i/t, \xi_n)$: probability for the respondent n of choosing the expression ***i*** in the frame ***t*** of the video sequence ***o,*** given ξ_n

- $\epsilon_{i,t,o,n} \sim EV(0, \mu)$: mixture logit for panel data

$$P_{o,n}(i/t, \xi_n) = \frac{\exp(v_{i,t,o,n}(\xi_n))}{\sum_{j=1}^{E} \exp(v_{j,t,o,n}(\xi_n))}$$

How link $P_{o,n}(i)$ **with** $P_{o,n}(i/t, \xi_n)$ **?**

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model: measurement equation

- $P_{o,n}(i)$: probability for the respondent ***n*** of choosing the expression ***i*** to label the video sequence ***o***

- $P_{o,n}(t)$: probability for the respondent ***n*** of making his final expression choice for the video sequence ***o***, when watching at the frame ***t***

- $f(\xi_n)$: multivariate density function of ξ_n

$$P_{o,n}(i) = \int \sum_{t=1}^{T_o} P_{o,n}(i/t, \xi_n) P_{o,n}(t) f(\xi_n) d\xi_n$$

Model: measurement equation

Probability of choosing the expression ***i***, for the individual ***n***, watching the video sequence ***o***, in the frame ***t***

multivariate density function of ξ_n

$$P_{o,n}(i) = \int \sum_{t=1}^{T_o} P_{o,n}(i/t, \xi_n) P_{o,n}(t) f(\xi_n) d\xi_n$$

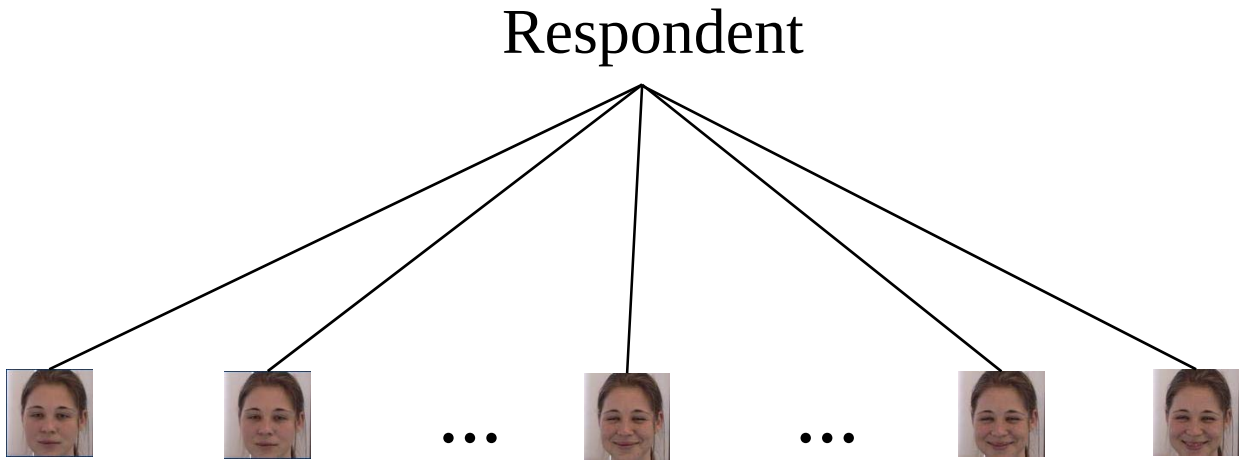
Probability of choosing the expression ***i***, for the individual ***n***, watching the video sequence ***o***

Probability for the individual ***n***, when watching at the video ***o*** to make his choice when faced to the frame ***t***

Model: measurement equation

- $P_{o,n}(t)$: probability for the respondent ***n*** of making his final expression choice for the video sequence ***o***, when watching at the frame ***t***

Discrete choice model



Model: measurement equation

- $\overline{v}_{t,o,n}$: utility measuring the dynamic of the frame ***t*** of the video sequence ***o***, watched by the respondent ***n***

$$P_{o,n}(t) = \frac{exp(\overline{v}_{t,o,n})}{\sum_{a=1}^{T_o} exp(\overline{v}_{a,o,n})}$$

→ **Derivatives** of features in $\overline{v}_{t,o,n}$

Model: likelihood function

- Estimation made by likelihood maximization
- $c_{i,o,n}$: indicator of choice equals to one if respondent ***n*** chose to label the video sequence ***o*** with the expression ***i***

$$l = \prod_{n=1}^{N} \prod_{o=1}^{O_n} P_{o,n}(i)$$

$$l = \prod_{n=1}^{N} \prod_{o=1}^{O_n} \left(\prod_{i=1}^{E} \int \sum_{t=1}^{T_o} P_{o,n}(i/t, \xi_n) P_{o,n}(t) f(\xi_n) d\xi_n \right)^{c_{i,o,n}}$$

Model: likelihood function

$$l = \prod_{n=1}^{N} \prod_{o=1}^{O_n} P_{o,n}(i)$$

$$l = \prod_{n=1}^{N} \prod_{o=1}^{O_n} \left(\prod_{i=1}^{E} \int \sum_{t=1}^{T_o} P_{o,n}(i/t, \xi_n) P_{o,n}(t) f(\xi_n) d\xi_n \right)^{c_{i,o,n}}$$

$$l = \prod_{n=1}^{N} \prod_{o=1}^{O_n} \left(\prod_{i=1}^{E} \int \sum_{t=1}^{T_o} \frac{\exp(v_{i,t,o,n}(\xi_n))}{\sum_{j=1}^{E} \exp(v_{j,t,o,n}(\xi_n))} \frac{\exp(\overline{v}_{t,o,n})}{\sum_{a=1}^{T_o} \exp(\overline{v}_{a,o,n})} f(\xi_n) d\xi_n \right)^{c_{i,o,n}}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model: specifications

- Discrete Choice Model framework
- Attributes

  - $\hat{V}_{t,o,n}$: **FACS**, **EDU**, **Texture**, **Socio-economics**

  → M. Sorci et al, “Static facial expression recognition”

  - $\overline{v}_{t,o,n}$: **Derivatives** of features

  → measure the frame dynamic

Conclusions and Perspectives

- Conclusion:
  - database of face video annotations
  - new model framework
  - estimation by likelihood maximization

- Perspectives:
  - implementation of the likelihood maximization
  - model estimation: find a satisfactory specification
  - model validation: measure the prediction power

Conclusions and Perspectives

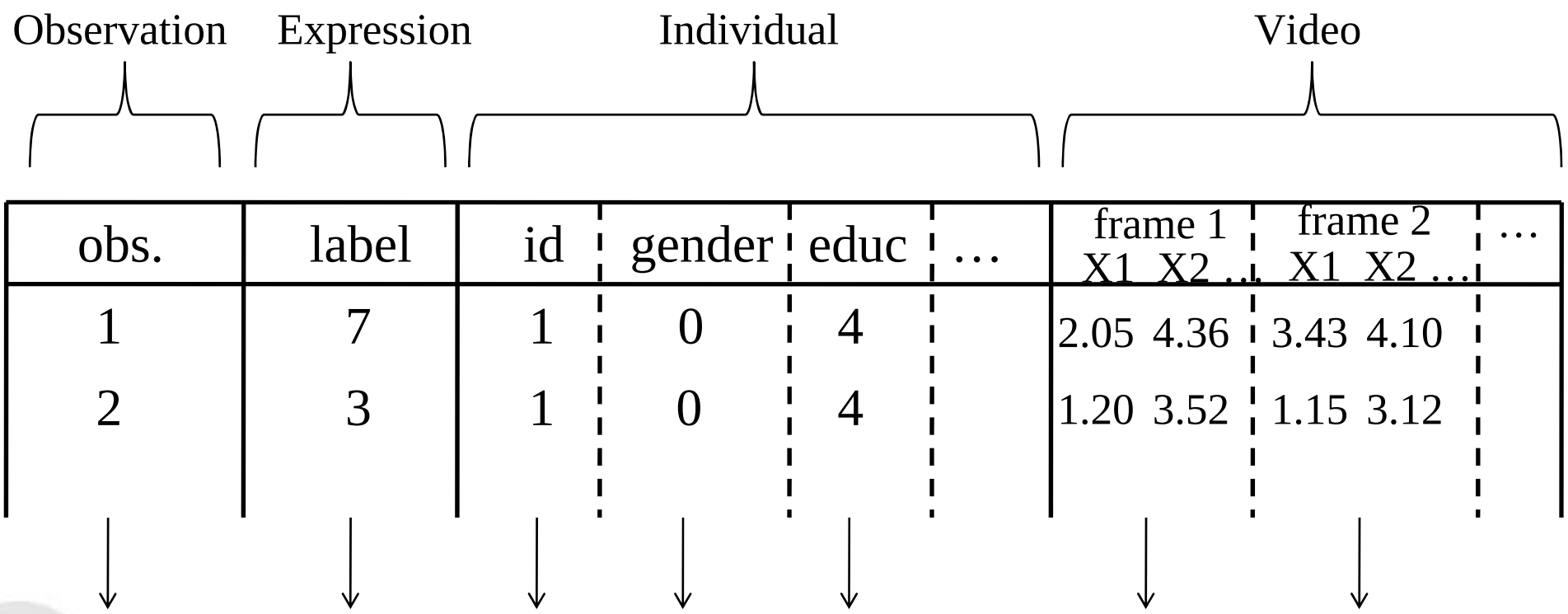
- Conclusion:
  - database of face video annotations
  - new model framework
  - estimation by likelihood maximization

- Perspectives:
  - implementation of the likelihood maximization
  - model estimation: find a satisfactory specification
  - model validation: measure the prediction power

Thank you for your attention

Data: data file

- Face video annotations data base → Data file for model estimation



Model: state transition process

- $s_{t,o,n}$: state associated with the frame t of the video sequence o watched by the respondent nÍ
- $\hat{V}_{t,o,n}$: vector of utilities characterizing the frame t of the video sequence o for the individual n (dimension E)
- $V_{t,o,n}$: vector of utilities associated with the state $s_{t,o,n}$ (dimension E)
- ξ_n : vector of error terms specific to the individual n, interfering in the transition process (dimension E)
- σ : vector of standard errors of ξ_n (dimension E)
- A: squared appreciation matrix of dimension E $\times$ E associated to the respondent n faced to the video sequence o

Model: measurement equation

- Link the observation choice $y_{o,n}$ with the states sequence $\{s_{t,o,n}\}_{t \leq T_o}$

  - $U_{t,o,n}$: vector of random utilities associated with $s_{t,o,n}$ (dimension E)
  - $\epsilon_{t,o,n}$: vector of unobserved attributes interfering in $U_{t,o,n}$ associated to $s_{t,o,n}$ (dimension E)
  - $\overline{v}_{t,o,n}$: utility associated with the frame t of the video sequence o for the individual n, measuring the frame dynamic
  - $y_{o,n}$: choice made by the respondent n when faced to the video sequence o