

# Visual Saliency and Perceptual Grouping in Multimodal Interactivity

Frédéric Landragin, Nadia Bellalem, Laurent Romary  
LORIA Laboratory, France

## General context

- Natural language and spontaneous gestures
- Interpretation of multimodal referring expressions in the visual context
- No dialogue history, no task model, no user model

⇒ focus on visual perception, through two notions:

- visual saliency
- perceptual grouping

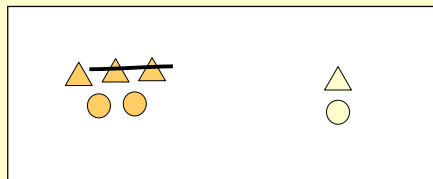
## Interaction context

- Examples adapted from a multimodal corpus
- No restriction on speech
- Restriction on gestures: use of a touch screen

⇒ 2-D trajectories, points or lines going on or between the percepts (graphical representation of the task objects)

## Problematics

- The use of perceptual grouping

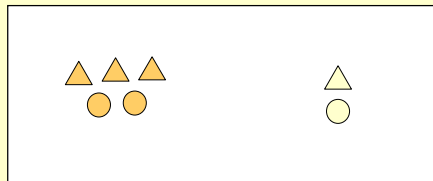


« these three objects »

→ {△, △, △}

## Problematics

- The use of perceptual grouping



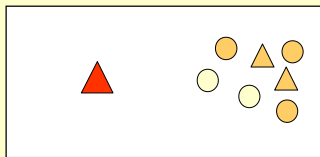
« these three objects »

→ {△, △, △}

« the two circles »

→ {○, ○}

- The use of salience

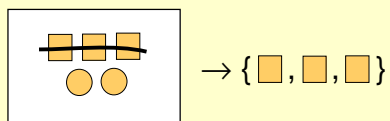


« the triangle »

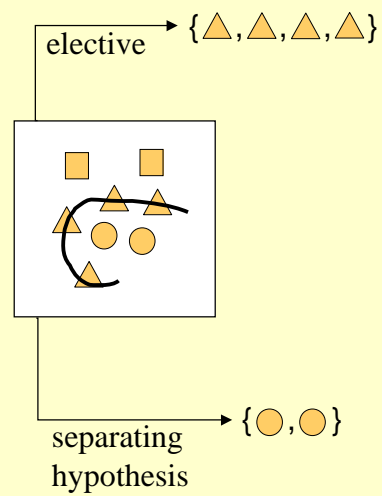
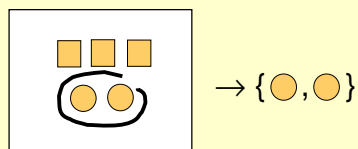
→ {▲}

## Gesture trajectories

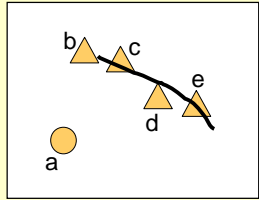
- Elective gestures



- Separating gestures

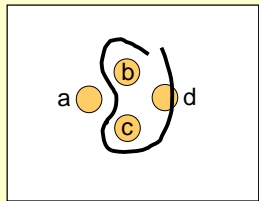


## Output of the gesture module



on the immediate continuity  
near the trajectory

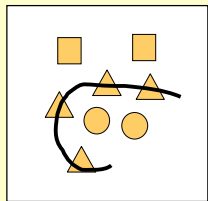
	a	b	c	d	e
elective:	0	.5	1	.5	1
separating:	<i>not relevant</i>				



covering ratio

	a	b	c	d
elective:	<i>not relevant</i>			
separating:	0	1	1	.8

## Output of the gesture module

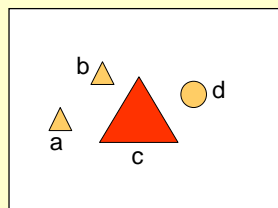


	□	△	○
elective:	0	1	0
separating:	0	0	1

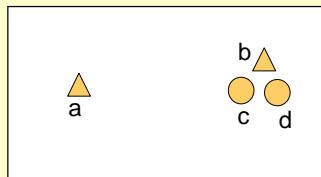
## Gesture and salience

- Gesture is the first way to make an object salient
- With no gesture, an object may be salient when it has a property that the other objects do not have:
  - being the only one of its category
  - being the only one of its size
  - being the only one of its colour
  - being isolated (and the others grouped)
  - ...

## Output of the salience module



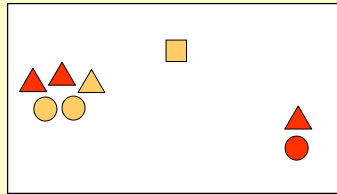
	category	size	colour	isolation	
a:	0	0	0	0	→ 0
b:	0	0	0	0	→ 0
c:	0	1	1	0	→ <b>.5</b>
d:	1	0	0	0	→ .25



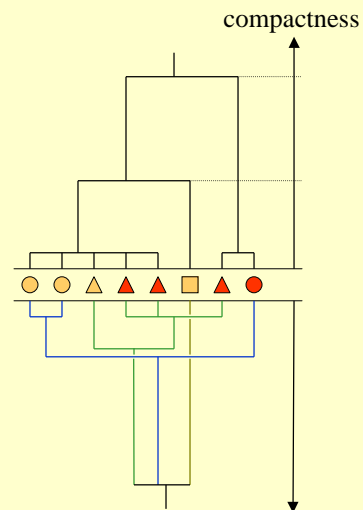
a:	0	0	0	1	→ <b>.25</b>
b:	0	0	0	0	→ 0
c:	0	0	0	0	→ 0
d:	0	0	0	0	→ 0

## Perceptual grouping

- Grouping by proximity



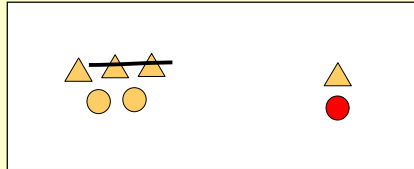
- Grouping by similarity



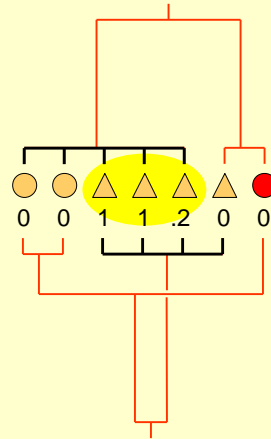
## Integration of the algorithms

- Build up the dendrograms for proximity, similarity...
- If a gesture is produced, link each object to the gesture scores. If no gesture is produced, link each object to the salience score.
- Taking the linguistic referring expression into account, build up the first group including the object with the biggest score.
- If it is possible, build up other groups, by going up in the dendrograms.

## Integration of the algorithms



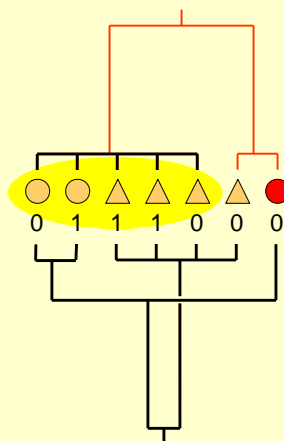
« these three objects »



## Integration of the algorithms



« these objects »



## Future work

- Integrate dialogue history, task and user model.
- Apply the algorithm to the generation of multimodal referring expressions.
- Integrate and validate this work in the MIAMM Project (Multimedia Information Access using Multiple Modalities)