



# Advanced Multimedia Interfaces for Flexible Communications



J. Allasia, A. C. Andrés, D. C. Barbu, I. Petre, V. Radut, U. Saaed & J. Urbain



hosted by *Boğaziçi University, Multimedia Group*



# Goal of the project

## Proof of concept

To exemplify a multimodal service (**vote registration**) in synchronous communications (**remote meeting**)



# Outline

## § Network

- ú Next Generation Networks
- ú Protocol for vote registration

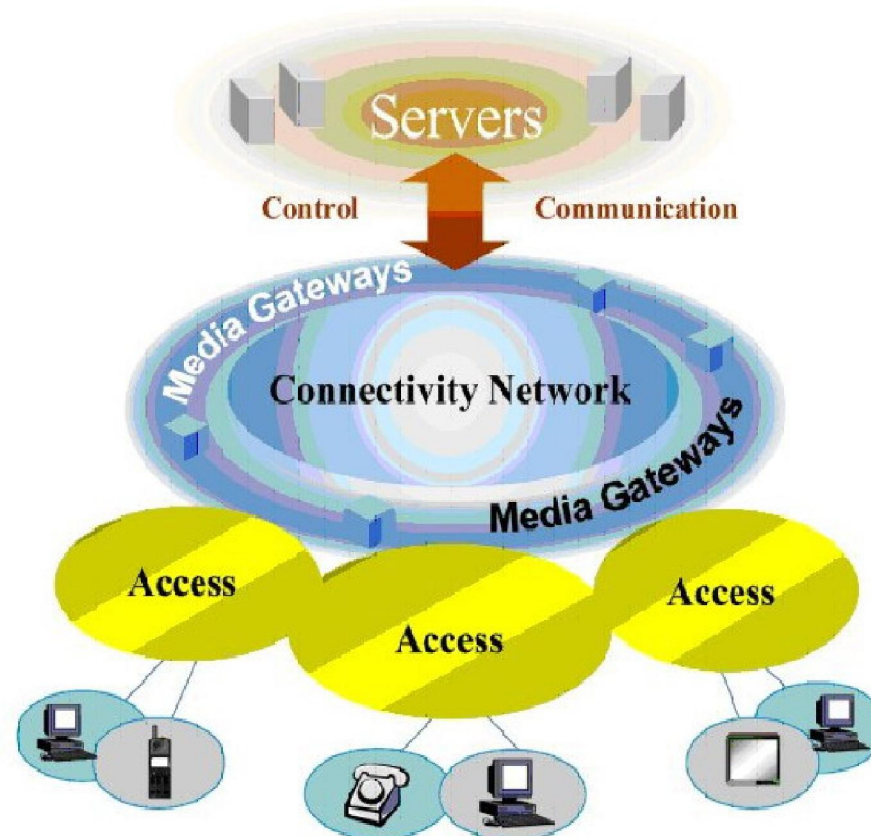
## § Vote-registration multimodal signal processing

- ú Text input-output – chat, sms, instant message
- ú Audio input-output – phone call
- ú Single-Face video input – home webcam
- ú Multiple-people video input – video conf input

# Next Generation Networks

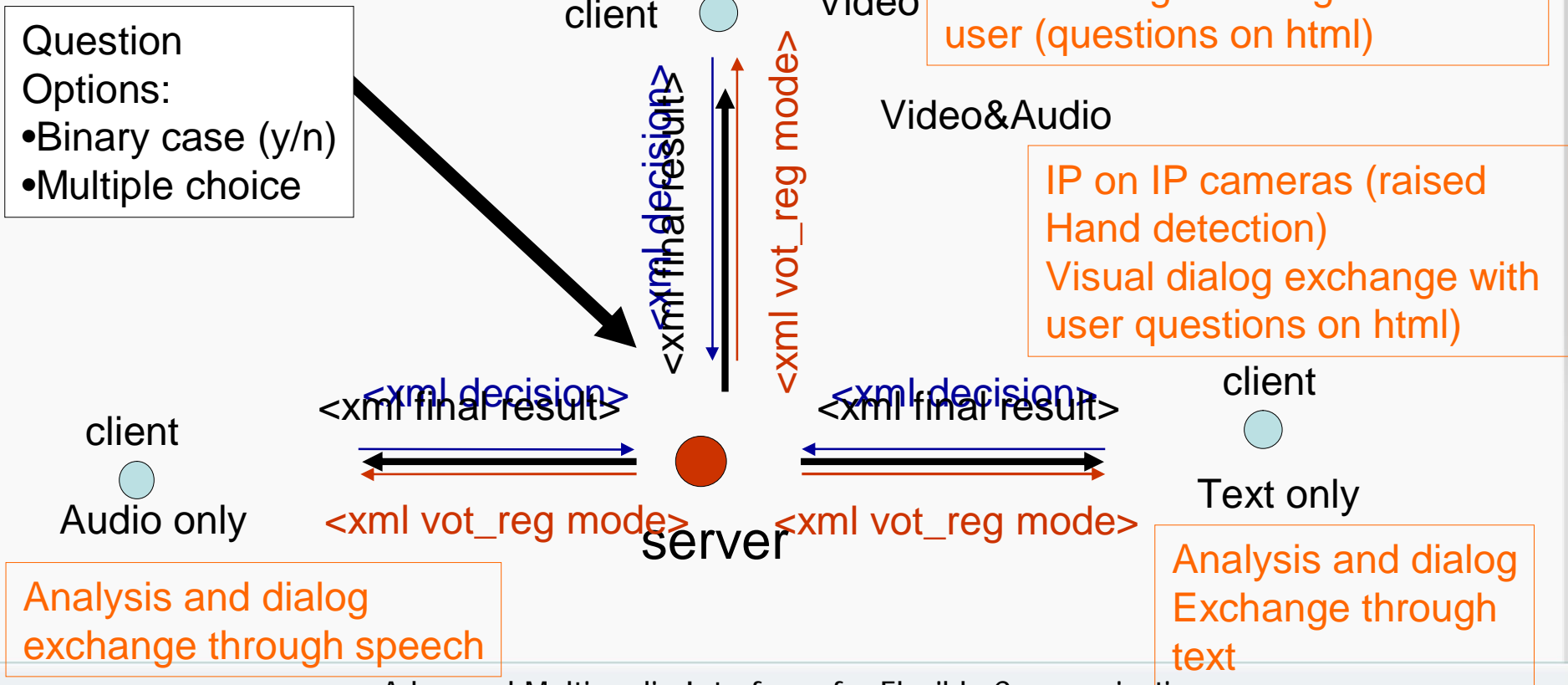


§ Advanced services for heterogeneous, platform-agnostic, asymmetric communications



# Network

## § Vote registration mode



# XML design



```
<?xml version="1.0" encoding="ISO-8859-1"?>
<context_profile>
  <meeting_id>100</meeting_id>
  <num_attendees>3</num_attendees>
  <participant>
    <participant_type>Organizer</participant_type>

    <communications_in>video_audio</communications_in>

    <communications_out>video_audio</communications_out>
    <communications_id>01</communications_id>
    <firstname>Usman</firstname>
    <lastname>Saeed</lastname>

    <company_name>EURECOM</company_name>
    <title>PhD Student</title>
  </participant>
</context_profile>
```

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<command_profile>
  <command_name>Start_chat</command_name>
  <num_param>2</num_param>
  <param_1>bosphorus_07</param_1>

  <param_2>bosphorus_08</param_2>
</command_profile>
```

# Text-Only integration into the architecture

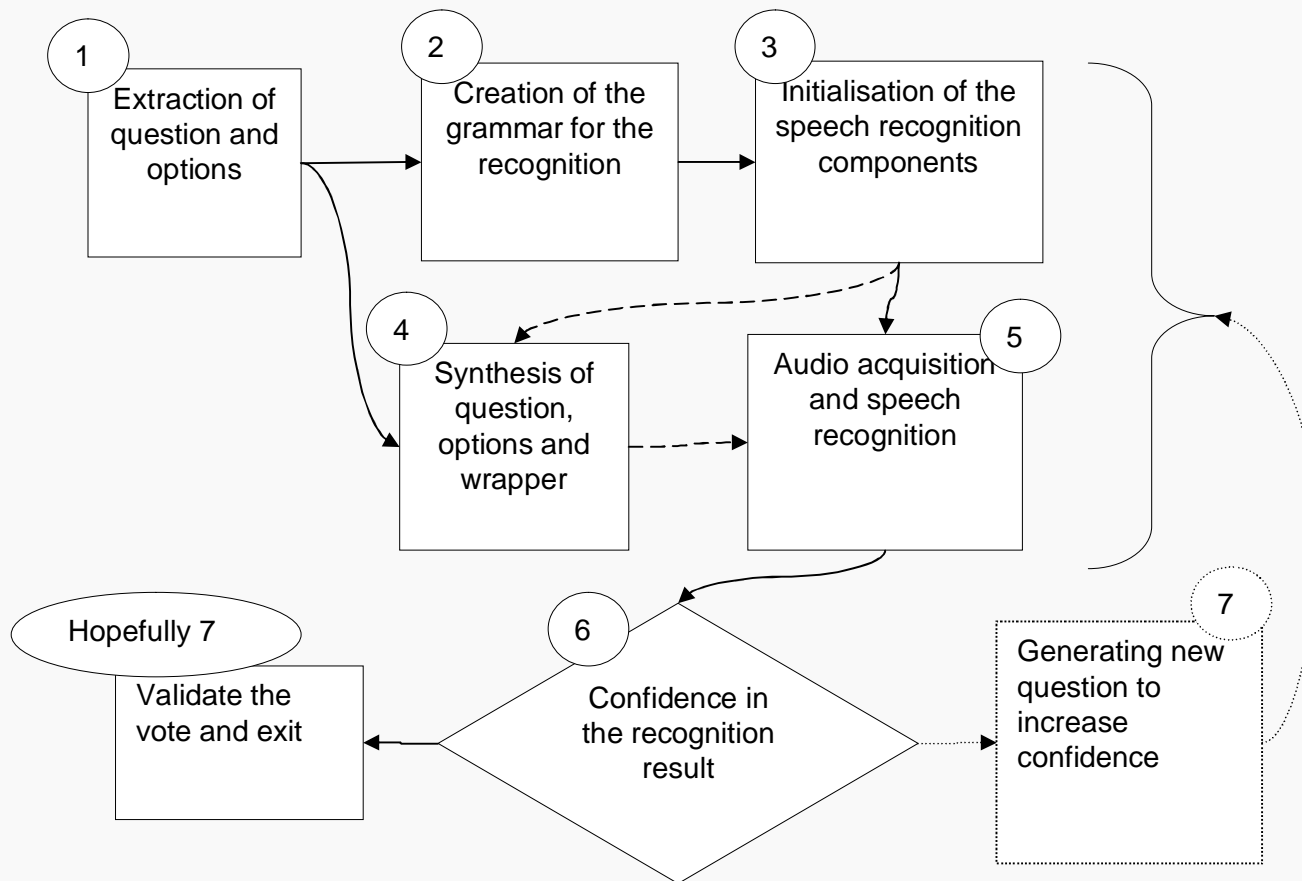


Demo

# Speech Processing: System Architecture

eNTERFACE '07

The SIMILAR NoE  
Summer Workshop  
on Multimodal Interfaces



# Speech Processing: Results

- q Experiment conducted to test the speech recognition:
  - Small vocabulary (5 words: car, plane, train, boat bus)
  - 12 different speakers
  - First in a quiet environment, then with music in background
  
- q Satisfying results:
  - Few mistakes in a quiet environment (7%), as expected some more with the added noise (18%)
  - Mistakes are in 85% accompanied with a medium ( $<0.5$ ) or small ( $<0.1$ ) confidence level, so it should be possible to correct these errors via the confirmation question
  - Most confusions between :
    - plane and train
    - boat and bus

# Head Gesture Based Yes/No Detection.

eNTERFACE '07

The SIMILAR NoE  
Summer Workshop  
on Multimodal Interfaces



## § Goal:

- ú User responds to Yes/No type questions by head gestures.

## § Method:

- ú Face Detection using Haar Object Detection.
- ú Selection of edge and corner points by Harris Operator.
- ú Tracking of facial points using Lucas Kanade Method.
- ú Decision of Yes / No based on amount of horizontal or vertical Motion.

## § Future Work:

- ú User responds to multiple choice questions with limited vocabulary.
- ú Lip detection and motion tracking to recognize 1,2,3,4. etc
- ú Integration in the communications network (client)

# Head Gesture Based Yes/No Detection.

eNTERFACE '07

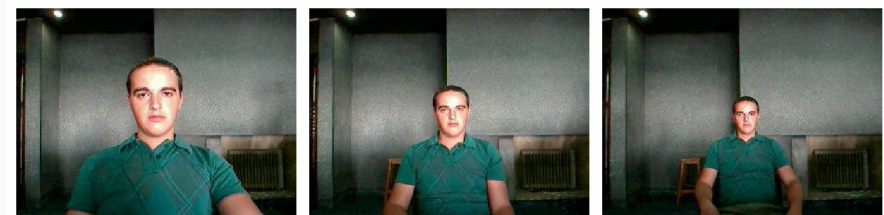
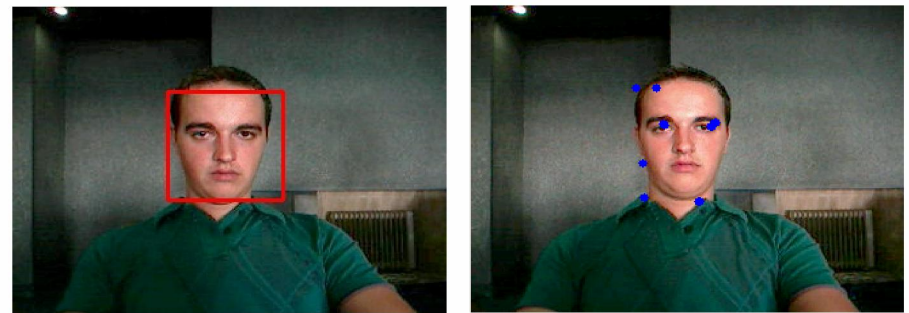
The SIMILAR NoE  
Summer Workshop  
on Multimodal Interfaces



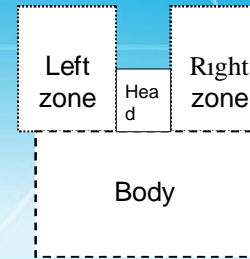
## § Experiments & Results

- ú 5 Subjects
- ú 5 Questions
  - Are You Male ?
- ú Scale and Illumination
- ú Sex/Facial Hair/Glasses
- ú Results: 92 %

- [Usman-Demo.avi](#)



# Voting system detect raised hands

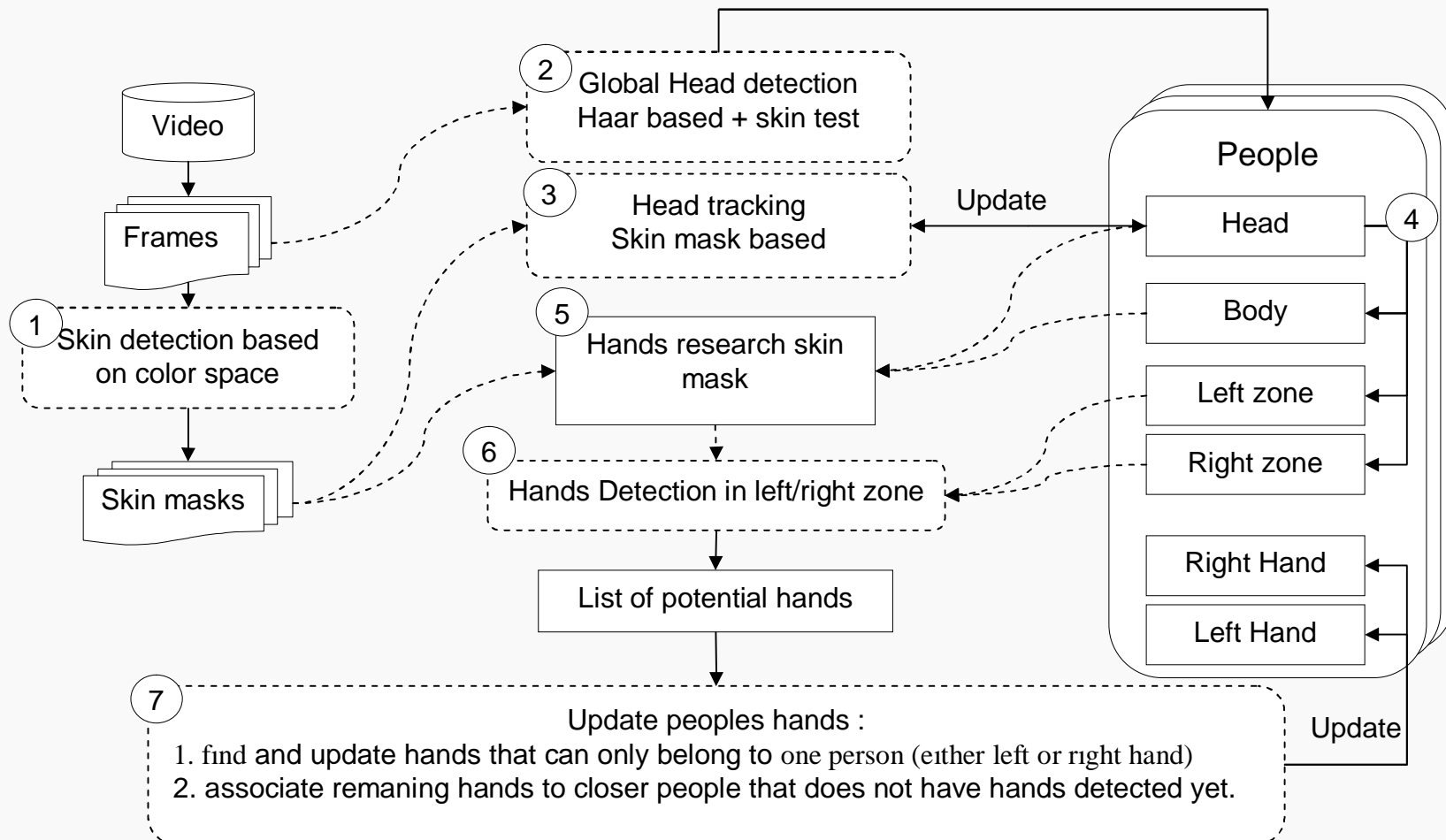


eNTERFACE '07

The SIMILAR NoE  
Summer Workshop  
on Multimodal Interfaces



Add new detected people





# Test and results

§ Single-user ([OneGoodCondition.xavi](#))

§ Multiple-user ([outputWithMax.xavi](#))

§ Bad lighting condition  
([outputVeryBadConditions.xavi](#),  
[outputMoreLight.xavi](#))



# Conclusions

- § Basic services can be implemented with high level of reliability
- § Integration becomes a complicated task:  
Expect NGN to ease integration
- § Research will also have to focus on building human-computer interaction techniques to compensate for not 100% reliable interfaces (NLP, AI, etc.)

# Any questions?

eNTERFACE '07

The SIMILAR NoE  
Summer Workshop  
on Multimodal Interfaces



Advanced Multimedia Interfaces for Flexible Communications