Interaction through natural language

(AC6030)

Mac Dang Khoa
Eric Castelli
TRAN Do Dat

SpeechCom department

International Research Institute MICA
Multimedia, Information, Communication & Applications
UMI 2954

Hanoi University of Science and Technology
1 Dai Co Viet - Hanoi - Vietnam

Course overview

International Research Institute MICA
Multimedia, Information, Communication & Applications
UMI 2954

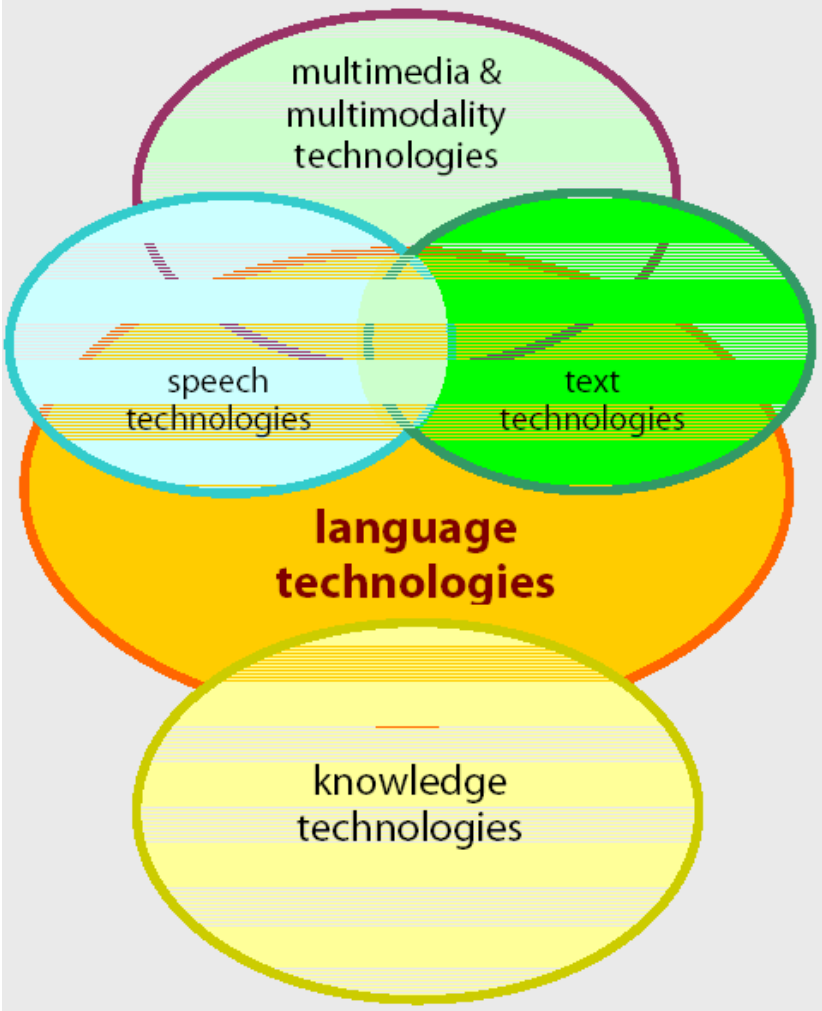
Hanoi University of Science and Technology
1 Dai Co Viet - Hanoi - Vietnam

Course information

- **Name: Interaction through natural language**
- **ID: AC6030**
- **Credits: 3**
- **Time: 9 lessons (9 weeks)**
- **Friday 13h30 – 17h00. Flexible**
- **Lecturers:**
  - Dr. Mac Dang Khoa (http://www.mica.edu.vn/perso/Mac-Dang-Khoa)
  - Pr. Eric Castelli (http://mica.edu.vn/perso/Eric-Castelli)
  - Pr. Pascal Nocera (Université d'Avignon, France)
  - Assoc. Pr. Tran Do Dat
- **Course webpage:**
  - http://www.mica.edu.vn/perso/Mac-Dang-Khoa/AC6030

Course objectives

- **Know**
  - Concepts: What?
  - Application: For what?
- **Understand**
  - Methods: How?
  - Techniques
- **Do**
  - Research
  - Practicing
  - Writing
  - Presentation



Interaction through natural language

Preliminaries required

- **Digital signal processing**
- **Interests in languages and linguistics**
- **Skills**
  - Document researching and management
  - Scientific writing
  - Software development/implement
  - Presentation

Evaluation

- **No examination**
- **Grading**

Category	Weight	Note
Answering during lecture	0.1	Average of 3 best answer
Mini presentations	0.3	Average of 3 best answer
Project	0.6	Topics given in week 5
- Report	0.2	
- Implement + Demo	0.2	
- Presentation	0.2	Week 9

References

- Varile, Giovanni Battista, and Antonio Zampolli. ***Survey of the state of the art in human language technology***. Vol. 13. Cambridge University Press, 1997.
- Benesty, Jacob, M. Mohan Sondhi, and Yiteng Huang, eds. ***Springer handbook of speech processing***. Springer Science & Business Media, 2007.
- Daniel Jurafsky, James H. Marti "***Speech and Language Processing***", Pearson Prentice Hall, 2009, xxxi+988 pp; ISBN 978-0-13-187321-6.

Report writing

- **References**
  - Searching
  - Management
  - Citation
  - Style : Springer (Author-Date)
- **Reports**
- **Presentation**
  - MICA templates

Course structure

- **9 weeks**
  - W1: Introduction
  - W2-3: Speech signal representation and analysis
  - W4-5: Speech synthesis and production
  - W5: Automatic speech recognition
  - W6: Speech understanding
  - W7-8: Advance topics on SPL
    - Spoken dialog system
    - Multilingual technology and speech translation
    - Expressive speech
    - Speech database
    - Deep learning in SLP
    - Speech Prosody
    - SLP for Under-resourced languages
  - W9: Student project final presentation

Introduction

Human-Machine Interaction through natural language

International Research Institute MICA
Multimedia, Information, Communication & Applications
UMI 2954

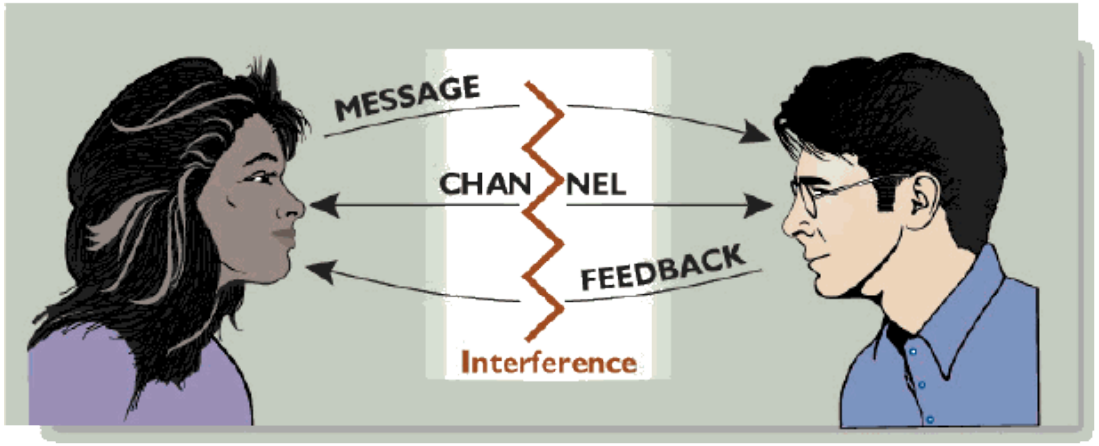
Hanoi University of Science and Technology
1 Dai Co Viet - Hanoi - Vietnam

Content

- ***Speech communication***
- ***Human machine interaction (HMI)***
  - Conventional HMI
  - Challenges
  - New HMI approaches
  - Voice based HMI
- **Applications using voiced based HMI**
- **Development of Voiced based HMI systems in Vietnam**

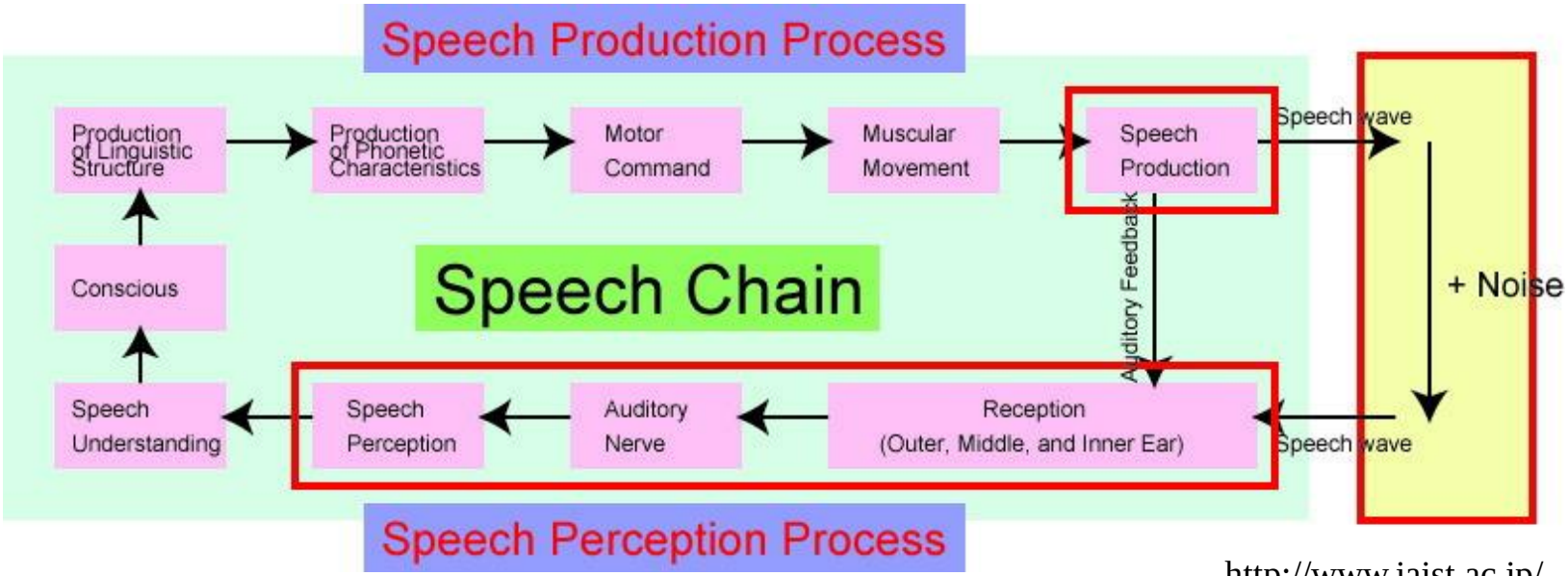
Human communication/interaction

- Face-to-face interaction



Gestures/facing/body languages

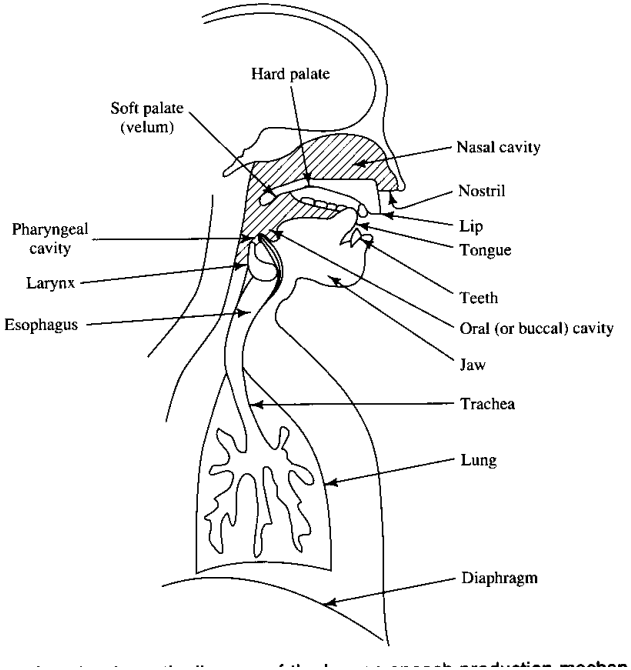
Speech communication



http://www.jaist.ac.jp/

Speech production

- **Human**
  - Blow air: Lung => Larynx =>Vocal tract => Tongue/teeth, lips

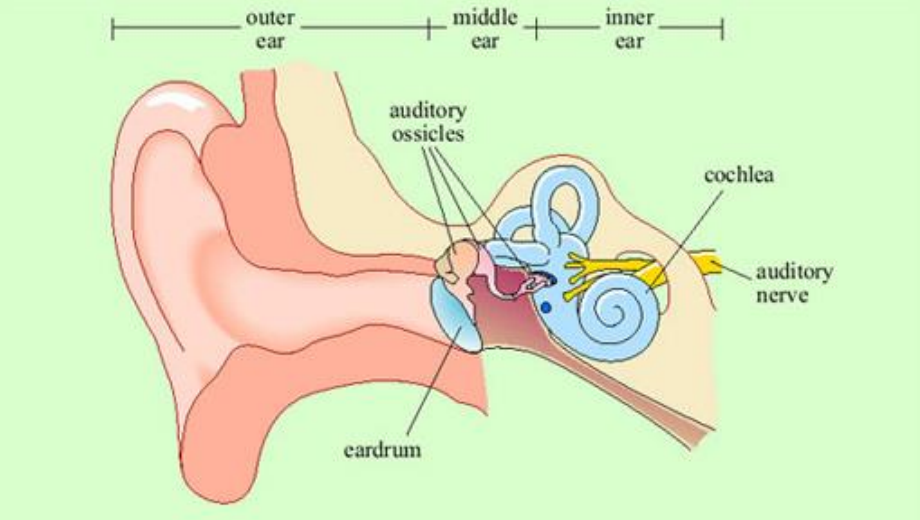


- **Computer**
  - Text to Speech (TTS)

Describe human speech production

Speech perception

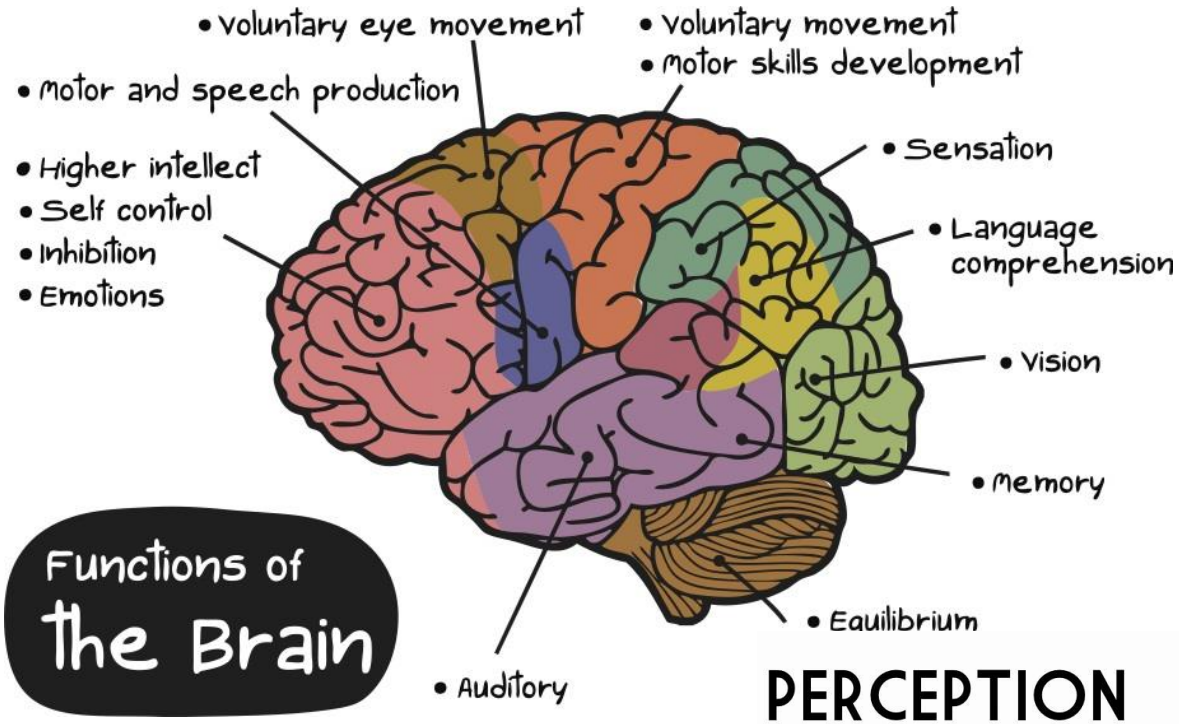
- **Human**
  - Sound wave: eardrum => cochlear => auditory nerve: frequency

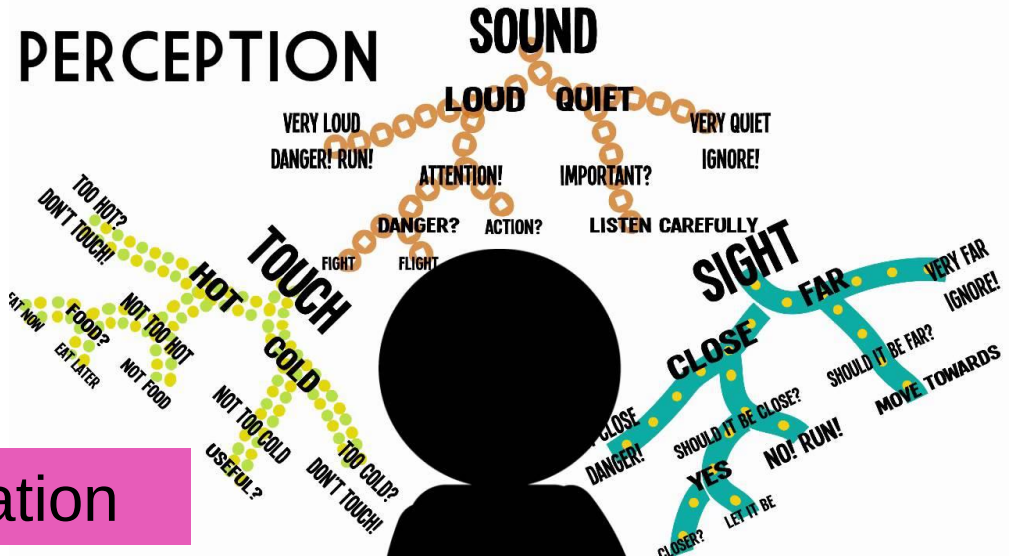


- **Computer**
  - Speech recognition

Describe human speech perception

Human speech understanding





How our brain process information

Content

- ***Speech communication***
- ***Human machine interaction (HMI)***
  - Conventional HMI
  - Challenges
  - New HMI approaches
  - Voice based HMI
- **Applications using voiced based HMI**
- **Development of Voiced based HMI systems in Vietnam**

The 1st machine

- **Machine vs Tool**

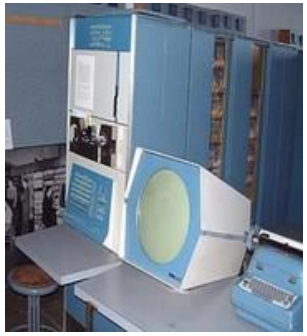
2m years BC

100 years BC

Computer

- 1940s
  - Switch
- 1950s
  - Transitor
  - FORTRAN, COBOL
- 1965
  - IC: Integrated Circuit
- 1972
  - Microprocessor
  - Keyboard

ENIAC (Electronic Numerical Integrator And Computer)



Conventional HMI methods

Inputs

- Button, control board
- Mouse
- Keyboard – commands

Outputs:

- Speakers: sound, speech
- LED board, desktop screen, TV.. : GUI

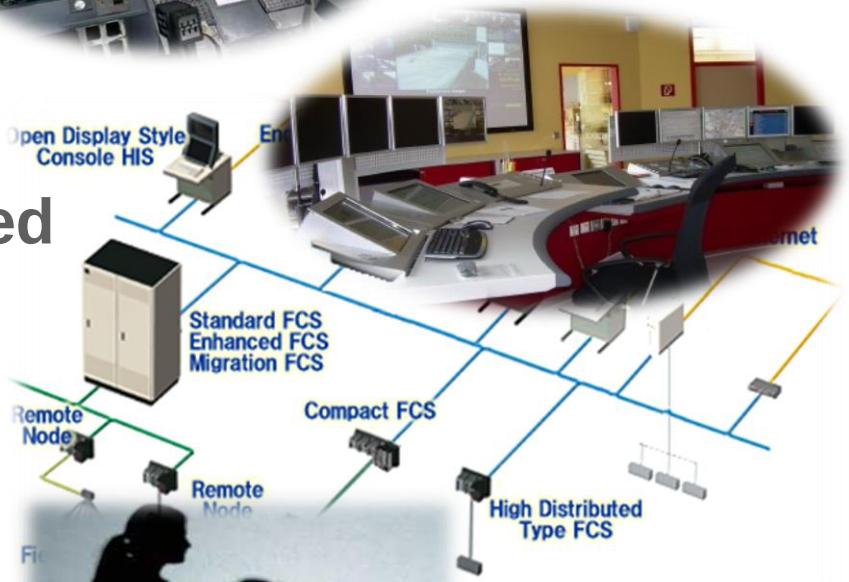
Challenges with conventional HMI methods

High speed command:

- Military
- Rescue and disaster

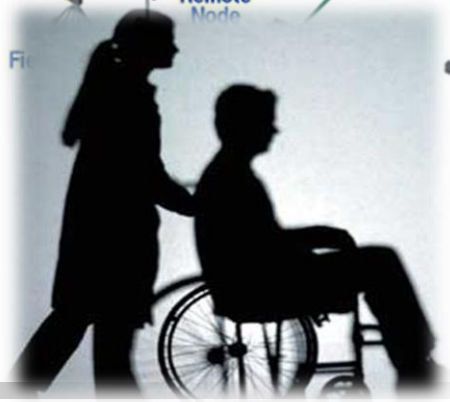
Many control parameters:

- Control center



Hand are busy or handicapped

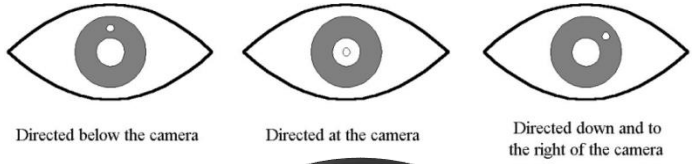
- Driving
- Handicapped



Voice based HMI
New HMI approaches

Brain computer interaction

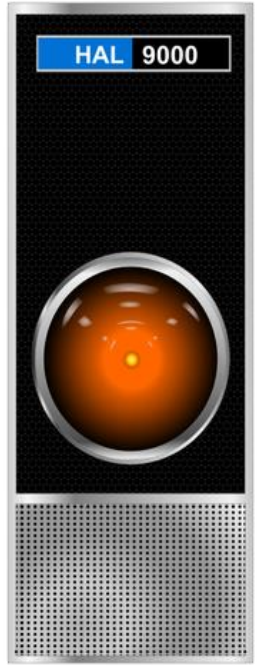
Eye movement based HCI

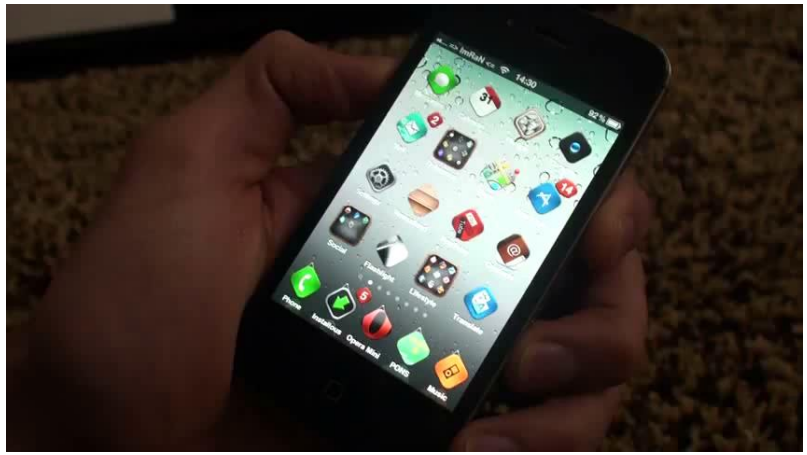


- Interaction by gestures

Human-Machine Speech interaction

- **From HAL 9000 (science fiction)**
  - **H**euristically programmed **AL**gorithmic Computer
  - 2001 A Space Odyssey - 1968 :
- **To Apple Siri (2011)**





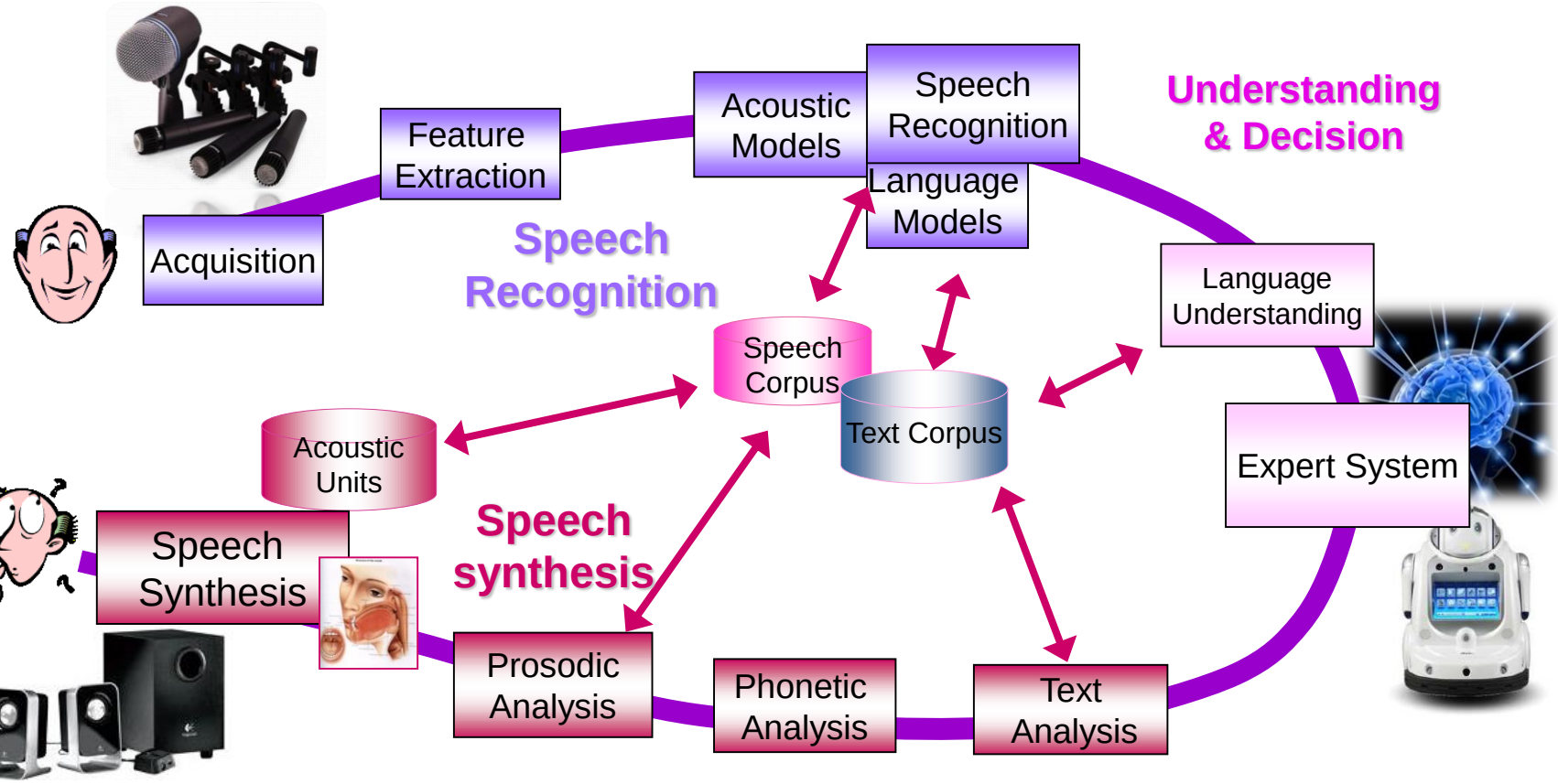
Voice based human machine interaction

Asimo

- Industrial Robot

- Daily life

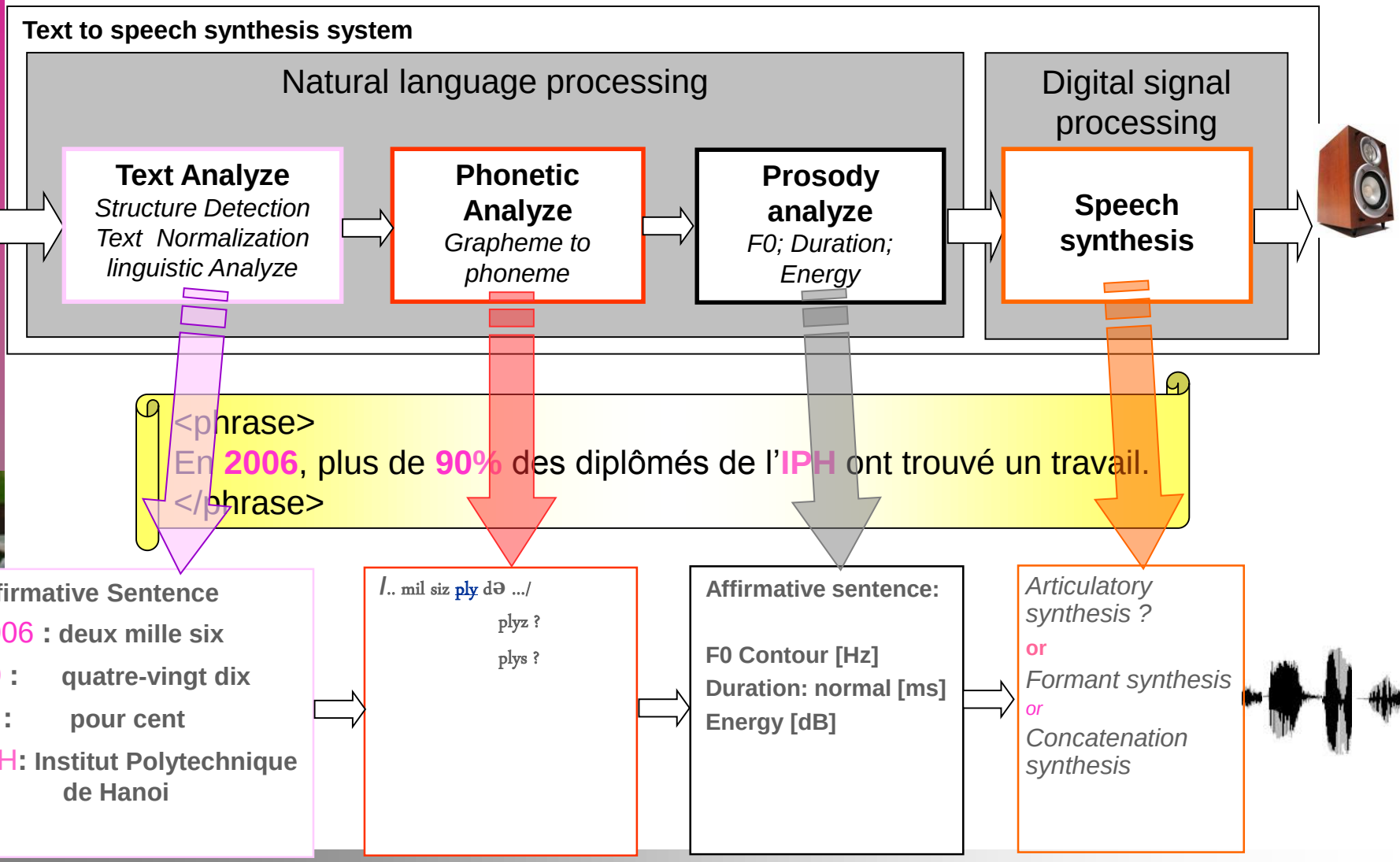
Voice based human machine interaction



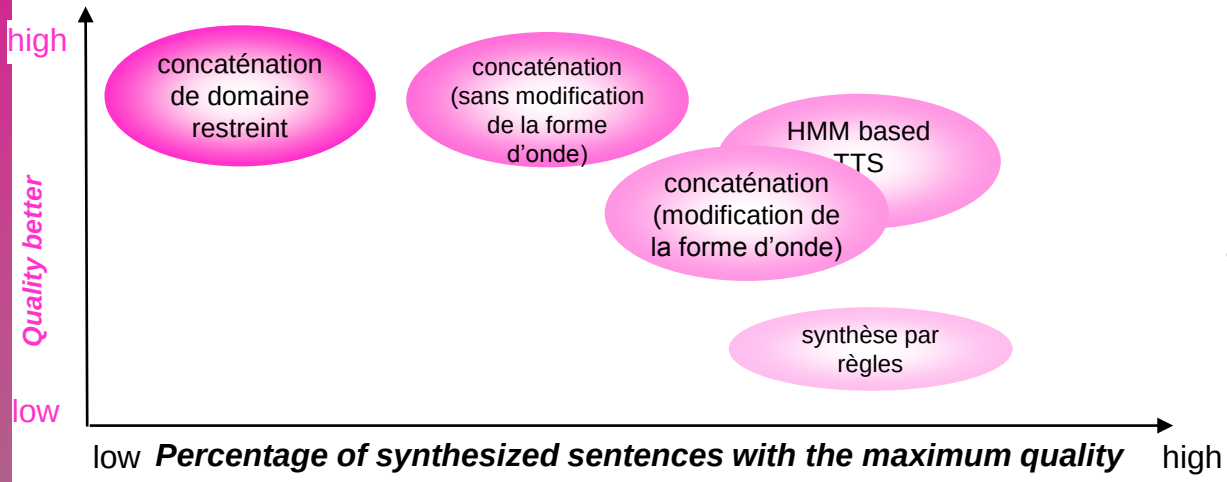
Speech is the best communication medium of human society

Voice based HMI

Speech synthesis

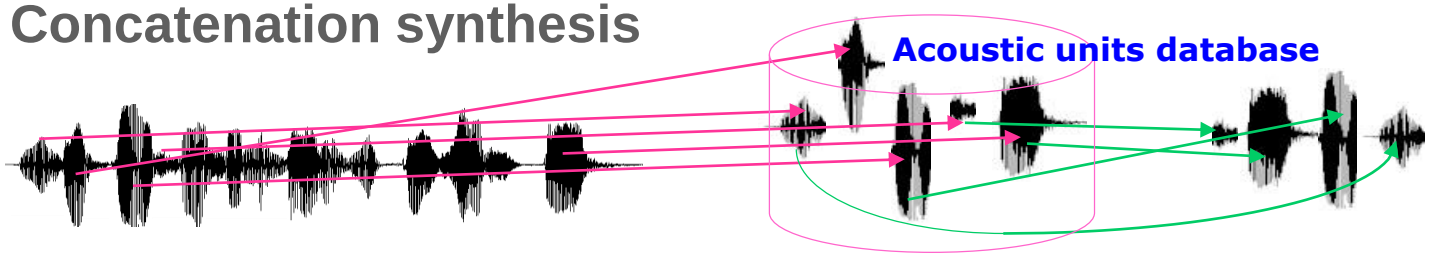


Voice based HMI
Speech synthesis



Quality and performance of speech synthesis techniques

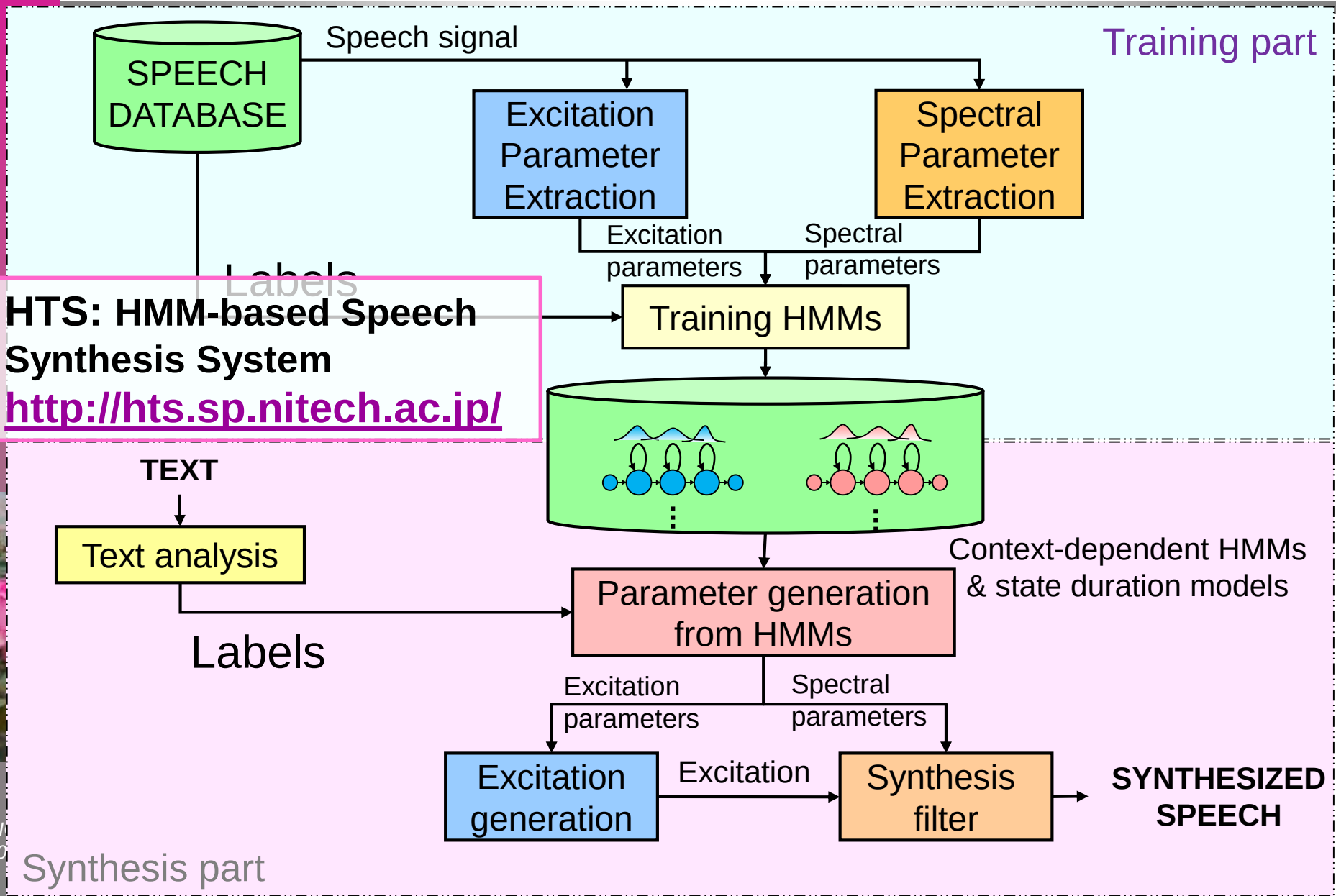
- **Concatenation synthesis**



- **PSOLA (Pitch-Synchronous OverLap-Add) (France Télécom - [Charpentier 1986])**
  - *TD-PSOLA, FD-PSOLA, LP-PSOLA* [Moulines 1990]
  - *MBR- PSOLA Multiband Resynthesis PSOLA* [Dutoit 1993]

Voice based HMI

Speech synthesis - HMM



Speech recognition

Information in speech?

Where is he/she from?

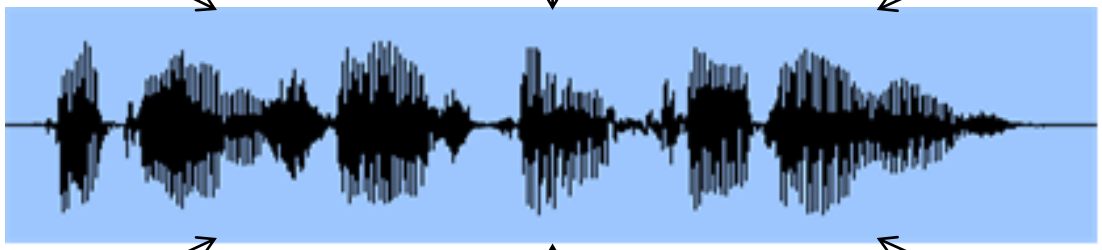
What language was spoken?

What was spoken?

Accent Recognition

Language Recognition

Speech Recognition



Emotion Recognition

Gender Recognition

Speaker Recognition

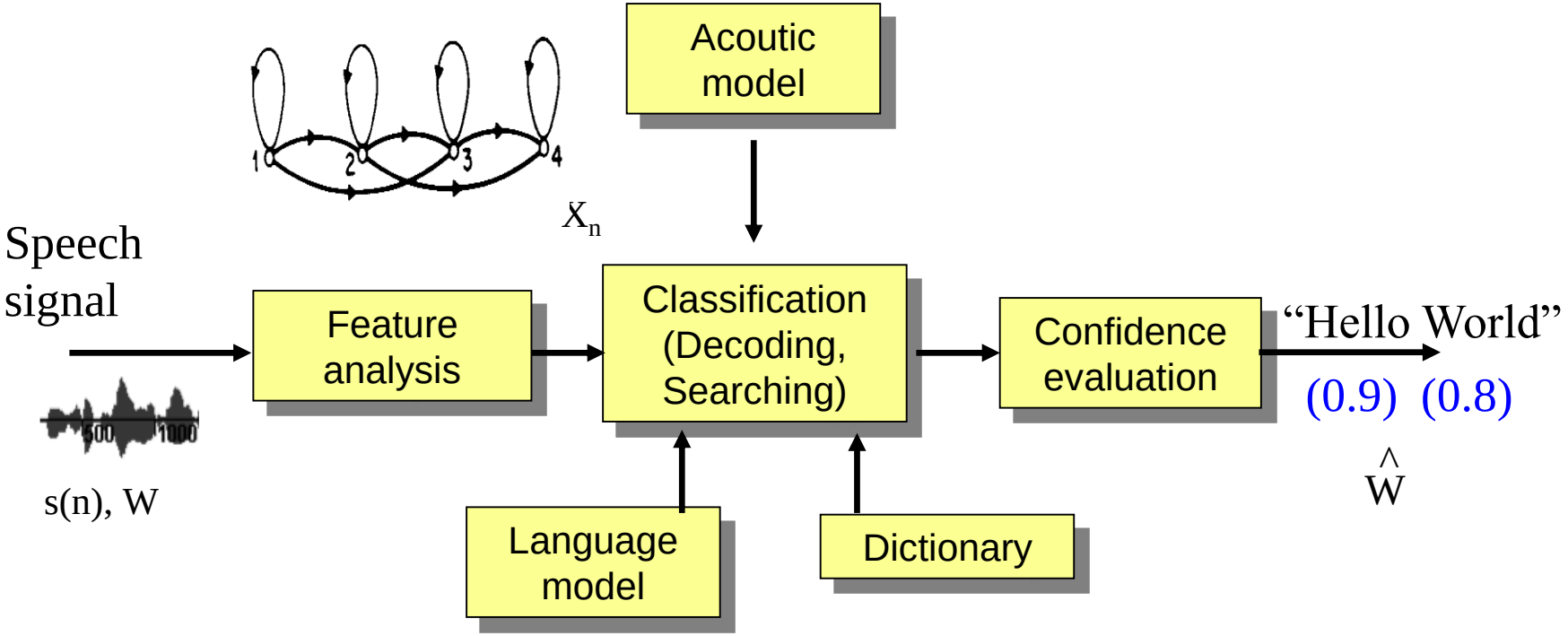
Positive? Negative?
Happy? Sad?

Male or Female?

Who spoke?

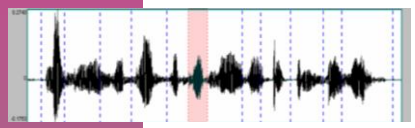
Voice based HMI

Speech Recognition



- **HMM based automatic speech recognition**
  - CMUSpinx toolkit http://cmusphinx.sourceforge.net/)
  - RASR : http://www-i6.informatik.rwth-aachen.de/rwth-asr/
  - HTK toolkit : http://htk.eng.cam.ac.uk/
  - Kaldi: http://kaldi-asr.org/
- **Artificial Neural Network (ANN) based ASR**

Voice based HMI
Speech Understanding



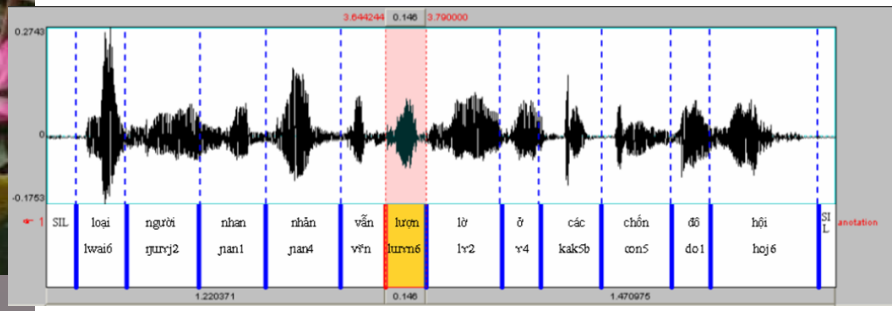
voice

Speech Recognition

Alignment text

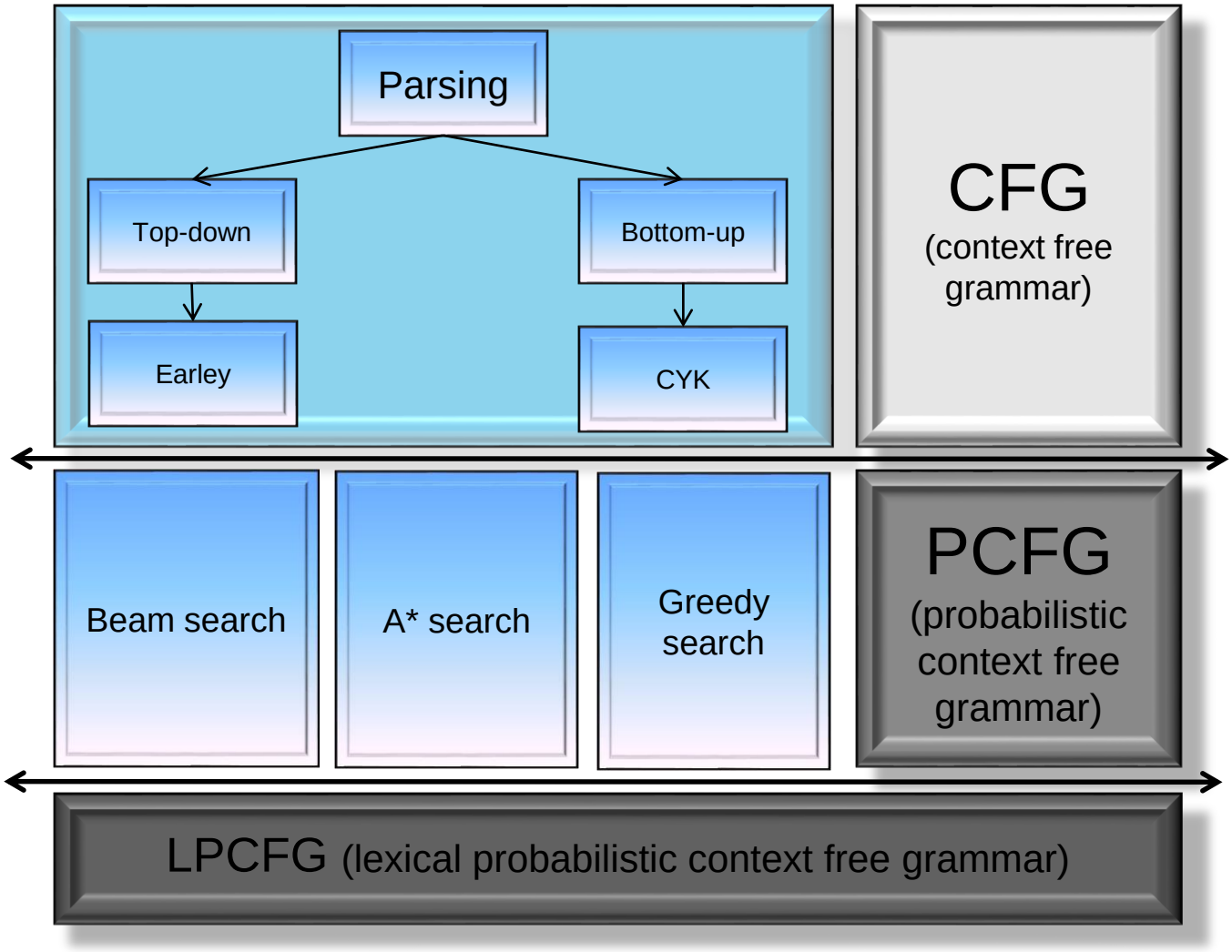
Speech Understanding

Speech Synthesis



Speech Understanding
Structure analysis

- Text parsing:



Representation of meaning

- Extract text information
  - Thematic roles

Thematic Role	Definition
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of an event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	A beneficiary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

Content

- Voiced based human machine interaction (HMI)
- *Applications using voiced based HMI*
- Development of Voiced based HMI systems in Vietnam

Speech based HMI application

- **Command and Control**
  - Automobile, Smarthome
- **Virtual agents (Assitants)**
  - Smartphone: Siri, Google Now
  - Virtual singer: Hatsune Miku
- **Speech to Speech Translation**
- **Transcription/Dictation**
  - MS Words
  - Nuance Dragon Dictation
- **Speaker Identification**
  - emotion/dialect/language
- **Language Learning**

Applications domain

Military domain

- Setting parameters in fighter aircraft [Weinstein 1990], [Englund 2004] (France, UK, USA, Russia Military).
  - *radio frequency,*
  - *steering-point coordinates*
  - *and weapons release parameters*

- Training system using voice user interface for interaction.

Applications using voiced based HMI

Industrial domain:

- Industrial robot controlled by voice [Pires 2005]
- Smart robot Asimo (Honda), Aichi robot (Kokoro)
- Voice user interface has been operated in cars (BMW, Honda, Ford) …:
  - *Navigation, setting functions…*

Personal assistant

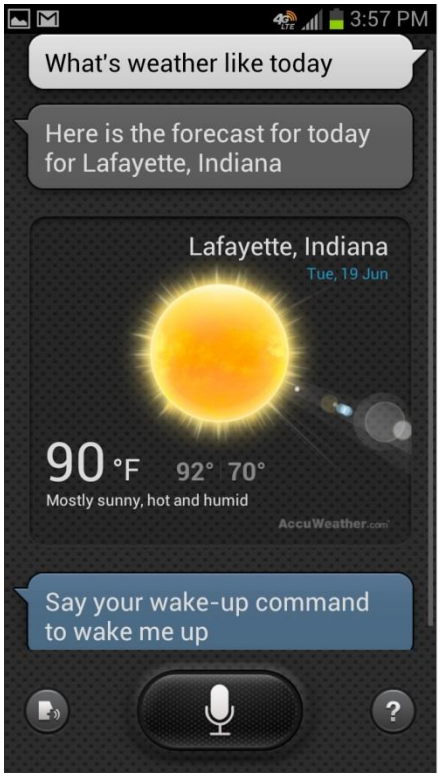
- Read out Web content, email on PC
- Used for Interactive Voice Response systems
- Used for information equipments, home electric appliances,

Usages of speech synthesis technology

Smartphone

SIRI

English, French, German, Japanese, Chinese, Korean, Italian, Spanish



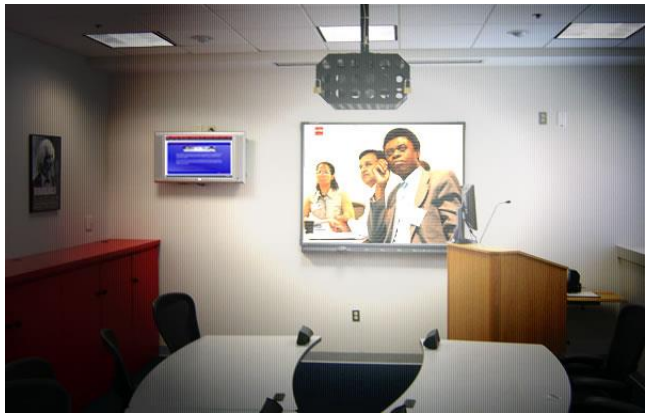
S-VOICE

English, Arabic, French, Spanish, Korean, Italian, German

- **users use their voice to :**
  - send messages,
  - schedule meetings,
  - place phone calls, and more.

“Hot” domain

- Smart home controled by voice
  http://www.smarthomeusa.com/Shop/Voice/
  - ★ Electric and electronic equipments controled by voice
- Speech to speech translation, video game, voice search
- Voice based controle system for handicaped



Content

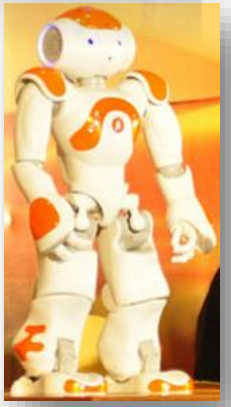
- Voiced based human machine interaction (HMI)
- Applications using voiced based HMI
- *Development of Voiced based HMI systems in Vietnam*

Development of Voiced based HMI systems in Vietnam

Robot interaction:

- Smart robot: guiding robot in Vietnam Museum of Ethnology (developped by MICA institute, HUST):
  - Interacted by voice
  - and gestures

- Home Robot controled by voice (IOIT – Vietnam)
- Smart robot - Smartoshin: (FPT) talks 3 languages



Development of Voiced based HMI systems in Vietnam

Vietnamese

SIRI

English, French, German, Japanese, Chinese, Korean, Italian, Spanish

iSago

Vietnamese

VIVA

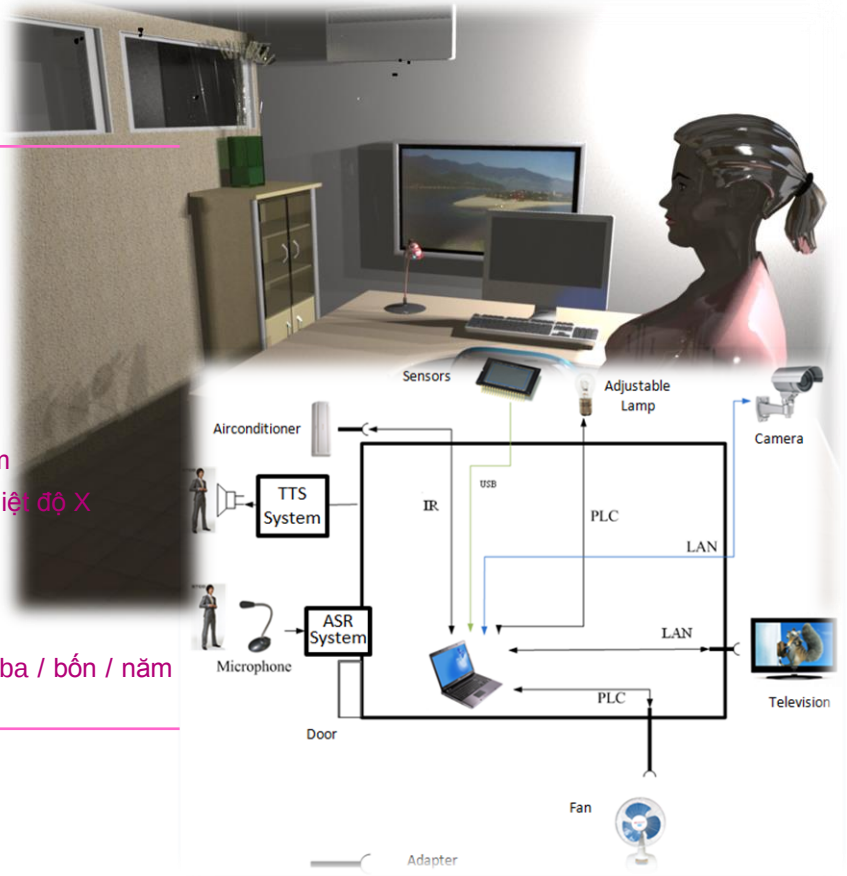
Vietnamese Voice Assistant
Developped by MICA Institute

Development of Voiced based HMI systems in Vietnam

DRONE controlled by voice
(in cooperation with PSI)

- **Intelligent room controlled by voice (KC03.15)**

	Commands
Door	cửa mở / đóng
Lamp	đèn bật / tắt
	đèn sáng lên / tối đi
Camera	camera bật / tắt
	camera phải / trái
Airconditioner	điều hòa bật / tắt
	điều hòa tăng / giảm
	điều hòa thiết lập nhiệt độ X
Television	tivi bật / tắt
	tivi tăng / giảm âm
	tivi tăng / giảm kênh
	tivi kênh một / hai / ba / bốn / năm
	…



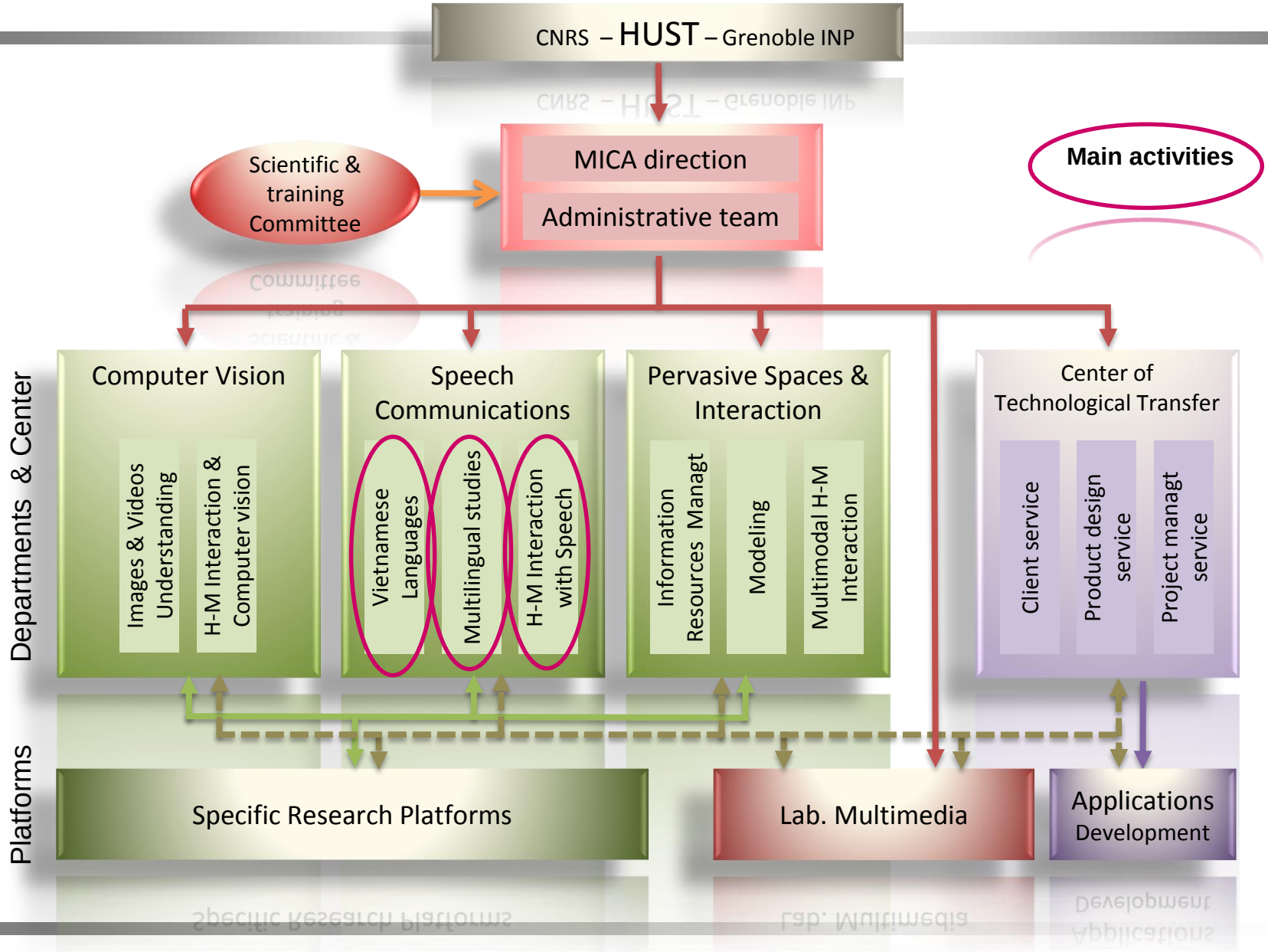
Speech Comunication Dept.

International Research Institute MICA

Multimedia, Information, Communication & Applications
UMI 2954

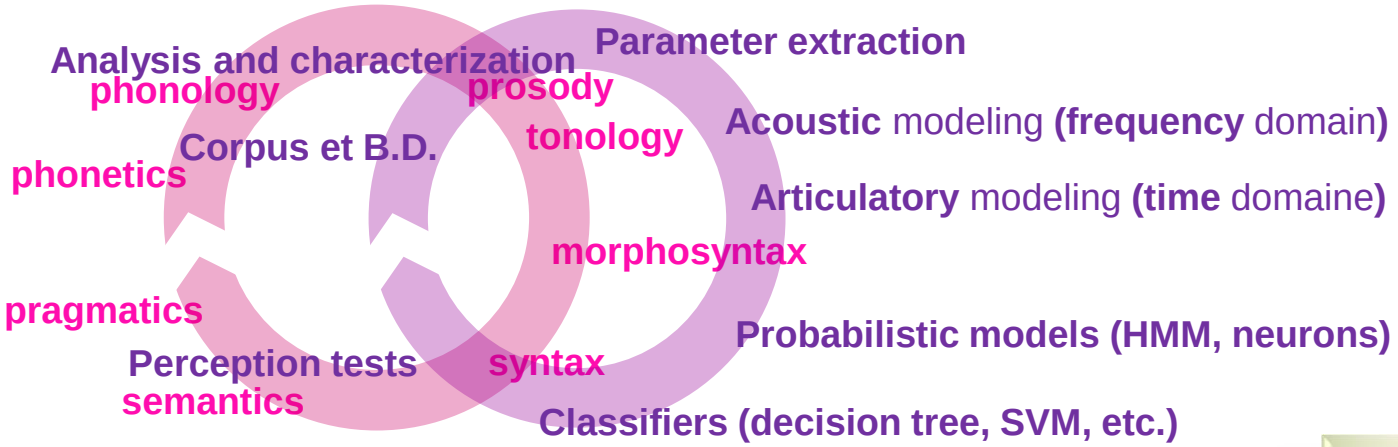
Hanoi University of Science and Technology
1 Dai Co Viet - Hanoi - Vietnam

MICA Institute structure

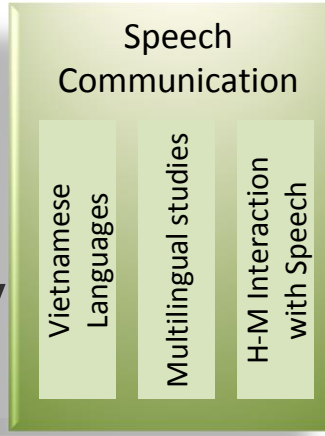


Research Direction

- **Two research activities are realized simultaneously:**
  - natural language processing
  - vocal technologies.

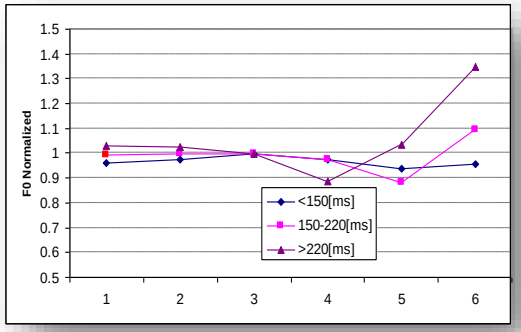


- **Research axes**
  - languages spoken in Vietnam
  - multilingual technology
  - voice-based human-machine interaction technology

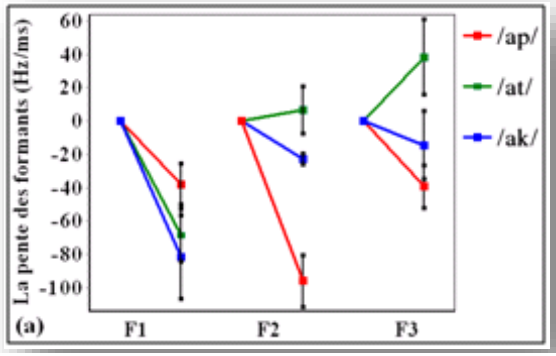


Research Direction

- **Study of languages in Vietnam**
  - Studying phonology, phonetics, prosody, and other linguistic topics
  - Modeling dynamic characteristics of the acoustic parameters
  - Developing research tools and methods to study and to transfer knowledge to under-resourced languages (Mopiu, Khmer, Lao…)
  - *Constructing speech databases*
  - Studying text processing, text parsing, etc.



Broken tone (Tone 3) in dynamic mode



Comparing formant transition slopes in VC sequences

Building Mopiu corpus

Research Direction

- **Vocal technologies & human - machine interaction by voice**
  - Development of speech technologies
    - speech synthesis
    - recognition and understanding
  - Development of applications on embedded systems
    - smartphones
    - DSP-based systems
    - service robots
  - Improving the quality of human-computer interaction systems *in cooperation with PSI, Multimedia lab*
    - Human-Machine Interaction by voice

Hoa Sen (recognition)

Hoa Sung (synthesis)

"Tôi tên là Mây"

"Vâng tôi múa đây"

"Robot tên là gì ?"

"Robot múa đi !"

Research Direction

- **Multilingual studies**
  - **Multilingual text corpus:** constructing bilingual dictionaries, bilingual parallel corpora, multilingual parallel corpora, etc.
  - **Machine Translation systems** for pairs of 'under-resourced' language and 'well-endowed' language (statistical approach)
  - **Speech translation systems** in combination with ASR system and TTS system
  - **Multilingual recognition systems**
    - ★ Adaptation of an ASR system from one language to others
    - ★ Building one multilingual model to recognize many languages

Parallel fragment between two articles

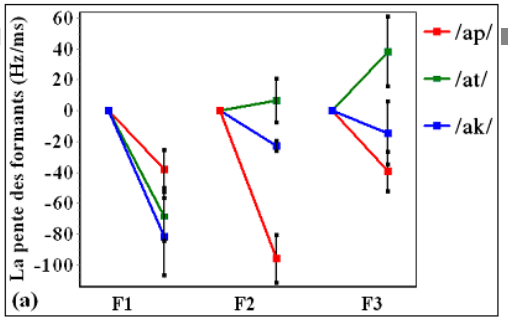
Vietnamese
Thai
Khmer
Lao
etc.

English
French
etc.

Main results

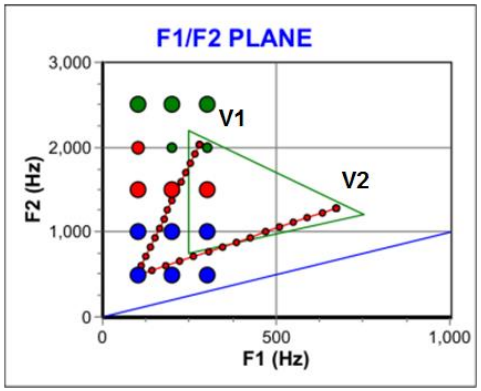
Study of languages in Vietnam

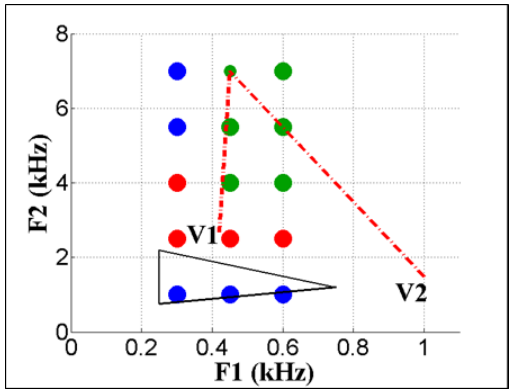
- **Phonetics phonology of Vietnamese language:** Analyze and understand dynamic characteristics
  - Consonants /p, t, k/ can be distinguished by formant transition slope (*dynamic characteristics*)



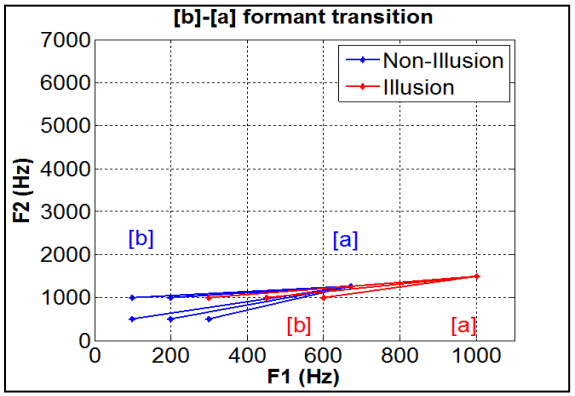
Formant transition slope of VC

[NGUYEN Viet-Son 2009 PhD]





Non-illusion & illusion experiments



Formant transition distribution tendency

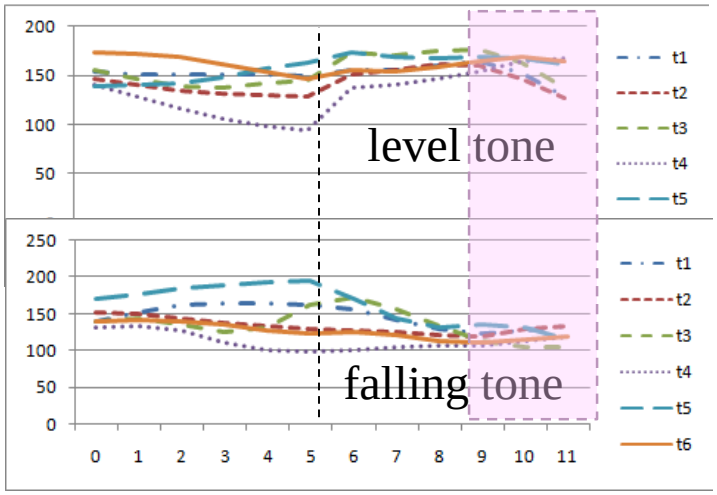
  - Consonants /ɓ, ɗ, ɣ/ can always be recognized both inside & outside the vocalic triangle
  - Relative formant variations & formant transition rate (*dynamic values)* play a significant role (Interspeech 2011, 2013, IALP 2012)

Main results

Study of languages in Vietnam

- **Tonology of Vietnamese language:** tonal co-articulation on Vietnamese particles (IALP 2012, WSSANLP2012, Interspeech2013)
- **Phonetics and tonology of Mopiu minority language** (SLTU 2010, IALP 2011, SpeechProsody 2011, LREC 2012, Interspeech 2013)

Influence of the function of particles on form of F0

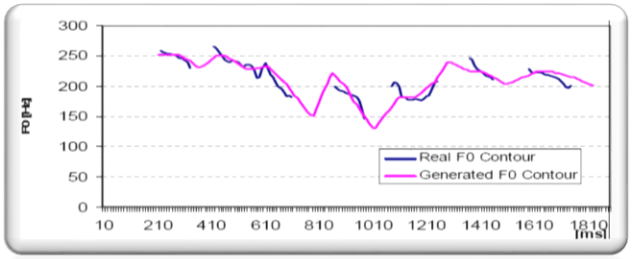


F0 contours of tone pairs (particle carries level/falling tone) with male voice

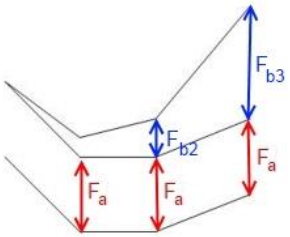
MISTRAL+ a tool for the analysis of expressive speech

Main results

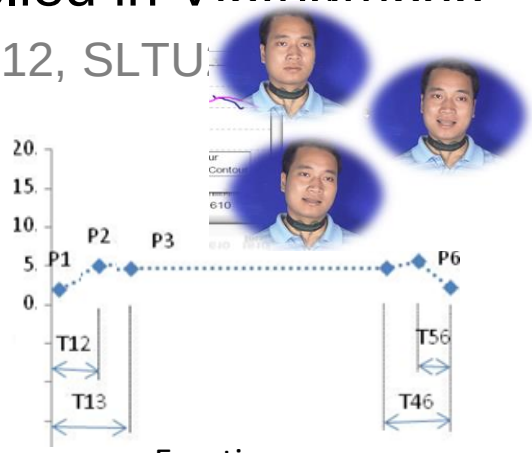
- **Synthesis for Vietnamese**
  - Model for generation of Vietnamese intonation (affirmative and question sentences + emotion), applied in Vietnamese TTS (SoICT2011, IALP2011, SpeechProsody2012, SLTU



Model for generation of Vietnamese intonation [TRAN 2007 - Thesis]

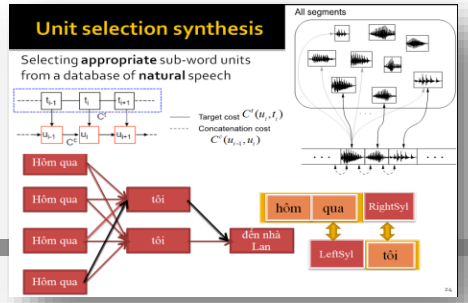


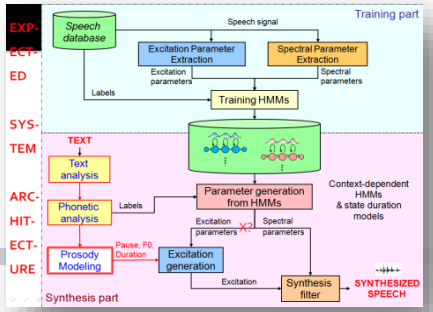
+ question sentence [TRAN 2011]



+ Emotion [MAC 2012 - Thesis]

  - Vietnamese TTS based on non-uniform unit concatenation or HMM model (SoICT2011, Interspeech2013)

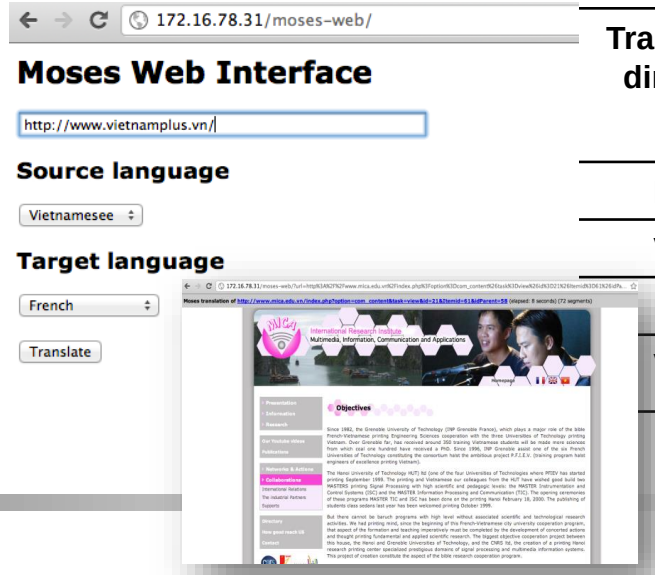




Main results

- **Speech recognition for Vietnamese, Khmer and Lao**
  - Robust Vietnamese ASR (SEATUC 2011, ICCE2010-2012, SoICT2012).
  - ASR for Khmer language (Interspeech2010, SLTU2010),
- **Automatic translation (VN-FR & VN-EN)**
  - Probabilistic text translation system (VN-FR & VN-EN) (TALN 2010, EAMT 2010, IWSLT 2010, IALP2011) [DO T.N.D 2011 Thesis]
- **Databases:**
  - Vietnamese, Mopiu (audio, video), Khmer

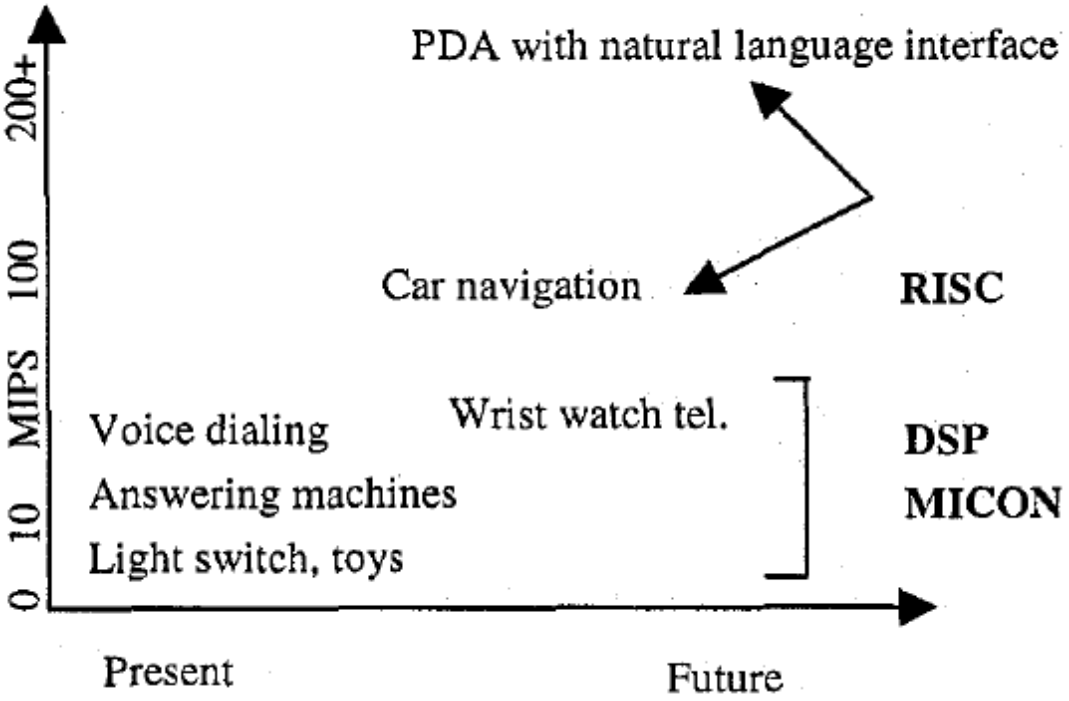
Human-Robot interaction by voice in Vietnam Museum of Ethnology



Translation direction	BLUEScore (%)	
	Our system	Google 02/2009
F→V	40,09	24,82
V→F	32,08	15,63
F→V	25,00	24,38
V→F	20,22	15,82

Một số bài toán
Nhận dạng tiếng nói

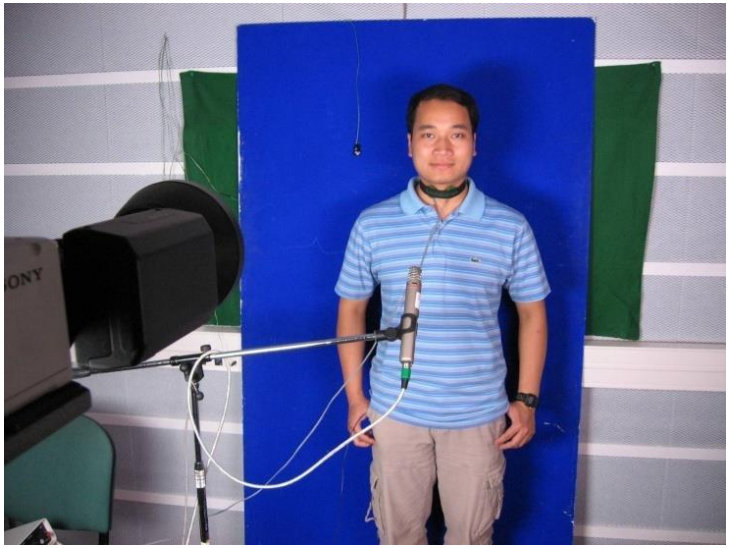
- *Kiến trúc hạ tầng của ứng dụng*



Phân loại các ứng dụng nhúng theo chủng loại bộ xử lý (Vi điều khiển, DSP, hay Vi xử lý (RISC)) và tốc độ xử lý.

Study of languages in Vietnam
Corpus (example)

- **Attitude/emotion corpus (audio+visual)**
  - 90 mn audio-visual signal
  - 2 speakers
    - 1 woman, 1 man



Maternel

Autorité

Séduction

Study of languages in Vietnam
Under-resourced languages

H-Mong minority

- **Aims: Collect data about ethnologic subjects**
  - Origins, technology, religion, **language**, social structure, ...

Mo Piu minority

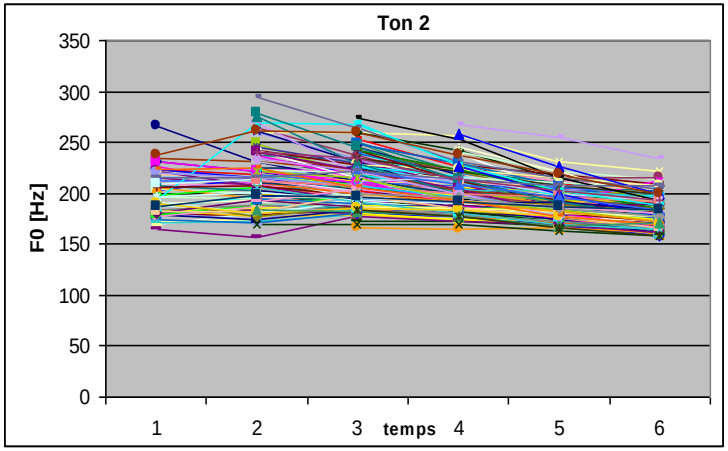
9 languages

Au Co project

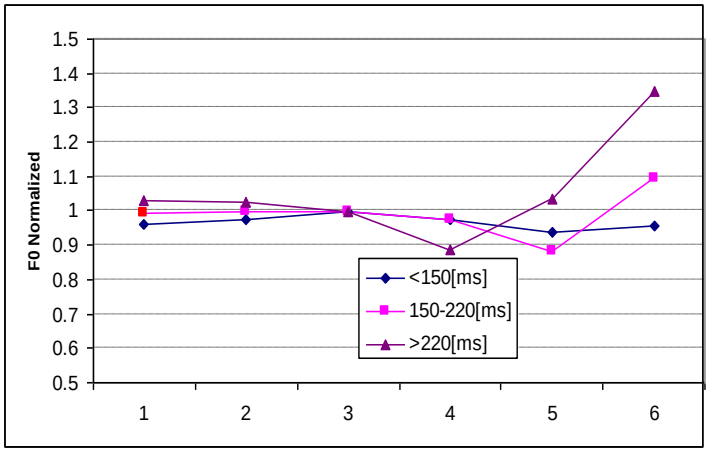
Study of languages in Vietnam

Results

- *Tone: dynamic characteristics - Variation of tones*



250 representatives of F0 contour of Tone 2



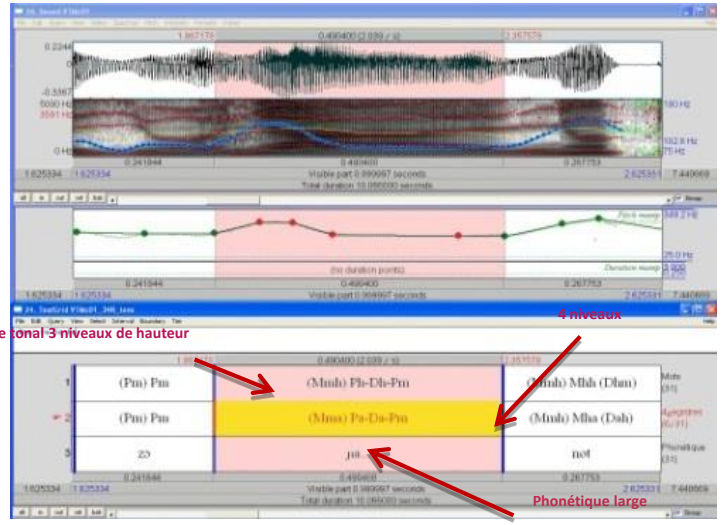
Three variants of Tone 3 with respect to syllable duration

- F0 contours are not stable (especially at the initial points).
- The initial points of tones change in a large range and the last points are quite stable.
- The variation of F0 contour is dependent on the carrying tone and duration of the syllable.

Study of languages in Vietnam
Results

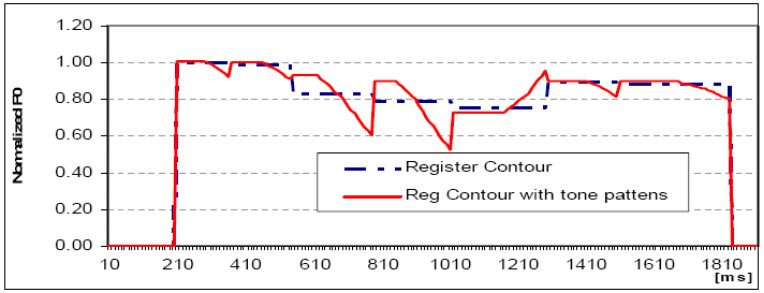
- **Database of Mo Piu language**
  - composed of 36 h films
    - 35 h speech, 1 h songs,
    - for a total of 4000 sound/video files

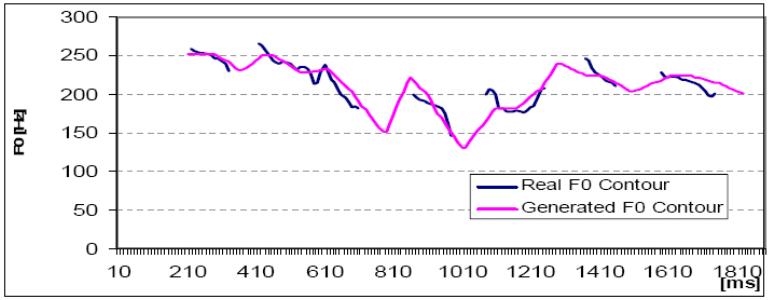
- **MISTRAL+ (Melody Intonation Speaker Tonal Range Analysis using variable Levels)**
  - a tool for the analysis of the expressive speech of any language
  - For example:
    - Tonal languages
    - Endangered languages
    - Etc.



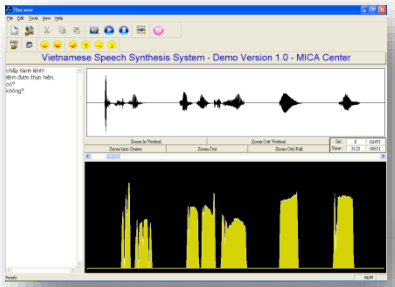
Vocal technologies & HMI
Results

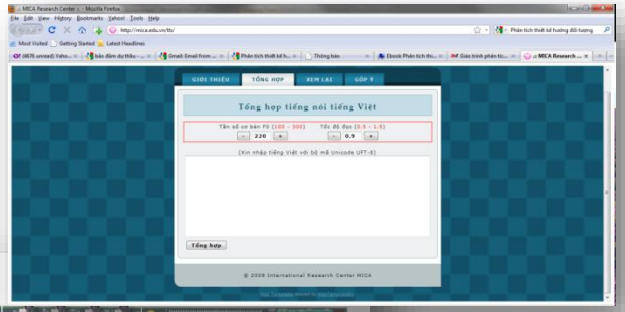
- **Model for generation of Vietnamese intonation (affirmative and question sentences), applied for Vietnamese TTS.**

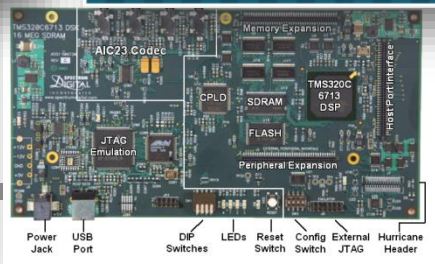


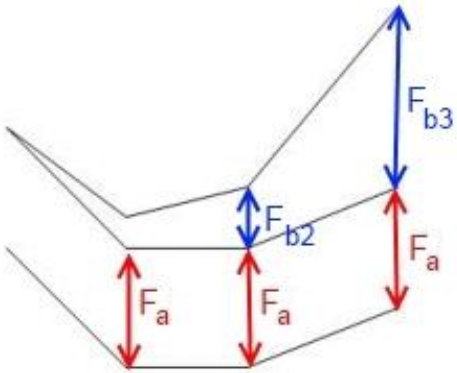


- **Synthesis for Vietnamese**
  - 3 versions : research, embedded, web



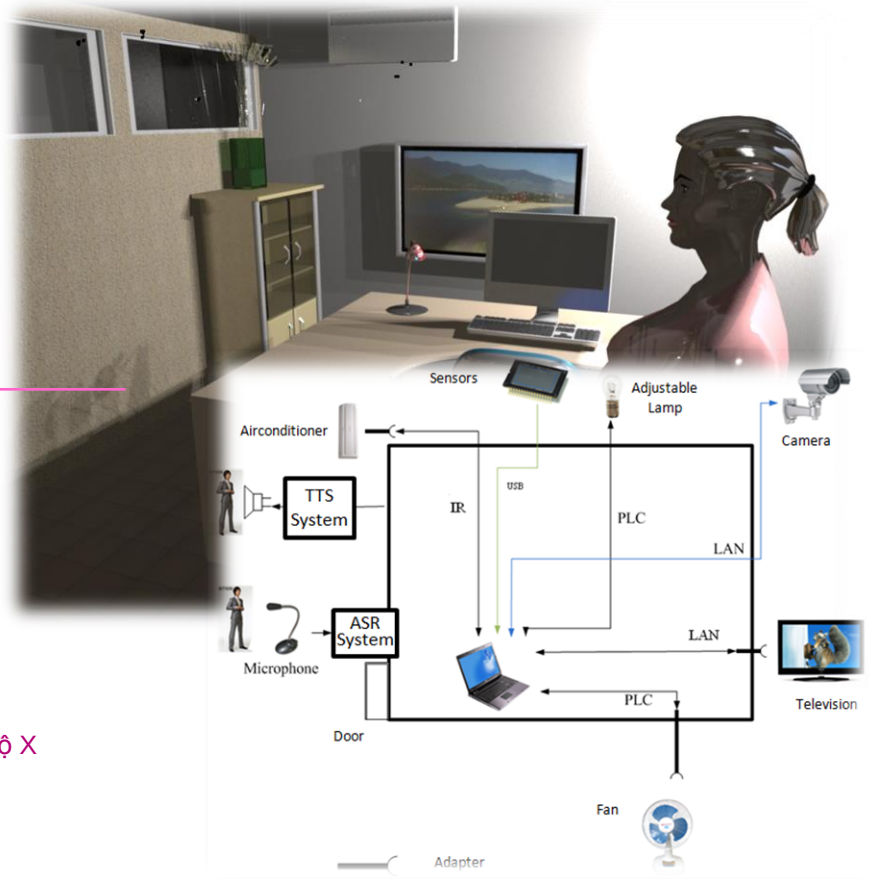




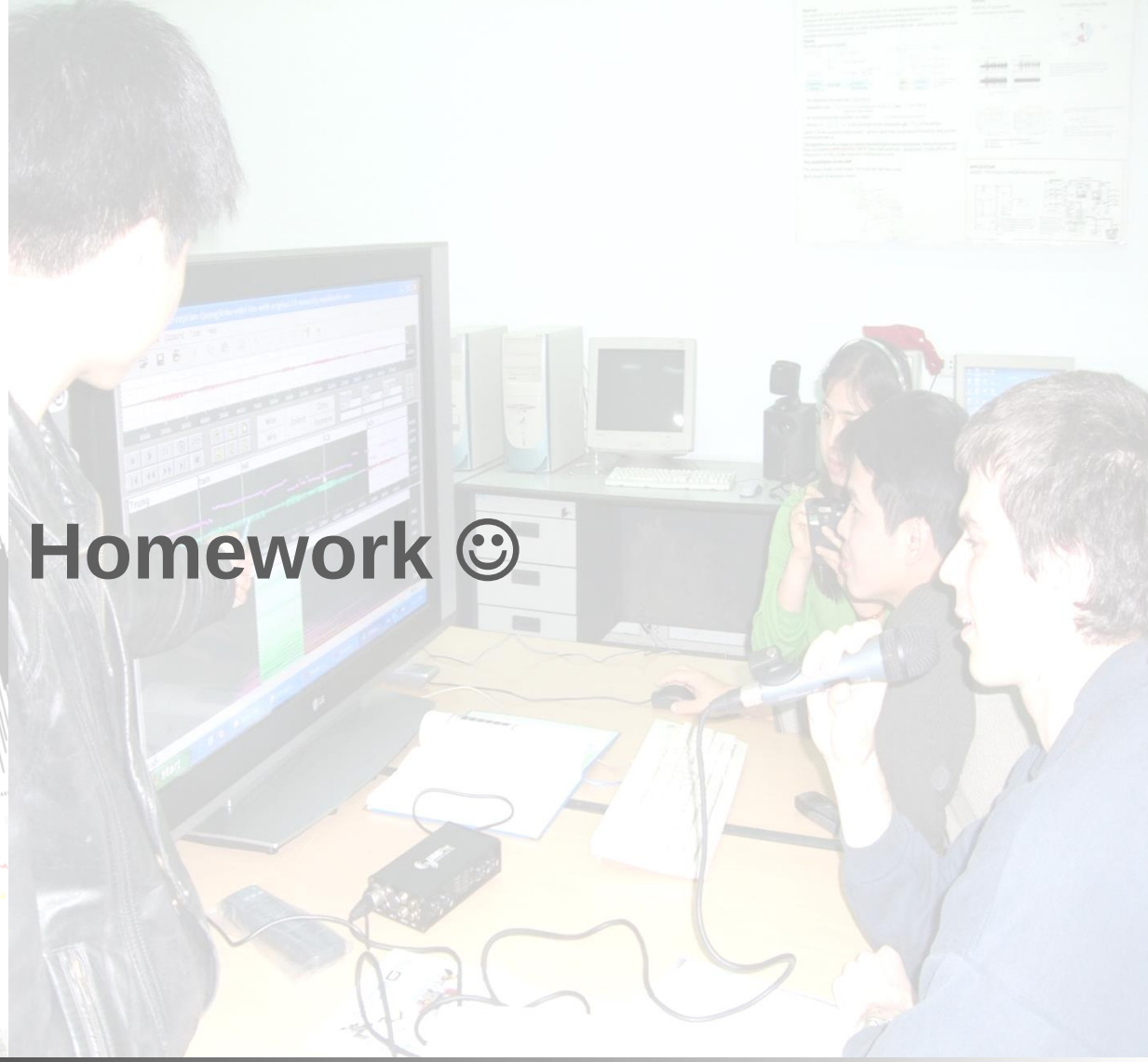


Vocal technologies & HMI
Results

- **Intelligent room controlled by voice (KC03.15)**



	Commands
Door	cửa mở / đóng
Lamp	đèn bật / tắt
	đèn sáng lên / tối đi
Camera	camera bật / tắt
	camera phải / trái
Airconditioner	điều hòa bật / tắt
	điều hòa tăng / giảm
	điều hòa thiết lập nhiệt độ X
Television	tivi bật / tắt
	tivi tăng / giảm âm
	tivi tăng / giảm kênh
	tivi kênh một / hai / ba / bốn / năm
	...



Homework ☺

International Research Institute MICA
Multimedia, Information, Communication & Applications
UMI 2954

Hanoi University of Science and Technology
1 Dai Co Viet - Hanoi - Vietnam

Next week: 1st mini presentation

- **Human speech**
  - Production
  - Perception
  - Understanding/Processing
- **Voice base application/product**
  - Intro
  - Techniques
  - Demo: Video, online demo etc.