Audio-Visual Attention: Eye-Tracking Dataset and Analysis ToolBox

Pierre Marighetto, *Université de Mons*
Antoine Coutrot, *University College London*
Nicolas Riche, *Université de Mons*
Nathalie Guyader, *Université de Grenoble*
Matei Mancas, *Université de Mons*
Bernard Gosselin, *Université de Mons*
Robert Laganière, *University of Ottawa*

Problem

- Auditive attention is hard to evaluate alone
- Dynamic saliency maps are difficult to plot and evaluate
- Audio-visual saliency database are rare

Proposition

- New eye tracking dataset in audio and non audio conditions
- ToolBox to help analysis on eye tracking data on videos
- Comparison of eye tracking data in audio and non audio conditions

Dataset

Experiment status	Participants	Stimuli	Source
Already existing	40	50 videos from professional movies	Antoine Coutrot, Nathalie Guyader, Gelu Ionescu, Alice Caplier, et al., “Influence of soundtrack on eye movements during video exploration,” Journal of Eye Movement Research, vol. 5, no. 4, 2012.
Already existing	72	60 videos	Antoine Coutrot, Nathalie Guyader, Gelu Ionescu, and Alice Caplier, “Video viewing: do auditory salient events capture visual attention?,” Annals of Telecommunications, vol. 69, no. 1, pp. 89–97, 2014.
Already existing	40	15 videos extracted from the AMI Meeting Corpus	Antoine Coutrot and Nathalie Guyader, “An Efficient Audiovisual Saliency Model to Predict Eye Positions When Looking at Conversations ,” in European Signal Processing Conference (EUSIPCO), Nice, France, 2015.
Original experiment	24	23 videos from Hollywood2	

Our Dataset

- 24 peoples watched 23 videos, with eye-tracking device pointing at them
  - Half of the videos were displayed with sound
  - Half of them were muted

- Example
  - https://youtu.be/Et8ea9nboE4

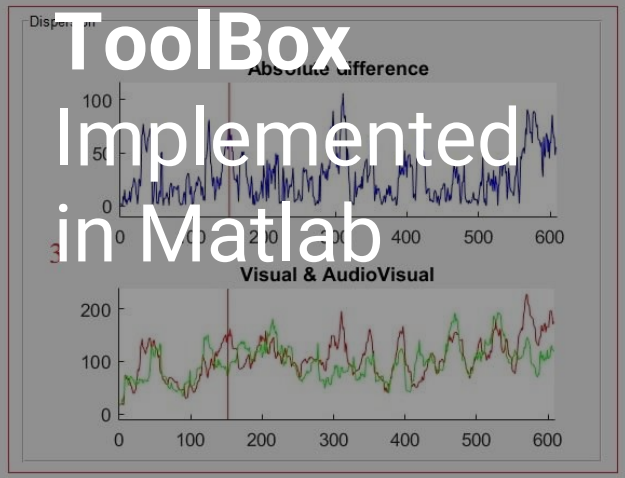
Dataset

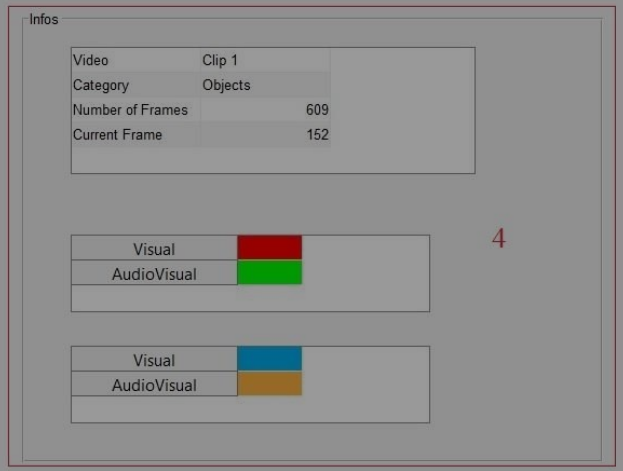
Final dataset contains 148 videos explored by a total of 176 participants, in different audio conditions

The new dataset was split into 3 visual categories

- Moving objects
- Landscapes
- Faces

ToolBox Implemented in Matlab





Give access to 5 metrics computed on every videos and each condition

Metrics are divided 3 categories

- Inter condition hybrid metrics
- Inter condition density comparisons
- Intra condition fixation comparisons

Metrics

Inter condition density comparisons

- Compares audiovisual and visual density maps
- Symmetric metrics (one result)
- 2 metrics
  - Pearson's Correlation Coefficient (CC)
  - Similarity

Inter condition hybrid metrics

- Compares density to fixation
- Asymmetrical metrics (2 results)
- 2 metrics
  - Kullback-Lieberg Divergence (KL-Divergence)
  - Normalized Scanpath Saliency (NSS)

Intra condition fixation comparisons

- Computes audiovisual and visual fixations
- Each condition is computed sparately (one result)
- 1 metric
  - Dispersion

ToolBox

Database Metrics

Stimulus

1

Gaze

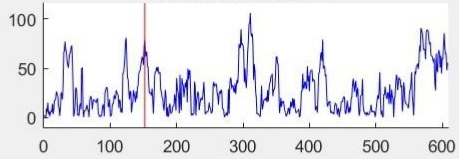
AudioVisual

2

Visual

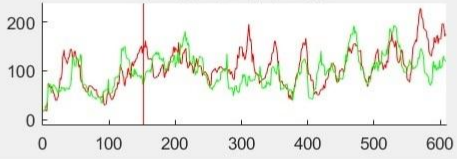
Dispersion

Absolute difference



3

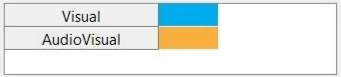
Visual & AudioVisual



Infos

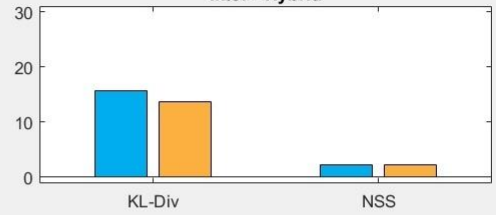
Video	Clip 1
Category	Objects
Number of Frames	609
Current Frame	152

4



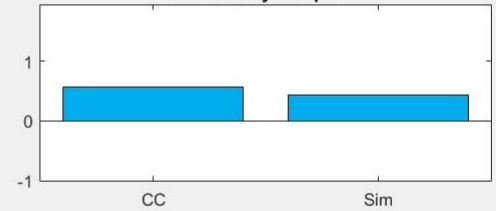
Metrics

Inter - Hybrid

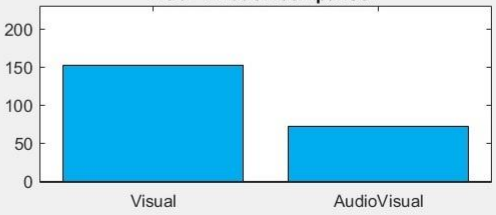


5

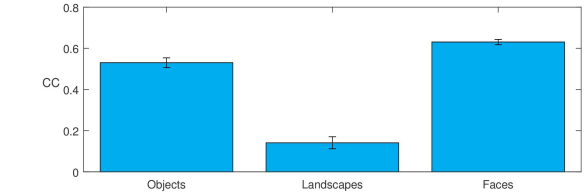
Inter - Density comparison

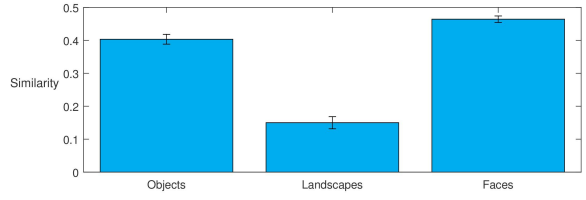


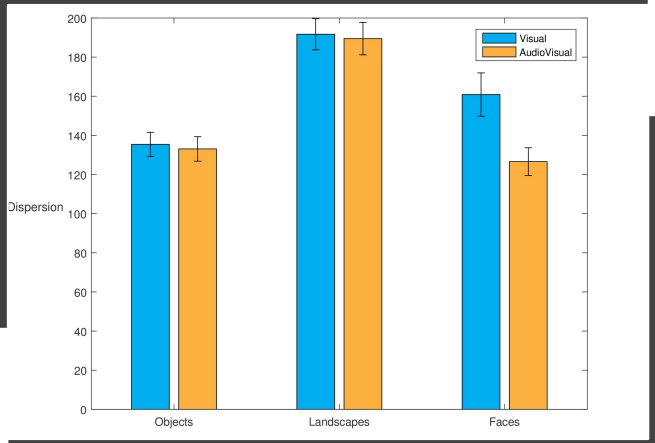
Intra - Fixation comparison

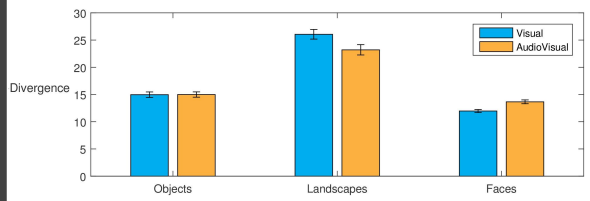


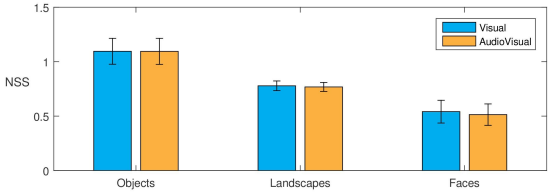
Analysis


CC
0.8
0.6
0.4
0.2
0
Objects
Landscapes
Faces


Similarity
0.5
0.4
0.3
0.2
0.1
0
Objects
Landscapes
Faces


Dispersion
200
180
160
140
120
100
80
60
40
20
0
Visual
AudioVisual
Objects
Landscapes
Faces


Divergence
30
25
20
15
10
5
0
Visual
AudioVisual
Objects
Landscapes
Faces


NSS
1.5
1
0.5
0
Visual
AudioVisual
Objects
Landscapes
Faces

Future works

- Add audio features extracted from the video
- Add video features extracted from the video
- Extend the ToolBox to different datasets

Thank you for your attention

UMONS
Université de Mons

- Contact
  - pierre.marighetto@gmail.com

uOttawa

- Project’s website
  - http://tcts.fpms.ac.be/attention/index.php?static06/projects

VIVA