Overview

• **Research domain:** Interpretation of natural language and spontaneous gestures.

• **Background:** A model of contextual interpretation of multimodal referring expressions in visual and task contexts.

• **Objective:** To show that our model can be extended to an interaction mode including tactile and kinesthetic feedback.

• **Context:** Conception phase of the IST-MIAMM European project, with DFKI, TNO, SONY & CANON (Multidimensional Information Access using Multiple Modalities).
Reference domains and visual context

- The use of perceptual grouping

  « these three objects »
  → \{△, △, △\}

  « the two circles »
  → \{○, ○\}

- The use of salience

  « the triangle »
  → \{△\}

Multimodal fusion architecture
Haptics and deixis

• Haptic gestures can take the three classical functions of gesture in man-machine interaction:
  – semiotic function: ‘select this object’
  – ergotic function: ‘reduce the size of this object’
  – epistemic function: ‘save the compliance of this object’

• How can the system identify the function(s)?
  – linguistic clues (referential expression, predicate)
  – task indications (possibilities linked to a type of objects)

• Deixis role: to make the object salient, whatever the function, in order to focus the addressee’s attention on it.

Haptics and perceptual grouping

• Interest: formalism for the focalization on a subset of objects

• Grouping factors:
  – objects which have similar tactile or haptic properties (shape, consistency, texture)
  – objects that have been browsed by the user (the elements of such a group are ordered)
  – objects that are stuck together, parts of a same object...
Haptics and perceptual domains

• Can visual and tactile perceptions work together?
  – simultaneous visual and tactile perception implies the same world of objects (and synchronized feedbacks)
  – a referring expression can be interpreted in visual context or in tactile context

• How can the system identify the nature of perception?
  – for immediate references, the visual context gives the reference domain and haptic gives the starting point in it
  – for references chains, each type of context can give the reference domain (so both hypotheses must be tested)

Haptics and dialogue history

• Interpretations that need an order within the reference domain: ‘the first one’, ‘the next one’, ‘the last one’
  – in visual perception, guiding lines can be helpful (if none, an order can always be built with the reading direction)
  – in haptic perception, the only criterion can be the manipulation order

• Some referring expressions that do not need an order may be interpreted in the haptic manipulation history
  – ‘the big one’ (in the domain of browsed objects)
  – ‘them’ (the most pressured objects)
Summary

• What does not change from deictic to haptic
  – the status of speech and gesture in the architecture
  – the repartition of information among speech and gesture
  – the need of reference domain
  – the use of salience and the use of orders in domains
  – the algorithms for the exploitation of all these notions

• What does change
  – some unchanged notions can have one more cause
  – objects must be browsed to be grouped in a haptic domain
  – one aspect of the architecture: the visual perception module becomes the visuo-tactile perception module
Future work

• Within the dialogue manager module, domains may be confronted, using a relevance criterion
  The way the linguistic contraints of the referring expression apply in the different domains may be such a criterion.

• Validation in the MIAMM framework
  The transition from deictic to haptic may not be an additional cost for the development of a dialogue system, both from the architecture point of view and the dialogue management point of view.