Speech and Sound Use in a Remote Monitoring System for Health Care

M. Vacher J.-F. Serignat S. Chaillol D. Istrate
V. Popescu

CLIPS-IMAG, Team GEOD
Joseph Fourier University of Grenoble - CNRS (France)

Text, Speech and Dialogue 2006
11th to 15th of September

Motivation
The Medical Remote Monitoring
Speech and Sound Corpora
The Real-Time Architecture
The Global System Organization
The Sound Analysis System
The Speech Recognizer RAPHAEL
Conclusion
The end

Outline

Motivation

- The Medical Remote Monitoring
- Speech and Sound Corpora

The Real-Time Architecture

- The Global System Organization
- The Sound Analysis System

The Speech Recognizer RAPHAEL

Conclusion

Motivation
The Medical Remote Monitoring
Speech and Sound Corpora
The Real-Time Architecture
The Global System Organization
The Sound Analysis System
The Speech Recognizer RAPHAEL
Conclusion
The end

Outline

Motivation
The Medical Remote Monitoring
Speech and Sound Corpora
The Real-Time Architecture
The Global System Organization
The Sound Analysis System
The Speech Recognizer RAPHAEL
Conclusion

Medical Remote Monitoring: various sensors

Position

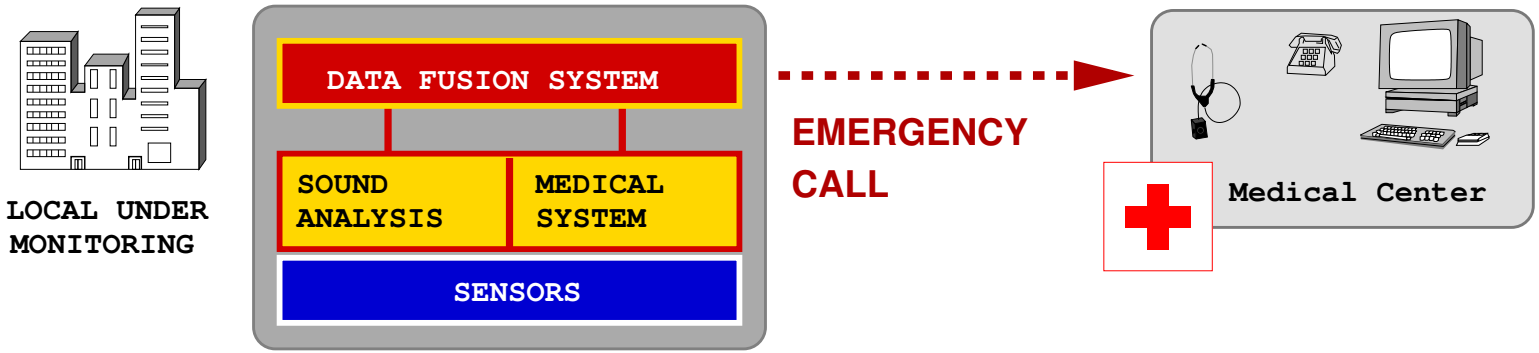
- infrared
- contact

Medical

- tensiometer
- heart rate

Sound

- microphones



Extracted informations through Sound Analysis :

- patient's **activity**: door lock, phone, "It's hot!"...
- patient's **physiology**: cough...
- **distress situation**: scream, glass breaking, "Help me!"...

Motivation
The Medical Remote Monitoring
Speech and Sound Corpora
The Real-Time Architecture
The Global System Organization
The Sound Analysis System
The Speech Recognizer RAPHAEL
Conclusion
The end

Outline

Motivation
The Medical Remote Monitoring
Speech and Sound Corpora

The Real-Time Architecture
The Global System Organization
The Sound Analysis System

The Speech Recognizer RAPHAEL

Conclusion

Speech Corpus

The *"Normal-Distress"* corpus :

- 21 speakers
  - 11 men, 10 women
  - 20 years$\leq$age$\leq$65 years
- French
- 64 normal situation sentences
  - *"Bonjour"* (Good morning)
  - *"Où est le sel?"* (Where is the salt)
- 64 distress sentences
  - *"Au secours!"* (Help me)
  - *"Un médecin vite!"* (Doctor quickly)
- 2,646 audio speech files
- duration: 38 mn
- sampling rate: 16 kHz

Sound Corpus

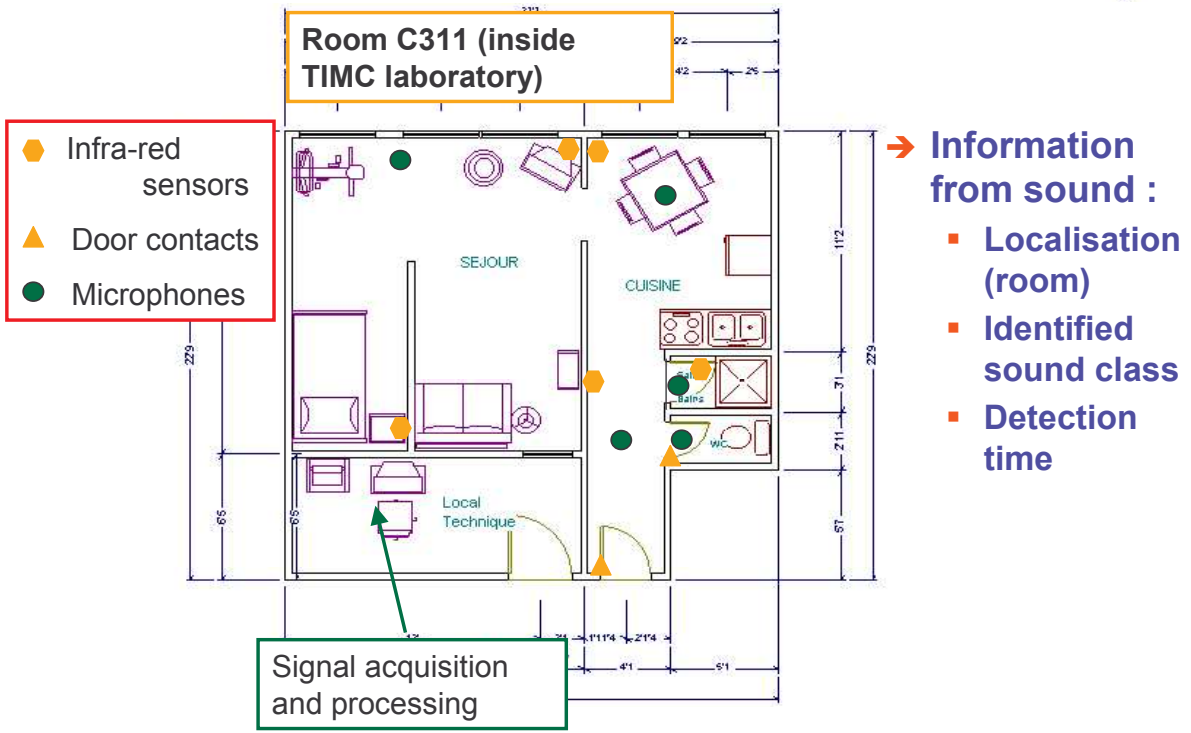
- 7 sound classes:
  - **normal** sounds: door clapping, phone ringing, dishes...
  - **abnormal** sounds: breaking glasses, falls, screams
- 1,577 audio sound files
- duration: 20 mn
- sampling rate: 16 kHz

Class of sound	% of the corpus (duration)	% of the corpus (number)	Average of duration
Dishes	6%	10.4%	402 ms
Door lock	1%	12.7%	36 ms
Door slap	33%	33.2%	737 ms
Glass breaking	6%	5.6%	861 ms
Ringing phone	40%	32.8%	928 ms
Scream	12%	4.6%	1,930 ms
Step sound	2%	0.8%	2,257 ms

Noisy Corpus

- Signal to noise ratio (SNR) : 0, +10, +20, +40dB
- Experimental HIS Noise :
  - non stationary
  - recorded inside experimental apartment

In collaboration with TIMC laboratory



Motivation
The Medical Remote Monitoring
Speech and Sound Corpora
The Real-Time Architecture
The Global System Organization
The Sound Analysis System
The Speech Recognizer RAPHAEL
Conclusion
The end

Outline

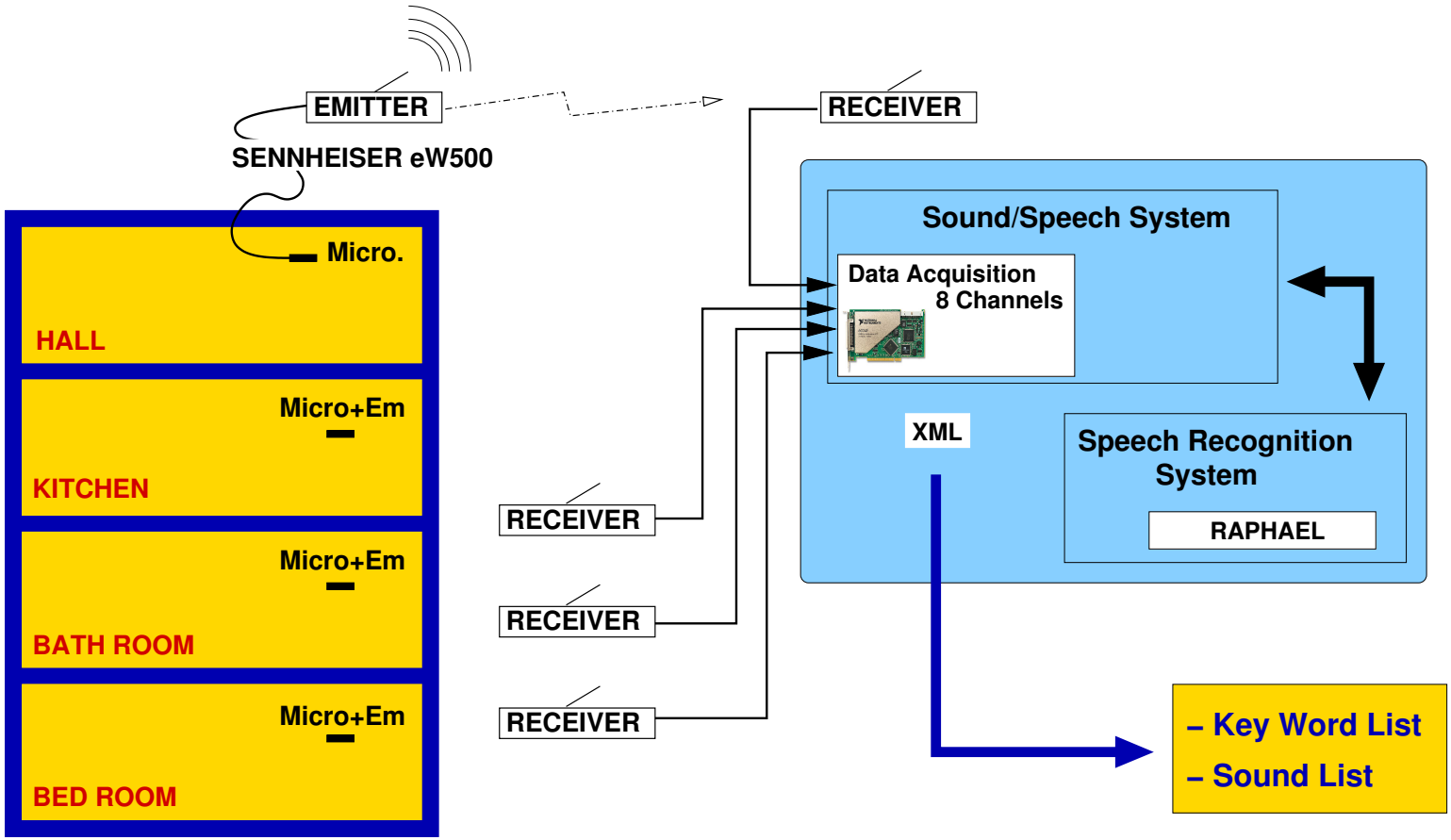
Motivation
The Medical Remote Monitoring
Speech and Sound Corpora

The Real-Time Architecture
The Global System Organization
The Sound Analysis System

The Speech Recognizer RAPHAEL

Conclusion

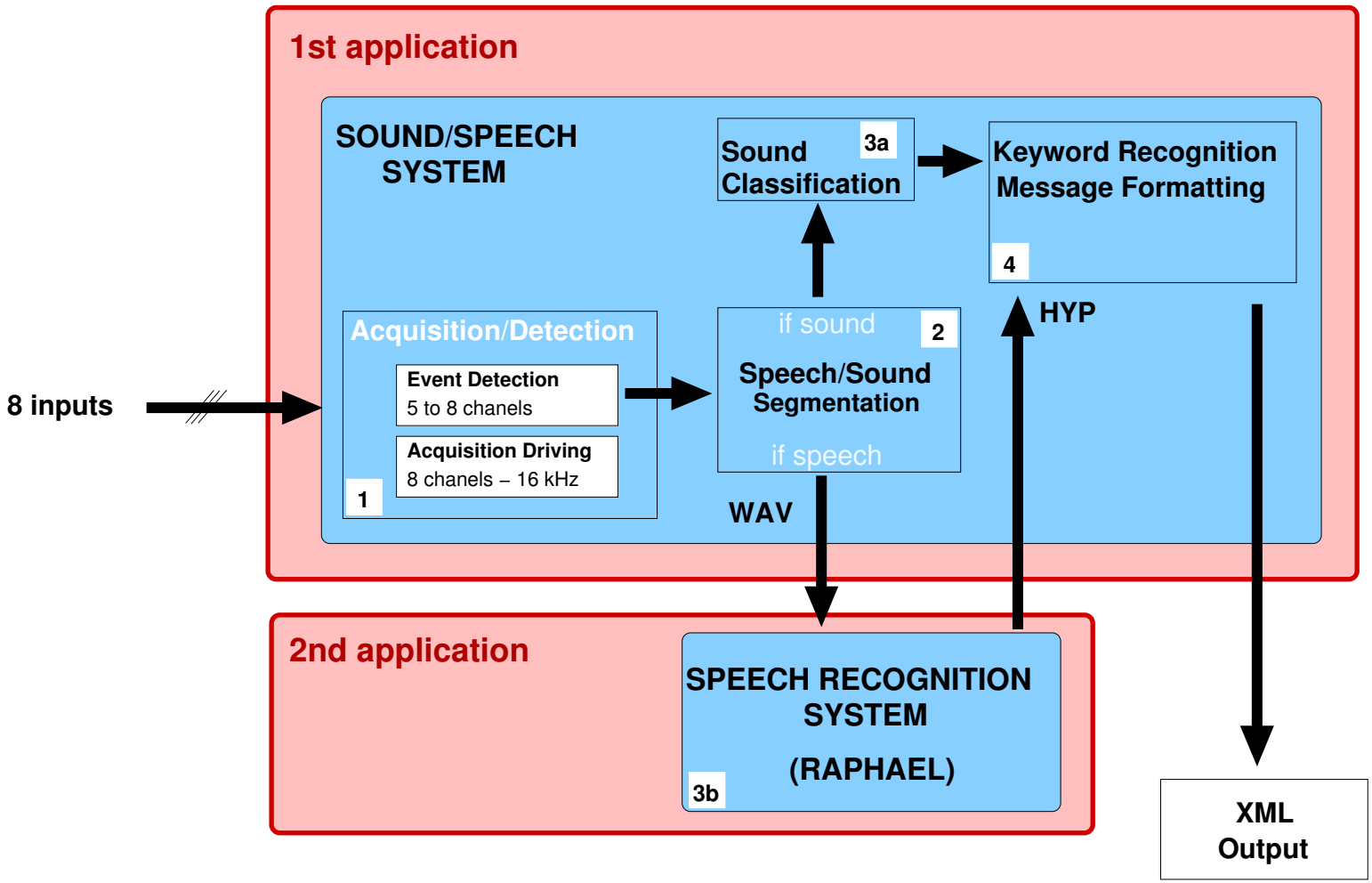
Global System Organization (1/2)



Data Acquisition Card: 200 ksamples/s – 8 differential channels

Sampling rate: 16 ksamples/s for each channel

Global System Organization (2/2)



Motivation
The Medical Remote Monitoring
Speech and Sound Corpora
The Real-Time Architecture
The Global System Organization
The Sound Analysis System
The Speech Recognizer RAPHAEL
Conclusion
The end

Outline

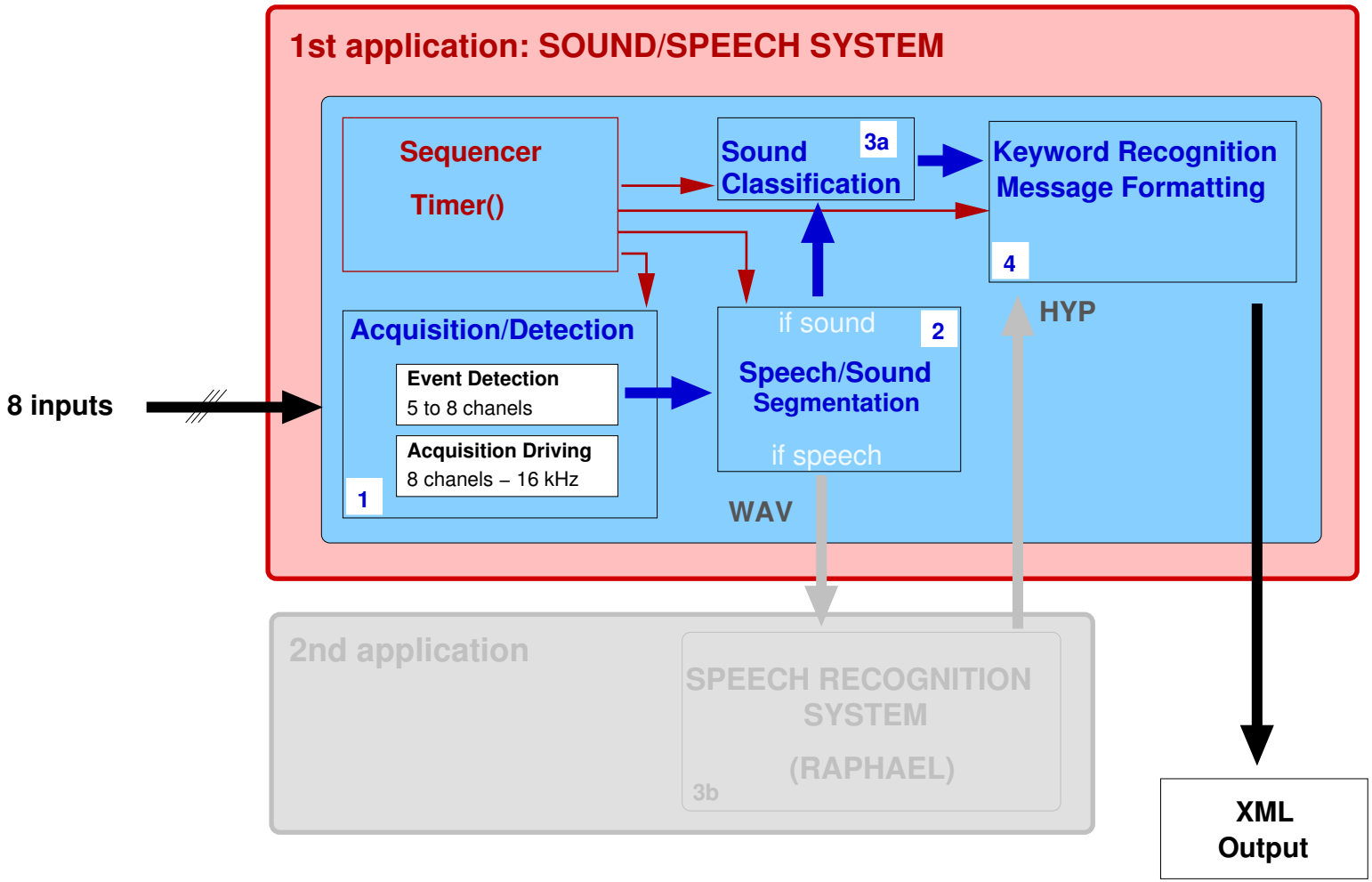
Motivation
The Medical Remote Monitoring
Speech and Sound Corpora

The Real-Time Architecture
The Global System Organization
The Sound Analysis System

The Speech Recognizer RAPHAEL

Conclusion

Sound Analysis System



Detection + Speech/Sound Segmentation

- Detection [1]:
  - Wavelet Tree Detection
  - Equal Error Rate: (ROC curves)
    $EER = 6.5\%$ if $SNR = 0dB$, 0% if $SNR \geq +10dB$
- Segmentation
  - frame 16 ms - overlap 8 ms
  - GMM, 24 Gaussian models
  - 16 LFCC coupled with energy
  - Error Segmentation Rate "cross validation protocol":

SNR	ESR gobal	Speech	Sound
$+40dB$	4.1%	0.8%	9.5%
$+20dB$	3.9%	0.8%	9.1%
$+10dB$	3.9%	1.5%	7.9%
$0dB$	14.5%	21.0%	3.6%

[1] M. Vacher et al., *Sound Detection and Classification through Transient Models using Wavelet Coefficient Trees*, EUSIPCO, 20th September 2004.

Sound Classification for Medical Remote Monitoring (1/2)

- a more complete sound corpus, a new class: object falls
- adapted to HMM training and classification: one sound per file
- 1,315 files, 25 mn

Class of sound	% of the corpus (duration)	% of the corpus (number)	Average of duration
Dishes	5.3%	12.4%	380 ms
Door lock	27.2%	12.5%	3,150 ms
Door slap	22%	26.2%	950 ms
Glass breaking	12.3%	8.2%	1,690 ms
Object falls	5.4%	5.5%	1,150 ms
Ringing phone	11.7%	19.4%	700 ms
Scream	13%	7.2%	2,110 ms
Step sound	3.7%	8.4%	450 ms

Sound Classification for Medical Remote Monitoring (2/2)

- analysis window 16 ms - overlap 8 ms
- 12 Gaussian models
- 16 LFCC with Δ and $\Delta\Delta$
- GMM or 3 state HMM
- "cross validation protocol"
- Error classification rate:

SNR	GMM	HMM
$\geq +50dB$	3.2%	2%
$+40dB$	10.2%	9%
$+20dB$	16.5%	10.8%
$+10dB$	12.6%	15.4%
$0dB$	23.6%	30%

XML Output

Speech: RAPHAEL analysis is initiated

<appli:segmentation description="appli audio">
</pièce>Cuisine</pièce>
<horodate>1−12−2005 à 15:19:20</horodate>
<résultat>parole</résultat>
<information>Probabilité de son=−20.2018, Probabilité de parole=−17.2258</information>
</appli:segmentation>

Module 2: segmentation

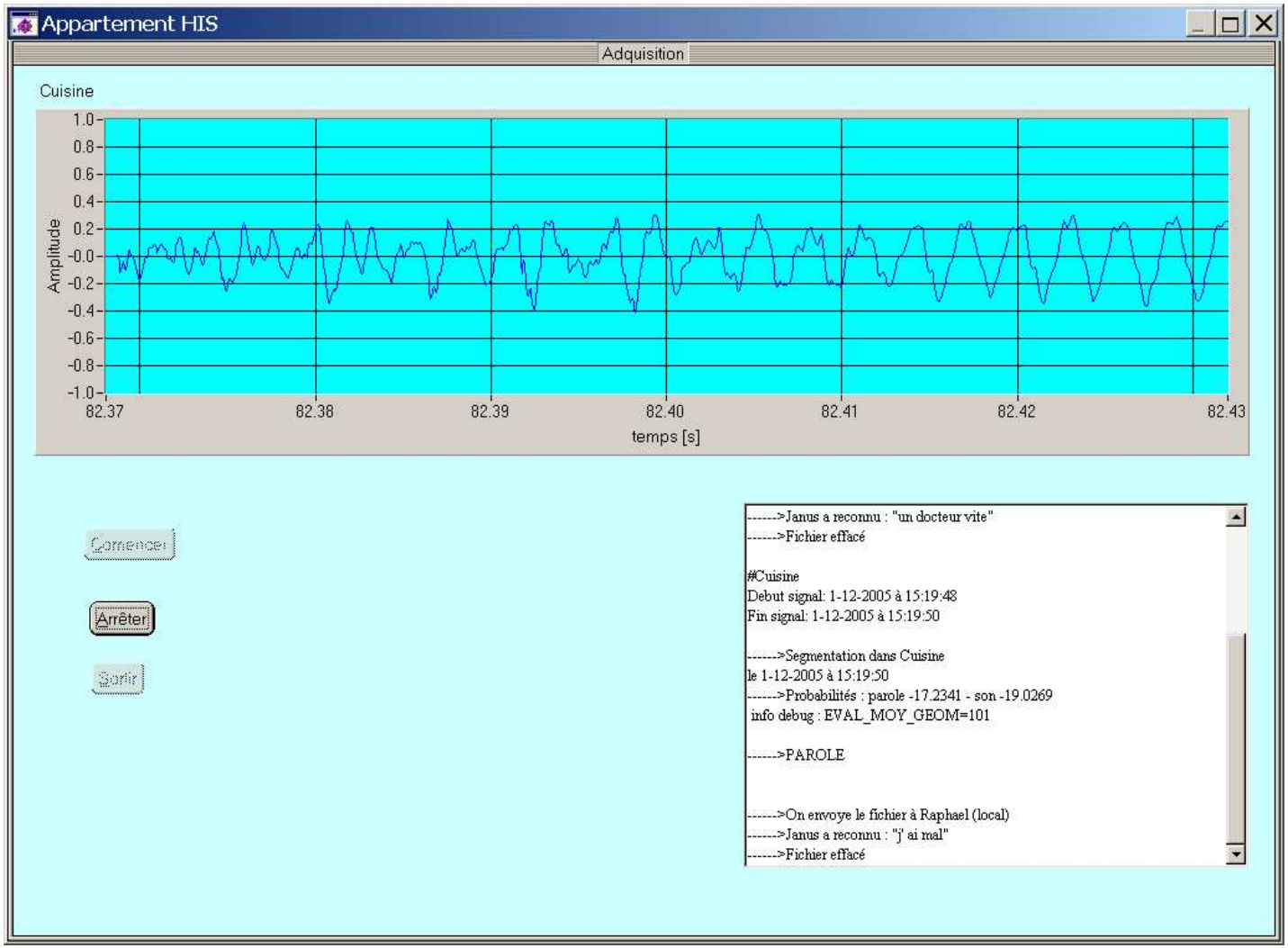
<appli:reconnaissance description="appli audio">
</pièce>Cuisine</pièce>
<horodate>1−12−2005 à 15:19:20</horodate>
<résultat>un docteur vite</résultat>
</appli:reconnaissance>

Module 3b: RAPHAEL

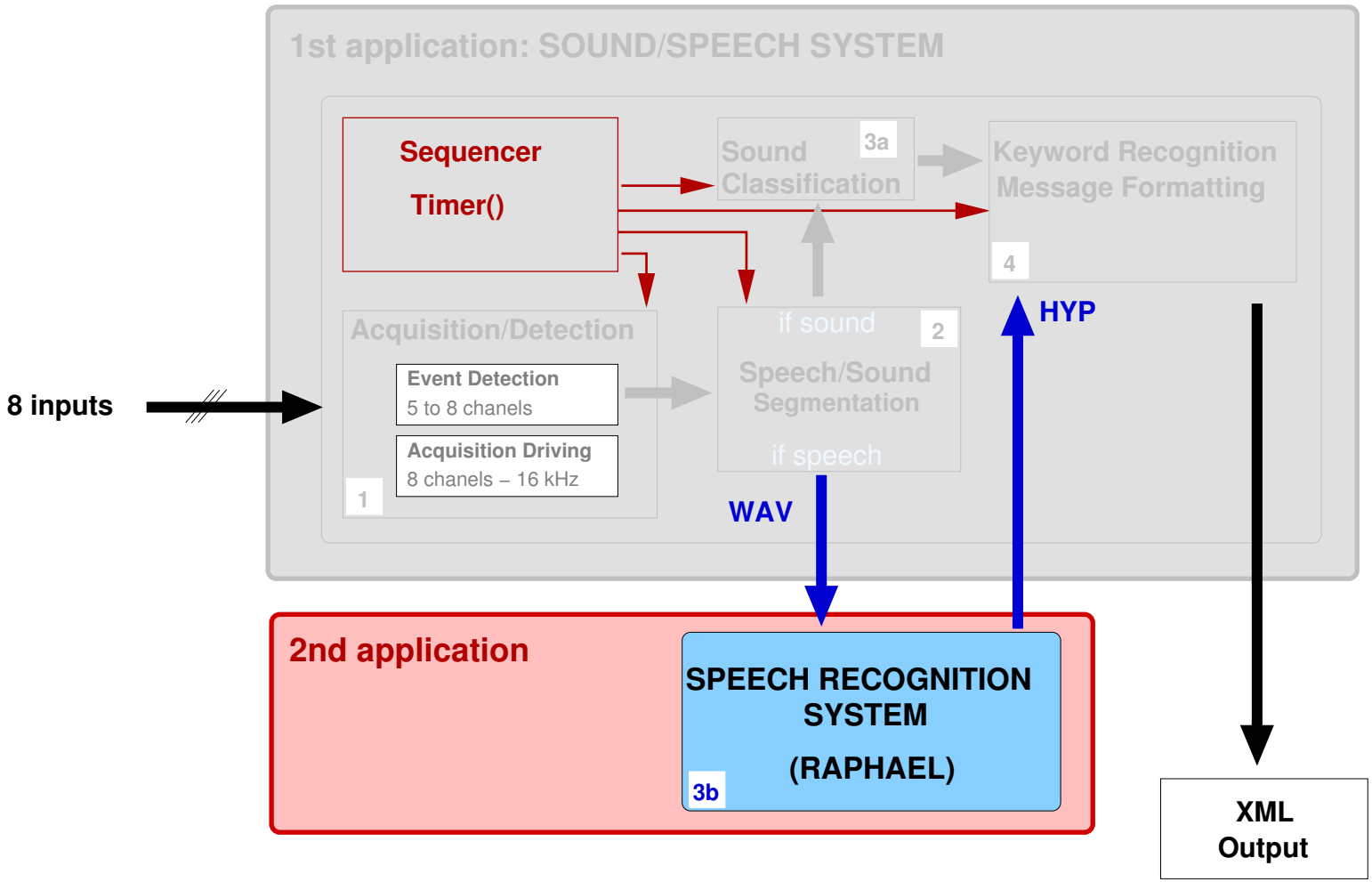
sentence recognized by RAPHAEL

Motivation
The Medical Remote Monitoring
Speech and Sound Corpora
The Real-Time Architecture
The Global System Organization
The Sound Analysis System
The Speech Recognizer RAPHAEL
Conclusion
The end

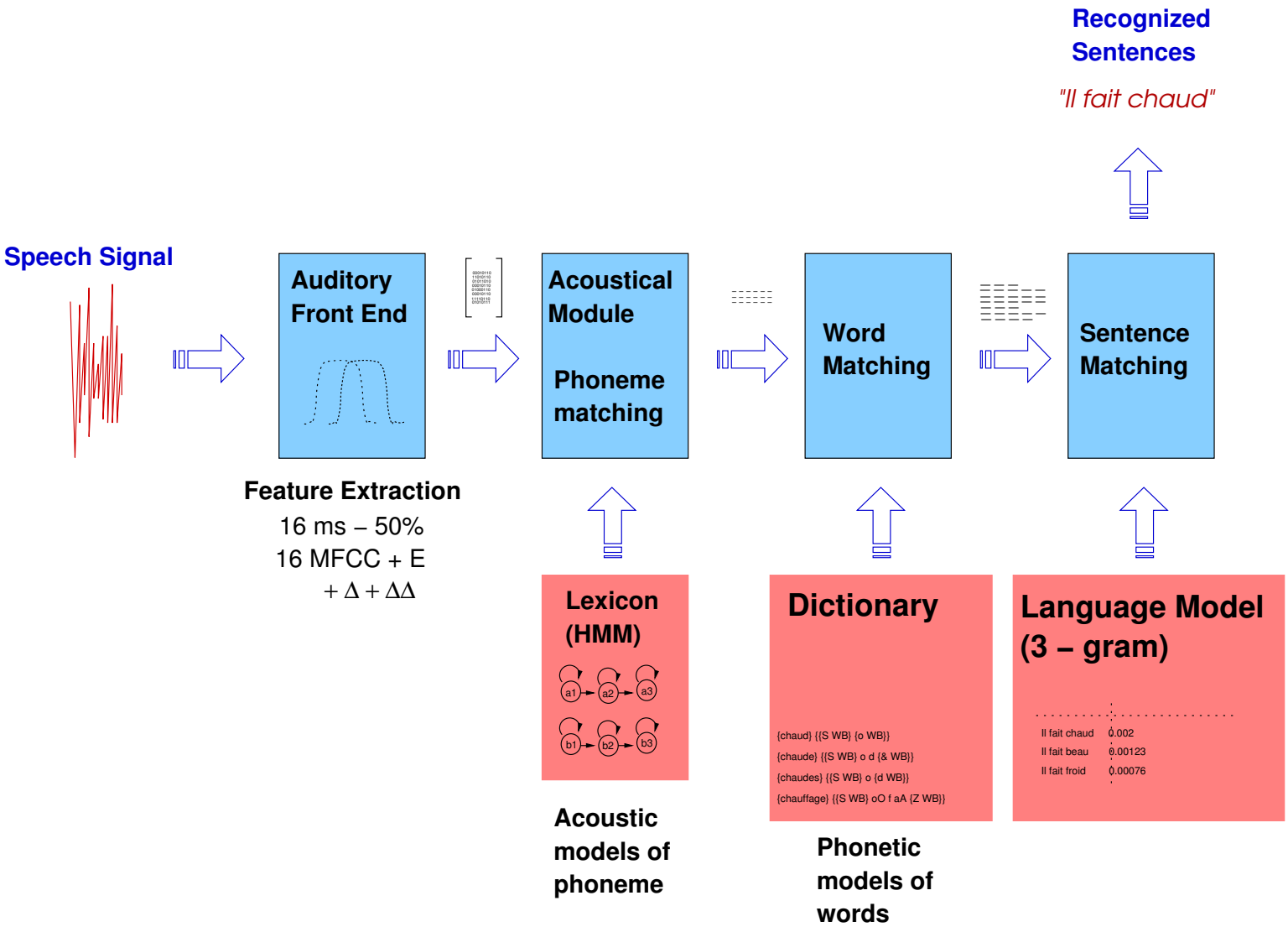
Acquiring Front Panel



RAPHAEL (1/3)


1st application: SOUND/SPEECH SYSTEM
Sequencer
Timer()
Sound Classification
3a
Keyword Recognition Message Formatting
4
Acquisition/Detection
Event Detection
5 to 8 chanels
Acquisition Driving
8 chanels – 16 kHz
1
if sound
2
Speech/Sound Segmentation
if speech
HYP
8 inputs
WAV
2nd application
SPEECH RECOGNITION SYSTEM
(RAPHAEL)
3b
XML Output

RAPHAEL (2/3)


Recognized Sentences
"Il fait chaud"
Speech Signal
Auditory Front End
Acoustical Module
Phoneme matching
Word Matching
Sentence Matching
Feature Extraction
16 ms – 50%
16 MFCC + E
+ Δ + ΔΔ
Lexicon (HMM)
a1
a2
a3
b1
b2
b3
Dictionary
{chaud} {{S WB} {o WB}}
{chaude} {{S WB} o d {& WB}}
{chaudes} {{S WB} o {d WB}}
{chauffage} {{S WB} oO f aA {Z WB}}
Language Model (3 – gram)
Il fait chaud 0.002
Il fait beau 0.00123
Il fait froid 0.00076
Acoustic models of phoneme
Phonetic models of words

RAPHAEL (3/3)

- Acoustic models:
  - training with large corpora
    Braf80, Braf100, Braf120
  - large number of French speakers
    more than 200
- Dictionary: 11,000 words (French)

  (Speech Assessment Methods Phonetic Alphabet)
- Language model: n - gram, $n \in \{1, 2, 3\}$
  - extraction from the WEB and "Le Monde" corpora
  - optimization for distress sentences

Recognition Results

- all the sentences of our corpus
- 5 speakers

Corpus	Recognition Error	
Normal Sentences	False Alarm Sentence "Quel temps fait-il dehors" / "Quel temps fait-il de mort"	6%
Distress Sentences	Missed Alarm Sentence "Faîtes vite" / "Equilibre" "A moi" / "Un grand"	16%

Motivation
The Medical Remote Monitoring
Speech and Sound Corpora
The Real-Time Architecture
The Global System Organization
The Sound Analysis System
The Speech Recognizer RAPHAEL
Conclusion
The end

Conclusion

- Sound classification for distress detection.
- Speech recognition allowing call for help.
- Real-Time operation on an operating system without real-time capacity.
- Speaker independence of the ASR system.
- For 10dB and upper: errorless detection and segmentation error below 5%.

- Outlook
  - Improvement of sound classification through HMM
  - Development of a complete acoustical analysis system for life-sized tests.

Detection and Speech/Sound Segmentation in a Smart Room Environment

Thank you for your attention.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

CLIPS
Communication Langagière et Interaction Personne-Système
Fédération IMAG
BP 53 - 38041 Grenoble Cedex 9 - France

IMAG

GEOD
Groupe d'Etude sur l'Oral et le Dialogue

For Further Reading I

D. Istrate et al.
Information Extraction From Sound for Medical Telemonitoring,
IEEE Trans. on Information Tech. in Biomedicine
Vol. **10**, issue 2, pp. 264-274, 2006.

M. Vacher, D. Istrate.
Sound detection and classification for medical telesurveillance,
BIOMED'2004 Proceedings, ACTA PRESS, pp. 395–398, 2004.

D. Istrate, M. Vacher.
Détection et classification des sons : application aux sons de la vie courante et à la parole,
20th GRETSI Proceedings, Vol. **1**, pp. 485-488, 2005.

Acknowledgement I

This study is a part of the DESDHIS-ACI "Technologies pour la Santé" project of the French Research Ministry. This project is a collaboration between the CLIPS laboratory, in charge of the sound analysis, and the TIMC laboratory, in charge of the medical sensor analysis and data fusion.
CLIPS and TIMC are 2 laboratories of the IMAG institute. IMAG has funded the implementation of the habitat used for our studies through the RESIDE-HIS project.