

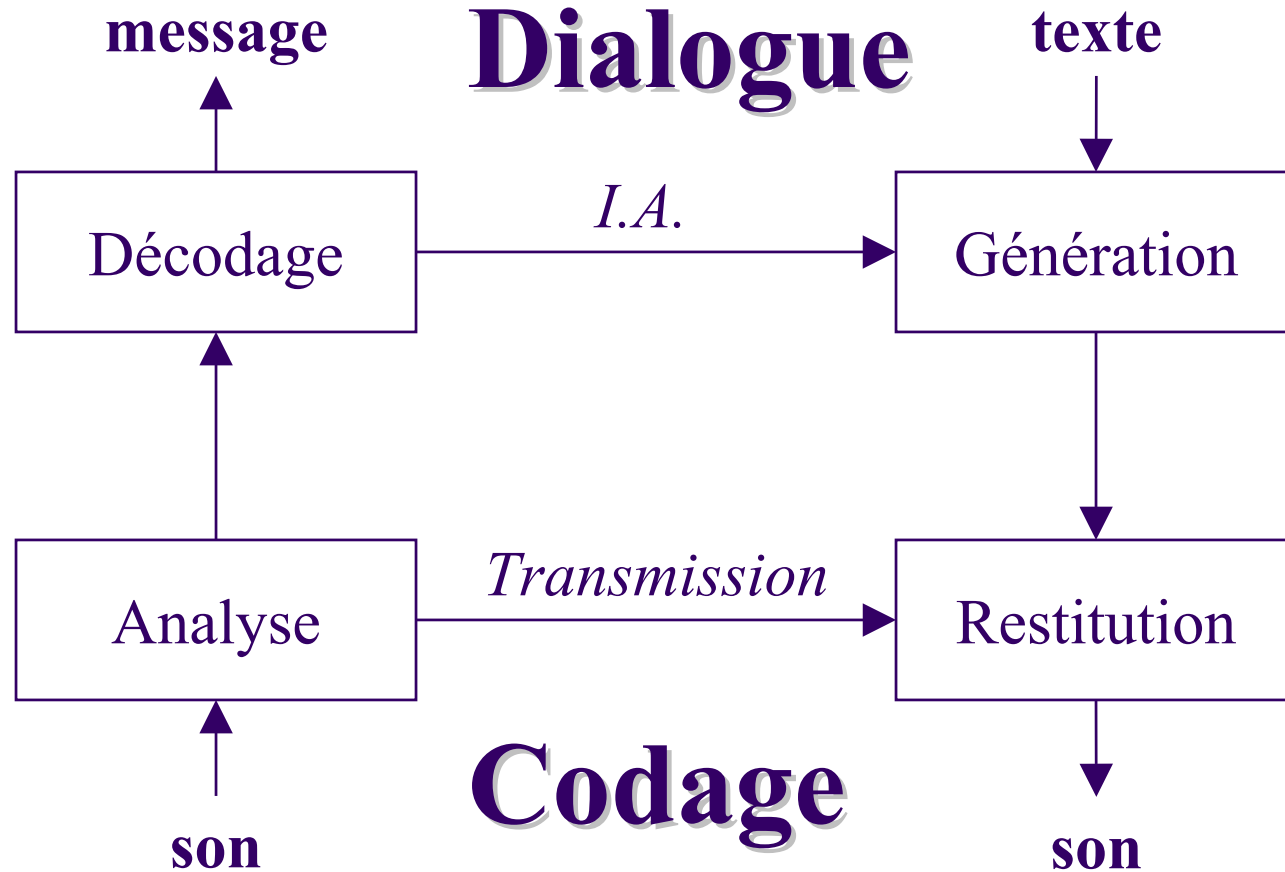
La détection Parole/Non-Parole dans les applications de reconnaissance vocale Environnements bruités - Parole continue

Arnaud MARTIN

22 octobre 2002

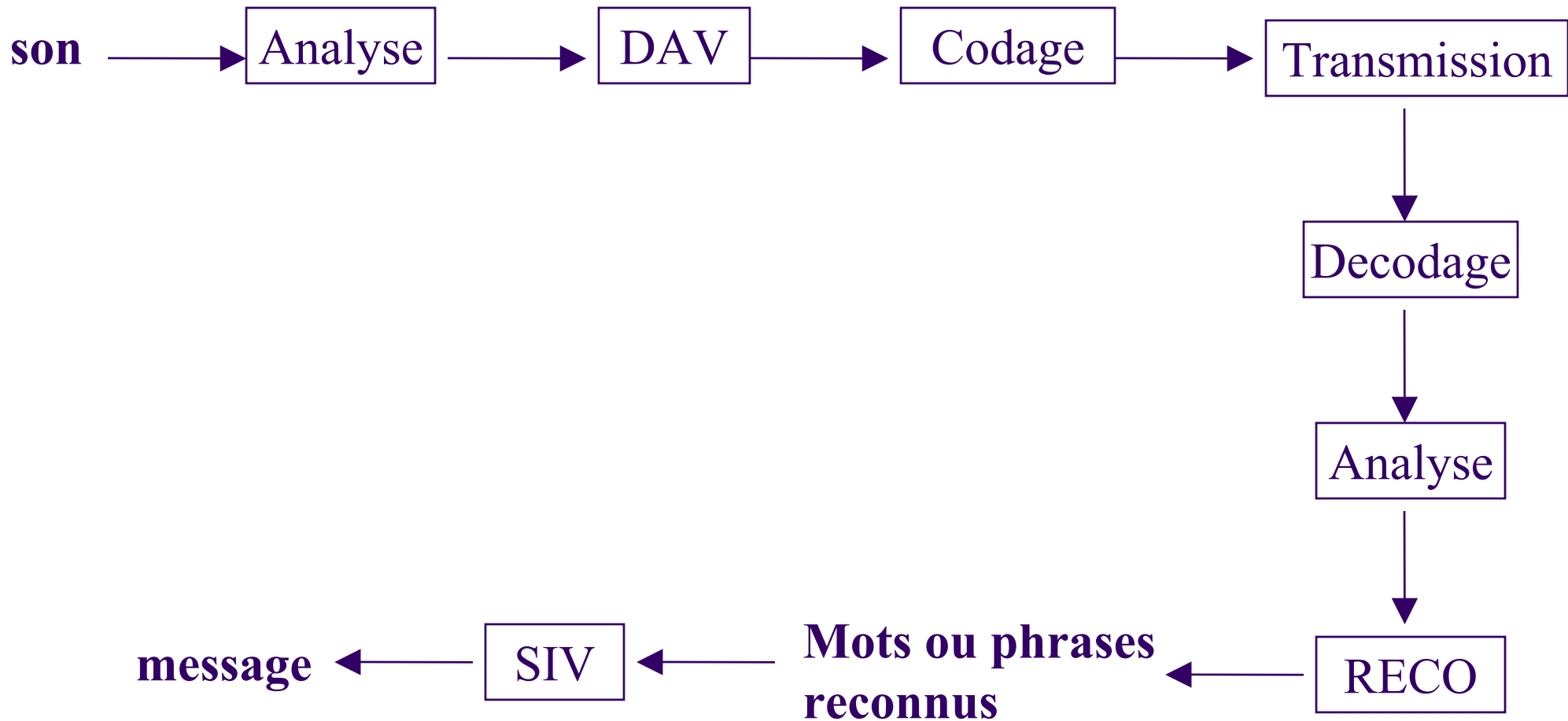
Domaines du Traitement de la Parole

Reconnaissance

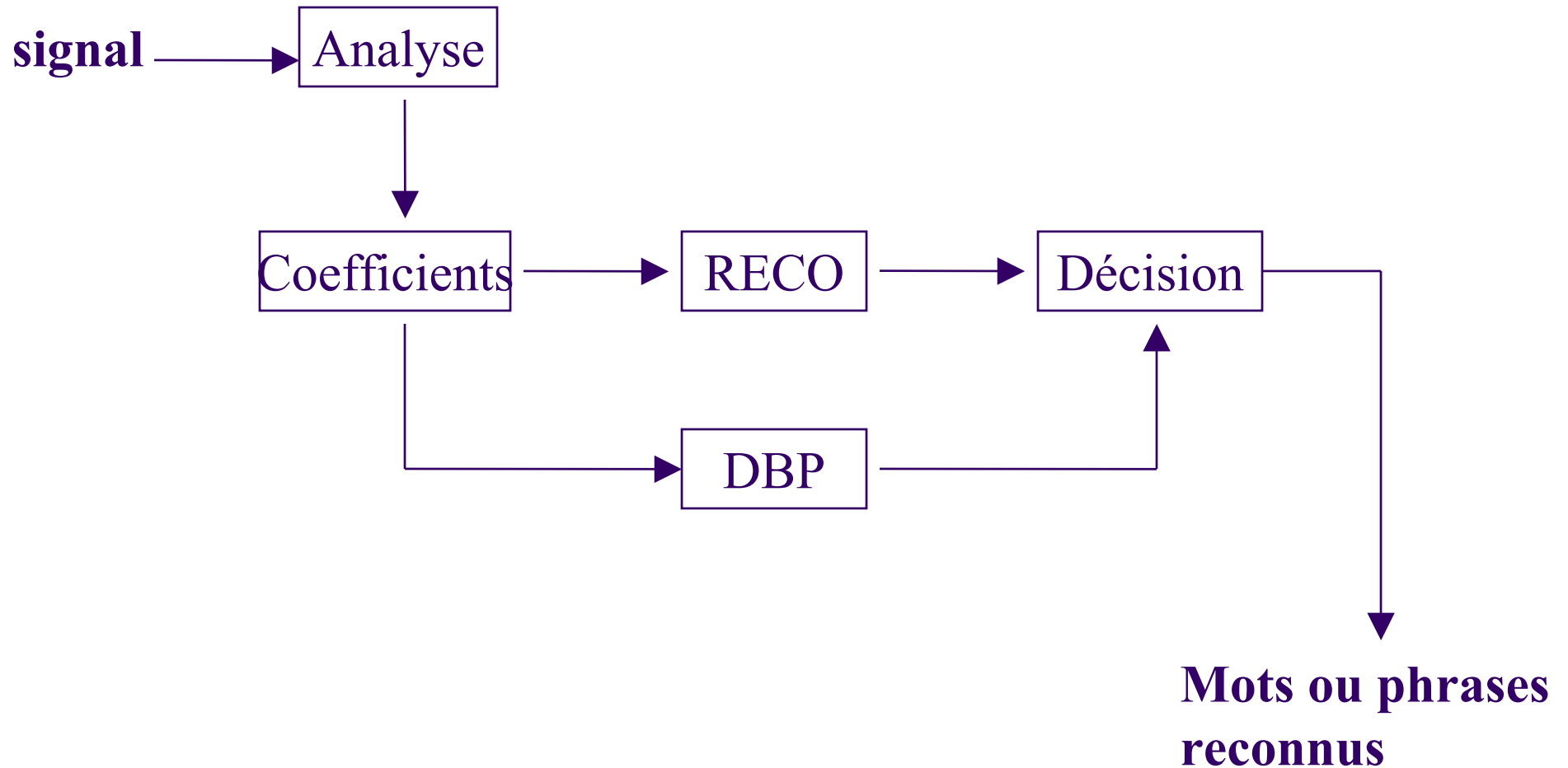


Synthèse

Appel à un Système Interactif Vocal (SIV)



Module de Reconnaissance



Différents contextes de la détection de parole

→ Transmission - Codage

→ Détection d'Activité Vocale (DAV)

faible délai – omissions non acceptées – fausses alarmes tolérées

→ Reconnaissance

→ Détection Bruit / Parole (DBP)

délai plus grand – compromis omissions/fausses alarmes
dépendant de la reconnaissance

Plan

- La reconnaissance de la parole
 - Approche reconnaissance des formes
 - Approche déterministe/statistique
 - Modèle à base de chaînes de Markov cachées

Plan

→ La reconnaissance de la parole

→ La détection de Parole

- Quelques caractéristiques acoustiques
- Quelques méthodes statistiques
- Problèmes actuels

Plan

- La reconnaissance de la parole
- La détection de Parole

- Principe d'évaluation de la détection
 - Évaluation du module de détection
 - Évaluation de la détection dans le système de reconnaissance

Plan

- La reconnaissance de la parole
- La détection de Parole
- Principe d'évaluation de la détection

- Approches précédemment développées dans le laboratoire
 - Automate Bruit/Parole
 - 3 critères utilisés
 - Évaluation des 3 critères

Plan

- La reconnaissance de la parole
- La détection de Parole
- Principe d'évaluation de la détection
- Approches précédemment développées dans le laboratoire

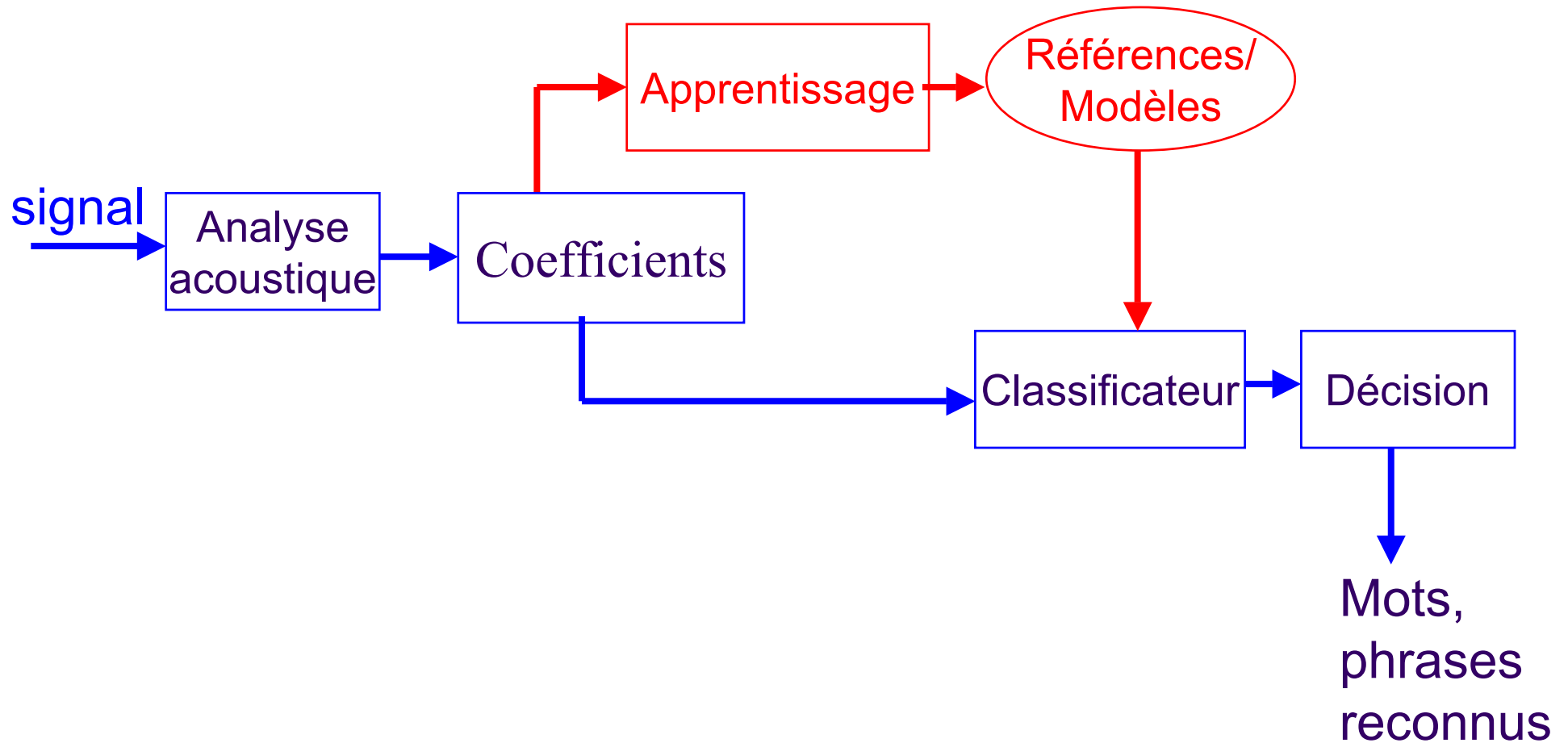
- Amélioration du module de détection
 - Prise en compte d'une nouvelle condition
 - 3 voies envisagées
 - Les statistiques d'ordre supérieur
 - Un paramètre de voisement
 - L'analyse factorielle discriminante
 - Utilisation d'un module de débruitage

Plan

- La reconnaissance de la parole
- La détection de Parole
- Principe d'évaluation de la détection
- Approches précédemment développées dans le laboratoire
- Amélioration du module de détection

- Conclusion

Approche reconnaissance des formes



Approche reconnaissance des formes

→ Caractéristiques :

- ⑦ performances du système dépendent de la quantité de données d'apprentissage.
- ⑦ Les formes de références dépendent des conditions de prise de son et de transmission.
- ⑦ Pas de connaissance spécifique requise.
- ⑦ Coût proportionnel au nombre de références.
- ⑦ Système indépendant de la modélisation acoustique choisie.
- ⑦ Peut intégrer facilement des contraintes syntaxiques, sémantiques...

Reconnaissance de la parole :

Approche déterministe/statistique

→ Approche par DTW (déterministe) :

⑦ Vecteurs moyens difficiles à définir

⑦ Néglige l'effet de la dispersion autour des moyennes

→ Approche statistique :

⑦ Prise en compte de la dispersion autour des vecteurs moyens par l'introduction de distributions de probabilité

⑦ Outil : modèles de Markov cachés (*HMM* : Hidden Markov Model)

Formulation déterministe (DTW)

$$\hat{W} = \arg \min_W d(X, Y_W)$$



Distance entre les formes
 X et Y_W

X = Séquence de trames (signal à reconnaître)

Y_W = Séquence de trames (forme de référence) du mot W

\hat{W} = Mot reconnu

Formulation probabiliste (HMM)

$$\hat{W} = \arg \max_W \Pr(W | X)$$

$$= \arg \max_W \Pr(X | W) \Pr(W)$$

Modèle acoustique

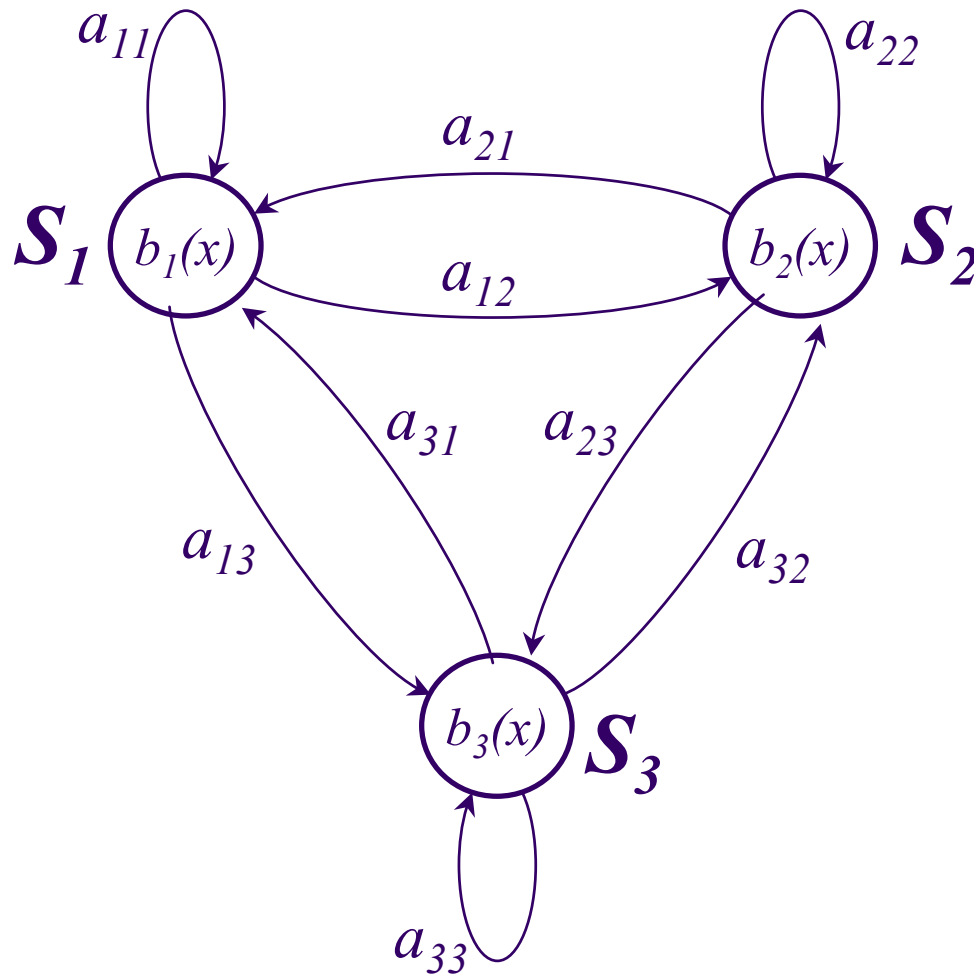
Modèle de langage

X = Séquence de trames (signal à reconnaître)

W = Mots du vocabulaire

\hat{W} = Mot reconnu

Chaînes de Markov cachées (HMM)



Paramètres λ :

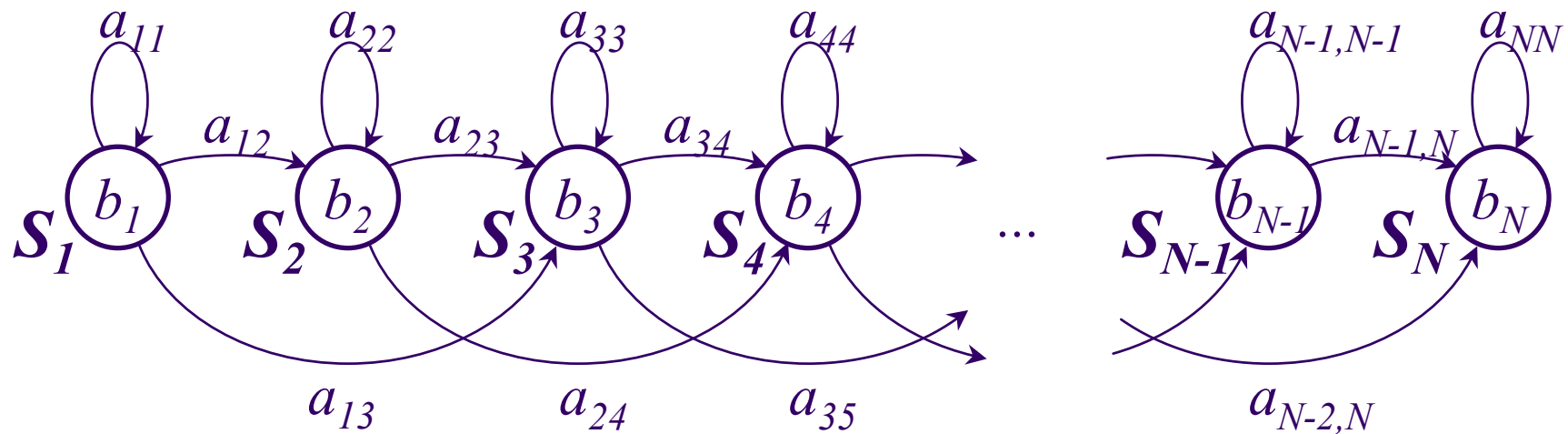
S_1, S_2, S_3 : états

π_i : probabilité d'être initialement dans l'état S_i

a_{ij} : probabilité de transition de l'état S_i à l'état S_j

$b_i(x) = P(x | S_i)$: probabilité d'émission de x dans l'état S_i

Modèle gauche-droite : modèle de Bakis



Modélise les variabilités fréquentielles et temporelles du signal de parole

Emission d'une observation

→ Un processus markovien permet « d'émettre » une observation :

$$X = x_1 x_2 x_3 \dots x_T$$

→ A partir de X , impossible de retrouver la séquence d'états Q

$$Q = q_1 q_2 q_3 \dots q_T$$

→ Probabilité d'émission de X par une séquence Q :

⑦ Pour une trame : $b_{q_i}(x) = P(x | q_i)$

⑦ Pour X : $P(X, Q | \lambda) = P(X | Q, \lambda) \cdot P(Q | \lambda)$

$$P(X | Q, \lambda) = b_{q(1)}(x_1) \cdot b_{q(2)}(x_2) \dots b_{q(T)}(x_T)$$

$$P(Q | \lambda) = \pi_{q(1)} \cdot a_{q(1),q(2)} \cdot a_{q(2),q(3)} \dots a_{q(T-1),q(T)}$$

HMMs : 3 problèmes fondamentaux

→ Trois problèmes fondamentaux :

- ⑦ Calculer la vraisemblance d'une observation $P(X | \lambda)$
- ⑦ Étant donnée une observation, trouver la séquence d'états optimale
- ⑦ Comment ajuster au mieux les paramètres du modèle (apprentissage)

HMMs : problème 1

$$P(X | \lambda) = \sum_Q P(X, Q | \lambda) = \sum_Q \pi_1 \cdot a_{12} \cdot P(x_1 | q_1) \cdots a_{T-1, T} \cdot P(x_T | q_T)$$

avec $X = x_1 x_2 x_3 \dots x_T$

→ Calcul direct infaisable : trop de séquences d'états Q à parcourir...

→ Calcul par l'intermédiaire des probabilités avant ou arrière :

$$\alpha_t(i) = P(x_1, \dots, x_t, q_t = i | \lambda)$$

$$\beta_t(i) = P(x_{t+1}, \dots, x_T | q_t = i, \lambda)$$

$$\alpha_{t+1}(j) = \sum_{i=1}^N \alpha_t(i) a_{ij} b_j(x_{t+1})$$

$$\beta_t(i) = \sum_{j=1}^N \beta_{t+1}(j) a_{ij} b_j(x_{t+1})$$

$$P(X | \lambda) = P(x_1, \dots, x_T | \lambda) = \sum_{i=1}^N \alpha_T(i) \quad P(X | \lambda) = P(x_1, \dots, x_T | \lambda) = \sum_{i=1}^N \beta_1(i)$$

Avec $N =$ nombre d'états du modèle λ

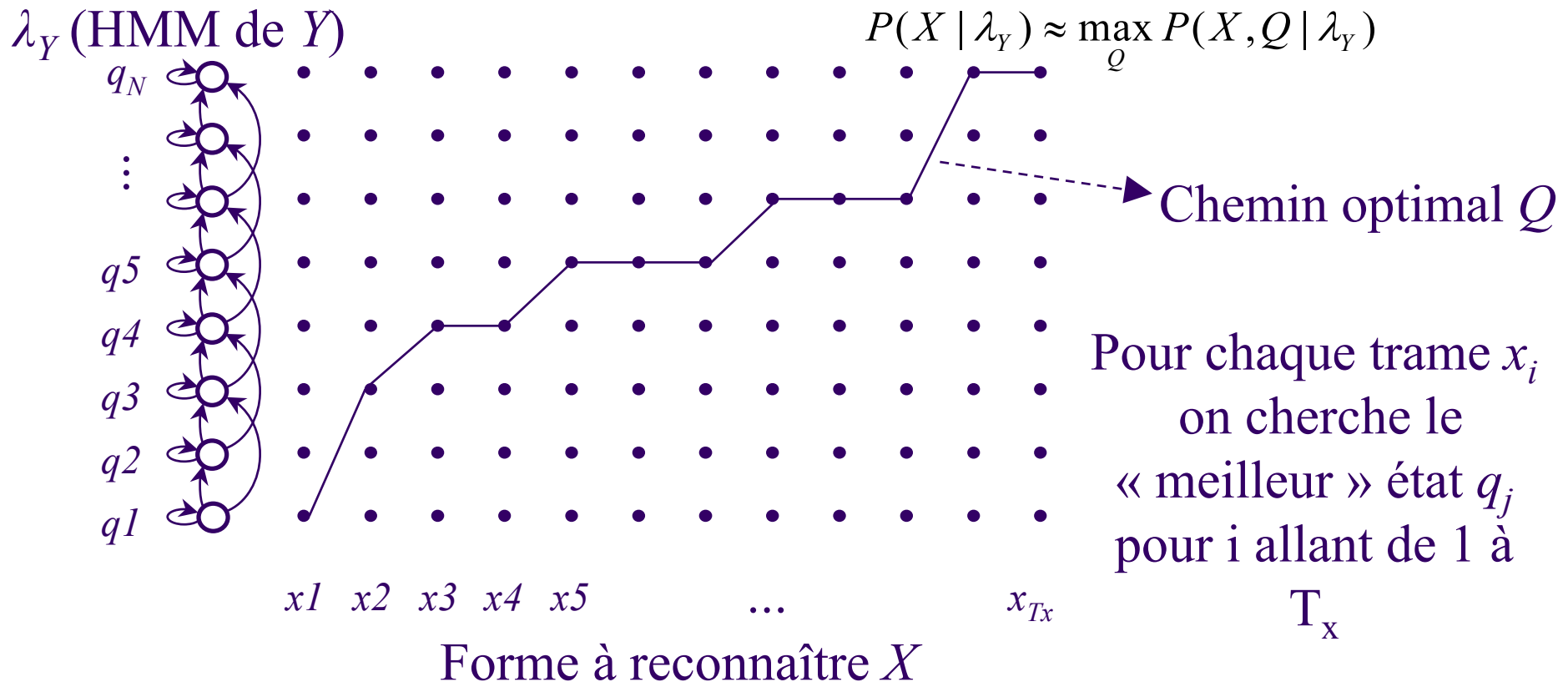
HMMs : problème 2

→ Séquence la plus probable :

$$\begin{aligned}\hat{Q} &= \arg \max_Q P(Q | X, \lambda) \\ &= \arg \max_Q P(X, Q | \lambda)\end{aligned}$$

Problème 2 : Programmation dynamique

➔ Recherche du chemin optimal dans le réseau de points :



Problème 2 : Viterbi

→ Algorithme de Viterbi (programmation dynamique):

⑦ Initialisation : $\delta_1(i) = \pi_i b_i(x_1) \quad 1 \leq i \leq N$

⑦ Récurrence : $\delta_t(j) = \max_i [\delta_{t-1}(i) a_{ij}] b_j(x_{t-1})$

⑦ Fin : $\hat{P}(X, Q | \lambda) = \max_i \delta_T(i)$

⑦ Chemin optimal : états sauvegardés à chaque itérations

→ Passage dans le domaine logarithmique : évite les multiplications et le calcul d'une exponentielle

HMMs : problème 3

→ Trouver le jeu de paramètres qui maximise la vraisemblance des données :

$$\hat{\lambda}_W = \arg \max_{\lambda_W} P(X | \lambda_W)$$

Les données d'apprentissage X sont des prononciations du mot W qu'on désire modéliser par le HMM λ_W

→ Solution 1 : Apprentissage de Viterbi

→ Solution 2 : Algorithme de Baum-Welch

Problème 3 : Viterbi

→ Initialisation

- ⑦ Définir un alignement initial (= correspondances vecteurs acoustiques x / états q)
- ⑦ En déduire les probabilités de transition a_{ij} (fréquence d'utilisation) et les distributions d'émission $b_i(x)$ (à partir des vecteurs associés à chaque état)

→ Itérations

- ⑦ Calcul par Viterbi du meilleur alignement pour chaque donnée d'apprentissage
- ⑦ Recalcul des probabilités et distributions d'émission en fonction des nouveaux alignements

Problème 3 : EM / Baum-Welch

→ EM = *Expectation Maximisation*

→ 2 étapes :

⑦ étape **estimate** : calcul de la fonction

$$Q(\lambda', \lambda) = \sum_Q P(X, Q | \lambda') \log(P(X, Q | \lambda))$$

⑦ étape **maximize** : calcul du maximum de la fonction Q, par rapport à λ .

HMMs : apprentissage

→ Apprentissage avec plusieurs répétitions :

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmax}} p(\mathbf{X}^1, \dots, \mathbf{X}^R | \lambda) = \underset{\lambda}{\operatorname{argmax}} \prod_{r=1}^R p(\mathbf{X}^r | \lambda)$$

→ La fonction Q globale est une somme sur l'ensemble des répétitions de fonctions Q.

→ Etape E : calcul des probabilités a posteriori de transitions et d'occupations des états.

→ Etape M : maximisation de la fonction Q.

La détection de parole

- Quelques caractéristiques acoustiques
- Quelques méthodes statistiques
- Problèmes actuels

Quelques caractéristiques acoustiques pour la détection

→ L'énergie

Très discriminant en environnement peu bruité

→ Représente bien les voyelles

→ Le taux de passage par zéro [Rabiner *et al.* 1975] [Junqua *et al.* 1994]

→ Détecter les fricatives faibles aux frontières de mots
(Important si modèle de rejet mauvais)

→ L'abscisse curviligne [Puel 1997]

→ Représente l'amplitude du signal et sa fréquence

Quelques caractéristiques acoustiques pour la détection

- La fréquence fondamentale [Hamada *et al.* 1990] [Ramana Rao *et al.* 1996]
 - représente la périodicité du signal
- L'entropie [Abdallah *et al.* 1997]
 - détecter périodes d'instabilité (changement d'état)

Quelques méthodes statistiques

→ La décision

- La décision bayésienne : test d'hypothèses [Karray et Monné 1998] [Arslan et Hansen 1998] [Singh et al. 2001]
 - hypothèses sur les distributions
- Opérateurs logiques (ET, OU) : [Rabiner et Sambur 1977]
 - combine 2 distances LPC
 - Équivalent à une mesure de probabilité
- La logique floue [Mwangi et Xydeas 1985] [Cavallaro et al. 1998]
 - définir des règles
 - Revient à ordonner des connaissances heuristiques
- Arbres de décision [Shin et al. 2000]
 - fusion des décisions de 6 caractéristiques
 - Équivalent à fusion en entrée

Quelques méthodes statistiques

→ La fusion en entrée

- Mesures de probabilité [Atal et Rabiner 1976] [Hörmann et Rozinaj 1998]

→ Matrice de covariance

- Fusion d'un grand nombre de données

- Réseaux neuronaux [Bendiksen *et al.* 1990] [Cohn 1991]

→ Classificateur

- En entrée : l'énergie dans différentes fréquences
- En sortie : voisé / non-voisé

Problèmes actuels

→ 3 problèmes principaux

- La détection de parole dans le bruit
- La détection de parole pour la reconnaissance de parole continue (Applications dialogue Homme-Machine)
 - silence inter-mots peut être long
 - rejet des détections de bruits moins performant
- La robustesse du seuil de détection

→ But de cette étude

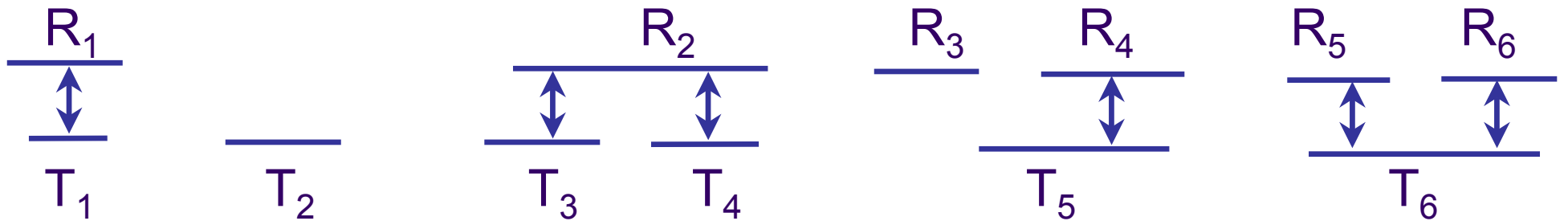
Trouver un module de détection Bruit/Parole qui résout ces problèmes

Principe d'évaluation de la détection de parole

- Évaluation du module de détection
Comparaison à la segmentation manuelle
- Évaluation de la détection dans le système de reconnaissance
(module de reconnaissance + module de détection)
- Bases de données utilisées

Évaluation du module de détection

→ Comparaison à la segmentation manuelle

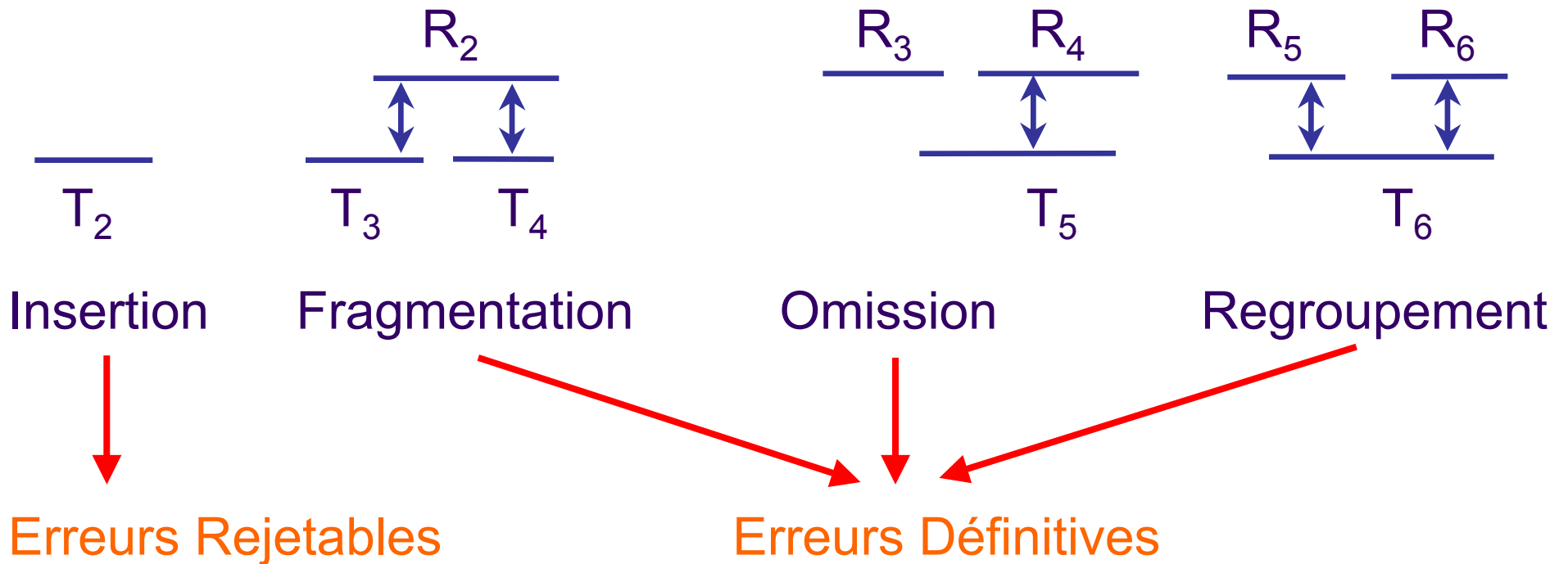


→ Règle de mise en relation

$$R_i \leftrightarrow T_j \text{ si } \frac{\min(L(R_i), L(T_j))}{2} \leq \|L(R_i \cap T_j)\|$$

→ Comptage des erreurs

Évaluation du module de détection



Détections imprécises (segments tronqués ou élargis)

Évaluation de la détection dans le système de reconnaissance

	Mots isolés	Parole continue
Rejet à tort	détection d'un mot du vocabulaire rejetée par le modèle de rejet + <u>parole non détectée</u>	détection d'une requête rejetée par le modèle de rejet + <u>parole non détectée</u> (exprimé en omission de mots)
Fausse Acceptation	détection de parole hors vocabulaire ou non parole reconnu comme un mot du vocabulaire	
Insertion		insertion d'un mot dans une requête
Substitution	un mot du vocabulaire pris pour un autre	un mot du vocabulaire pris pour un autre
Omission		omission de mot dans un requête

Bases de données

Bases	Réseau	Vocabulaire	Segments de parole
RTC_A	RTC	25	7144
GSM_A	GSM	53	22977
GSM_T RSB<18 dB RSB>18 dB	GSM	65	26366 11328 15038
AGORA	RTC	1633	2520 énoncés 12635 mots

Approches précédemment développées dans le laboratoire

→ Automate Bruit/Parole

→ 3 critères utilisés

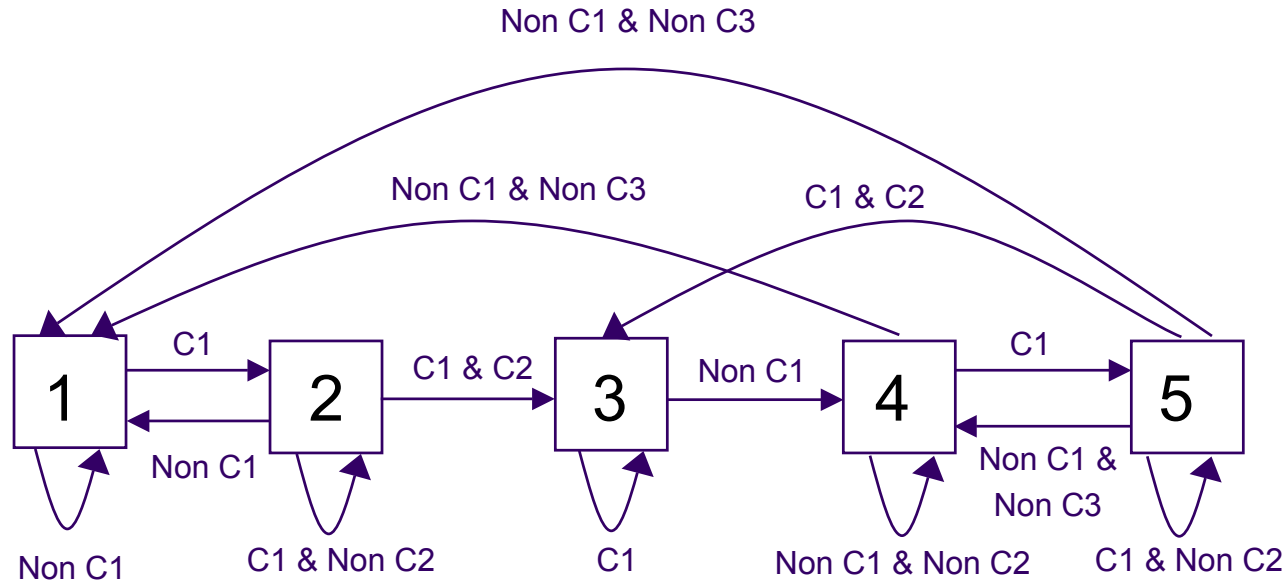
- Critère LCT
- Critère SB
- Critère SBP

→ Évaluation des 3 critères

Automate Bruit/Parole

C1
E > Seuil_détection

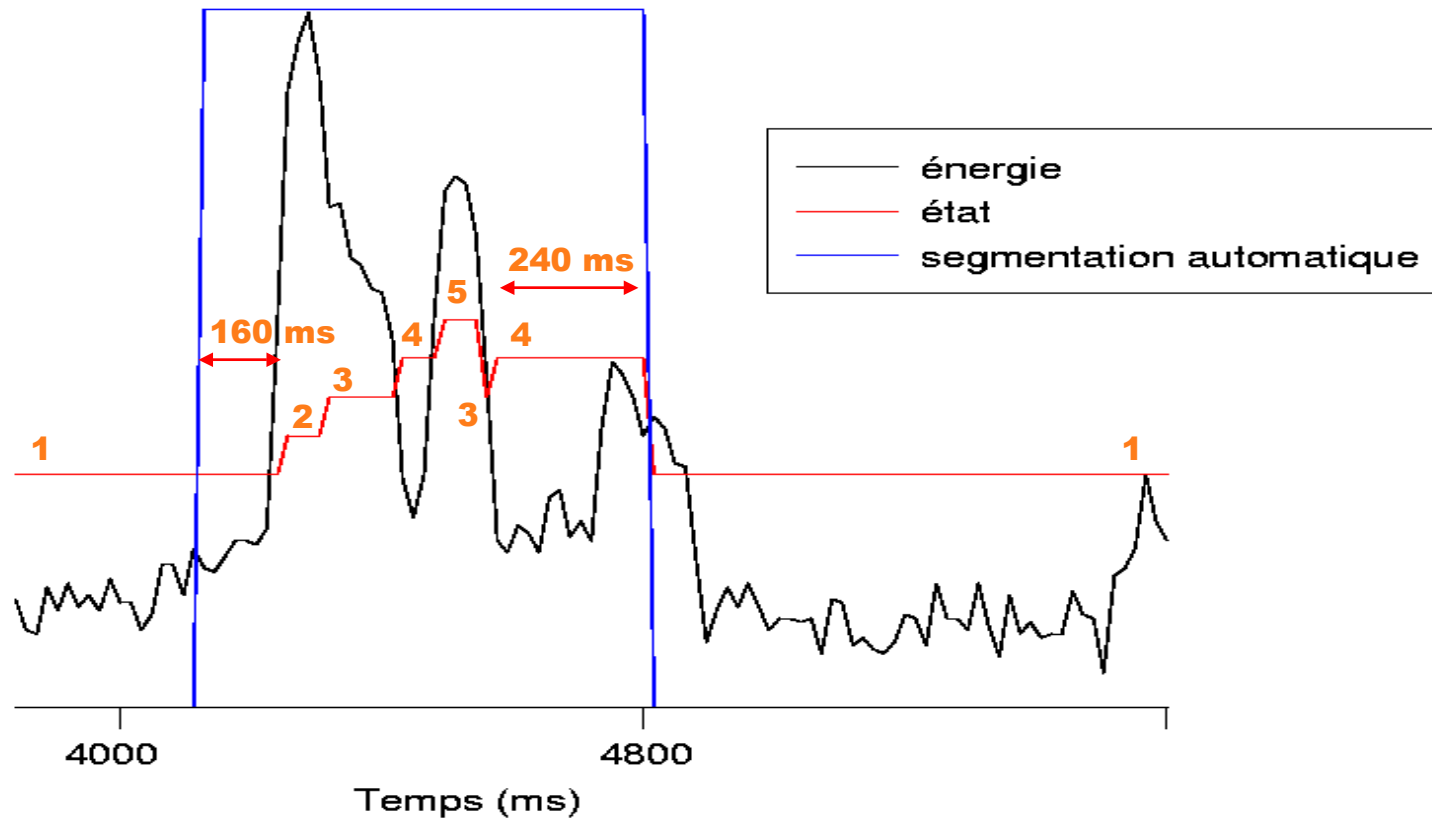
Contraintes de durées
C2 : Durée parole > Parole min
C3 : Durée silence > Silence fin



BRUIT OU SILENCE PRESOMPTION DE PAROLE PAROLE PLOSIVE OU SILENCE REPRISE POSSIBLE DE PAROLE

Fonctionnement de l'automate sur un mot

“Validation”

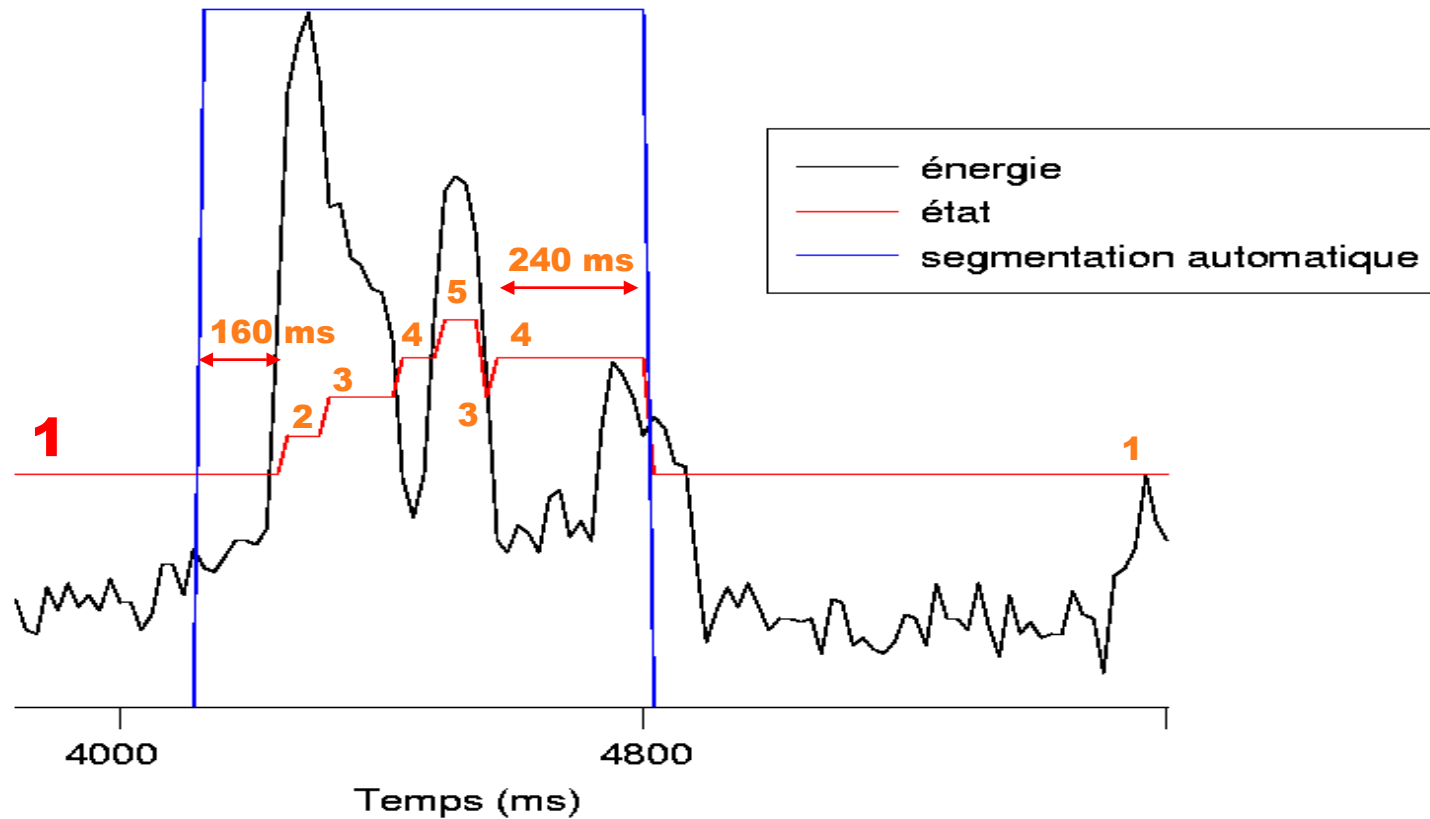


Fonctionnement de l'automate sur un mot

“Validation”

État 1

Bruit ou Silence

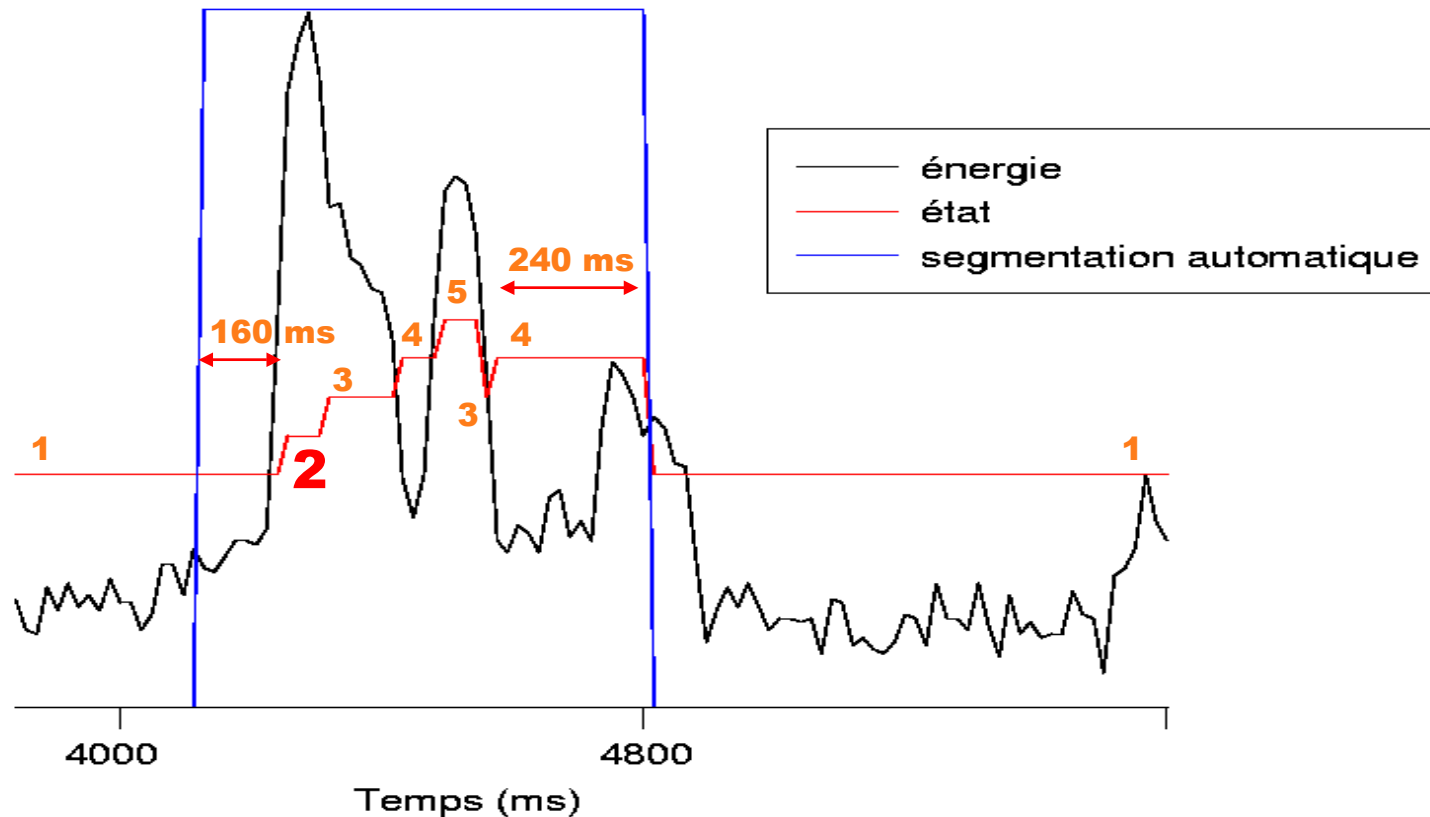


Fonctionnement de l'automate sur un mot

“Validation”

État 2

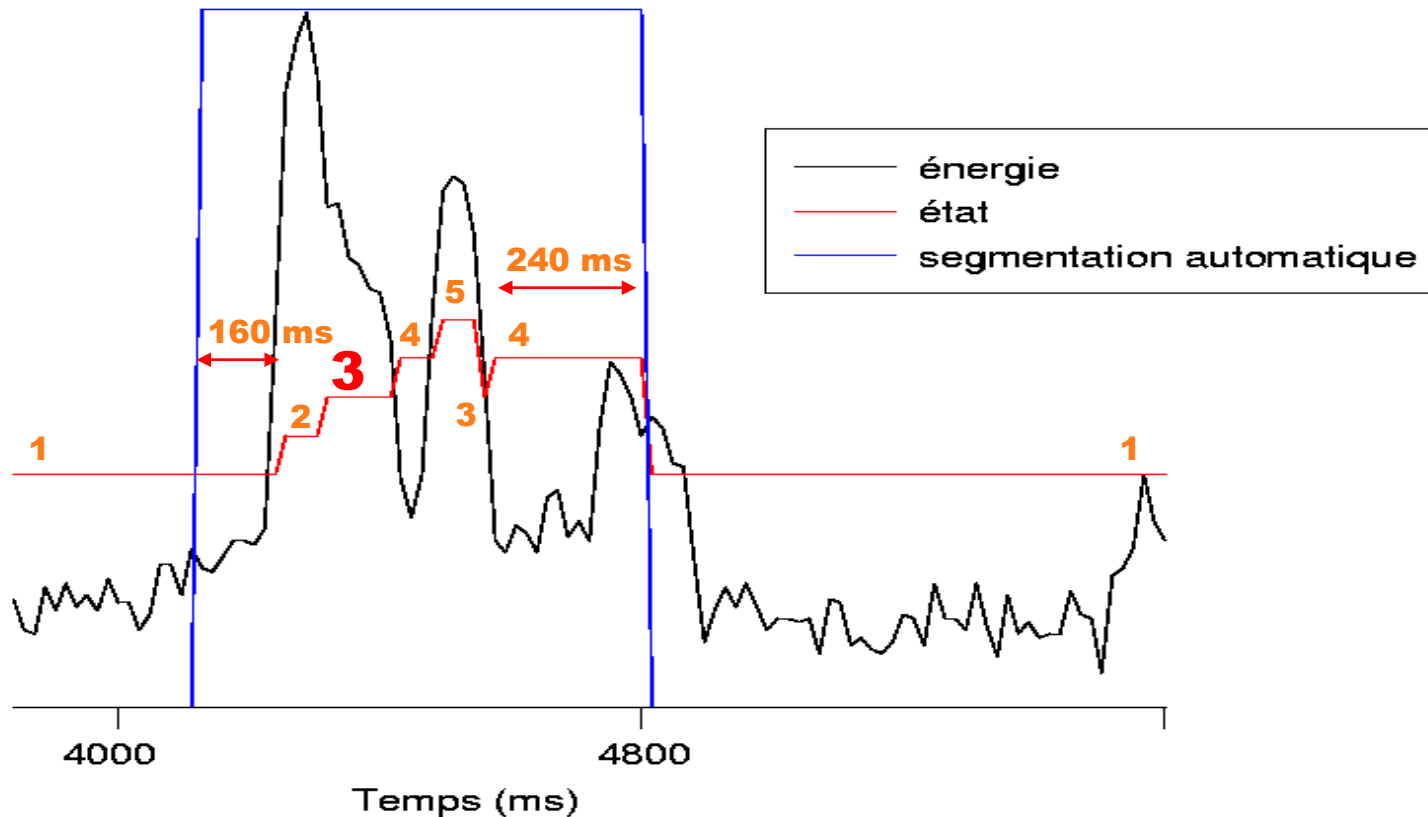
Présomption de
Parole



Fonctionnement de l'automate sur un mot

“Validation”

État 3
Parole

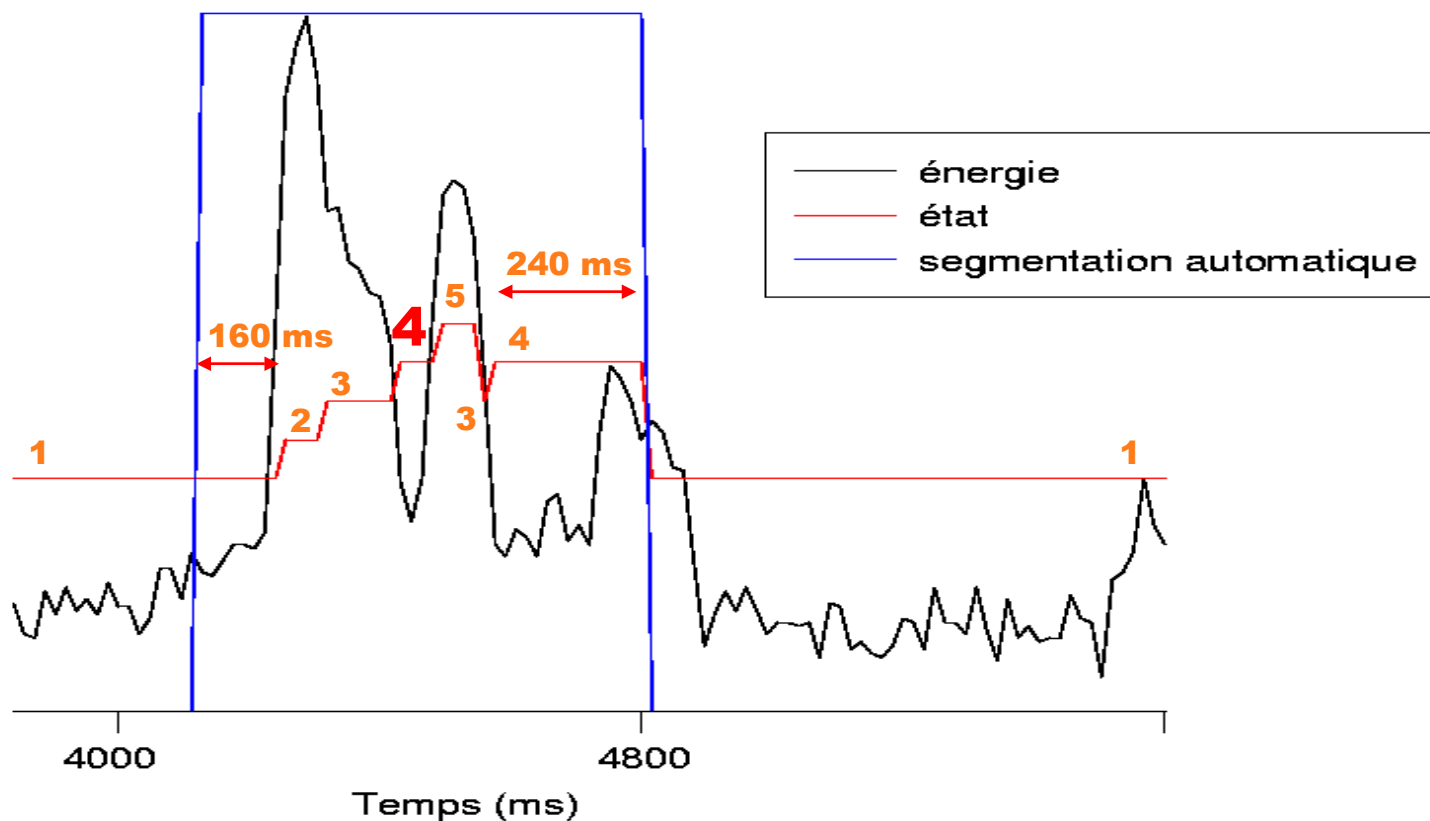


Fonctionnement de l'automate sur un mot

“Validation”

État 4

Plosive ou silence

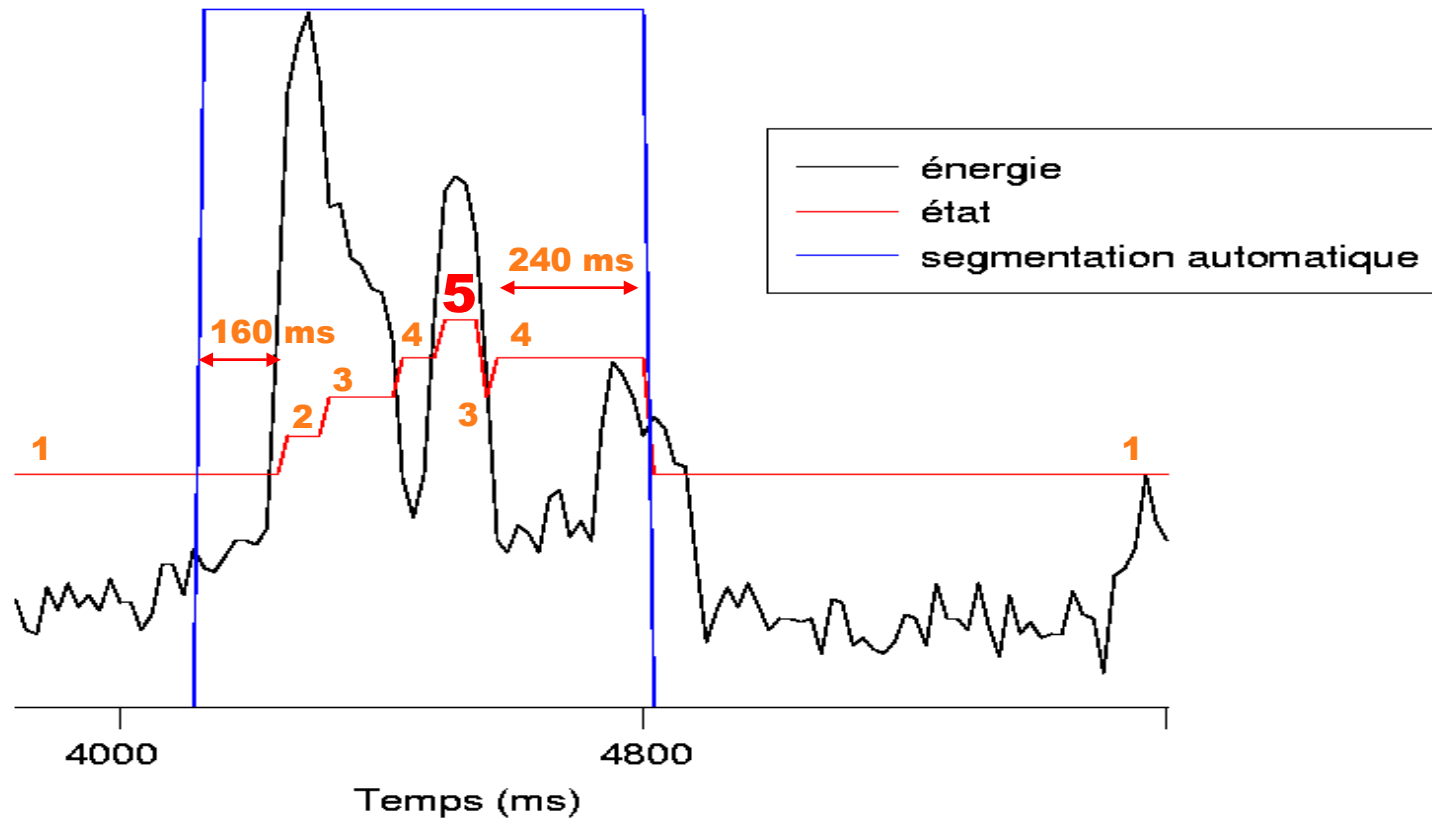


Fonctionnement de l'automate sur un mot

“Validation”

État 5

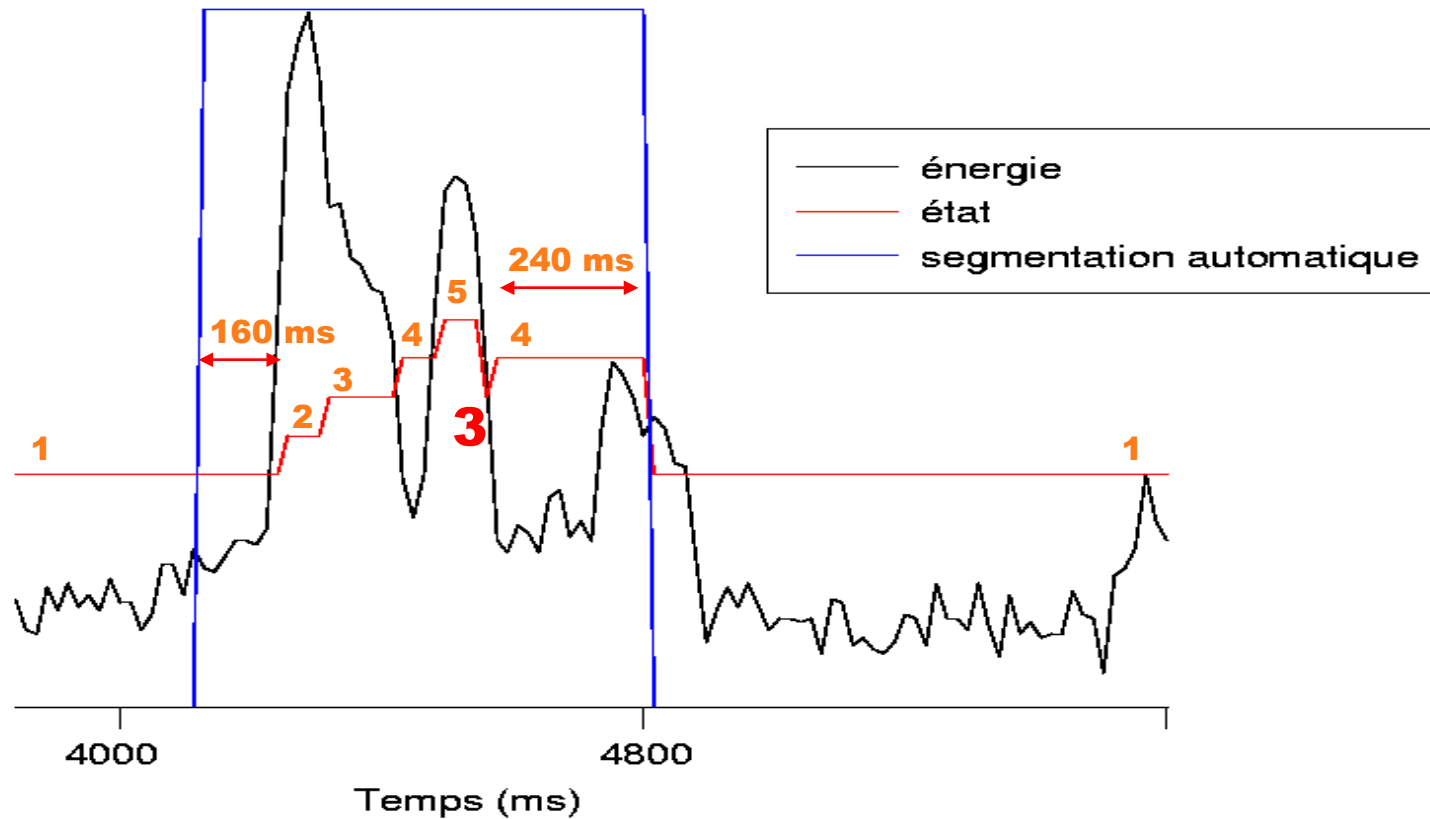
Reprise possible
de parole



Fonctionnement de l'automate sur un mot

“Validation”

État 3
Parole

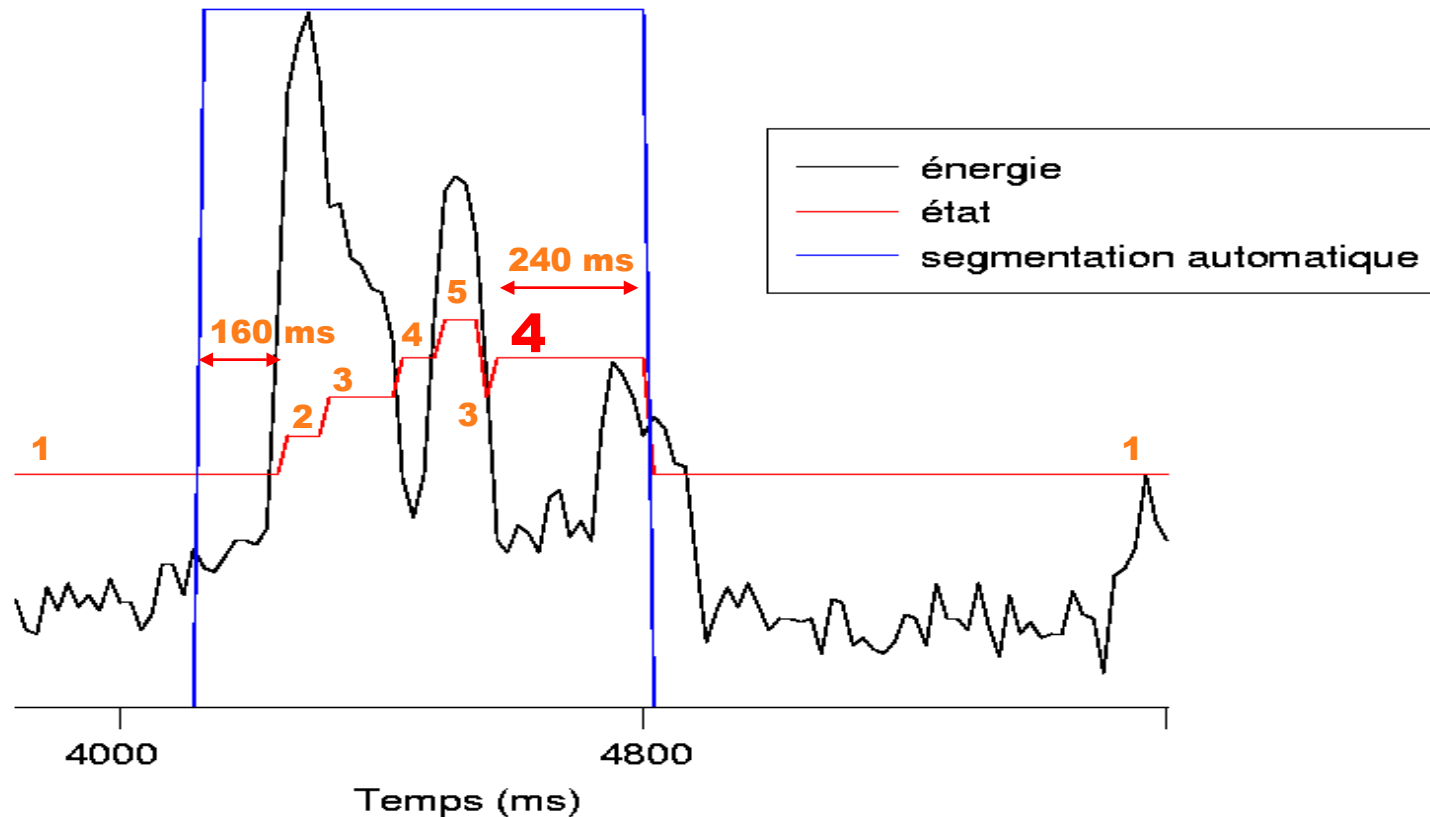


Fonctionnement de l'automate sur un mot

“Validation”

État 4

Plosive ou silence

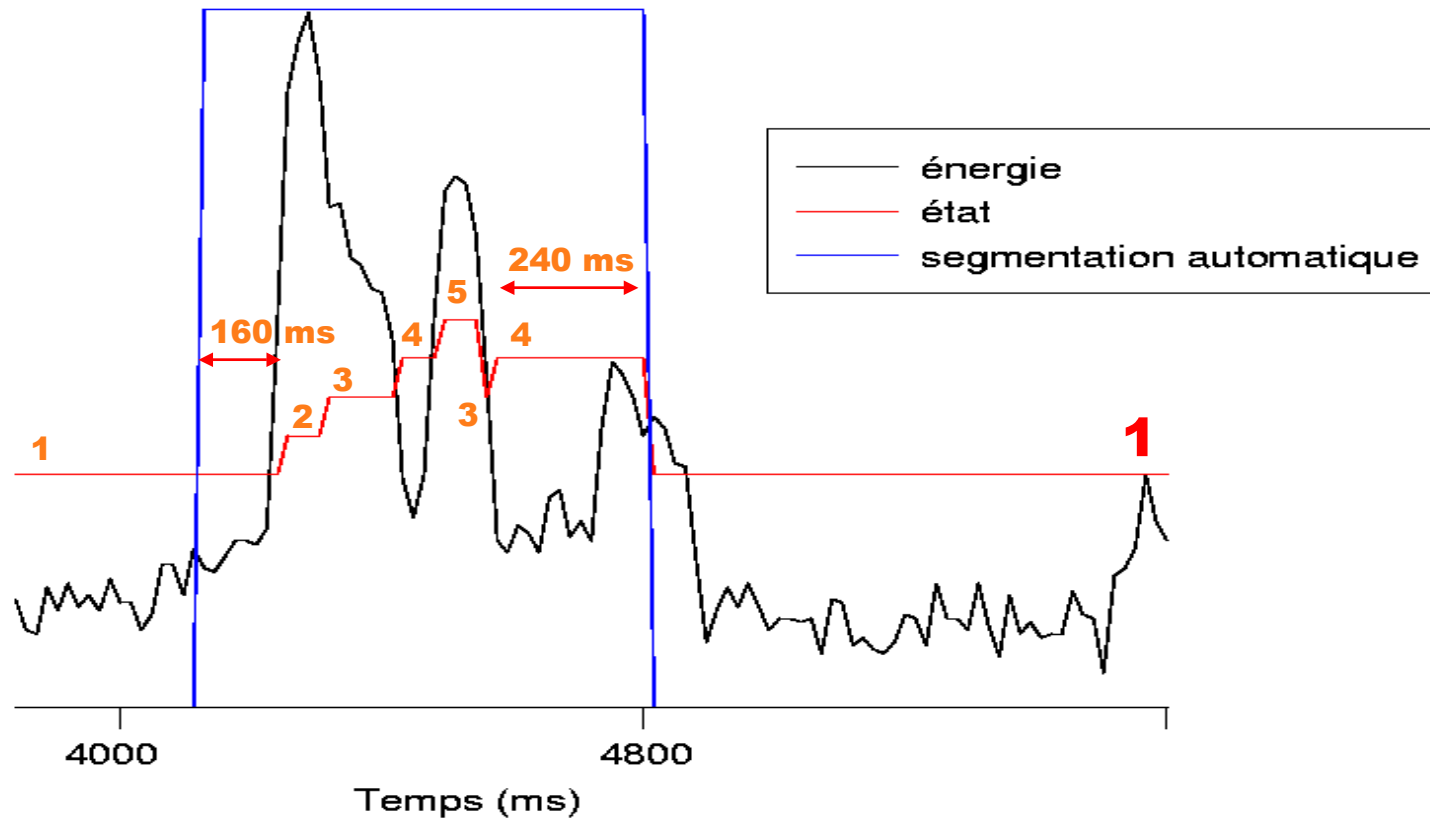


Fonctionnement de l'automate sur un mot

“Validation”

État 1

Bruit ou Silence



3 critères utilisés – Critère LCT

→ Critère fondé sur le "rapport signal à bruit"

Estimation de l'énergie à long terme calculée dans l'état bruit ou silence

$$ELT(n+1) = ELT(n) + (1-\lambda)(E(n) - ELT(n))$$

où $\lambda = 0.99$ (1600 ms)

→ C1 : $E(n) - ELT(n) > \text{Seuil_détection}$

3 critères utilisés – Critère SB

→ Critère fondé sur les statistiques du bruit

Hypothèse : énergie du bruit suit loi Gaussienne

Estimation des statistiques du bruit (moyenne μ et variance σ) dans état bruit ou silence

Calcul du rapport critique

$$r_{SB}(E(n)) = \frac{E(n) - \hat{\mu}(n)}{\hat{\sigma}(n)}$$

→ C1 : $r_{SB}(E(n)) > \text{Seuil_détection}$

3 critères utilisés – Critère SBP

- Critère fondé sur les statistiques du bruit et de la parole
- Approche bayésienne

H_0 : la trame observée est une trame de bruit (ou non-parole)
 H_1 : la trame observée est une trame de parole

Rapport de vraisemblance

$$r_{\text{SBP}}(E(n)) = \frac{P(E(n)/H_0)}{P(E(n)/H_1)}$$

- Estimation des statistiques du bruit dans état bruit ou silence
Estimation des statistiques de la parole dans état parole

- **C1 : $E(n) > \text{Seuil_détection}$** avec $\text{Seuil_détection} = s.a$, où s est une solution de $r_{\text{SBP}}(x) = 1$ et a est un facteur d'interpolation

Évaluation des 3 Critères

→ Différences significatives

- Sur RTC_A : $SB=LCT > SBP$
- Sur GSM_A et GSM_T : $SB=SBP > LCT$
- Avec bruits : $SB > SBP=LCT$

→ Robustesse du seuil “optimal”

- Au changement de base : $SB=SBP > LCT$
- Au niveau de bruit : $SB=SPB > LCT$
- Au réseau d'appel : $SBP > SB=LCT$

→ Critère retenu : **critère SB**

Améliorations du module de détection

→ Prise en compte d'une nouvelle condition

3 voies envisagées

- Intégration des moments d'ordre 3 sur l'énergie
- Intégration d'un paramètre de voisement
- Intégration de l'analyse factorielle discriminante sur les coefficients MFCC

→ Utilisation d'un module de débruitage

Intégration d'une nouvelle condition

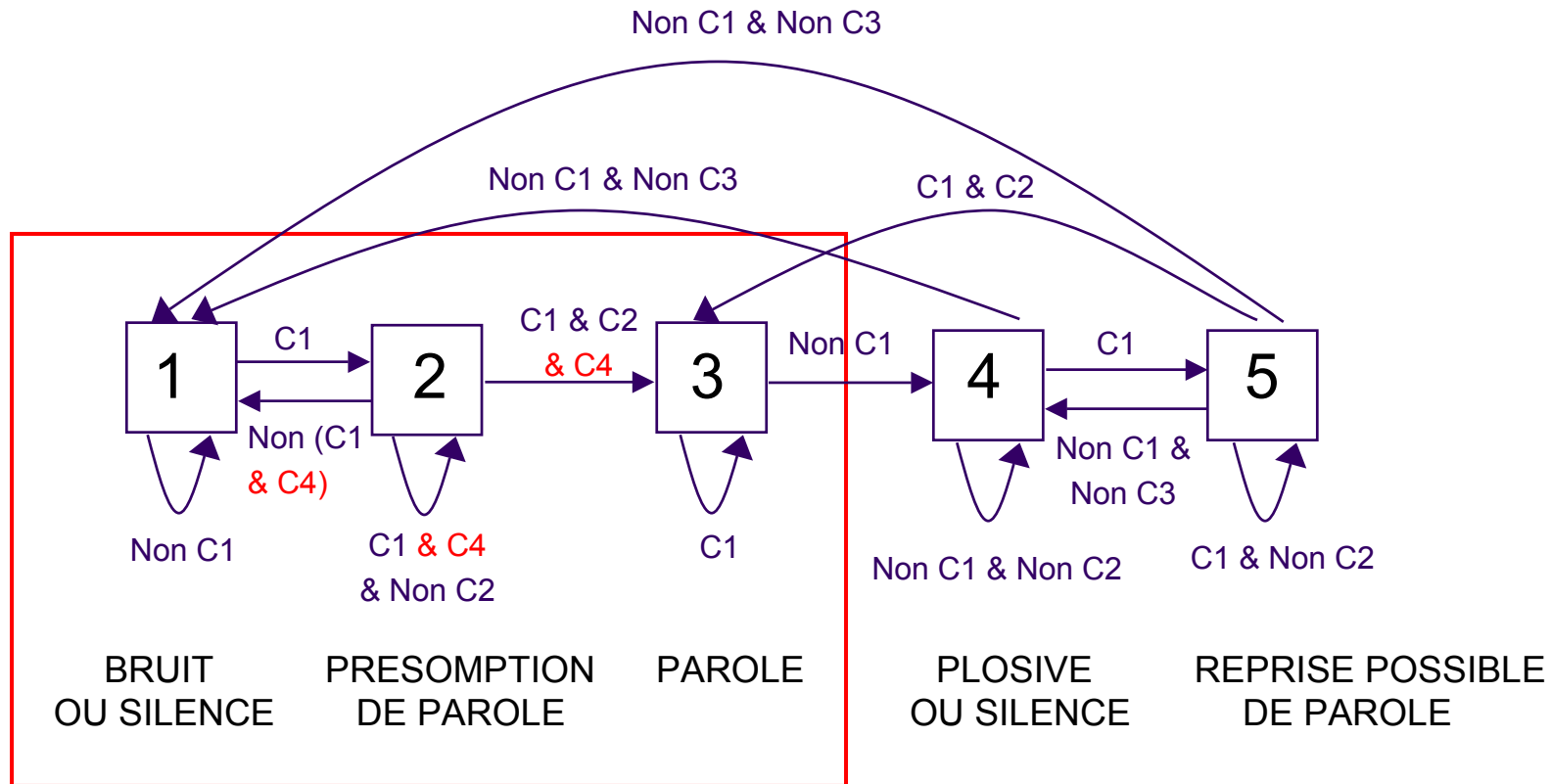
→ But : améliorer les transitions de l'automate

→ Ajout d'une nouvelle condition : **C4**
→ Fusion de décision

→ Différentes possibilités envisageables

→ But : diminuer les insertions sans augmenter les erreurs définitives
→ baisse du nombre total d'erreurs

Intégration d'une nouvelle condition



Statistiques d'ordre supérieur

- But : apporter une précision à l'estimation de la distribution de l'énergie
- Supprimer l'hypothèse de gaussianité faite sur les distributions de l'énergie [Jacovitti *et al.* 1991]
 - moment et cumulant d'ordre 3 et 4
- Estimation des moments d'ordre k avec fenêtres exponentielles

$$\hat{\mu}_k(n) = (1-\lambda) \sum_{i=1}^n \lambda^{n-i} x_i^k$$

Intégration du moment d'ordre 3

→ Moment d'ordre 3, non centré, normalisé

$$\hat{m}_3(n) = \frac{\hat{\mu}_3(n)}{(\hat{\sigma}(n))^3}$$

→ Rapport court-terme / long-terme

$$\text{rap}(n) = \frac{\hat{m}_{3\text{ct}}(n)}{\hat{m}_{3\text{lt}}(n)}$$

- Court-terme : estimation arithmétique sur $N=10$ trames (160 ms)
- Long-terme : $\lambda_{\text{lt}}=0.995$ (3200 ms) dans l'état silence

→ C4 : $\text{rap}(n) < \hat{T}(n)$

Étude des variations du rapport

Seuil adaptatif

$$\hat{T}(n+1) = \lambda_T \hat{T}(n) + (1 - \lambda_T) C \cdot \text{rap}(n)$$

$$\lambda_T = 0.99, C = 3$$

Paramètre de voisement

→ Caractérisation du signal en sons voisés/non voisés

→ Sons voisés

⑦ Signal périodique

● Cause : vibration des cordes vocales

● Une fréquence fondamentale F_0 et ses harmoniques

⑦ Phonèmes voisés : Voyelles : /a/, /i/, /ε/..., Consonnes nasales : /m/, /n/..., Fricatives voisées : /v/, /z/..., Plosives voisées : /b/, /d/, /g/...

→ Sons non voisés

⑦ Pas de périodicité du signal

⑦ Propriétés spectrales proches du bruit

⑦ Phonèmes non voisés : Plosives non voisées : /p/, /t/, /k/, ..., Fricatives non voisées : /f/, /s/, /ch/, .

→ But : discriminer les bruits énergétiques et la parole

Utilisation de “F0”

- Calcul de “F0” sur signal voisé ou non
- F_0 toutes les 4 ms
- Calcul de la médiane sur des sous-frames n de 4ms

$$\text{med}(n) = \text{med}(F_0(n-1), F_0(n), F_0(n+1))$$

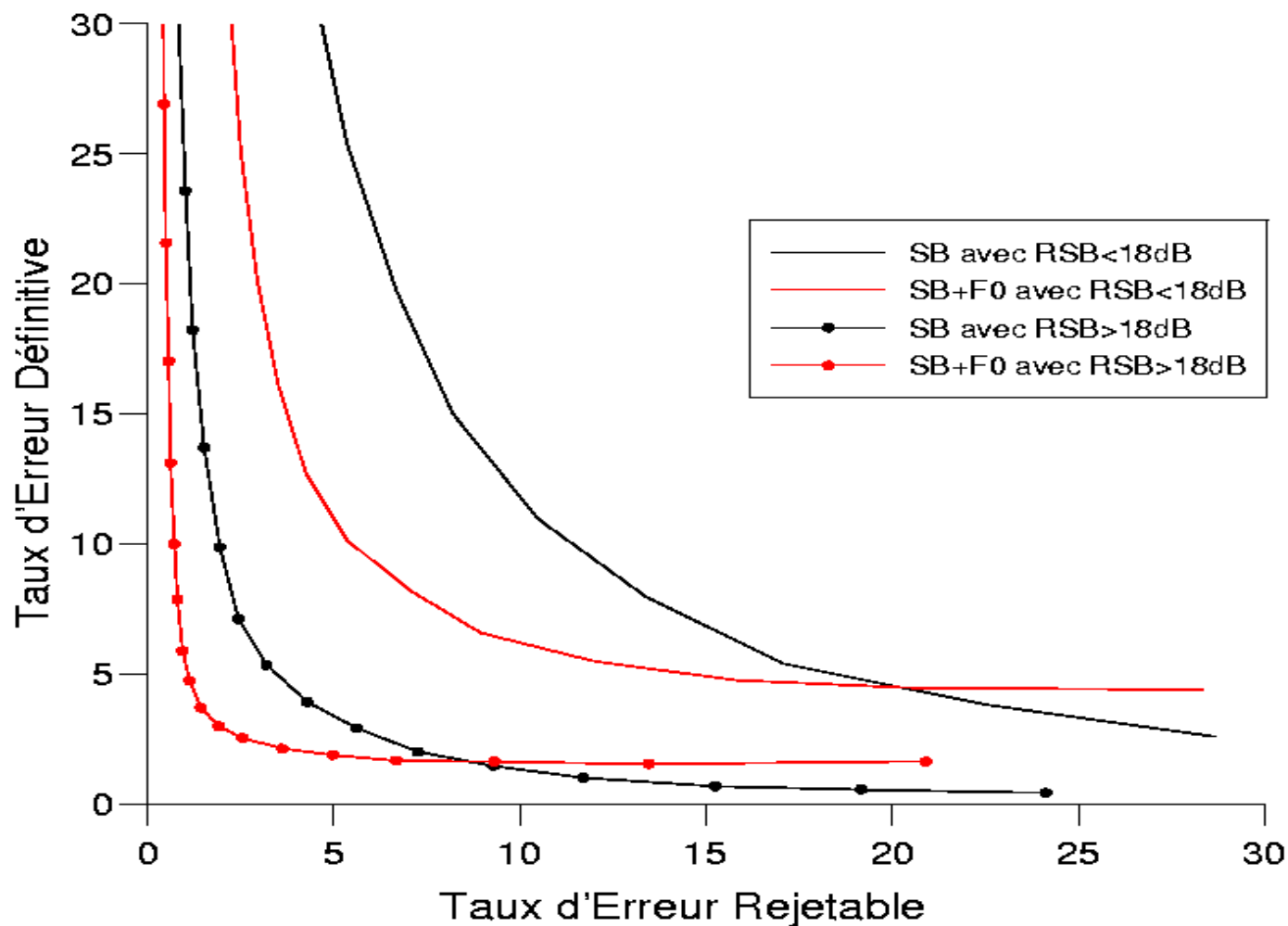
- Estimation de la variation moyenne de la médiane

$$\delta \text{med}(n) = \frac{1}{N} \sum_{m=n-N}^n |\text{med}(m) - \text{med}(m-1)|$$

- C4 : $\delta \text{med}(4m) < \text{seuil}_{\delta \text{med}(4m)}$ où $m \in \mathbb{N}^*$

Résultats de détection

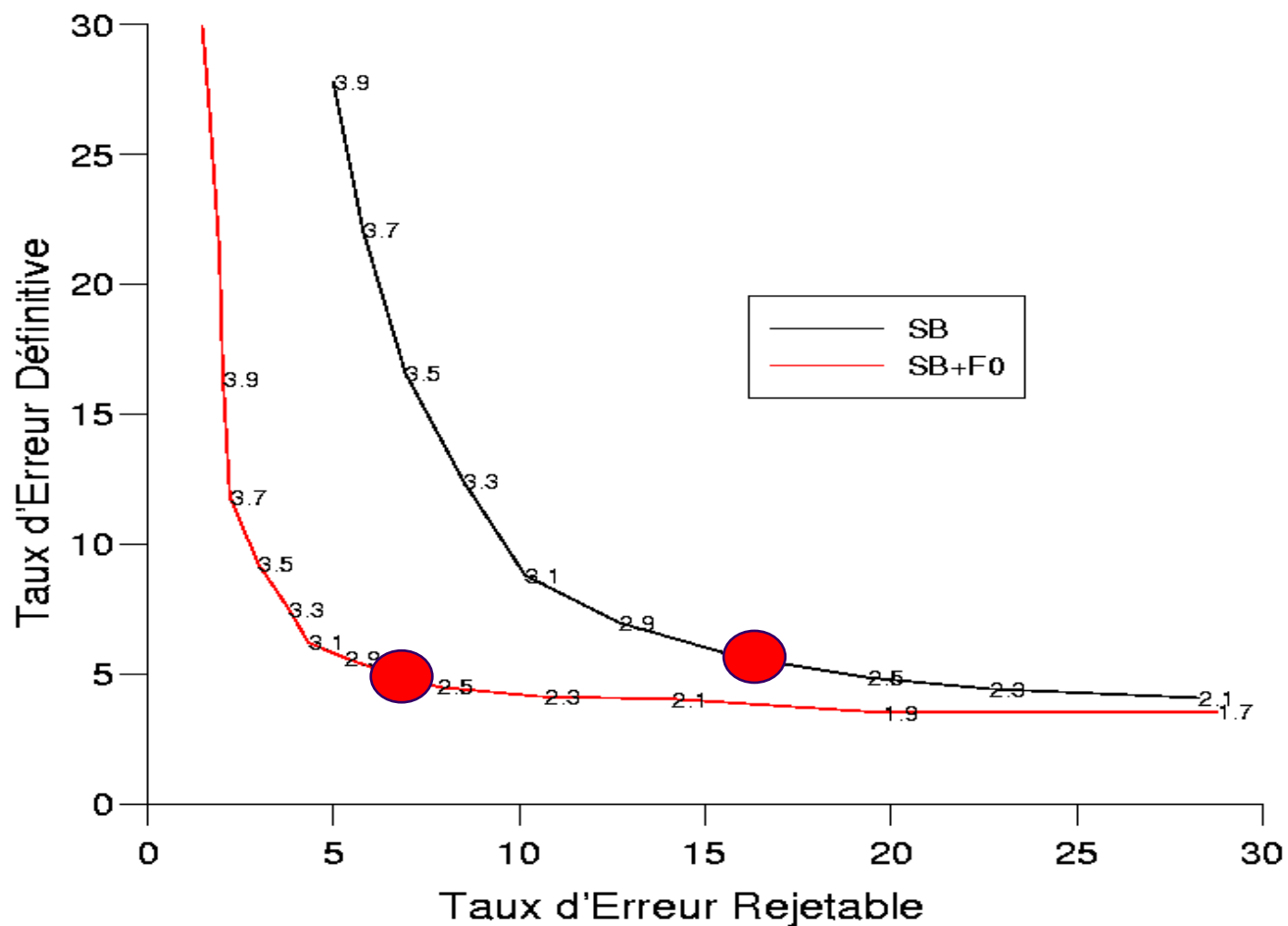
Base GSM_T (mots isolés)



Améliorations significatives quel que soit le RSB

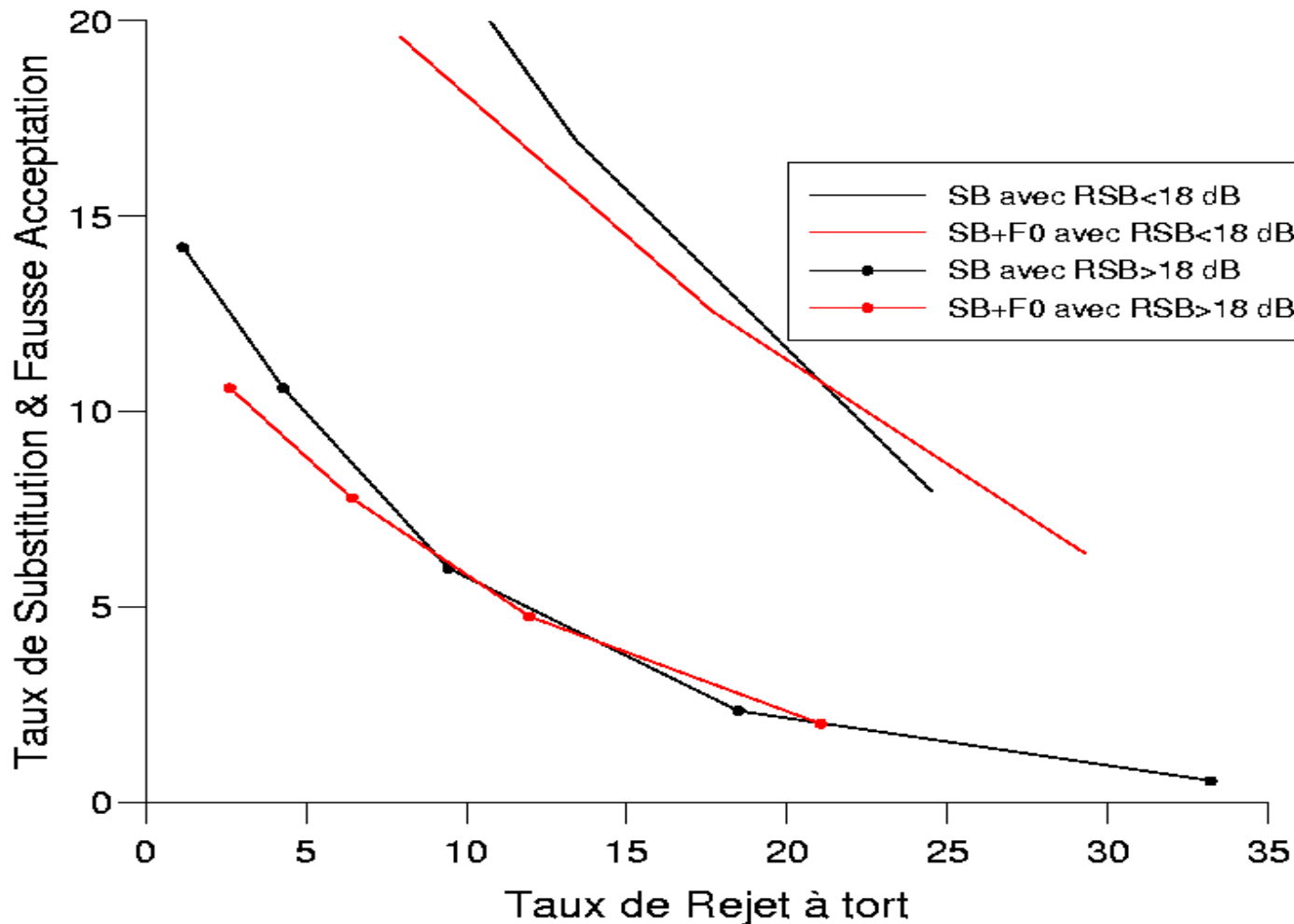
Résultats de détection

Base AGORA (parole continue)



Améliorations
significatives

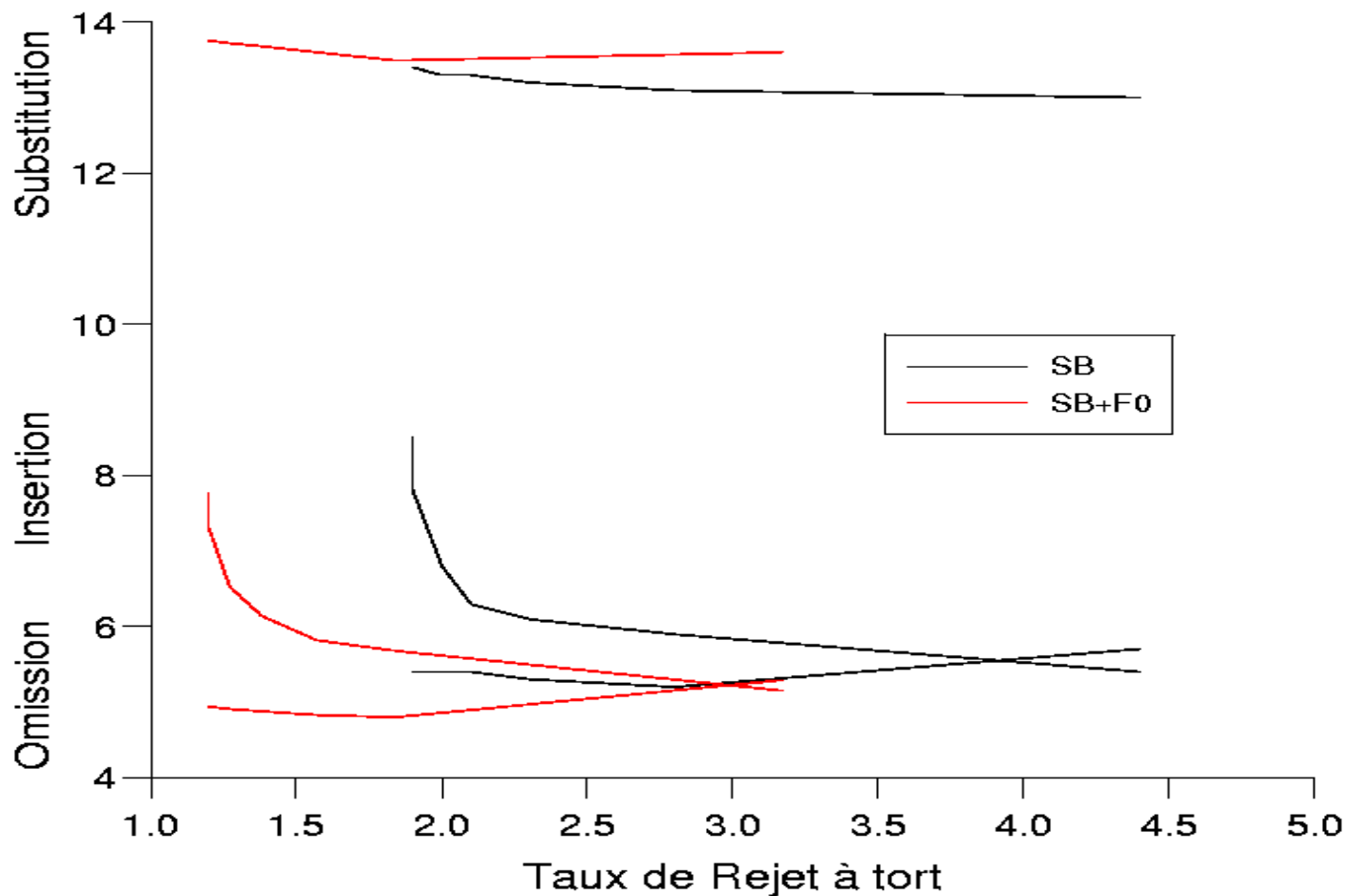
Résultats de reconnaissance Base GSM_T (mots isolés)



Améliorations
quel que soit
le RSB pour
faible rejet à
tort

Améliorations
significatives
sur RSB<18dB

Résultats de reconnaissance Base AGORA (parole continue)



Améliorations significatives sur les erreurs d'insertion, d'omission et de rejet à tort

L'analyse factorielle discriminante

→ But : fusionner un grand nombre de coefficients

→ Principe : trouver a

minimisant la covariance intra-classe (matrice D) et maximisant la covariance inter-classe (matrice E)

→ Dans le cas de 2 classes

$$E=cc^* \quad \text{avec} \quad c = \frac{\sqrt{n_P n_B}}{n_P + n_B} (\bar{X}_B - \bar{X}_P)$$

et si $T = D + E$ alors $a = T^{-1}c$

a est l'unique vecteur propre [Lebart *et al.* 1995]

Intégration de l'analyse factorielle discriminante

→ Calcul du vecteur propre avec MFCC sur les bases RTC_A et GSM_A

- Classe Parole : segments de parole

- Classe Bruit : segments de bruit

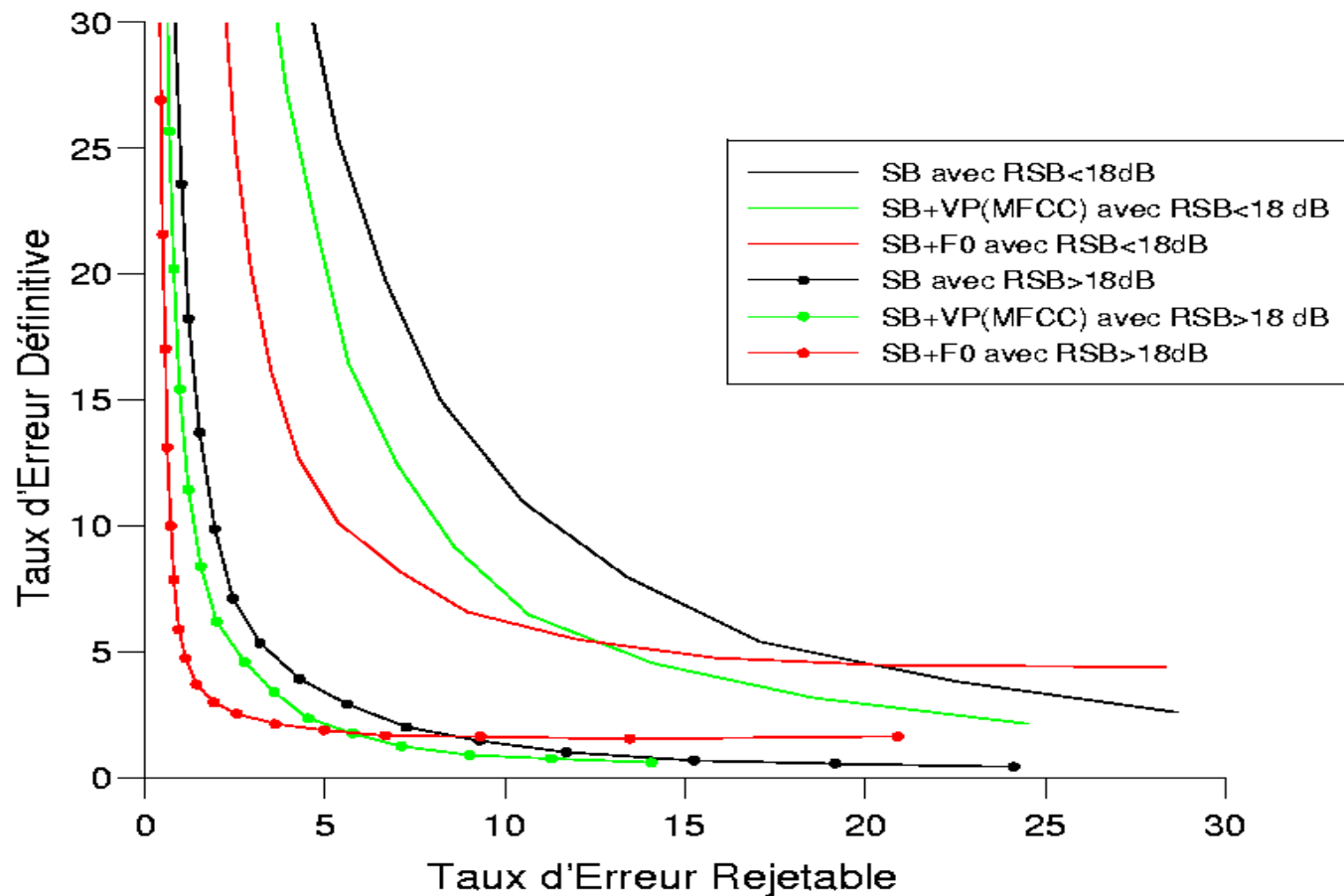
$a=(7.5, 4.3, 2.0, 1.4, 0.9, 0.8, 0.6, 0.5)$

→ Calcul de la combinaison linéaire des MFCC : $a.X$

→ $C4 : a.X < \text{seuil fixe}$ (optimisé sur bases RTC_A et GSM_A)

Résultats de détection

Base GSM_T (mots isolés)

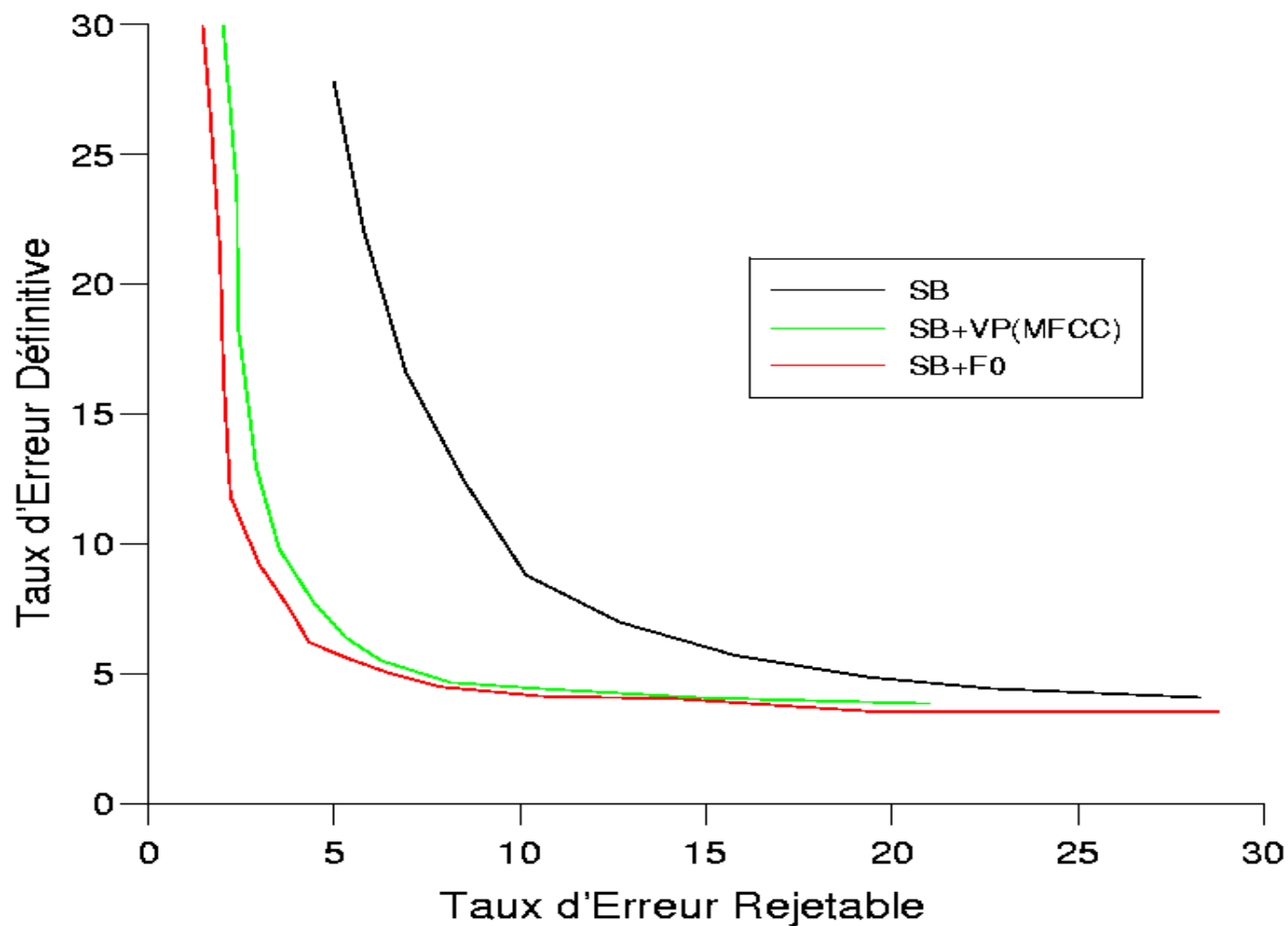


Améliorations significatives par rapport à SB

SB+F0 > SB+VP(MFCC)

Résultats de détection

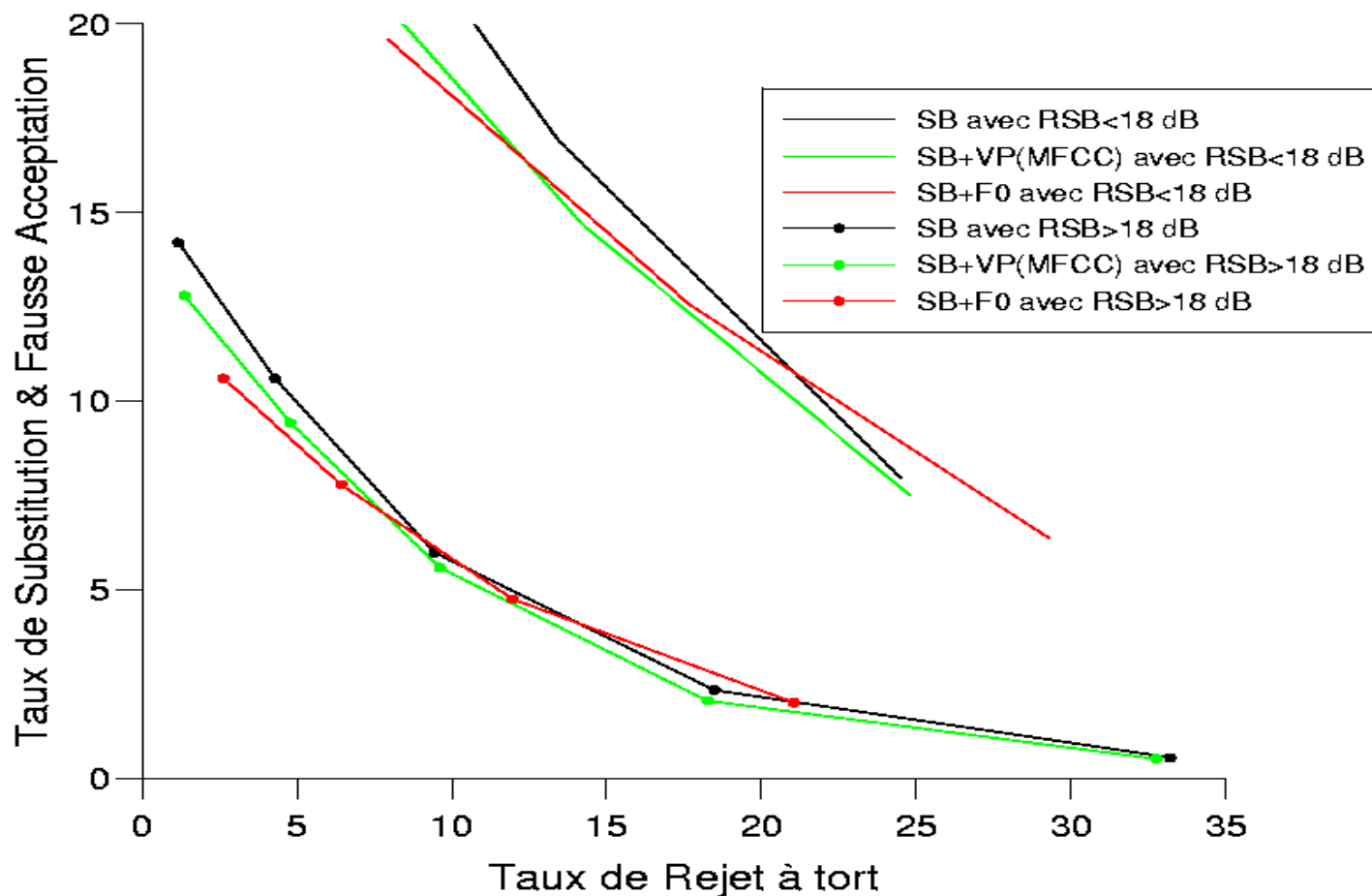
Base AGORA (parole continue)



Améliorations
significatives par
rapport à SB

SB+F0 >
SB+VP(MFCC)

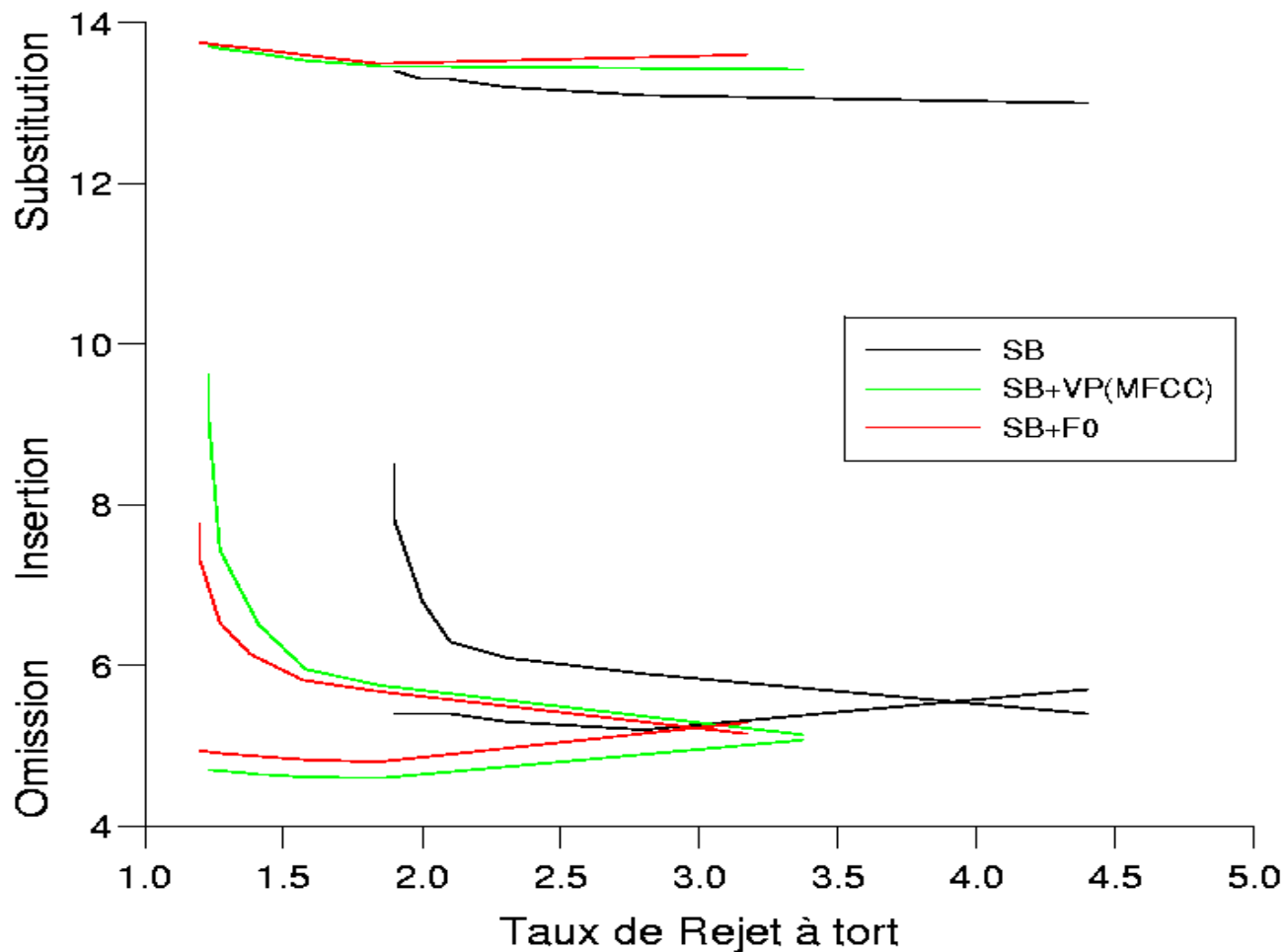
Résultats de reconnaissance Base GSM_T (mots isolés)



SB+F0 et
SB+VP(MFCC)
équivalentes

Améliorations
significatives sur
RSB < 18 dB

Résultats de reconnaissance Base AGORA (parole continue)



Améliorations significatives sur l'ensemble des erreurs

Omission :

$SB+F0 <$

$SB+VP(MFCC)$

Insertion :

$SB+F0 >$

$SB+VP(MFCC)$

Substitution :

$SB+F0 =$

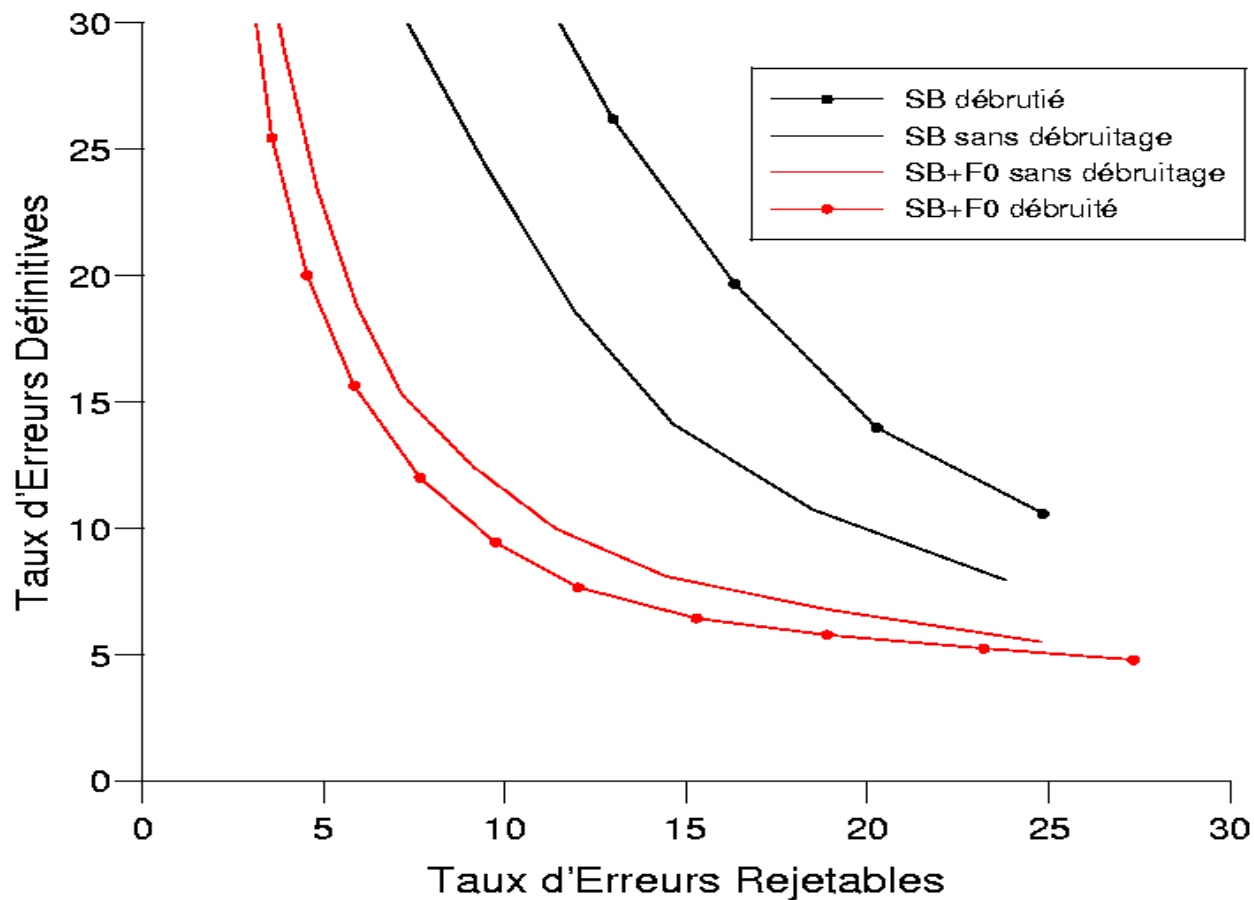
$SB+VP(MFCC)$

Utilisation d'un module de débruitage

- But : diminuer l'influence du bruit sur la détection et la reconnaissance
- Méthode de débruitage optimisée pour le système de reconnaissance
méthode d'atténuation spectrale fondée sur l'estimation de la densité spectrale de la parole non bruitée, dans le domaine temporel [Noé *et al.*, 2001]
- Résultats dans le mémoire de thèse : améliorations importantes pour les bruits stationnaires
- Pour les bruits non-stationnaires : RSB amplifié → insertion de bruit

Résultats du critère SB+F0 avec débruitage

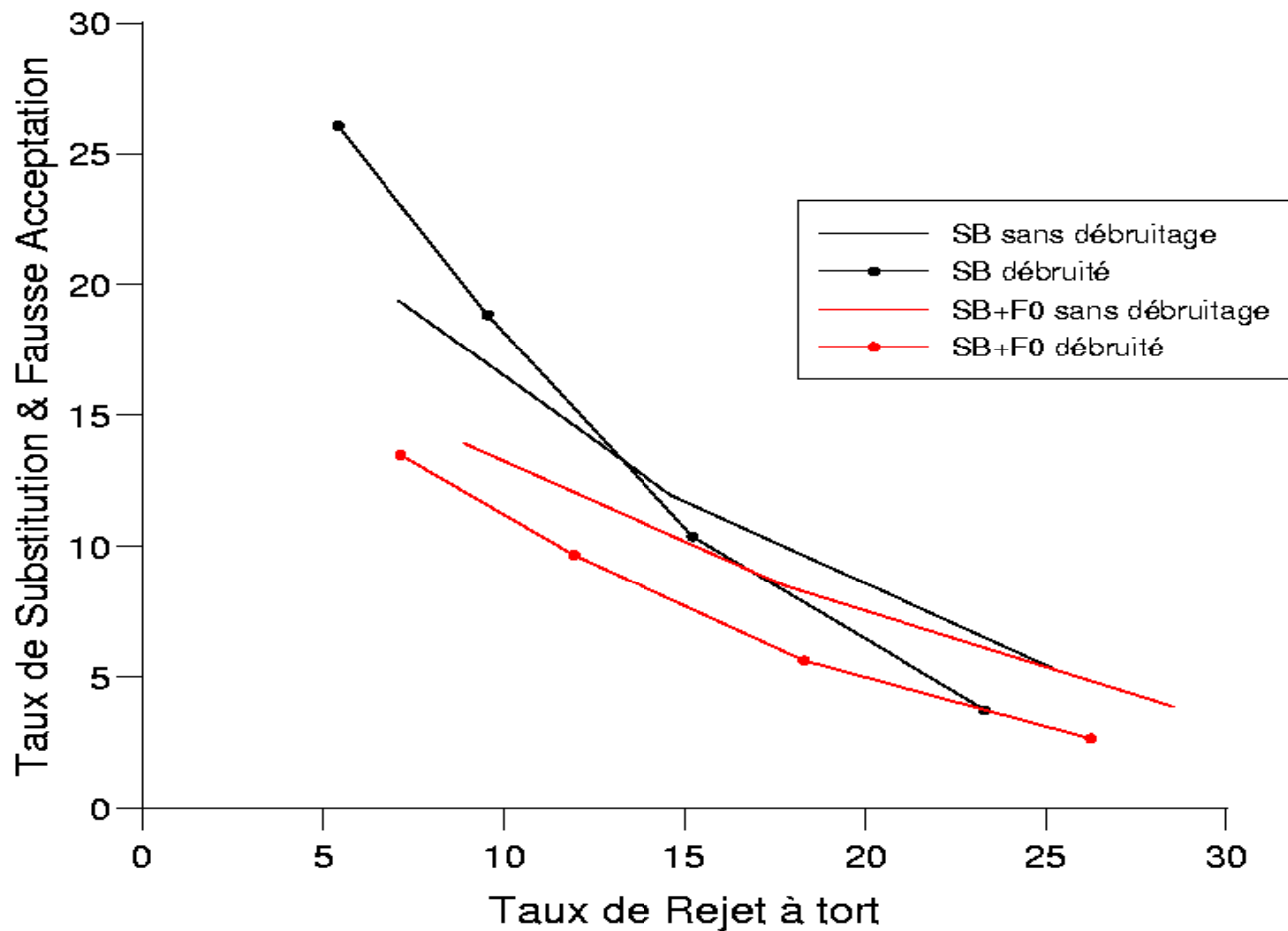
Résultats de détection
(GSM_A RSB<18 dB)



SB : débruitage dégrade
SB+F0 : amélioration

Résultats du critère SB+F0 avec débruitage

Résultats de reconnaissance
(GSM_A RSB<18 dB)



SB : pas d'amélioration
SB+F0 : amélioration

Conclusion

→ Meilleur critère : SB+F0

- Améliore les performances pour la détection de parole dans le bruit
 - baisse du coût CPU du système de reconnaissance
- Améliore les performances pour la détection de parole pour la reconnaissance de parole continue
- Augmente légèrement la robustesse du seuil de détection au niveau de bruit

Conclusion

→ Problèmes restants

- Interdépendance du module de détection et du module de reconnaissance
 - difficulté de déterminer les améliorations à apporter au module de détection
- Réglage de l'automate dépend du réseau d'appel
- Nouveau critère augmente légèrement les segments tronqués à gauche

Quelques références

- *La parole et son traitement automatique* CALLIOPE ed. Masson, 1989.
- *Signaux Aléatoires* B. Pincibono **Dunod 1995.**
- *Elements of Information Theory* T. M. Cover and J.A. Thomas **Wiley Interscience 1991.**
- *Fundamentals of Speech Recognition* L. Rabiner and B.-H. Juang **Prentice Hall Signal Processing A. V. Oppenheim Series Editor, 1993.**
- *Statistical Methods for Speech Recognition* F. Jelinek **MIT Press 1998.**