

Détection et reconnaissance des sons pour la surveillance médicale

Dan Istrate

le 16 décembre 2003

Directeur de thèse : Eric Castelli

Co-Directeur : Laurent Besacier

Thèse mené dans le cadre d'une collaboration avec le laboratoire TIMC



CLIPS

Communication Langagière et
Interaction Personne-Système

Fédération IMAG

BP 53 - 38041 Grenoble Cedex 9 - France



GEOD

Groupe d'Etude sur l'Oral et le Dialogue

Plan de la présentation



- Contexte de la thèse
- Objectif et problématiques de recherche
- Détection de sons
- Classification de sons
- Couplage entre la détection et la classification
- Évaluation du système
- Conclusions et perspectives

Contexte de la thèse – *Espaces perceptifs*



Salles de réunions



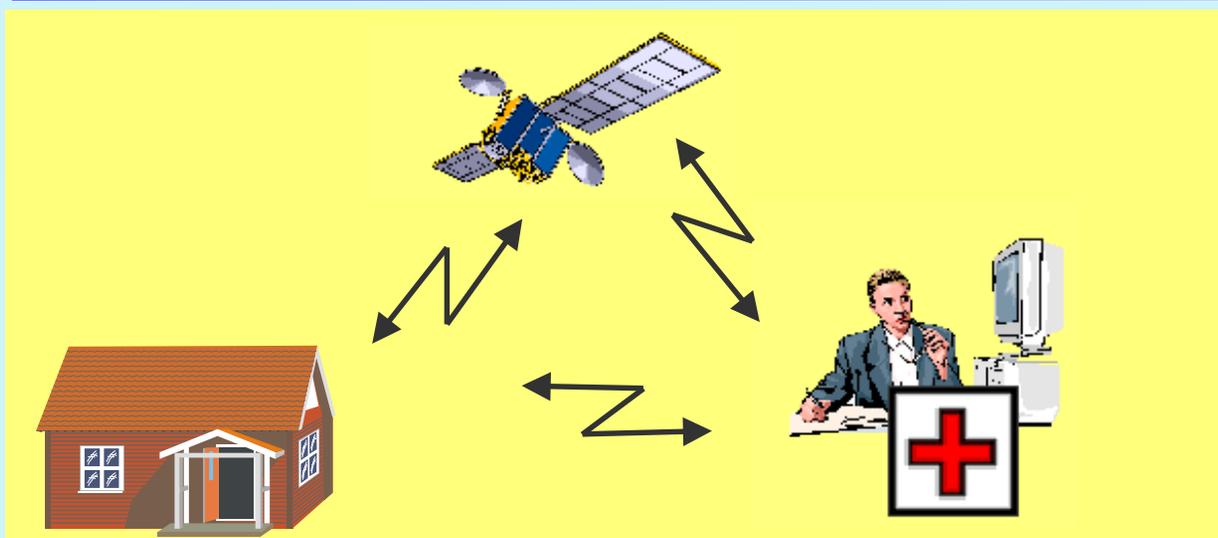
→ Multimodalité :

- Vidéo
- Son
- Gestes
- Paramètres de l'environnement
- Position de la personne

→ Problématiques :

- Standardisation des bases de données
- Reconnaissance de la parole / du son avec contraintes :
 - Bruit, réverbération, plusieurs locuteurs
 - Temps réel
 - Traitement en continu sur plusieurs voies

Contexte de la thèse - *Télémédecine*



→ Principales applications :

- Transfert et gestion d'informations médicales par réseau
- **Télésurveillance en temps réel ou en différé**
- Contrôle à distance des procédures médicales

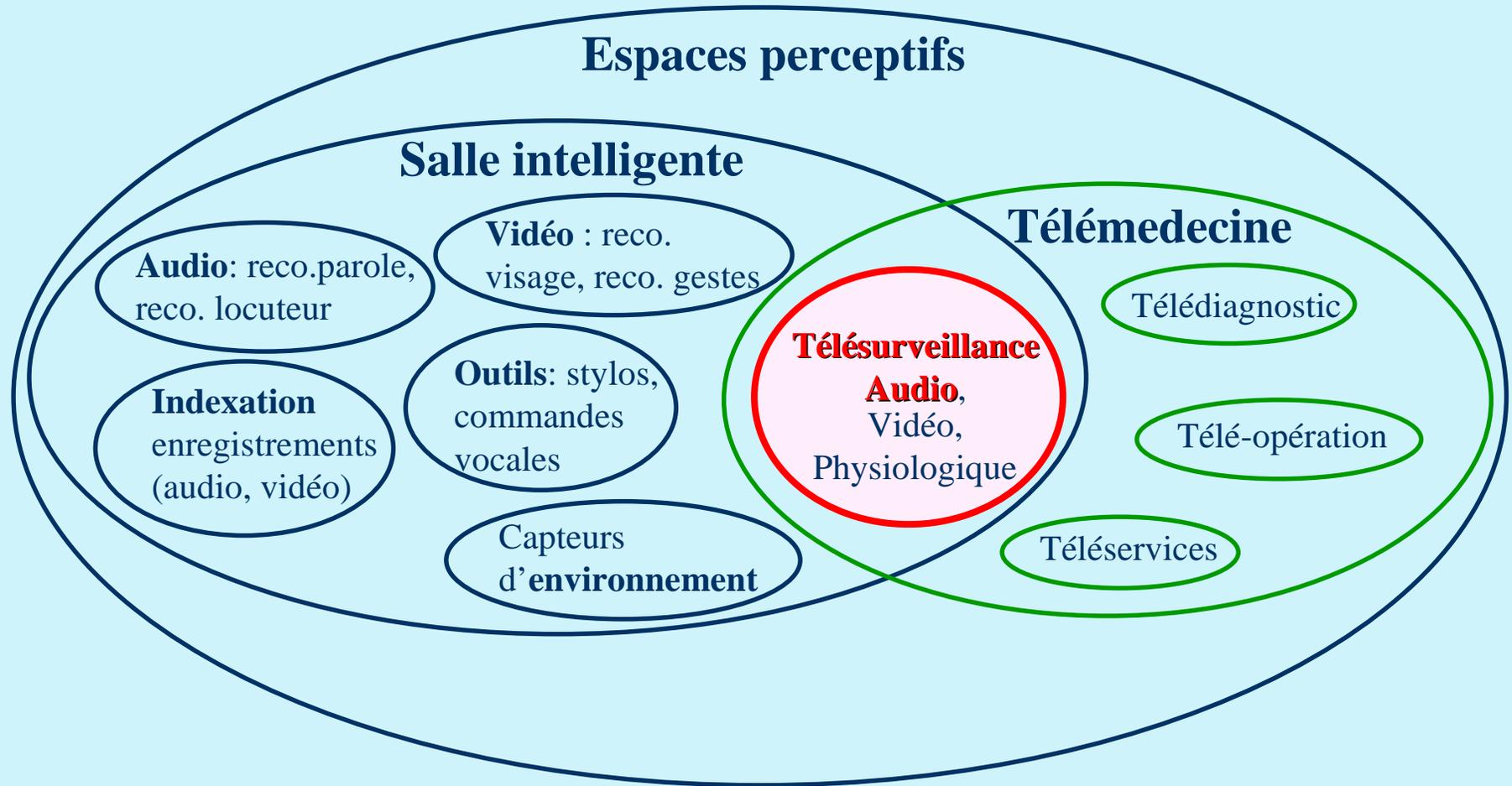
→ Contraintes :

- Fonctionnement 24h/24h
- Fiabilité des systèmes

→ Problématiques :

- Évaluation des algorithmes
 - Corpus de tests
- Capteurs portés par le patient

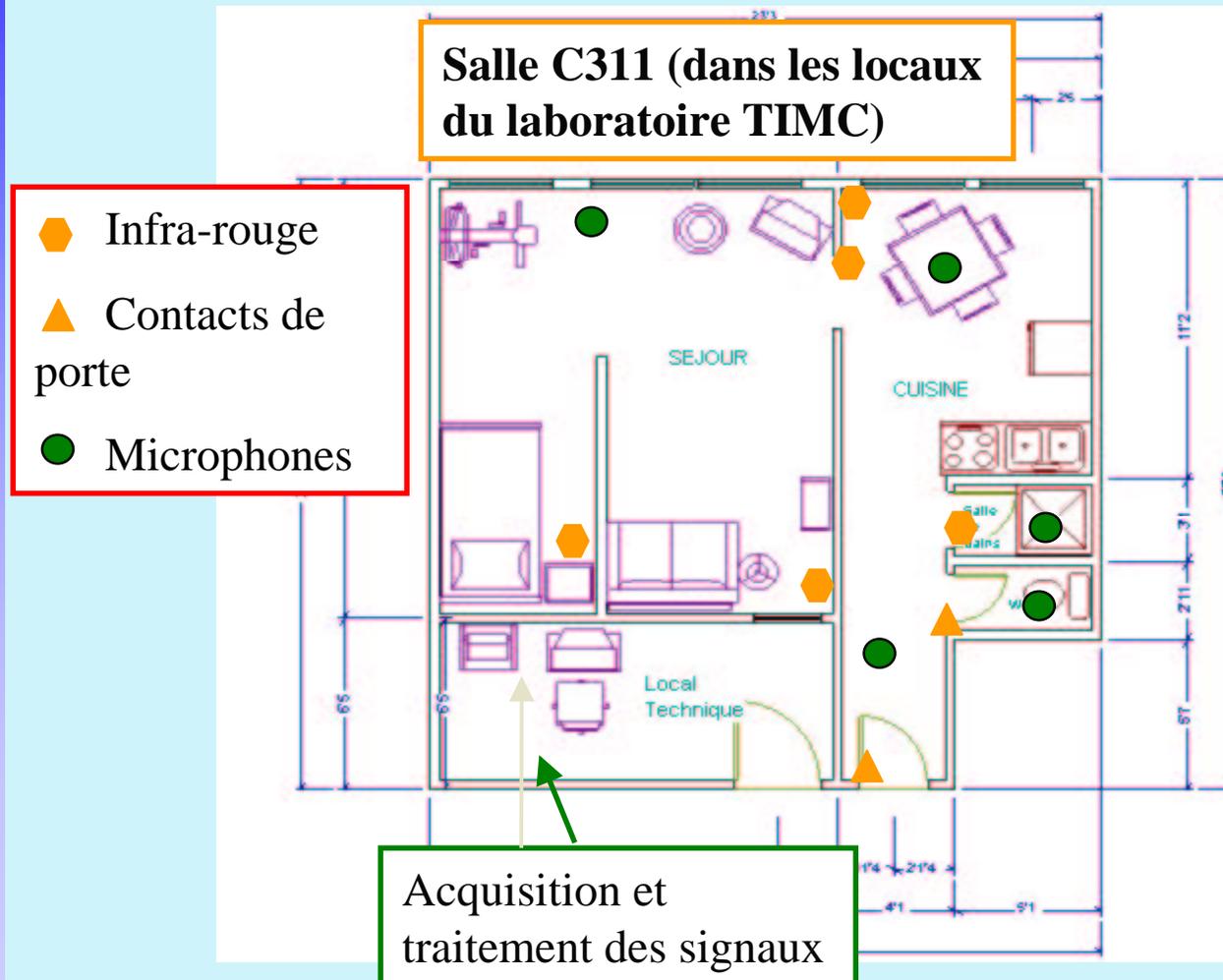
Contexte de la thèse – *Espaces perceptifs & télésurveillance médicale*



Contexte de la thèse – *Projet RESIDE - HIS*



Collaboration entre les laboratoires CLIPS et **TIMC**



- Information extraite du son :
- Localisation (pièce)
 - Classe de son identifiée
 - Instant de détection

Plan de la présentation



- Contexte de la thèse
- Objectif et problématiques de recherche
- Détection de sons
- Classification de sons
- Couplage entre la détection et la classification
- Évaluation du système
- Conclusions et perspectives

Objectif



L'étude et la validation d'un système d'extraction et de classification de sons de la vie courante

→ Contraintes :

- Application médicale, erreurs -> graves conséquences
- Fonctionnement 24h/24h
- Temps réel
- Difficultés pour obtenir des enregistrements spécifiques de l'application

Problématiques de recherche



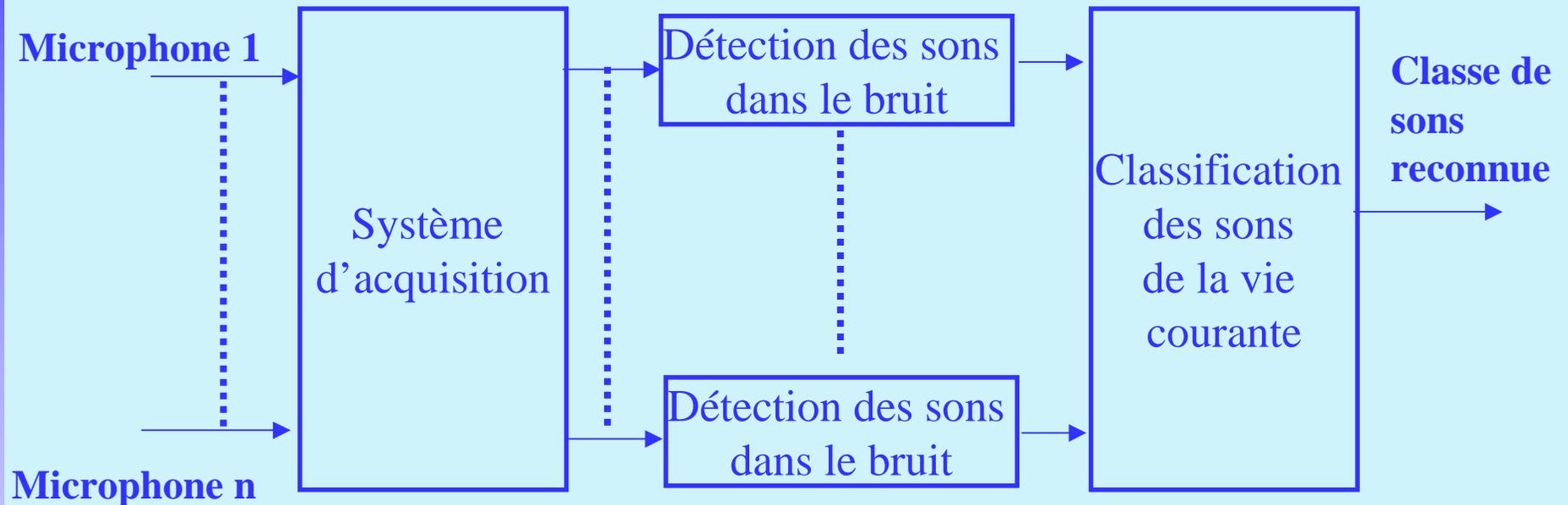
- Détection de signaux impulsionnels et non-impulsionnels dans des conditions bruitées
- Garantir une faible complexité des algorithmes (volume d'informations à traiter, multicanaux, 24h/24h)
- Méthodologie d'évaluation du système
- Adaptation des techniques de reconnaissance de la parole/du locuteur à la classification des sons de la vie courante

Plan de la présentation



- Contexte de la thèse
- Objectif et problématiques de recherche
- Détection de sons
 - Introduction
 - Algorithme fondé sur la transformée en ondelettes
- Classification de sons
- Couplage entre la détection et la classification
- Évaluation du système
- Conclusions et perspectives

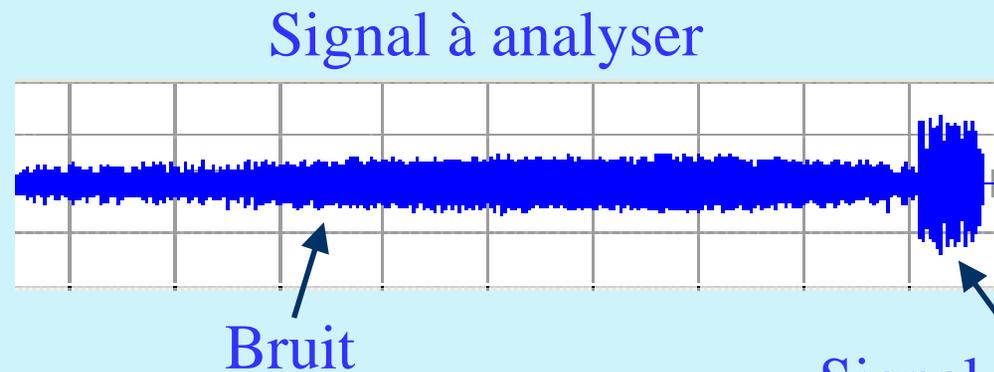
Structure du système proposé



Détection de sons - Introduction



- Identification de l'instant d'apparition d'un signal recherché dans un environnement bruité

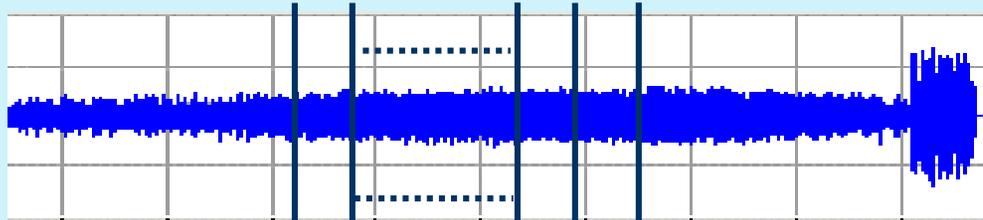


Signal à détecter

- H_0, H_1 - les hypothèses
- $o(t)$ - signal analysé
- $b(t)$ - bruit additif
- $s(t)$ - signal à détecter

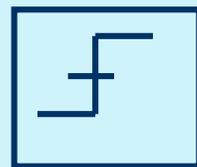
$$\begin{cases} H_0 & o(t) = b(t) \\ H_1 & o(t) = s(t) + b(t) \end{cases}$$

Détection de sons - Introduction (2)

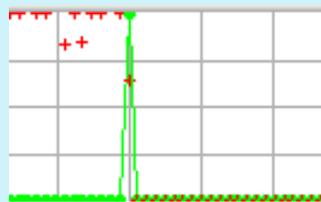


→ Algorithme de détection :
 H_0 contre H_1

Analyse
du signal



Seuillage (seuil λ)



Signal de détection

Détection de sons

- *Contraintes*



- Fonctionnement temps réel sur 5 canaux simultanément
- Sons audibles impulsionnels mais aussi non-impulsionnels
- Bruit important ($RSB \geq 0$ dB)
- Large dynamique du signal
- Différents types de bruit : bruit de l'environnement, écoulement de l'eau

Détection de sons

- Algorithmes évalués



- Trois algorithmes de l'état de l'art avec seuillage sur :
 - la variance de l'énergie du signal
 - la différence entre l'énergie et l'énergie filtrée par un filtre médian conditionné
 - l'énergie du signal avec seuil adaptatif

- Avantages :
 - Bonnes performances dans le bruit blanc
 - Temps de calcul réduit

- Désavantages
 - Non-adaptés à notre bruit environnemental
 - Plusieurs seuils à régler
 - Seuils dépendants du RSB

Détection de sons

- *Algorithmes proposés*



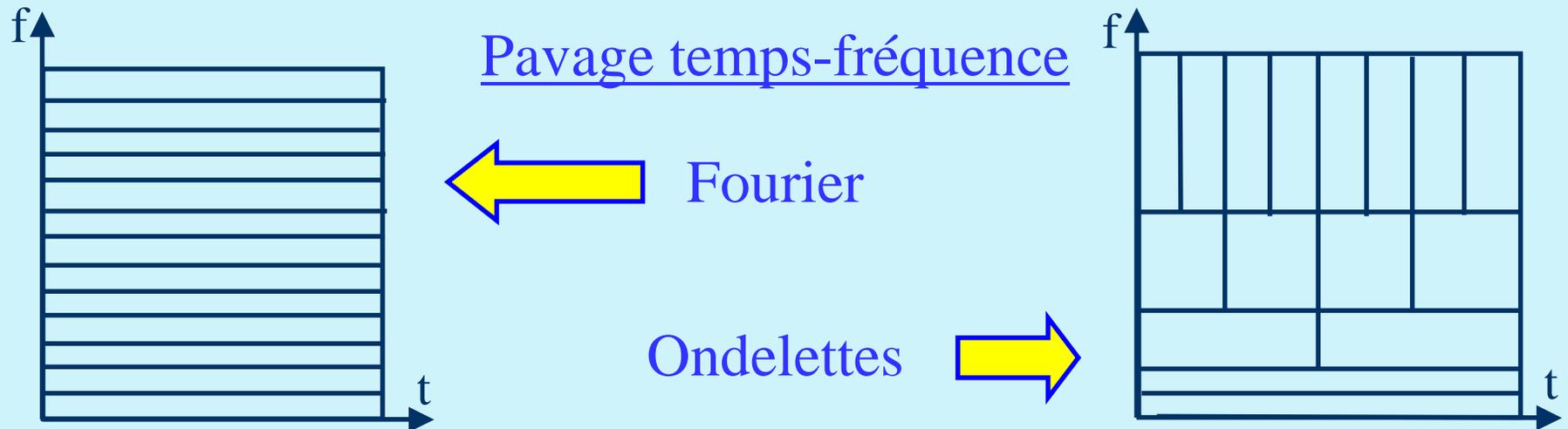
- Trois algorithmes proposés avec seuillage sur :
- le maximum de la fonction d'intercorrélation des deux fenêtres successives du signal
 - l'erreur de prédiction de l'énergie par interpolation SPLINE
 - **l'énergie de certains coefficients de la transformée en ondelettes**

Détection de sons - *Algorithme fondé sur la transformée en ondelettes*



→ Transformée en ondelettes :

- ondelette mère (Daubechies)
- facteur d'échelle (dilatation et décalage temporel)

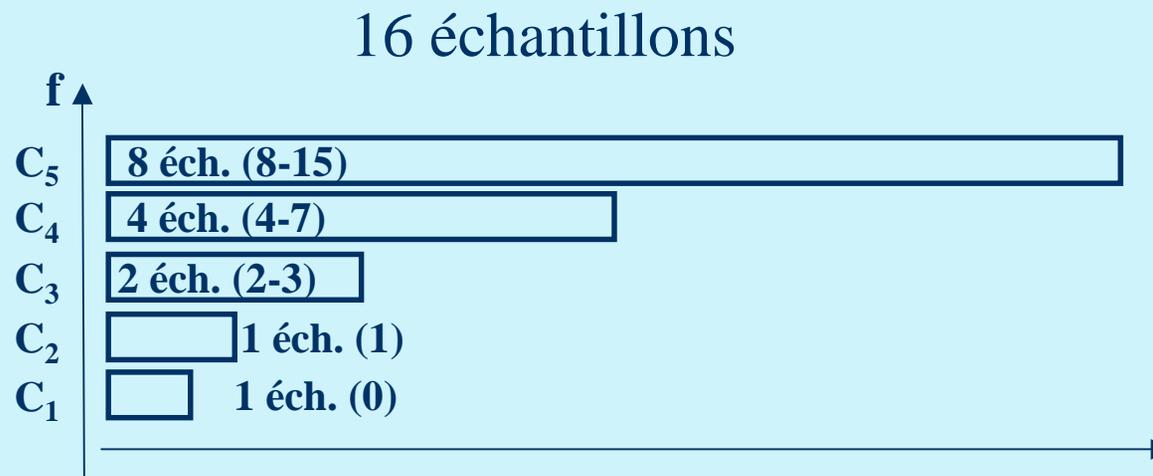


Détection de sons - *Algorithme fondé sur la transformée en ondelettes*



→ Transformée en ondelettes :

- ondelette mère (Daubechies)
- facteur d'échelle (dilatation et décalage temporel)



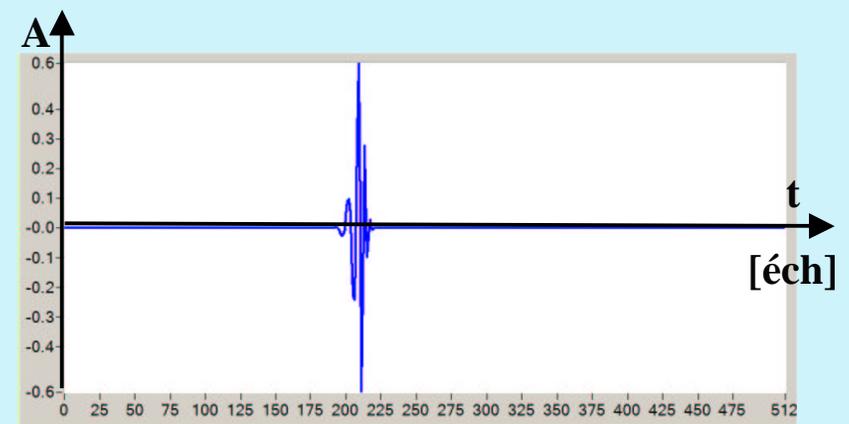
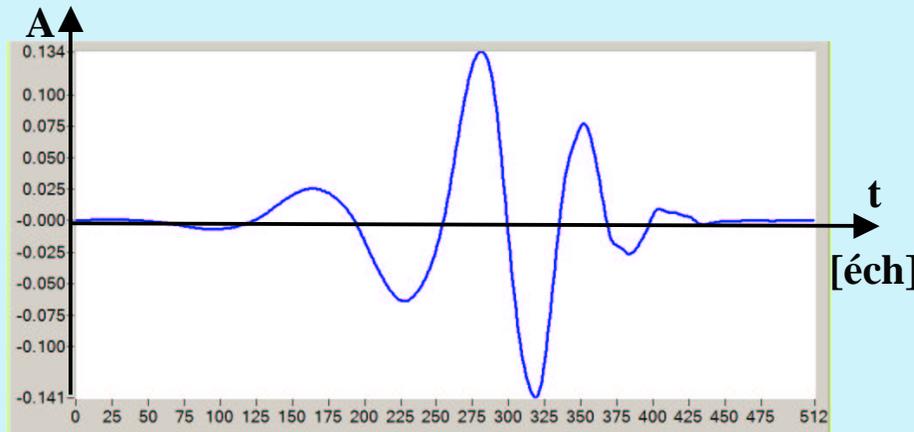
$C_1 - C_5$ coefficients de la transformée en ondelettes

Détection de sons

- Algorithme fondé sur la transformée en ondelettes (2)



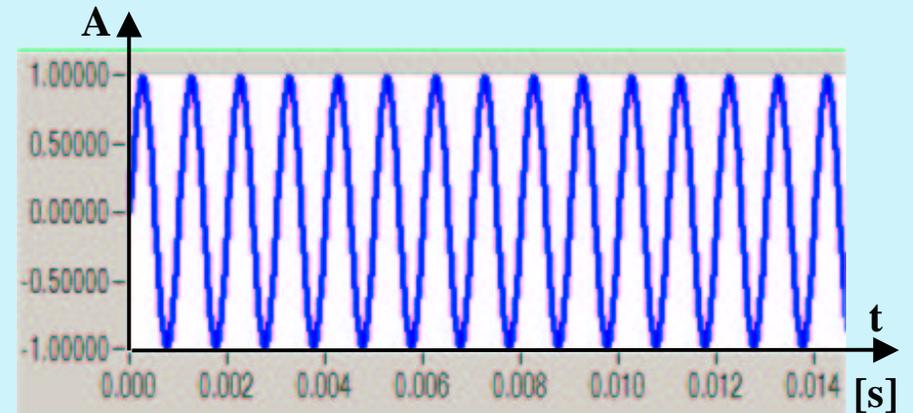
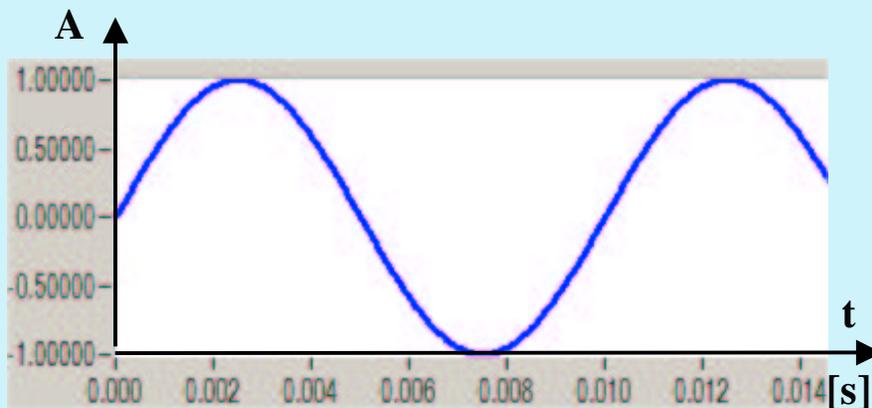
Ondelettes de Daubechies à 10 moments nuls



Ondelette – signal fini

Sinusoïde – signal périodique infini

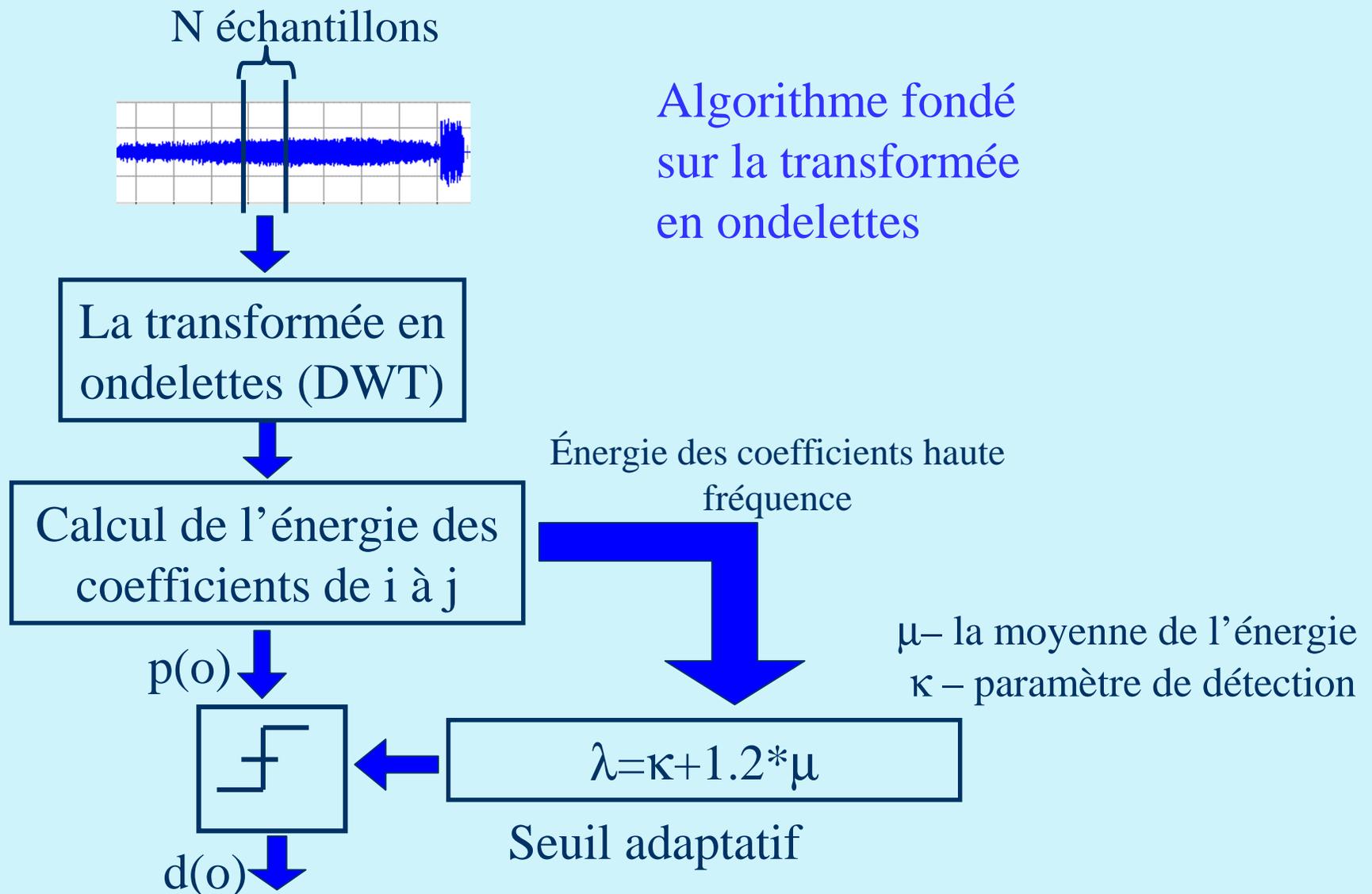
Sinusoïde de 100 Hz et 1000 Hz



Détection de sons



- Algorithme fondé sur la transformée en ondelettes (3)

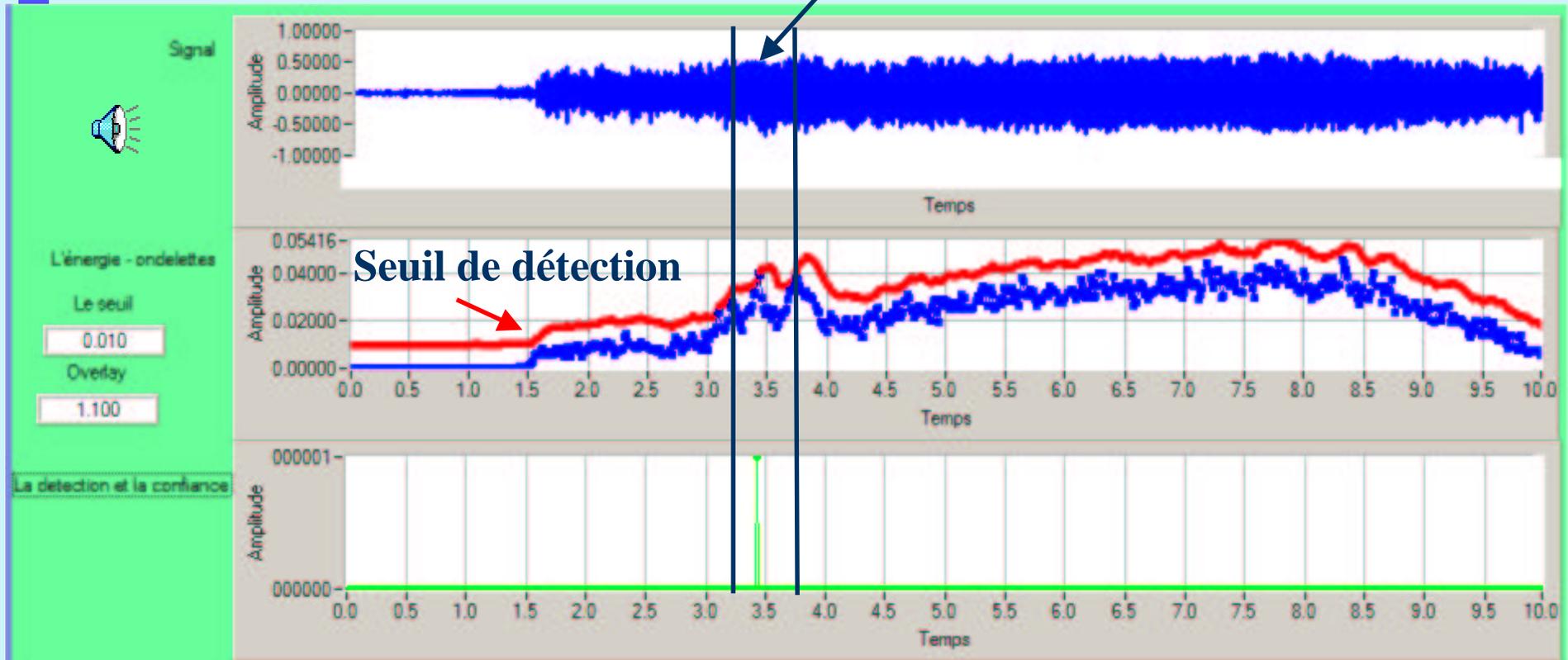




Détection de sons

- *Algorithme fondé sur la transformée en ondelettes (4)*

Sonnerie de téléphone dans un bruit d'eau



Plan de la présentation



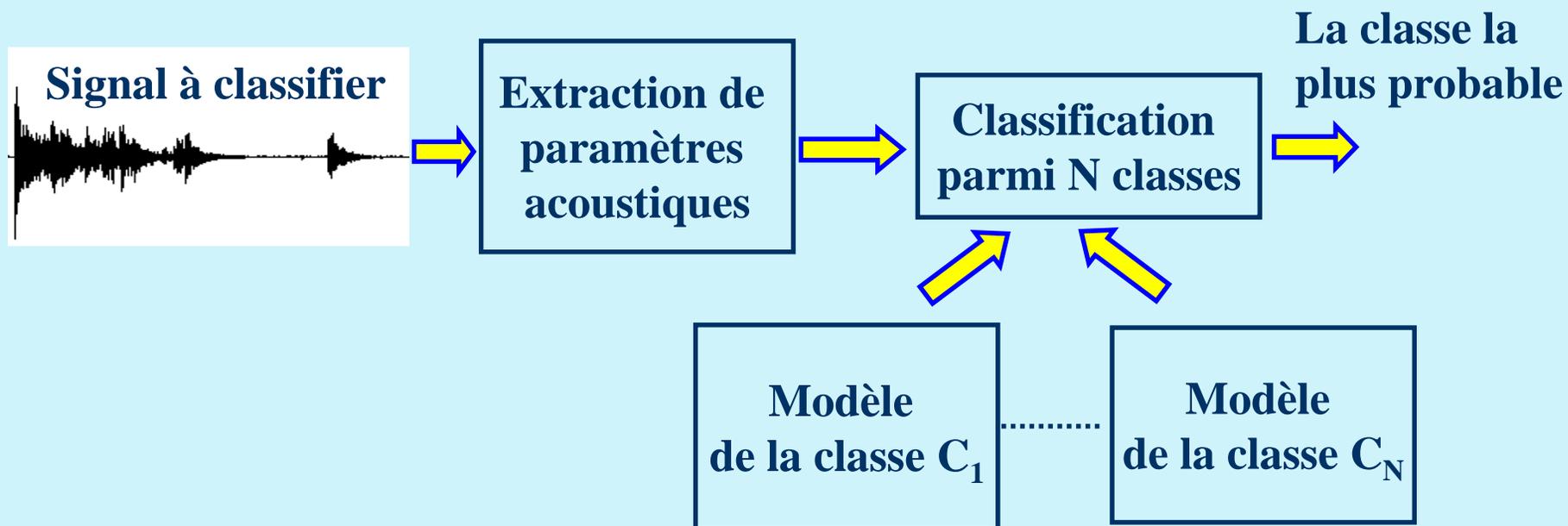
- Contexte de la thèse
- Objectif et problématiques de recherche
- Détection de sons
- Classification de sons
 - Introduction
 - Paramètres acoustiques proposés
- Couplage entre la détection et la classification
- Synthèse des résultats d'évaluation du système
- Conclusions et perspectives

Classification de sons

- Introduction



- Trouver l'appartenance d'un son, issu de la détection, à l'une des classes prédéfinies



Classification de sons

- *Choix de la méthode de classification*



- Méthode statistique :
 - Modèle de mélange de gaussiennes (GMM)
- Apprentissage supervisé

- Technique de modélisation statistique relativement indépendante du type de signal
 - Adaptation des paramètres acoustiques à l'application
- Sensibilité réduite à la coupure des signaux

- Corpus de taille réduite - > étiqueté

Classification de sons

- GMM



- La modélisation d'une classe de sons se fait par une somme pondérée de gaussiennes
- Apprentissage :
 - Algorithme des K-moyennes
 - Algorithme EM (Expectation- Maximization)
- Classification :
 - Probabilité d'appartenance d'une trame à la classe w_k :
$$p(x_i | w_k)$$
 - Probabilité d'appartenance d'un son à la classe w_k (la moyenne géométrique des probabilités des trames)
 - La classe reconnue = la vraisemblance la plus grande

Classification de sons

- *Extraction des paramètres acoustiques*



- Extraire les caractéristiques du signal
- Éliminer les redondances qui existent dans la représentation temporelle du signal
- Paramètres issus de la reconnaissance de la parole / du locuteur :
 - MFCC, LFCC
 - LPC, LPCC
 - Bancs de filtres
 - Énergie

Classification de sons

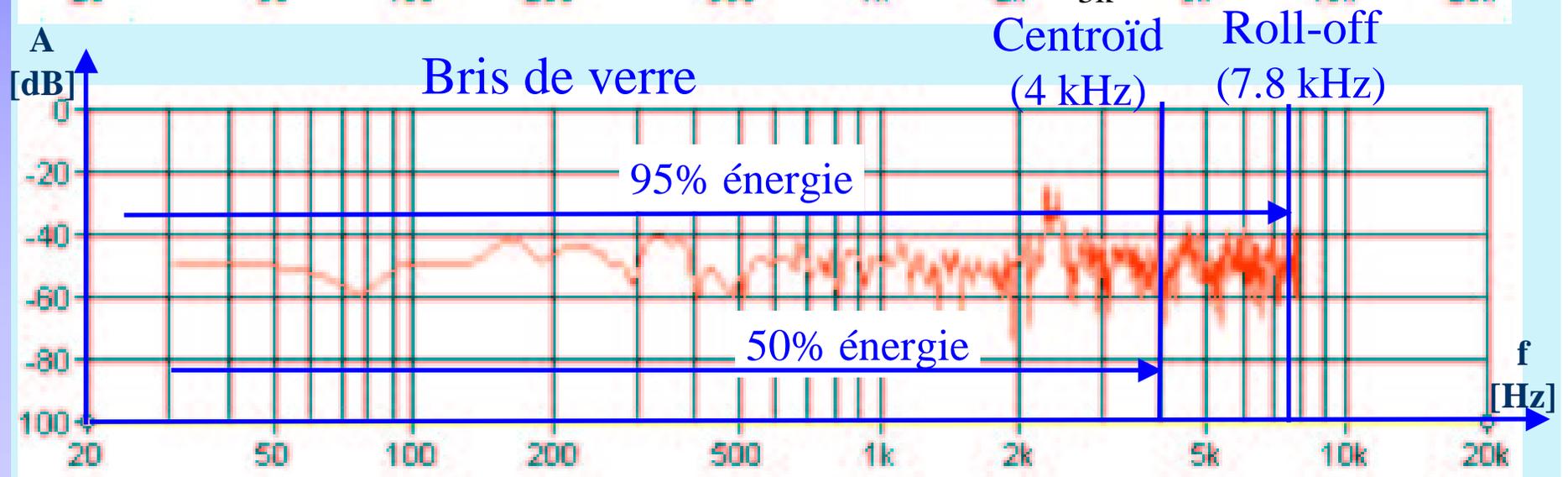
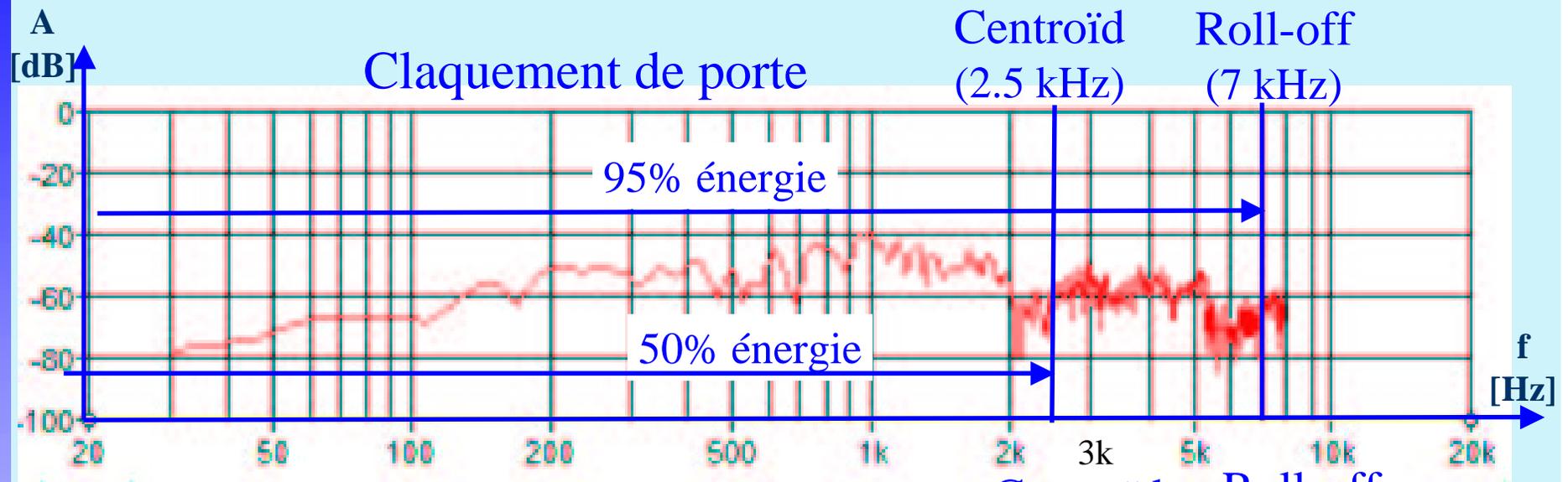
- *Paramètres acoustiques*



- Paramètres provenant de l'indexation parole/bruit/musique :
- Nombre de passages par zéro
 - « Roll-off Point » (fréquence au-dessous de laquelle se situe 95% de l'énergie)
 - « Centroid » spectral (fréquence qui partage le spectre en deux parties d'égale énergie)

Classification de sons

- Paramètres acoustiques (3)

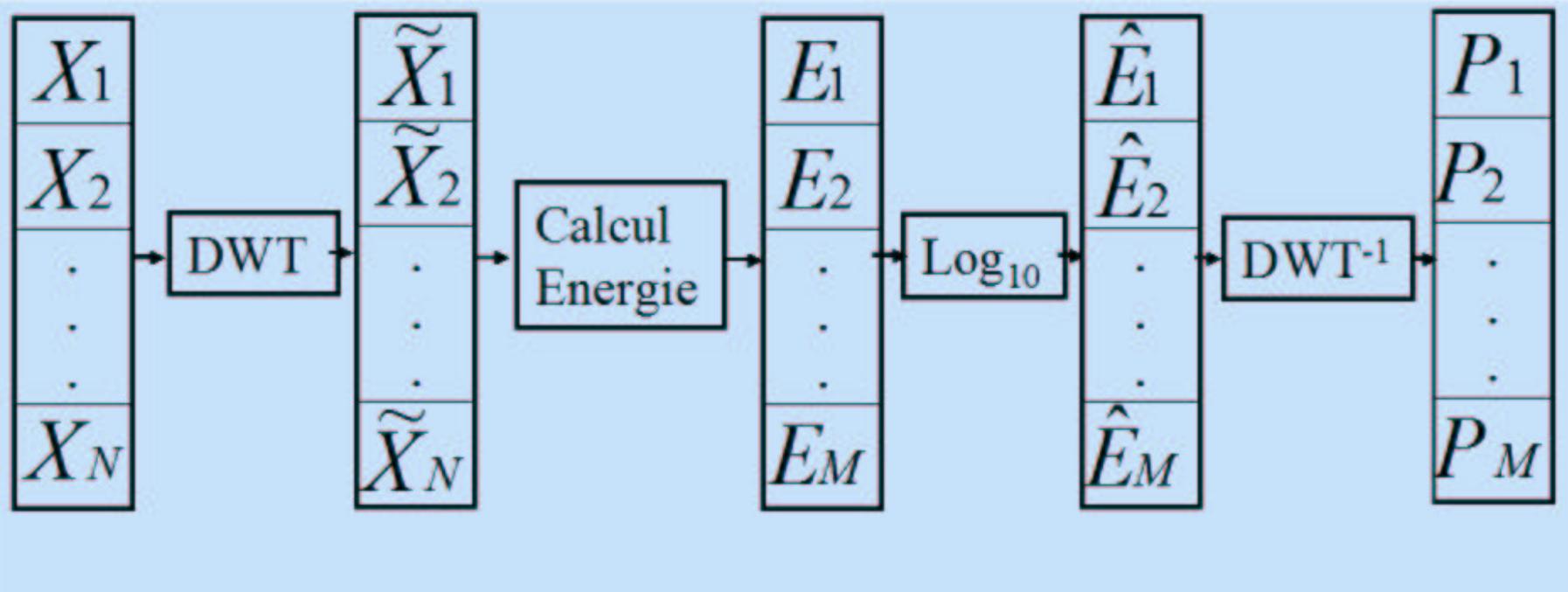


Classification de sons

- Paramètres acoustiques (3)



- Paramètres provenant de la transformée en ondelettes :
- Ecart-type, « skewness », « kurtosis », énergie
 - DWTC « Cepstraux » (analogie avec les MFCC)



Classification de sons

- *Sélection des paramètres acoustiques*



→ Critère de Fisher :

$$FDR = \frac{\sum_{i=1}^k \sum_{j=1}^k (\overline{x[i]} - \overline{x[j]})^2}{\sum_{i=1}^k Var(x)[i]}$$

- $\overline{x[i]}$ la moyenne du paramètre x pour la classe i
- $Var(x)[i]$ – la variance du paramètre x pour la classe i

Classification de sons

- Sélection des paramètres acoustiques (2)



Paramètre	FDR	Paramètre	FDR
MFCC1	2.72	MFCC14	3.61
MFCC2	16.07	MFCC15	3.26
MFCC3	10.33	MFCC16	4.41
MFCC4	10.02	ZCR	17.99
MFCC5	2.01	RF	16.70
MFCC6	2.91	Centroid	23.75
MFCC7	3.36	Energy	2.54
MFCC8	3.60	DWTC3	2.89
MFCC10	3.34	DWTC4	4.54
MFCC11	2.88	DWTC5	6.02
MFCC12	3.20	DWTC6	8.69

Classification de sons

- *Sélection du nombre de gaussiennes*



→ Utilisation du critère BIC (Bayesian Information Criterion) :

$$BIC_{m,K} = \underline{-2 \cdot L_{m,K}} + \underline{\nu_{m,K} \ln(n)}$$

- $L_{m,k}$ – maximum logarithmique de la vraisemblance
- $\nu_{m,K}$ – le nombre de paramètres libres
- n – nombre de trames

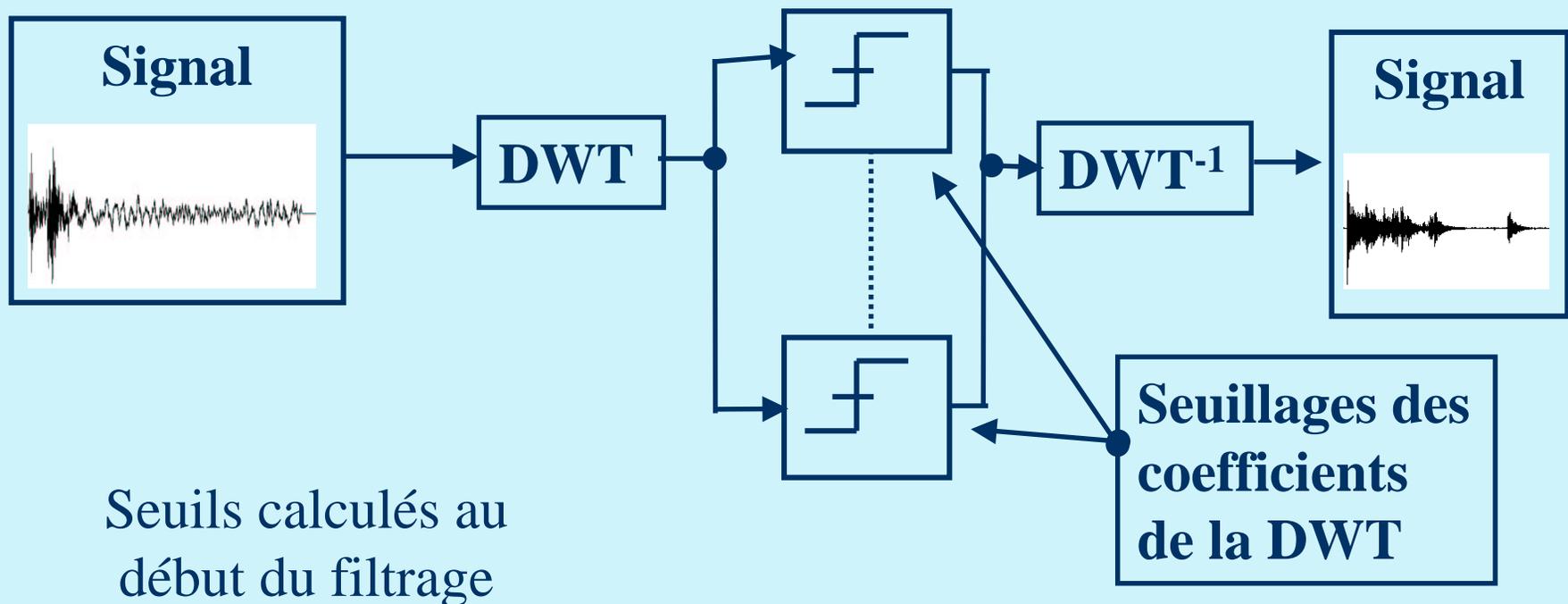
Nombre de gaussiennes	2	3	4	5	8
BIC	11043	10752	10743	10757	13373

- 16 coefficients MFCC, classe - Serrure de porte

Classification de sons - Atténuation du bruit

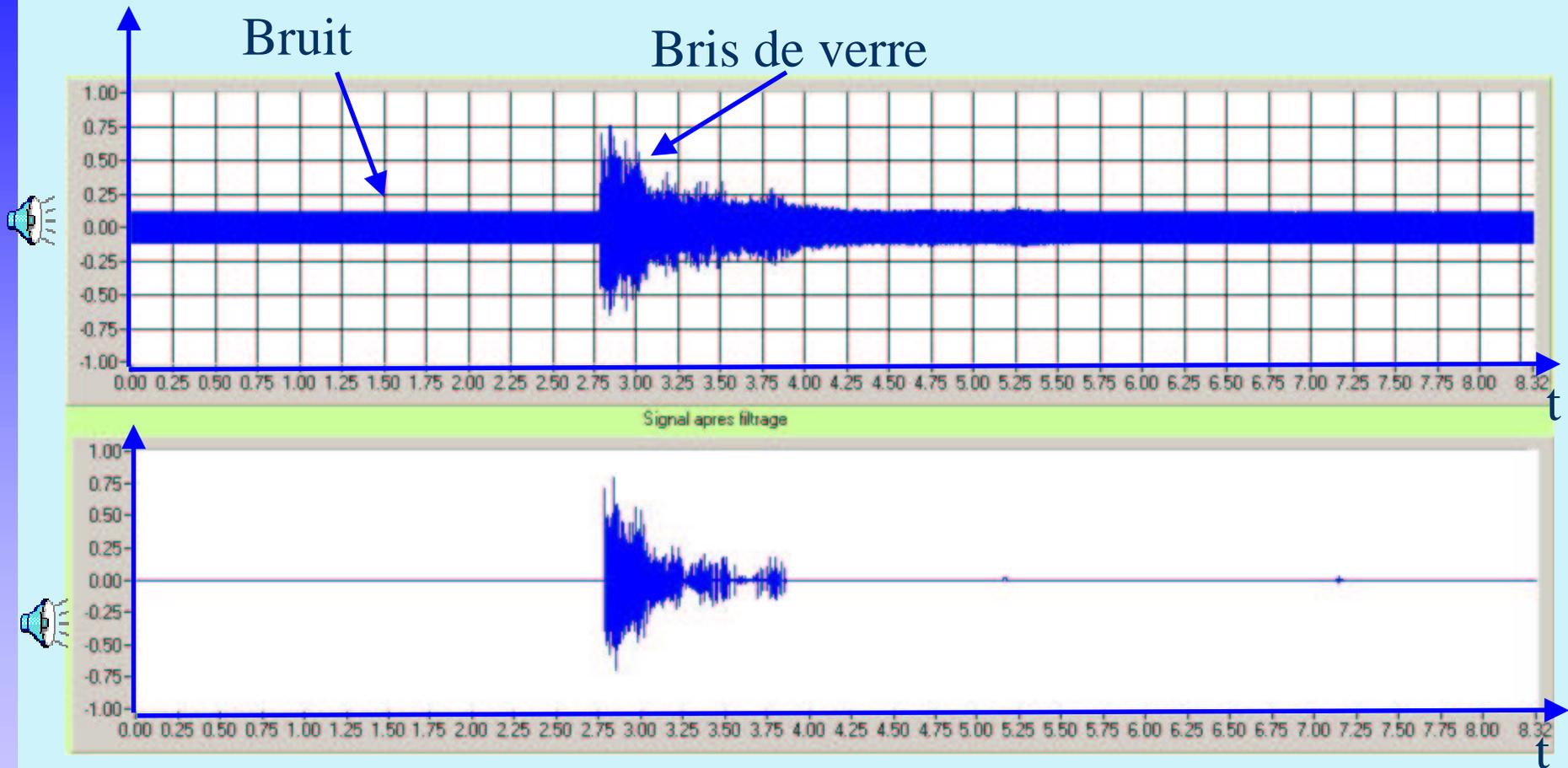


$$\text{Seuil} = \begin{cases} 1.2 * \text{Amplitude max du bruit} & \text{pour coefficients 1 à 4} \\ 0.9 * \text{Amplitude max du bruit} & \text{pour coefficient 5} \\ 0 & \text{pour coefficients 6 à 9} \end{cases}$$



Classification de sons

- Atténuation du bruit (2)

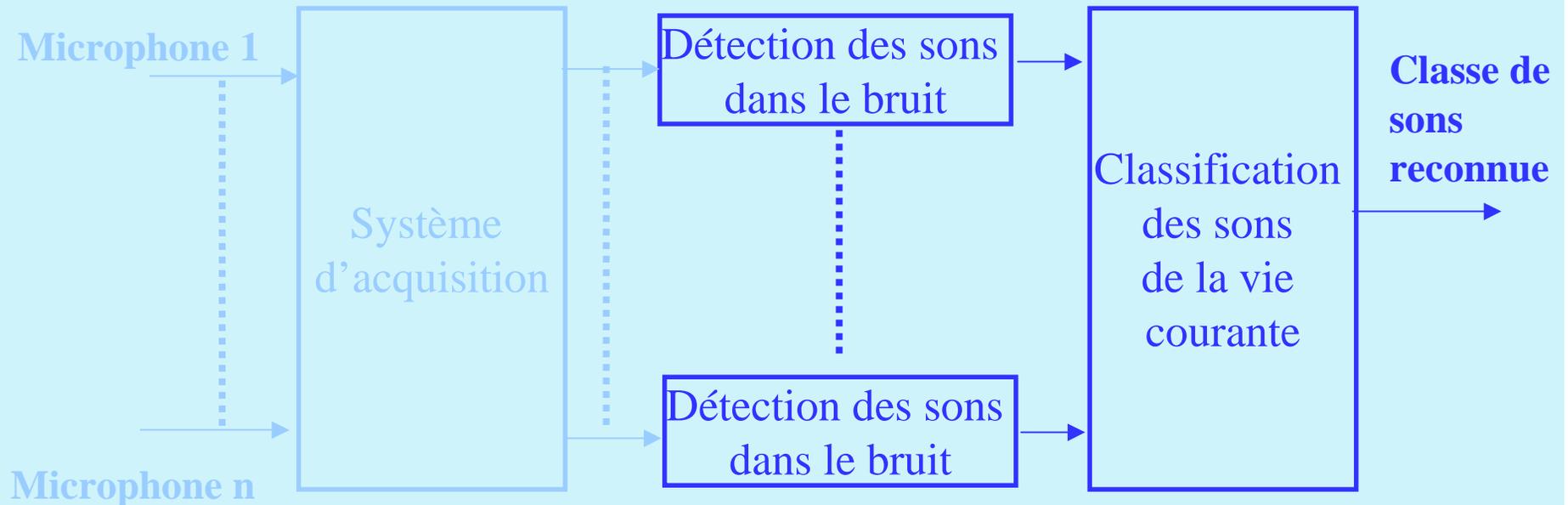


Plan de la présentation



- Contexte de la thèse
- Objectif et problématiques de recherche
- Détection de sons
- Classification de sons
- Couplage entre la détection et la classification
- Synthèse des résultats d'évaluation du système
- Conclusions et perspectives

Couplage entre la détection et la classification



→ Erreurs de la détection avec influence sur la classification :

- Détection avec retard
- Fausse alarme

→ Détection de la fin du signal :

- Même algorithme de détection sur le signal inversé dans le temps

Plan de la présentation



- Contexte de la thèse
- Objectif et problématiques de recherche
- Détection de sons
- Classification de sons
- Couplage entre la détection et la classification
- Évaluation du système
 - Corpus de validation
 - Protocoles de test
 - Synthèse de résultats
- Conclusions et perspectives

Évaluation du système

- *Corpus des sons de la vie courante*



→ Corpus des sons de la vie courante :

- enregistrements effectués dans le laboratoire CLIPS
- sons enregistrés par le laboratoire ATR (Japon)
- CD d'effets pour films
- 3354 sons, 1h35, 650 Mo
- $F_e=44.1$ kHz et 16 kHz

→ Bruit environnemental :

- enregistrement dans l'appartement d'étude
- par le système d'acquisition

Évaluation du système

- *Corpus de validation*



Différents corpus élaborés pour la validation du système

Tâche / Conditions	Détection	Classification	Couplage
Non Bruité	-	✓	-
Bruité (simulation du mélange)	✓	✓	✓
Bruité (mélange réel)	✓	-	-

Évaluation du système

- *Classes de sons de la vie courante*



→ Télésurveillance médicale :

- Sons pouvant indiquer une situation de détresse :
 cris, bris de verre ou de vaisselle, chute
- Sons liés à la physiologie :
 toux, éternuement
- Sons caractéristiques de l'activité :
 pas, claquement de porte

Évaluation du système

- Méthodologie d'évaluation de la détection



$$TDM = \frac{\text{Le nombre de détections manquées}}{\text{Le nombre de signaux à détecter}}$$

$$TFA = \frac{\text{Le nombre de fausses alarmes}}{\text{Le nombre de fausses alarmes} + \text{Le nombre de signaux à détecter}}$$

$$TEE = TDM \Big|_{TDM=TFA}$$

Évaluation du système

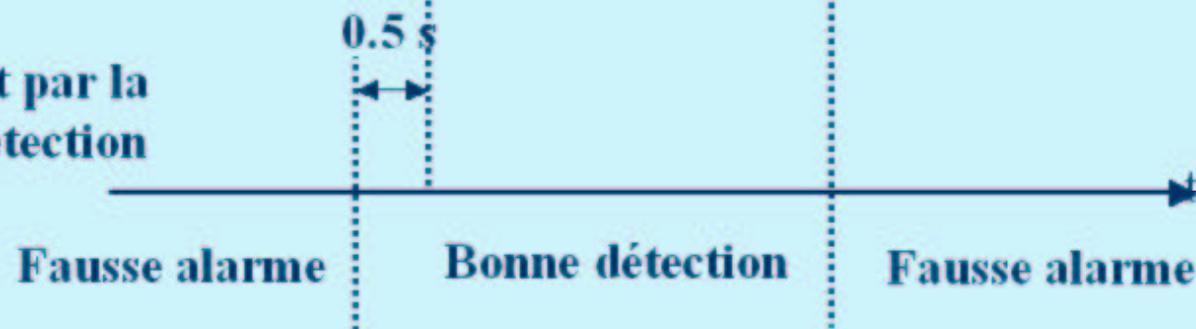
- Méthodologie d'évaluation de la détection (2)



Étiquetage de référence



Étiquetage fait par la méthode de détection

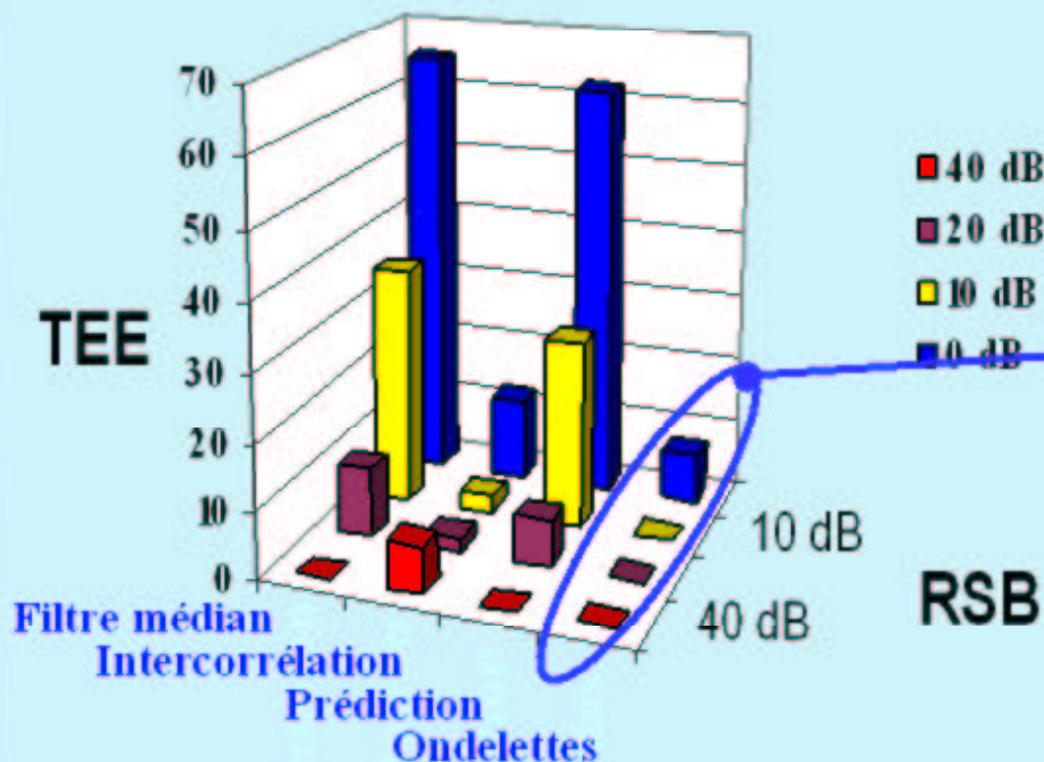


Réponse du système	Entrée	H_0 vraie	H_1 vraie
	H_0		«Non détection» correcte
H_1		Fausse alarme	Détection correcte

Évaluation du système - Résultats détection



Les TEE pour bruit HIS



Meilleur compromis, seuil adaptatif, temps de calcul réduit

Algorithme	Temps de calcul [ms/s]
Filtre médian	2 (\approx 1 MIPS)
Intercorrélation	55 (\approx 30 MIPS)
Prédiction	2 (\approx 1 MIPS)
Ondelettes	10 (\approx 5.5 MIPS)

Évaluation du système

- *Protocole pour la classification de sons*



→ Corpus de taille réduite

→ Protocole Leave-one-out :

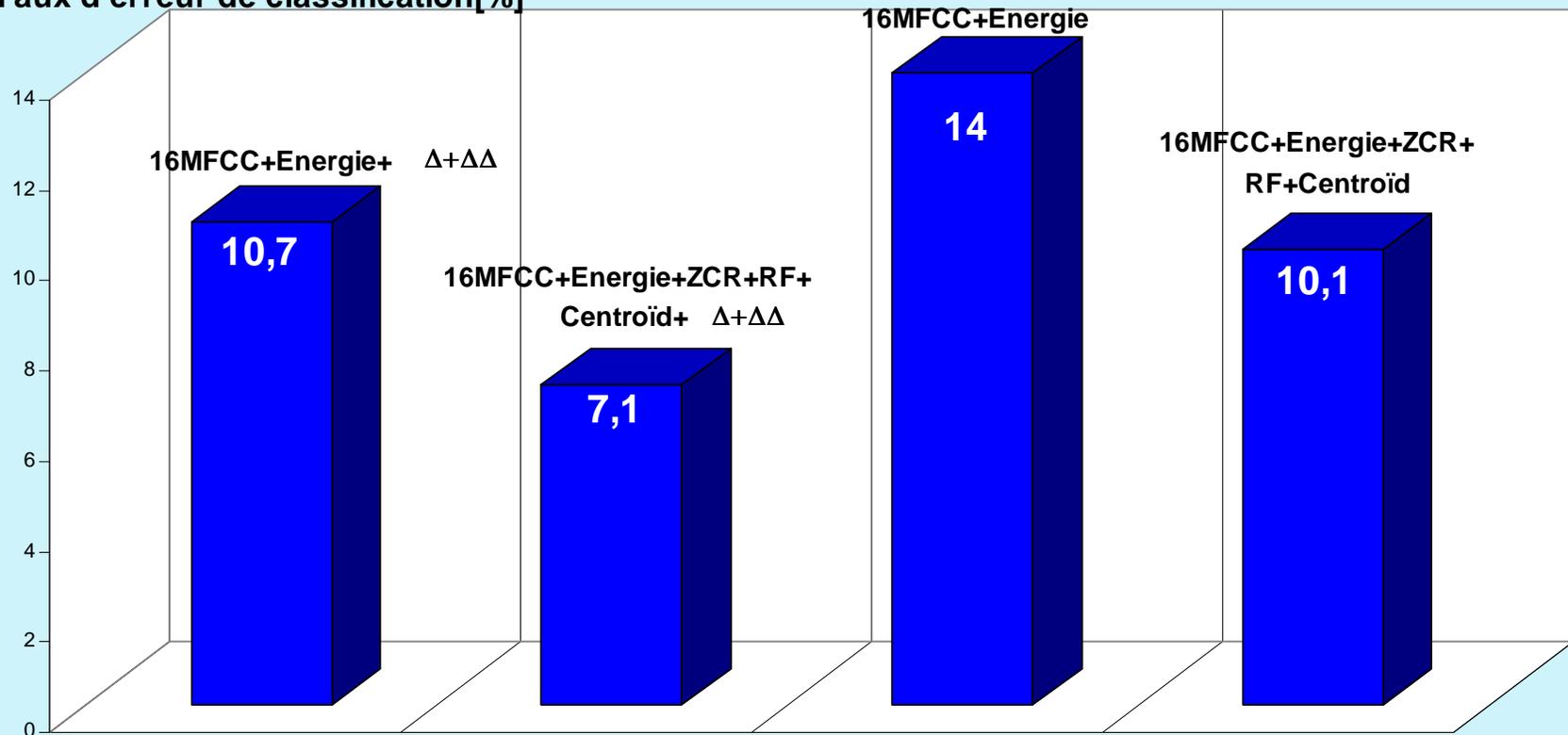
- dans une classe, le son utilisé pour le test est éliminé de l'ensemble d'apprentissage
- l'étape précédente est itérée pour tous les sons d'une classe, puis, pour toutes les classes

Taux d'erreur de classification = $\frac{\text{Nombre de sons non reconnus}}{\text{Nombre total de sons à reconnaître}}$

Évaluation du système - Résultats classification des sons



Taux d'erreur de classification[%]



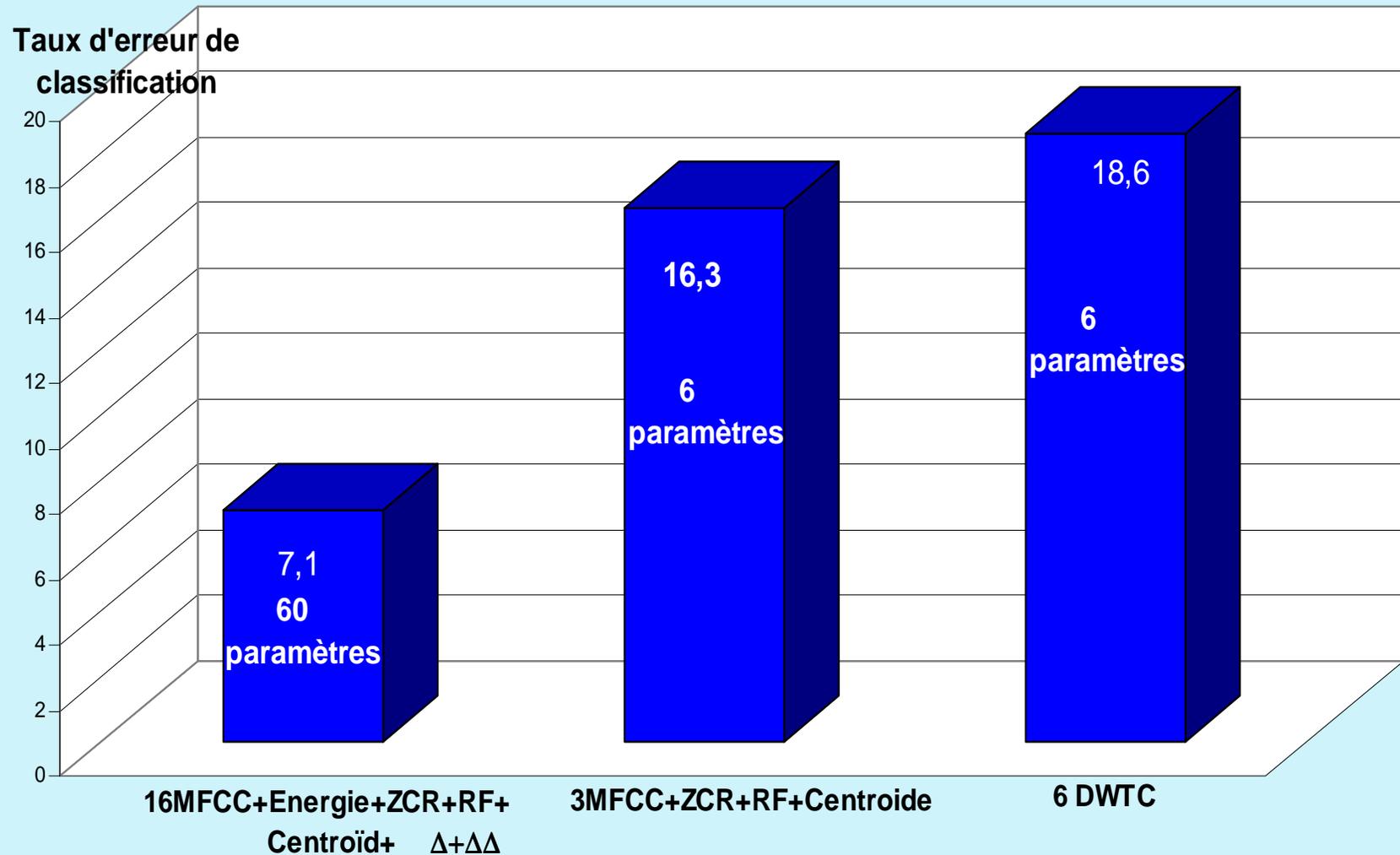
Leave-one-out, 4 gaussiennes, 1577 tests

Évaluation du système

- Résultats classification des sons (2)



Application embarquée = nombre réduit de paramètres



Évaluation du système

- Résultats atténuation du bruit



Seuillage des coefficients de la transformée en ondelettes

Taux d'erreur de classification

Avec/Sans filtrage	RSB				Moyenne
	0 dB	10 dB	20 dB	40 dB	
Avec	40.0%	20.5%	14.6%	11%	21.5
Sans	48.3%	27.2%	13.1%	11%	25

Gain de 8% en absolu

Évaluation du système - Résultats couplage



Pertes des performances dues au couplage ($\approx 6\%$)

	RSB $\in [10,20]$ dB Erreur moyenne de classification [%]	RSB $\in [0,40]$ dB Erreur moyenne de classification [%]
Détection idéale Durées réelles des signaux	21.5	22.7
Détection réelle Durées estimées des signaux	27.7	25.5

Évaluation du système - *Résultat global*



Classe en entrée \ Classe reconnue		\bar{A}				A		
		C1	C3	C4	C7	C2	C5	C6
\bar{A}	Claquements de porte (C1)	✓	-	-	-	DM		
	Sonneries de téléphone (C3)	-	✓	-	-			
	Sons de pas (C4)	-	-	✓	-			
	Serrures de porte (C7)	-	-	-	✓			
A	Bris de verre (C2)					✓	-	-
	Cris (C5)					-	✓	-
	Vaisselle (C6)					-	-	✓

→ TDM = 3%

- Assez important pour une application médicale (30 situations sur 1000)
- Fusion de données entre le système sonore et les autres capteurs médicaux

→ TFA = 12%

Plan de la présentation



- Contexte de la thèse
- Objectif et problématiques de recherche
- Détection de sons
- Classification de sons
- Couplage entre la détection et la classification
- Évaluation du système
- Conclusions et perspectives

Points forts et limitations du système



- Détection des sons en présence du bruit (7.6% d'erreur pour RSB=0 dB)
- Détection de la fin du signal
- Classification de 7 classes de sons (taux d'erreur de 7.1% sans bruit)
- TDM global de 3% et TFA global de 12%
- Classification difficile en présence de bruit fort (taux d'erreur de 46% à 0 dB)
- Absence de certaines classes (chute, toux, éternuement)

Conclusions et perspectives : Bilan



- Création d'un corpus de sons de la vie courante
- Utilisation des propriétés de la transformée en ondelettes tant pour la détection que pour la classification
 - filtrage du bruit
 - proposition de nouveaux paramètres acoustiques
- Combinaison entre les paramètres acoustiques issus de reconnaissance de la parole et ceux de segmentation parole/bruit/musique
- Proposition d'une méthode de détection de la fin du signal
- Proposition d'une méthodologie d'évaluation d'un système de télésurveillance médicale

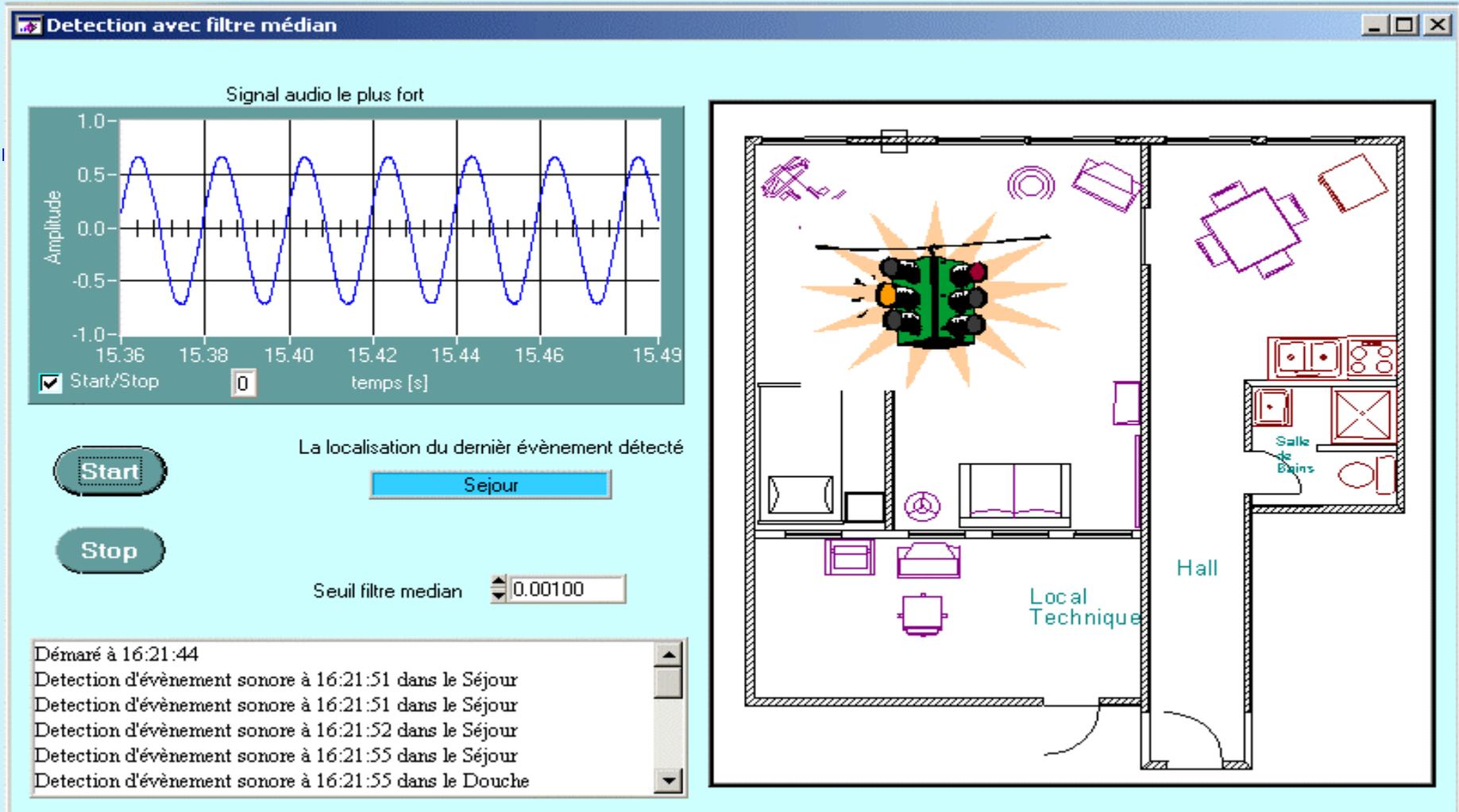
Conclusions et *Perspectives* :



- Évolution du système en 2 étapes vers un système à une seule étape : détection de sons clés
 - une première ébauche d'un tel système est encourageante
- Utilisation conjointe des signaux des voies sonores
- Fusion de données

- Extension de l'application :
 - Aide aux handicapés (COST 219ter)
 - Recherche de sons dans des bases de données (évaluation TREC)
 - Système de surveillance de bâtiments

L'interface du logiciel de détection en temps réel



CLIPS

Communication Langagière et
Interaction Pers onne-Sys tème

Fédération IMAG

BP 53 - 38041 Grenoble Cedex 9 - France

Dan Istrate

le 16 décembre 2003

Directeur de thèse : Eric Castelli

Co-Directeur : Laurent Besacier



GEOD

Groupe d'Etude sur l'Oral et le Dialogue