Project 3
A Multimodal Framework for the Communication of the Disabled

Lale Akarun, Murat Saraçlar, Dimitrios Tzovaras

Oya Aran, Savvas Argyropoulos, Ismail Ari, Pavel Campr, Erinç Dikici, Marek Hruz, Deniz Kahramaner, Harun Karabalkan, Alexey Karpov, Byungjun Kwon, Konstantinos Moustakas, Siddika Parlak, Thanos Tsakiris, Giovanna Varni

Objective

§ Enable the communication of people with different kind of disabilities

§ Build a multimodal framework that combines visual, aural and haptic interaction with

  ú gesture, audio-visual and text recognition

  ú sign language recognition/synthesis

Challenges

Modality Replacement

§ Develop a cross-modal transformation framework

  ú Combine modalities from an individual

  ú Recognize the transmitted message

  ú Translate in a perceivable form by the impaired people

§ Explore correlation among modalities to enhance the perceivable information of an individual who cannot access all incoming modalities.

Modality Replacement
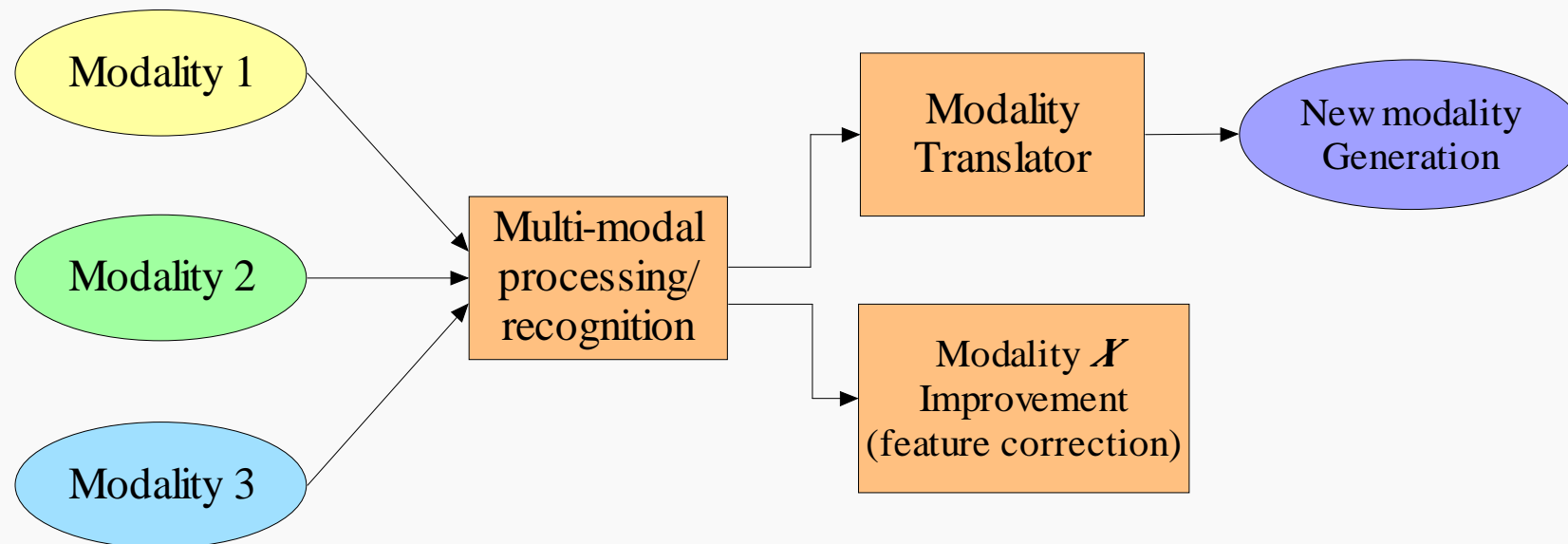
Applications

§ A collaborative VR game (Group 1)

§ Video news for the hearing impaired (Group 2)
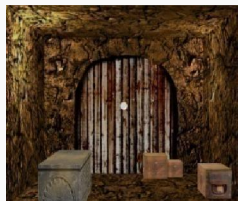
Application 1 – VR Game

The village

Step 1: Find the
red closet (blind)

Step 2: Go to the town
hall (deaf-mute)
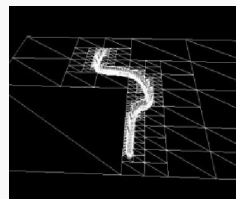
Step 5: "Go to the
catacombs" (blind)

Step 4: "The dead will
save the city" (deaf-mute)

Step 3: Go to the
temple ruins (blind)

Step 6: Sketch the path to
the treasure (deaf-mute)

Step 7a: Follow the path
to the forest (blind)

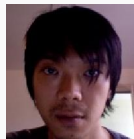Step 7b: Find the
treasure (blind)

Application 1 – VR Game

§ Game Setup

§ Modalities
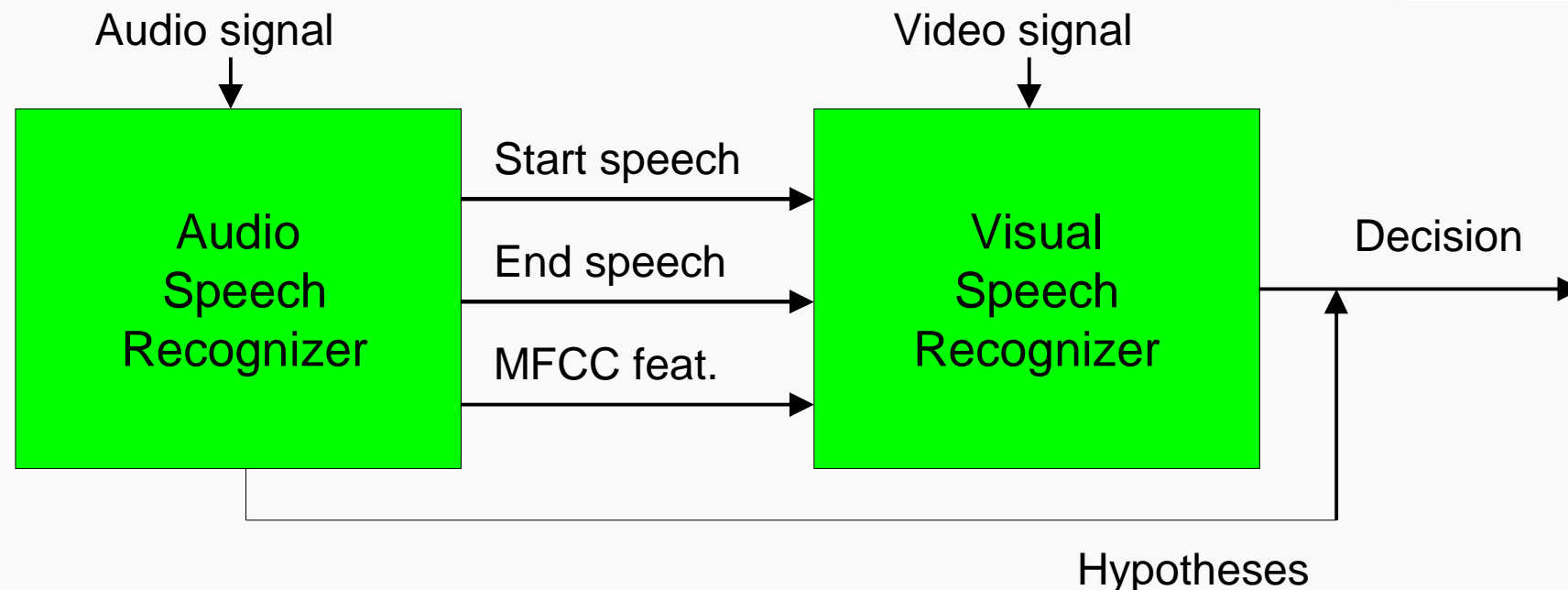
ú Speech

ú Lips/Fusion

ú Gesture

§ eNTERFACE 2007:

- Multimodal tools and interfaces for the communication between visually impaired and "deaf-and-mute" people

- Alternative tools and interfaces to disabled people to enable their communication and their interaction with the computer

Automatic Speech Recognition

Audio signal → **Audio Speech Recognizer**

Video signal → **Visual Speech Recognizer**

Audio Speech Recognizer → Start speech → Visual Speech Recognizer

Audio Speech Recognizer → End speech → Visual Speech Recognizer

Audio Speech Recognizer → MFCC feat. → Visual Speech Recognizer

Visual Speech Recognizer → Decision

Hypotheses

- § Speaker-dependent audio-visual speech recognition system
- § Phoneme-viseme based recognition
- § 12 MFCCs + 1-st derivative + 2-nd derivative are extracted from audio
- § HTK 3.4 Toolkit was used for audio speech recognition
- § Intel AVCSR system for lip motion tracking

Automatic Speech Recognition

§   Audio-visual speech database recorded by the web-camera Philips SPC900

§   Audio data format: 11025 Hz, 16-bit, mono, SNR ~ 15-30 db

§   Video data format: 640 x 480, 25 fps, RGB color

§   320 utterances for training the HMMs and 100 utterances for testing

§   30% of the speech training data were manually labeled on phonemes by WaveSurfer

§   16 voice commands in English to communicate with the game process and GUI interface (for a blind person)

Vocabulary of voice commands			
START-GAME	ENTER	NORTH	CATACOMBS
STOP-GAME	DOOR	SOUTH	EXIT
HELP	OPEN	EAST	INSCRIPTION
CLICK	GO	WEST	RESTART

Visual Feature Extraction

§ Visual feature extraction

$\rightarrow$ | Video Sequence | $\rightarrow$ | Face Detection | $\rightarrow$ | Mouth Region Extraction | $\rightarrow$ | 2-D DCT | $\rightarrow$ | LDA | $\rightarrow$ Feature Vector

Multimodal Fusion

Observable node

Audio

Hidden State
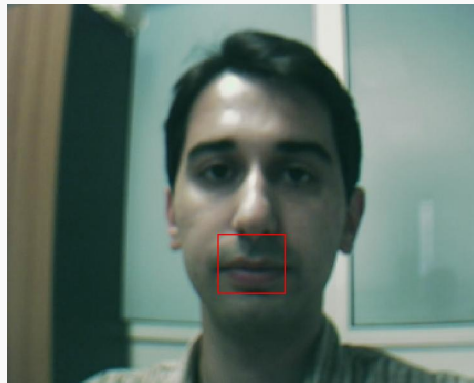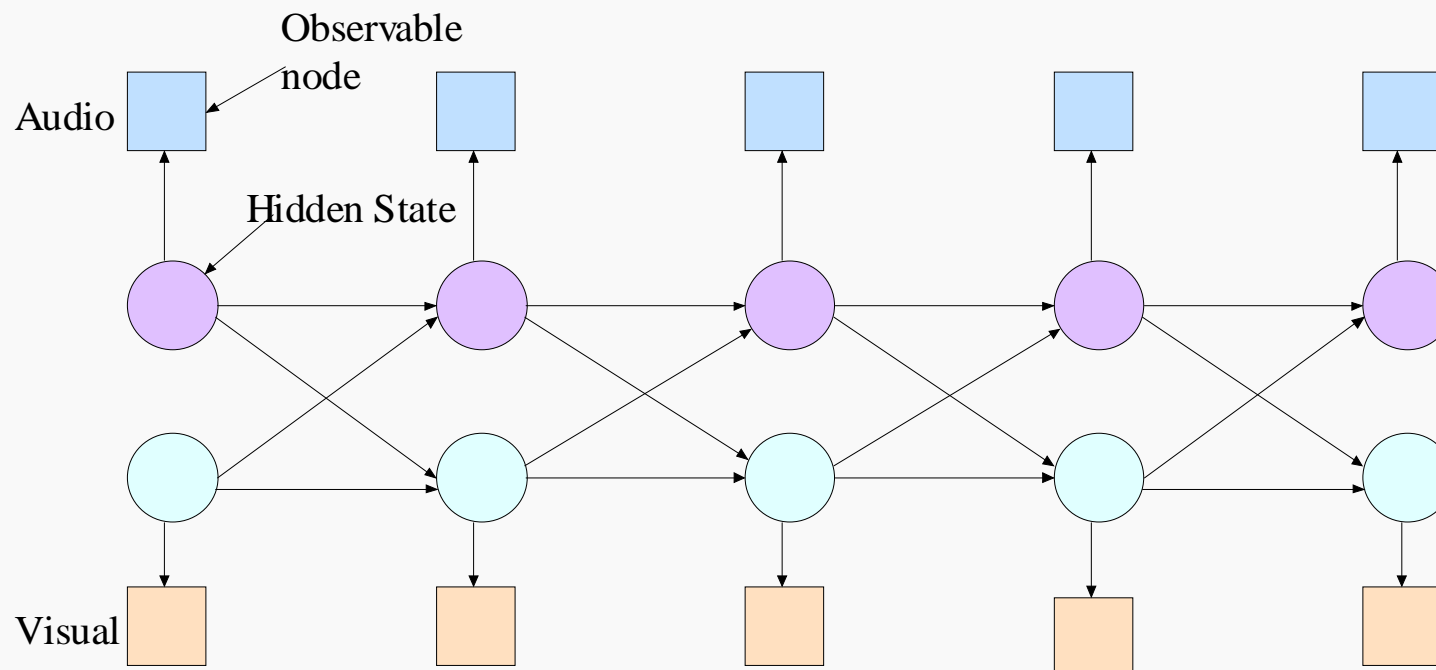
Visual

Coupled HMM for fusion of audio-visual information

Gesture Recognition

Gesture of deaf-mute player	→	Hands Tracking	→	Sketching of trajectories on the map
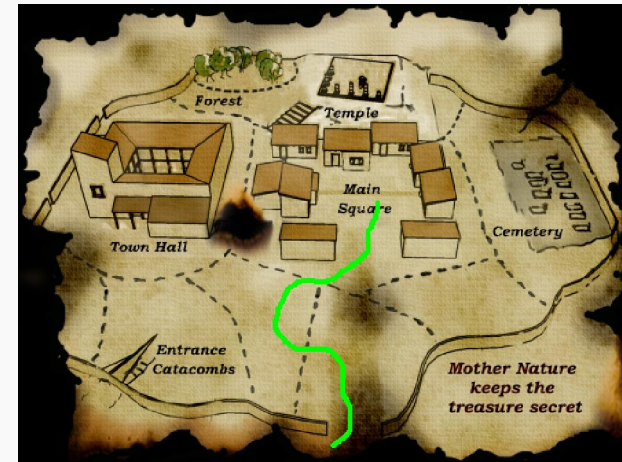
Gesture Recognition

§ Techniques:

   ú Color blob tracking (yellow and blue gloves)

   ú Kalman filter
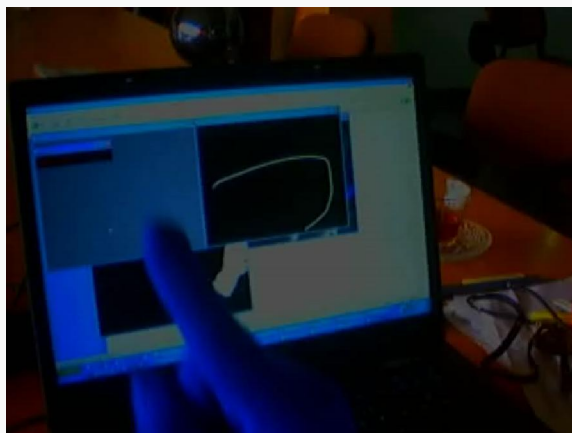
Sign Language Recognition
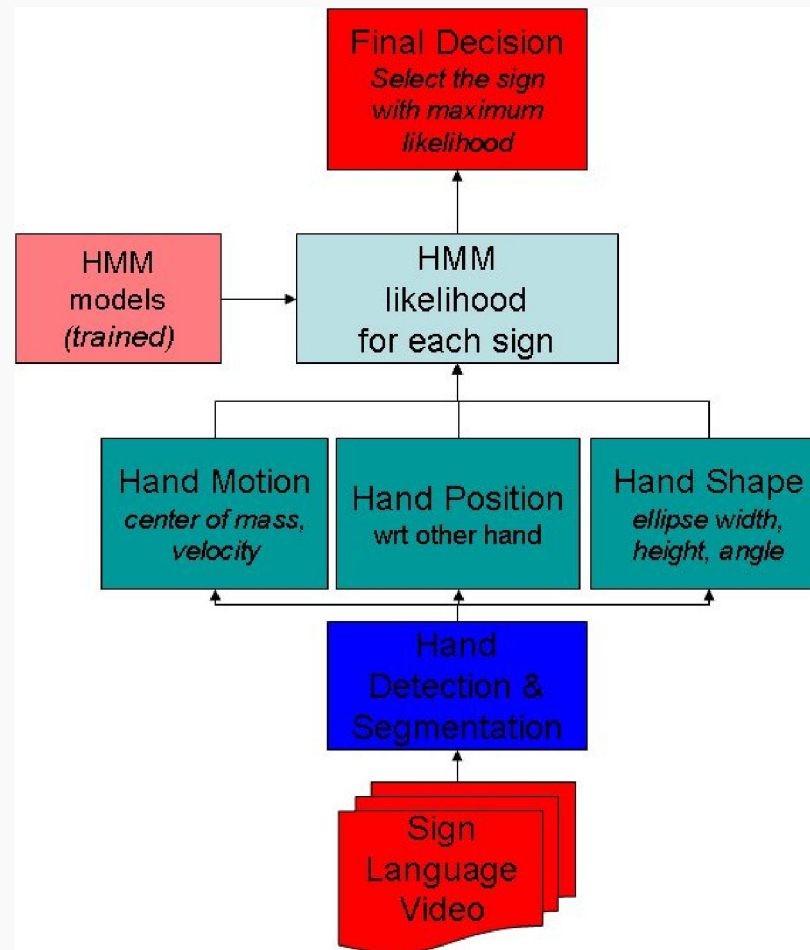
Future Work – VR Game

§ Evaluation of the developed unimodal modules

§ Multimodal fusion of audio-visual signals

   ú Improve recognition rate using cHMMs

§ Integrate gestures and audio-visual information

§ Integration with the VR Game