

ECOLES NATIONALES POLYTECHNIQUES

DEPARTEMENT D'ELECTRONIQUE

المدرسة الوطنية المتعددة التقنيات  
BIBLIOTHEQUE — المكتبة  
Ecole Nationale Polytechnique

# THESE DE MAGISTER

EN ELECTRONIQUE APPLIQUEE

Option : acquisition et traitement de l'information

Présentée par:

**M<sup>me</sup> K. KASDI-DJOUADI**

## THEME

**ETUDE DES TECHNIQUES  
D'INTELLIGENCE ARTIFICIELLE  
APPLICATION A LA REALISATION  
D'UN SYSTEME EXPERT  
DE RECONNAISSANCE AUTOMATIQUE  
DE LA PAROLE CONTINUE MULTILOCUTEUR**

Soutenu devant le Jury composé de :

<b>M. MEHENNI</b> — Maître de conférences	<b>:</b>	<b>Président</b>
<b>B. BOUSSEKSOU</b> — Chargé de cours	<b>:</b>	<b>Rapporteur</b>
<b>D. BERKANI</b> — Maître de conférences	<b>:</b>	<b>Examineur</b>
<b>M. GUERTI</b> — Docteur d'Etat	<b>:</b>	<b>Examinatrice</b>
<b>L. HAMMAMI</b> — Chargée de cours	<b>:</b>	<b>Examinatrice</b>

**Janvier 1996**



*Je tiens à exprimer toute ma gratitude à M. Bousseksou pour avoir accepté de diriger ce travail, pour les conseils qu'il m'a prodigués et pour m'avoir fait découvrir ce passionnant domaine qu'est la reconnaissance automatique de la parole.*

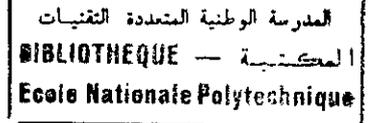
*Je remercie vivement M. Mehenni d'avoir bien voulu accepter de présider le jury de ma soutenance et d'avoir émis des remarques constructives.*

*Cette thèse a également bénéficié des suggestions, critiques et conseils judicieux de M. Berkani, M<sup>me</sup> Hammami et M<sup>me</sup> Guerti. Aussi, je tiens à les remercier chaleureusement pour leur disponibilité, leur soutien et leur aide précieuse.*

*Que toutes les personnes qui ont contribué à l'élaboration de cette thèse trouvent ici l'expression de mes vifs remerciements.*

---

# SOMMAIRE



Introduction générale .....	2
Première partie : Etudes théoriques .....	4
Chapitre I : La parole et sa reconnaissance .....	5
I - Introduction à la parole .....	5
II - Reconnaissance automatique de la parole .....	7
III - Décodage acoustico-phonétique .....	8
IV - Analyse lexicale .....	10
V - Conclusion .....	12
Chapitre II : Intelligence artificielle et systèmes experts .....	13
I - Principales techniques d'intelligence artificielle .....	13
II - Les systèmes experts .....	17
III - Représentation des connaissances .....	20
IV - Critères de différenciation des moteurs d'inférences .....	21
V - Conclusion .....	24
Deuxième partie : Conception et réalisation du système .....	25
Chapitre I : Position du problème .....	26
Chapitre II : Représentation interne du signal .....	29
Chapitre III : Module de prétraitement .....	32
Chapitre IV : Module d'étiquetage .....	41
Chapitre V : Module de segmentation .....	50
Chapitre VI : Module d'identification phonétique .....	58
Chapitre VII : Module d'analyse lexicale .....	66
Chapitre VIII : Organisation et structure des logiciels .....	68
Troisième partie : Tests et résultats .....	70
Test 1 .....	73
Test 2 .....	81
Conclusions et perspectives .....	89
Bibliographie .....	90

## INTRODUCTION GENERALE

L'intelligence artificielle a pour but d'analyser les comportements humains dans les domaines de la perception, de la compréhension et de la décision afin de les reproduire sur un ordinateur.

Elle s'intéresse principalement aux processus cognitifs mis en œuvre par l'être humain lors de l'accomplissement de tâches réputées intelligentes. Un programme d'intelligence artificielle résout un problème ou prend une décision au vu de la description d'une certaine situation. Il trouve lui-même la méthode de résolution à employer grâce à une vaste gamme de mécanismes de résolution qui lui ont été incorporés.

Cette conception est une évolution importante par rapport à l'informatique classique où le raisonnement est fait par l'être humain et où l'ordinateur est utilisé surtout pour ses capacités de calcul.

La première application vraiment opérationnelle de la recherche en intelligence artificielle correspond aux systèmes experts.

Les systèmes experts sont des logiciels d'intelligence artificielle qui agissent sur des domaines précis. Leur but est de modéliser le comportement d'un "expert humain" accomplissant une tâche intellectuelle dans un domaine précis où la méthodologie est souvent difficile à modéliser sous forme d'algorithmes sûrs et définitifs.

La reconnaissance automatique de la parole semble être un des domaines auquel la méthodologie des systèmes experts est susceptible d'apporter des solutions.

En effet, le travail d'un nombre important de chercheurs, depuis plus de quarante ans, n'a pas suffi à atteindre le but ultime de la reconnaissance de la parole à savoir: la communication en langage naturel avec la machine.

En 1969, un chercheur des Bell-Laboratories, découragé sans doute par la complexité de la tâche, déclarait dans un article retentissant: «Il est impossible de reconnaître la parole».

Depuis, des progrès notables ont été réalisés et confortent les chercheurs dans l'idée que le problème est soluble, du moins partiellement.

L'apparition sur le marché de matériels de reconnaissance vocale, bien que de performances limitées, est un résultat encourageant.

Cependant, il semble à présent, que le chemin à parcourir soit encore important.

Quels sont donc ces problèmes qui font que la reconnaissance de la parole est difficile et non encore résolue totalement à l'heure actuelle?

Un premier pas vers la solution a été de les répertorier et d'en cerner la nature:

- La parole présente une très grande variabilité, interlocuteur (voix de femme, d'homme, d'enfant, accents régionaux, etc.) et intra-locuteur (selon le débit, la façon de parler, etc.);
- L'effet de coarticulation dans la parole fait que chaque son est altéré par les sons voisins;
- Il n'y a pas de silence entre les mots, comme les blancs dans un texte écrit. Par contre, des silences peuvent exister à l'intérieur des mots.
- Il y a plusieurs niveaux d'informations transmis par le même signal: phonétique, lexical, sémantique, etc., et il n'existe pas de règles précises pour formaliser l'information à chaque niveau;
- Il est nécessaire d'avoir des connaissances dans les domaines très différents: physiologie, phonétique, linguistique, traitement de signal, informatique, etc.

Au vu des différents problèmes cités, la solution à la reconnaissance de la parole ne peut venir que d'un long et laborieux travail pluridisciplinaire.

Notre travail s'inscrit dans une contribution à cette délicate tâche en tentant de mettre au service de la reconnaissance de la parole les puissantes performances de l'intelligence artificielle, voie en pleine expansion.

L'objectif principal de notre recherche est d'étudier et de concevoir un système expert limité de reconnaissance automatique de la parole continue multilocuteurs.

La première partie de ce document se compose d'un chapitre où nous présentons le domaine particulier de la parole et sa reconnaissance et d'un chapitre où nous étudions les principales techniques d'intelligence artificielle et les systèmes experts.

Dans la deuxième partie de ce document, nous présentons l'ensemble des logiciels conçus pour la réalisation de notre système de reconnaissance et décrivons tous les modules que nous avons réalisé pour chaque étape de la reconnaissance.

Nous présentons par la suite, les tests effectués sur notre système et les résultats obtenus.

Nous terminons ce document par les conclusions que nous auront inspiré les tests effectués.

---

# **PREMIERE PARTIE ETUDES THEORIQUES**

## **Chapitre I**

### **La parole et sa reconnaissance**

## **Chapitre II**

### **Intelligence artificielle et systèmes experts**

# CHAPITRE I

## LA PAROLE ET SA RECONNAISSANCE

### I - INTRODUCTION A LA PAROLE

#### I.1 - Structure du signal de parole

La parole est une suite continue de sons dont la principale fonction est d'établir des distinctions entre des unités de signification.

Les phonèmes sont les éléments sonores les plus brefs qui permettent d'opérer ces distinctions. Ils sont identifiés par l'oreille, par des caractéristiques acoustiques liées aux mécanismes de leur production par l'appareil phonatoire humain.

En effet, pour chaque position articuloire, l'appareil vocal prend une configuration particulière qui détermine le spectre du phonème produit.

Mais la grande variabilité de réalisation physique d'un phonème donné, en fonction du locuteur, du débit d'élocution ou du contexte, ne permet pas d'associer à un phonème une représentation acoustique unique.

L'oreille semble cependant extraire du signal vocal des caractéristiques acoustiques pertinentes: «indices acoustiques» lui permettant d'identifier un phonème, malgré la variabilité de ses représentations.

L'analyse acoustique des phonèmes tente de mettre en évidence ces indices et de les associer à des traits articuloires particuliers, pour établir des corrélations entre un phonème et sa représentation spectrale<sup>(10)</sup>.

#### I.2 - Classification des phonèmes en traits phonétiques

Les phonèmes d'une langue se regroupent en catégories dont les éléments partagent des traits phonétiques distinctifs (voir tableau 1).

##### I.2.1- Classification des voyelles

Les voyelles produites par la vibration des cordes vocales se distinguent des autres phonèmes par leur périodicité et sont classées selon quatre traits articuloires.

##### a. La nasalité

Les voyelles nasales ont toutes en commun la mise à contribution des fosses nasales. Elles se distinguent des voyelles orales par la présence d'une anti-résonance.

##### b. Le degré d'ouverture du conduit vocal

Les voyelles ouvertes se distinguent des voyelles fermées par l'augmentation de l'ouverture articuloire.

##### c. Le lieu d'articulation

Les voyelles antérieures se distinguent des voyelles postérieures par la position moyenne de la langue.

##### d. L'arrondissement des lèvres

On distingue les voyelles labialisées des voyelles non labialisées par l'arrondissement des lèvres<sup>(11)</sup>.

I.2.2 - Classification des consonnes

Les consonnes forment une classe hétérogène et sont regroupées généralement selon 3 traits articulatoires.

a. Le voisement

Les consonnes voisées se distinguent des non voisées par la présence de vibrations des cordes vocales.

b. Le mode d'articulation

On distingue cinq modes d'articulations :

\* Occlusif : Les occlusifs se distinguent par trois phases de réalisation:

- un silence, qui correspond à l'occlusion totale du conduit vocal,
- une explosion, qui correspond à un écartement brutal des articulateurs,
- un bruit de friction.

\* Fricatif : Les consonnes fricatives sont des bruits résultant d'une turbulence aérodynamique.

\* Nasal : Les consonnes nasales présentent une occlusion dans la partie buccale du conduit vocal.

Mais elles se caractérisent par une structure périodique et ne possèdent que peu ou pas de bruit. Elles appartiennent à la sous-classe des sonantes.

\* Liquide.

\* Glissant.

c. Le lieu d'articulation

On classe aussi les consonnes selon la position de la constriction principale:

- Labiales,
- Dentales,
- Vélo-palatales.

Mode d'articulation		Consonnes			Lieu d'articulation ←
		Labiales	Dentales	Velo-palatales	
<b>Occlusives</b>					
	Voisées	[ b ]	[ d ]	[ g ]	
	Non voisées	[ p ]	[ t ]	[ k ]	
<b>Fricatives</b>					
	Voisées	[ v ]	[ z ]	[ ʒ ]	
	Non voisées	[ f ]	[ s ]	[ ʃ ]	
<b>Nasales</b>		[ m ]	[ n ]	[ ŋ ]	
<b>Glissantes</b>		[ w ]	[ j ]	[ j ]	
<b>Liquides</b>			[ l ]	[ R ]	
		<b>Voyelles</b>			
<b>Orales</b>		Antérieures		Centrales	Postérieures
		Non arrondies		Arrondies	Arrondies
	Fermées	[ i ]	[ y ]	[ u ]	
		[ e ]	[ ø ]	[ o ]	
		[ ε ]	[ œ ]	[ ɔ ]	
	Ouvertes	[ a ]		[ ɑ ]	
<b>Nasales</b>					
	Fermées	[ ẽ ]		[ õ ]	
	Ouvertes			[ ã ]	

Tableau I : Classification des phonèmes en traits distinctifs <sup>[19]</sup>

### I.3 - Rôle de la classification des phonèmes

Les traits articulatoires d'un phonème ne spécifient pas complètement sa représentation acoustique, car tous les aspects de la parole n'ont pas pour but d'établir des distinctions de signification.

Ainsi, les phénomènes prosodiques liés aux variations de la fréquence du fondamental, de l'intensité et des durées servent souvent à indiquer l'attitude du locuteur vis-à-vis de son discours.

De même que le timbre de la voix joue un rôle dans l'identification du locuteur.

Au niveau acoustique, ces différentes fonctions sont imbriquées et il s'agit, pour la tâche de reconnaissance, d'extraire de la représentation acoustique les indices pertinents que l'on peut le mieux corrélérer aux traits articulatoires, pour l'identification phonétique<sup>[10]</sup>.

## II - RECONNAISSANCE AUTOMATIQUE DE LA PAROLE

### II.1 - Introduction

La nature très variable et continue du signal de parole rend sa reconnaissance difficile.

La variabilité «intra-locuteur» et «inter-locuteur», le phénomène de co-articulation, l'absence de silences entre les mots, les informations extra-linguistiques contenues dans le signal vocal... déterminent la complexité de la tâche de reconnaissance.

Néanmoins, des systèmes de reconnaissance automatiques de la parole, même à des objectifs limités, tentent de résoudre ces différents problèmes.

On peut classer ces systèmes selon la complexité de la tâche de reconnaissance qui leur est assignée<sup>[10]</sup>.

### II.2 - Facteurs de complexité

#### a - Reconnaissance mono-locuteur ou multi-locuteur

La reconnaissance multi-locuteur introduit une difficulté supplémentaire liée à la variabilité des locuteurs due à des différences anatomiques, à des accents spécifiques, etc.

#### b - Reconnaissance de mots isolés ou parole continue

Dans la reconnaissance de mots isolés, les frontières entre les mots sont bien déterminées. De plus, les entités à reconnaître peuvent être considérées de façon globale.

La reconnaissance de la parole continue impose de considérer le signal vocal comme une suite d'entités élémentaires.

#### c - Taille du vocabulaire et grammaire admise

Il est plus difficile de travailler sur un vocabulaire très étendu que sur un vocabulaire restreint. De même qu'il est plus simple de traiter une grammaire rigide et contrainte que d'admettre toute la souplesse que présente la langue naturelle.

#### d - Environnement

La présence de bruit dégrade généralement les performances des systèmes de reconnaissance.

Il est plus aisé de travailler dans un environnement calme<sup>[10]</sup>.

### II.3 - Méthodes et systèmes de reconnaissances

#### II.3.1 - Systèmes "auto-organiseurs" ou systèmes "fondés sur la connaissance"

Les systèmes auto-organiseurs utilisent des échantillons d'apprentissage pour extraire, par des méthodes mathématiques telles que la programmation dynamique et la modélisation markovienne, les paramètres nécessaires à leur reconnaissance.

Les systèmes fondés sur la connaissance tentent d'utiliser l'expertise humaine notamment les connaissances phonétiques, acoustiques et linguistiques pour déterminer les paramètres utiles à la reconnaissance.

Ces systèmes sont de type systèmes experts<sup>[10]</sup>.

## II.3.2 - Reconnaissance globale et reconnaissance analytique

Les méthodes dites globales, qui consistent à comparer des entités de la taille des mots à des références pré-enregistrées ont été appliquées avec succès à des objectifs limités, telle la reconnaissance de mots isolés ou connectés appliquée à quelques locuteurs. Mais ces méthodes deviennent inutilisables pour la parole continue multi-locuteurs. En effet, la quantité des informations à stocker et les temps de traitement sont inacceptables.

Les méthodes analytiques, qui décomposent le signal en entités élémentaires (par exemple les phonèmes) à identifier, sont plus adaptées à la parole continue<sup>[20]</sup>.

## II.4 - Etat de l'art dans le domaine de la reconnaissance<sup>[6],[18],[10]</sup>

Les outils fournis par l'intelligence artificielle ont donné naissance à la réalisation de plusieurs systèmes de reconnaissance ou de compréhension de la parole dont les plus connus sont:

- Les systèmes américains Harpy, Hearsay I et II et Hwin issus du projet ARPA initié aux Etats-Unis.

Ces systèmes utilisant des techniques d'intelligence artificielle à des résultats d'analyse acoustique très élémentaires sont considérés comme des demi-échecs. Leur apport le plus important s'est situé dans le domaine des architectures logicielles de systèmes complexes plutôt qu'en reconnaissance de la parole.

- Les systèmes français Keal, Esope et Myrtille I et II sont des systèmes de compréhension de la parole fondés sur la connaissance.

Nous retiendrons particulièrement les systèmes experts Sonex (lecture de spectrogrammes), Systexp, Serac et Aphodex fondés sur des connaissances pour le décodage acoustico-phonétique.

La place de plus en plus importante prise par la modélisation markovienne constitue un autre fait saillant des dernières années. Cette approche du type auto-organisation permet, dans sa forme la plus complète, d'attaquer le problème du décodage de la parole continue avec modélisation stochastique du langage.

Le système développé par Jelinek au centre de recherche IBM de Yorktown est entièrement basé sur cette approche probabiliste.

Depuis les années 80, les chaînes de Markov sont de plus en plus utilisées dans les systèmes de reconnaissance de mots.

On peut prédire que les systèmes à venir utiliseront conjointement les chaînes de Markov et les systèmes experts, méthodes présentées encore aujourd'hui comme concurrentes.

## II.5 - Conclusions

Les premiers systèmes commercialisés, généralement de type auto-organiseurs, utilisent des méthodes de reconnaissance globales de mots isolés ou enchaînés avec des vocabulaires limités s'appliquant à quelques locuteurs.

Les systèmes de reconnaissance de la parole continue multi-locuteur sont généralement fondés sur les connaissances et utilisent des méthodes analytiques. Nous nous intéressons pour notre part à ces derniers.

# III - DECODAGE ACOUSTICO-PHONETIQUE

## III.1 - Définition et rôle

Le décodage acoustico-phonétique est un passage obligé pour la reconnaissance analytique de la parole.

Il peut être défini comme la transformation du signal vocal en unités phonétiques identifiées.

Comme nous l'avons souligné précédemment, la difficulté de ce problème tient à la grande variabilité inter-locuteur et intra-locuteur et à la nature continue de la parole.

De plus, il n'existe pas de relations biunivoques qui permettent d'associer à une portion de signal donné une entité phonétique. Ceci pose clairement le problème de segmentation du signal.

Dans un système de reconnaissance automatique, le décodage acoustico-phonétique a pour fonctions:

- a. D'extraire, à partir du signal, les indices acoustiques pertinents.
- b. De segmenter le signal de parole en unités élémentaires clairement définies.
- c. D'identifier ces unités ou de leur attribuer des étiquettes.

Une des démarches envisageable est de se référer à une description des phonèmes ou d'événements phonétiques à l'aide de traits phonétiques articulatoires ou perceptifs, et de rechercher sur le signal, des indices acoustiques si possibles invariants permettant de localiser et de mettre en correspondance une portion du signal et un trait<sup>[12]</sup>.

### III.2 - Représentation paramétrique du signal vocal

La plupart des méthodes de traitement de signal, analysent un intervalle de parole de 20 ms et donnent toutes les 10 ms environ une représentation appelée échantillon centiseconde<sup>[10]</sup>.

Les chercheurs supposent que, sur l'intervalle analysé, le signal est stationnaire.

Le modèle d'audition humain nous incite à rechercher dans le signal le contenu fréquentiel et principalement les fréquences de résonance du conduit vocal (formants).

Les méthodes de traitement de signal utilisées sont:

- La transformée de Fourier rapide
- L'analyse spectrale par vocodeurs à canaux
- La prédiction linéaire
- L'analyse cepstrale.

Les indices acoustiques sont recherchés à partir du signal lui-même ou à partir de ses représentations diverses obtenues par le traitement de signal<sup>[19]</sup>.

### III.3 - Segmentation

La segmentation du signal de parole a pour but de détecter des segments correspondant à des parties stables ou instables, à des phonèmes ou à des syllabes<sup>[10]</sup>.

La plupart des méthodes de segmentation en unités phonétiques utilisent une fonction qui mesure les discontinuités locales du signal ou de son spectre au cours du temps.

Ces discontinuités ne sont pas toutes synchrones, selon le système de paramètres choisi et les critères retenus.

De plus, toutes les discontinuités acoustiques repérées sur le signal ne sont pas interprétables en termes articulatoires ou perceptifs. La tâche s'avère donc délicate, devant la complexité du problème.

La tendance actuelle est de séparer plus nettement la fonction acoustique de la fonction phonétique en traitant séparément les deux aspects de l'unité.

La notion de phones homogènes va dans ce sens.

Cette méthode regroupe en seul segment tous les échantillons spectraux homogènes vis-à-vis de certains paramètres. Ces segments sont infra-phonémiques par construction. Il y a donc sursegmentation vis-à-vis d'une découpe en phonèmes. Mais il devient possible d'insérer un modèle interprétatif sur ces segments pour reformer les phonèmes.

La segmentation consécutive à l'étiquetage des échantillons centisecondes entre dans cette catégorie.

Dans cette méthode, les indices acoustiques sont recherchés sur chaque échantillon centiseconde et une classe leur est attribuée. Les échantillons d'une même classe sont regroupés en un seul segment<sup>[10]</sup>.

### III.4 - Extraction des indices pertinents<sup>[10],[13]</sup>

Le processus de reconnaissance phonétique cherche à localiser sur une fenêtre de signal des indices de base dont l'utilisation dans des règles d'interprétation permettra d'identifier le trait articulaire correspondant.

Ces indices acoustiques sont calculés à partir des paramètres qui déterminent la représentation du signal.

Ce processus utilise plusieurs types de connaissances acoustiques:

- Des connaissances spectrales,
- Des connaissances sur l'évolution temporelle de paramètres tels que les formants ou les énergies,
- Des connaissances contextuelles prenant en compte les parties voisines du signal et éventuellement leurs étiquettes.

Les principaux paramètres utilisés pour l'extraction d'indices acoustiques sont:

- La fréquence du fondamental ou la présence d'énergie en basses fréquences pour la détection de l'indice de voisement,
- La présence d'un intervalle de silence pour la détection des sons plosifs,
- Le taux de passage par zéro du signal élevé (énergie haute-fréquence élevée) pour caractériser les sons fricatifs,
- La présence de formants nets pour la détection des sonantes et des voyelles,
- Les valeurs des formants F1 et F2 pour l'identification des voyelles,
- La valeur de l'énergie totale relative, et la recherche de maxima sur cette courbe pour la localisation des voyelles,
- Les variations de l'énergie dans certaines bandes de fréquences<sup>[3]</sup>.

## IV - ANALYSE LEXICALE

### IV.1 - Introduction

Le niveau lexical concerne la deuxième articulation après le niveau phonétique pour la structure du langage.

La reconnaissance lexicale, ou reconnaissance de mots, fonctionne à partir d'un lexique contenant les mots de la langue ou un sous-ensemble de ceux-ci.

Le lexique est un système dans lequel sont représentées toutes les informations concernant les mots de la langue.

Les problèmes liés à ce niveau sont principalement:

1. Quelle est la nature et l'organisation du lexique?
2. Quel est le format des représentations lexicales?
3. Quelles sortes de procédures rendent possibles l'accès à ces représentations?

### IV.2 - Organisation du lexique

Les mots du lexique sont décrits par les graphèmes qui les composent dans le domaine écrit et par leurs phonèmes dans le domaine oral.

Le passage de la forme graphémique à la forme phonémique peut être effectué directement lorsque les deux formes sont présentes et mises en correspondance dans le lexique (cas des petits vocabulaires) ou par un programme de transcription orthographe-phonétique utilisé pour la synthèse de la parole.

Les mots du lexique sont généralement classés selon une hiérarchie déterminée, ordre alphabétique notamment<sup>[20]</sup>.

### IV.3 - Format des représentations lexicales

Les mots décrivant le lexique peuvent être représentés sous leur forme complète avec toutes leurs déclinaisons ou par leur radical "lexème" et leur terminaison "morphème" (liée à la syntaxe) susceptible d'être déclinée (conjugaison, forme masculine-féminine, forme singulier-pluriel, etc.).

Le choix entre ces deux représentations correspond à un choix entre une quantité de mémoire importante ou un temps de calcul important pour obtenir les formes des mots lors de la reconnaissance<sup>[10]</sup>.

### IV.4 - Spécificité de l'analyse lexicale de la parole

Un système d'analyse lexicale de parole reçoit à son entrée un treillis d'éléments phonétiques correspondant à la transcription phonémique de la phrase prononcée. Ce treillis peut contenir un certain nombre d'erreurs issues de la phase de reconnaissance, ce qui introduit un indéterminisme important dans les traitements ultérieurs.

De ce treillis phonétique, une démarche ascendante permet d'extraire un treillis de mots possibles résultant de diverses combinaisons acceptables en fonction du lexique de l'application. Le rôle de l'analyseur lexical peut être de trouver un chemin cohérent à travers ce treillis de mots en utilisant les diverses sources d'informations disponibles.

Ce processus purement ascendant s'avère insuffisant pour des langages d'une certaine complexité et doit dans ce cas être complété par un processus de haut niveau (syntaxique, sémantique, pragmatique).

### IV.5 - Accès au lexique

La procédure d'accès au lexique est fortement liée à la nature et à l'organisation de ce dernier.

Les modèles d'accès "autonomes" n'utilisent que les informations infra-lexicales issues du traitement de signal ou du décodage phonétique pour accéder au lexique.

Les modèles "interactifs" utilisent, en plus des informations infra-lexicales, des informations de plus haut niveau comme des connaissances syntaxiques ou sémantiques.

Cette opposition entre les modèles autonomes et interactifs d'accès au lexique joue un rôle important pour la prise en compte du contexte lors de l'identification d'un mot.

Une représentation sous forme arborescente permet de diminuer l'occupation mémoire et d'avoir accès rapidement à tous les mots ayant le même radical, ce qui est utile dans un décodage de mots de la gauche vers la droite.

Cela correspond de plus au fonctionnement du décodage humain<sup>[10]</sup>.

### IV.6 - Principaux modèles d'accès au lexique spécifiques à la parole

#### IV.6.1 - Le modèle de la "cohorte" de Marslen-Wilson

Ce modèle s'appuie sur les propriétés de base de l'entrée "parole". Une propriété essentielle de cette dernière est sa durée relativement importante (par rapport à l'entrée "visuelle" de l'écrit). L'information nécessaire pour le traitement lexical d'un mot n'est pas disponible d'emblée mais doit être accumulée dans le temps.

La deuxième propriété de base du signal de parole est sa linéarité, les sons se succédant les uns aux autres sur l'axe temporel.

Le modèle de la "cohorte" envisage le signal de parole comme fournissant au receveur un continuum de possibilités de choix parmi les mots de son répertoire. L'accumulation de l'information extraite du signal va donner lieu à une diminution progressive du nombre d'éventualités au fur et à mesure de la réalisation temporelle de l'entrée.

L'hypothèse principale dans ce modèle est qu'un mot peut être identifié à l'instant "t" à partir duquel ce mot est le seul compatible avec la séquence des phonèmes préalablement reçue. Cet instant t est dit point d'identification du mot. Généralement, ce point d'identification précède la fin du mot.

Le système d'identification exige une mise en correspondance permanente de la chaîne phonémique avec l'ensemble des mots du lexique. L'ensemble des mots compatibles avec la chaîne phonémique à un moment donné de la réalisation du signal constitue la "cohorte". Un mot devient identifiable lorsque la "cohorte" ne contient qu'un candidat.

Le mot à reconnaître est alors dégagé à la suite d'un processus de réduction progressive de la "cohorte". Dans le modèle de Marslen-Wilson, il n'existe pas d'exécuteur central qui contrôle ce processus, les candidats incompatibles s'éliminent d'eux-mêmes.

Dans ce modèle, l'information contextuelle (de niveau supérieur, syntaxique ou sémantique) peut accélérer le processus de sélection des candidats de la "cohorte" en éliminant ceux d'entre eux qui ne sont appropriés au contexte. Mais ce modèle donne, dans tous les cas, la priorité aux informations infra-lexicales.

Par ailleurs, le fonctionnement de ce système est strictement orienté de gauche à droite.

#### **IV.6.2 - Le modèle de perception d'Elman et McClelland**

Ce modèle constitue une tentative d'implémentation du système de la "cohorte" de Marslen-Wilson et vise à rendre compte de la perception des mots parlés.

Dans le système actuellement implémenté, la parole est spécifiée directement en termes de traits distinctifs.

Chaque trait, chaque phonème et chaque mot est représenté par:

- a - Un niveau d'activation variable dans le temps;
- b - Un seuil au delà duquel le nœud est dit actif et transmet un signal d'activation à d'autres nœuds du système en modifiant leurs niveaux d'activation;
- c - Un niveau de repos qui correspond à l'état du nœud en l'absence d'activation.

La transmission de signaux entre les nœuds de différents niveaux peut se faire vers le bas ou vers le haut. Le caractère interactif du modèle suggère la prise en compte de l'effet de contexte. Toutefois, à l'heure actuelle, le mode de traitement des niveaux supérieurs syntaxique et interprétatif n'a pas été élaboré.

## **V - CONCLUSION**

Les notions que nous avons dégagées dans ce chapitre ne sont pas exhaustives mais vont nous permettre d'aborder l'étape de conception de notre système avec une maîtrise des différents concepts liés au domaine de la parole en général et de sa reconnaissance automatique en particulier.

## CHAPITRE II

# INTELLIGENCE ARTIFICIELLE ET SYSTEMES EXPERTS

## I - PRINCIPALES TECHNIQUES D'INTELLIGENCE ARTIFICIELLE

### I.1 - Introduction

Les méthodes classiques de programmation utilisent des suites «hiérarchiques» d'ordres spécifiés à l'avance, ces méthodes algorithmiques ne peuvent traiter, de ce fait, que des problèmes dont la méthode de résolution est standard et à priori connue.

L'intelligence artificielle s'intéresse à tous les domaines où le traitement ne peut être ramené à une méthode algorithmique.

Elle a pour but d'analyser puis de tenter de reproduire sur ordinateur les capacités de raisonnement de l'intelligence humaine<sup>[6]</sup>.

### I.2 - Position du problème

La plupart des problèmes d'intelligence artificielle peuvent se poser comme suit:

- Partant d'un certain état initial (données du problème);
- Ayant pour objectif un certain état final (but à atteindre);
- Disposant d'un certain nombre d'actions possibles (opérateurs de changement d'état).
- Comment procéder pour atteindre le but ?

Cette approche fait apparaître trois composantes:

1. **Les faits:** composés de l'état initial et l'état final;
2. **Les règles:** opérateurs de changement d'état dont on dispose;
3. **Le contrôle:** stratégie d'application des opérateurs sur les faits répondant au «comment?».

L'ensemble faits et règles représente les connaissances que l'on possède sur le domaine.

Le contrôle représente la stratégie d'application de ces connaissances ou stratégie de «raisonnement».

En effet, il est nécessaire pour résoudre un problème d'essayer un certain nombre de règles, pour trouver la suite de règles qui conduit à la solution.

Plusieurs démarches peuvent être envisagées, c'est ce que montrent les principales méthodes de résolution étudiées ci-après<sup>[18]</sup>.

### I.3 - Principales méthodes de résolution

#### I.3.1 - Résolution par espace d'état

Dans cette méthode, à chaque état, on applique toutes les règles possibles et on garde trace en mémoire de toutes ces applications et de tous les états qui en résultent.

On construit ainsi un «graphe d'états» dans lequel on recherchera l'existence d'une solution.

Ce graphe est construit de la façon suivante:

- Soit S l'état initial «nœud sommet»;
- On applique sur S toutes les règles possibles dont on dispose. Cette opération s'appelle «développer un nœud».

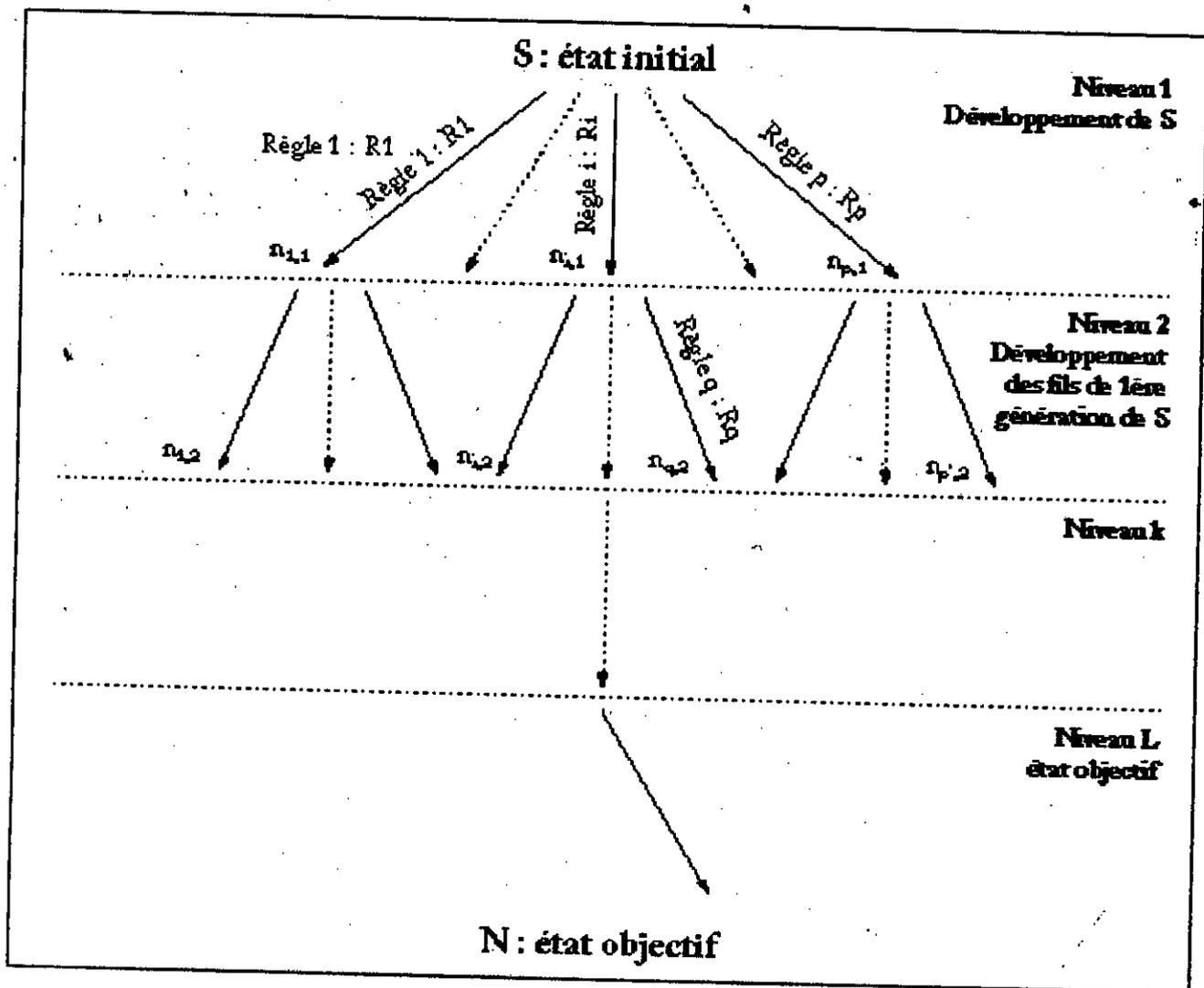
Le développement de S donnera naissance à de nouveaux états «nœuds fils»;

- Les nœuds fils sont à leur tour développés jusqu'à obtention d'un nœud objectif N;
- A partir du nœud N (état objectif), on remonte le graphe des états pour établir la suite des règles qui a conduit à N. Cette suite de règles est la solution du problème<sup>(27)</sup> (voir graphe I).

Cette méthode représente l'énorme inconvénient de nécessiter le développement de tous les nœuds engendrés et la mémorisation du graphe complet des états.

En pratique, une partie seulement du graphe est développée.

Selon les critères de choix sur les nœuds à développer en priorité, plusieurs stratégies sont envisageables:



Graphe I : GRAPHE D'ETATS

## 1.3.1.1. Méthodes aveugles<sup>(27)</sup>

Dans ces méthodes, le choix du nœud à développer ne dépend pas des connaissances que l'on peut avoir sur la position du nœud objectif. Ces méthodes utilisent des stratégies sans «heuristiques».

### a. Stratégie en «largeur d'abord»

Dans cette méthode, on développe en priorité les nœuds les moins profonds, c'est à dire les plus anciennement engendrés.

#### \* Principe

On développe S, puis tous les fils de S de première génération, puis tous ceux de deuxième génération... jusqu'à obtention de N.

#### \* Propriétés

- Si l'espace complet des états comporte une solution, cette méthode en trouvera une avec un chemin minimal;
- Cette méthode est pénalisante en temps, car elle impose de vérifier à chaque étape que l'on ne retombe pas sur un nœud déjà rencontré (cas des graphes non arborescents);
- Cette méthode s'applique à des graphes quelconques.

### b. Stratégie en «profondeur d'abord»

Dans cette méthode, on développe en priorité les nœuds les plus profonds, c'est à dire les plus récemment engendrés.

#### \* Principe

- On développe S, puis le premier fils  $n_1$  de S puis le premier fils  $n_{1,1}$  de  $n_1$ ... jusqu'à obtention de N.
- Pour éviter une recherche inutile sur des nœuds de plus en plus profonds on se fixe un niveau de profondeur maximal L au delà duquel les nœuds de niveaux inférieurs sont abandonnés.

#### \* Propriétés

- Si l'espace complet des états comporte une solution à un niveau de profondeur inférieur à la valeur limite fixée, cette méthode en trouvera une.
- Le chemin produit n'est pas a priori minimal;
- Cette méthode ne peut s'appliquer qu'à des graphes de type arborescent, car l'existence d'un nœud déjà exploité ne peut conduire qu'à une boucle sans fin.
- Le temps et l'espace mémoire occupé sont considérablement réduits.

### c. Méthode du «coût uniforme»

Dans cette méthode, on développe en priorité les nœuds dont la distance au nœud sommet est la plus petite. Cette distance étant calculée par le cumul des coûts de toutes les opérations qui ont conduit du sommet à ce nœud.

#### \* Principe

On définit la distance d'un nœud  $n_i$  au sommet S par la fonction  $g(n_i)$  telle que:

$$g(S) = 0$$

$$g(n_i) = g(s) + C_{s,i} \text{ où } C_{j,i}: \text{ Coût de l'opération qui transforme } n_j \text{ en } n_i.$$

On développe S, puis le fils  $k$  de S dont  $C_{s,k}$  est minimal. A tous les fils ainsi engendrés on associe la fonction  $g(n)$  et on développe le nœud dont  $g(n)$  est minimal... jusqu'à obtention de N.

#### \* Propriétés

- Si l'espace complet des états comporte une solution, cette méthode en trouvera une avec un chemin minimal et un coût des opérations minimal.
- L'espace mémoire et le temps sont réduits.

I.3.1.2. Méthodes ordonnées (utilisation d'heuristiques)<sup>(27)</sup>

Dans certains cas, il est possible de réduire la recherche en utilisant une information sur la position du nœud objectif. Cette information est appelée «heuristique».

Dans ces méthodes, à chaque nœud, on associe une valeur représentant la «promesse» du nœud et on développe en priorité les nœuds les plus prometteurs.

Soit  $f$  la fonction d'évaluation d'un nœud et  $f(n)$  la valeur promesse du nœud  $n$ . On convient qu'un nœud est d'autant plus prometteur que sa valeur  $f(n)$  est petite.

\* Principe

Une façon de se donner une fonction d'évaluation qui rende compte de la promesse d'un nœud est de la choisir de telle sorte qu'elle approche le coût minimal du chemin allant du sommet à l'état objectif en passant par ce nœud.

**I.3.2 - Résolution par espace de sous-problèmes**

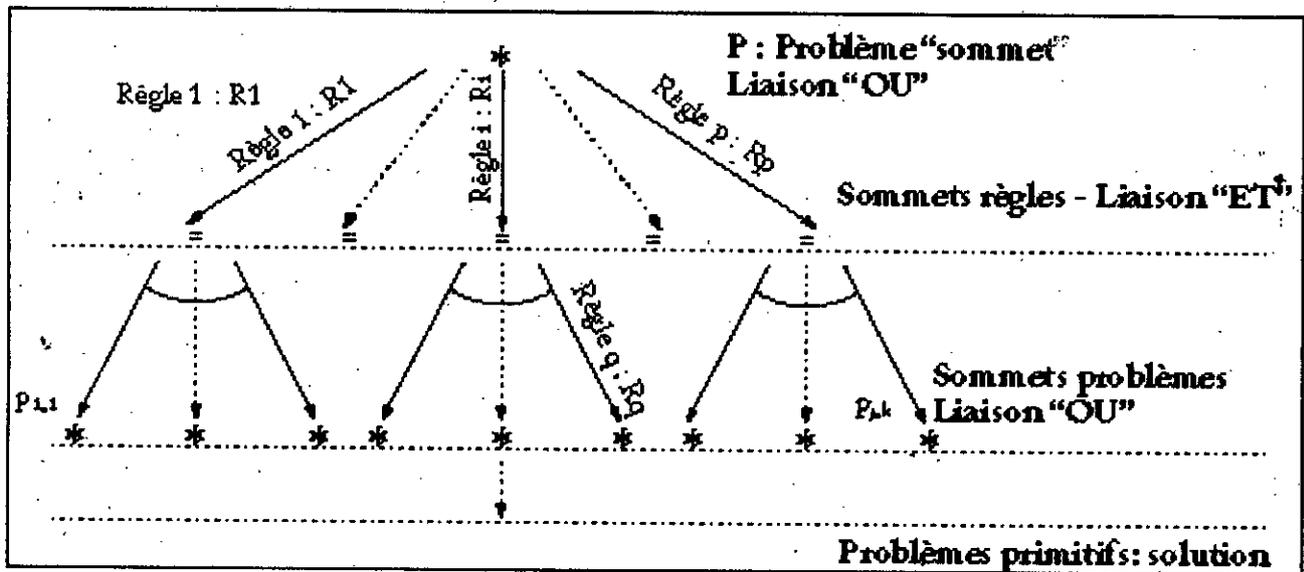
Dans certains cas, les règles dont on dispose sont des opérateurs de décomposition de problèmes en sous-problèmes. L'espace des sous-problèmes est alors construit de la façon suivante:

- Soit  $P$  le problème initial: problème «sommet»
- On applique à  $P$  toutes les règles possibles dont on dispose. Chacune de ces règles va donner un ensemble de sous-problèmes fils.

Ces fils sont à leur tour exploités par les règles jusqu'à obtention de problèmes primitifs considérés comme étant résolus. L'espace des sous-problèmes est alors représenté par un graphe de type ET-OU.

Chaque sommet symbolise soit un problème  $P_i$ , soit une application de règle  $R_j$ . Si  $R_1, \dots, R_p$  sont toutes les applications des règles qui décomposent  $P_i$ , l'ensemble des arcs  $P_i R_j$  détermine une liaison «OU».

Si  $P_1, \dots, P_q$  sont tous les sous-problèmes de  $P_i$  introduits par l'application de  $R_j$ , l'ensemble des arcs  $R_j P_i$  détermine une liaison «ET»<sup>(27)</sup> (voir graphe II).



Graphe II : Graphe ET-OU

## II - LES SYSTEMES EXPERTS

### II.1 - Introduction

La recherche sur les mécanismes généraux de résolution de problèmes en Intelligence Artificielle a abouti à des logiciels qui se sont avérés insuffisamment puissants dans des domaines particuliers. Il manquait à ces programmes la connaissance approfondie du domaine concerné.

Pour résoudre ce compromis entre généralité et efficacité, on s'est intéressé à des logiciels agissant sur des domaines précis. Ce sont les systèmes experts.

### II.2 - Définition

Les systèmes experts sont des logiciels destinés à modéliser le comportement d'un expert humain dans les domaines où l'expertise est:

- Insuffisamment explicite pour être ramenée à une méthode algorithmique;
- Sujette à évolution<sup>[18]</sup>.

### II.3 - Impératifs fondamentaux

Le système expert devant recevoir de l'humain les connaissances cognitives du domaine concerné, il doit, en plus de ses capacités de raisonnement, fournir une possibilité d'expression de ces connaissances.

Un système expert doit donc répondre aux impératifs fondamentaux suivants:

**Impératif 1:** Permettre l'expression la plus aisée des connaissances.

**Impératif 2:** Supporter l'ensemble des connaissances et prévoir leur évolution.

**Impératif 3:** Reasonner sur l'ensemble de ces connaissances pour en déduire d'autres et rendre compte parfois de la manière de procéder<sup>[9]</sup>.

### II.4 - Composantes principales (voir figure I)

Pour pouvoir répondre aux impératifs cités précédemment, un système expert doit comporter:

1. **Un langage d'expression des connaissances** qui permet une représentation symbolique des connaissances du domaine et ce, de la manière la plus proche possible de leur forme d'émergence, pour satisfaire l'impératif 1.

2. **Une base de connaissances** qui contient la connaissance spécifique du domaine. Celle-ci doit être organisée de façon dynamique et évolutive pour satisfaire l'impératif 2. Cette base est en fait constituée de deux bases distinctes:

a. **Une base de faits** décrivant des états et représentant le «savoir» du système.

b. **Une base de règles** indiquant les actions ou les conséquences à tirer au vu des faits présents. Cette base représente le «savoir-faire» du système.

3. **Un moteur d'inférences** capable de raisonner sur les connaissances pour satisfaire l'impératif 3. C'est un programme d'intelligence artificielle, mettant en œuvre des mécanismes généraux de combinaisons des faits et des règles.

Selon différentes stratégies, le moteur sélectionne des règles, les actionne, les enchaîne pour déduire de nouveaux faits jusqu'à satisfaire un but prédéterminé.

Pour rendre compte de la manière de procéder, le moteur peut assurer des fonctions supplémentaires de dialogue et d'explications<sup>[14]</sup>.

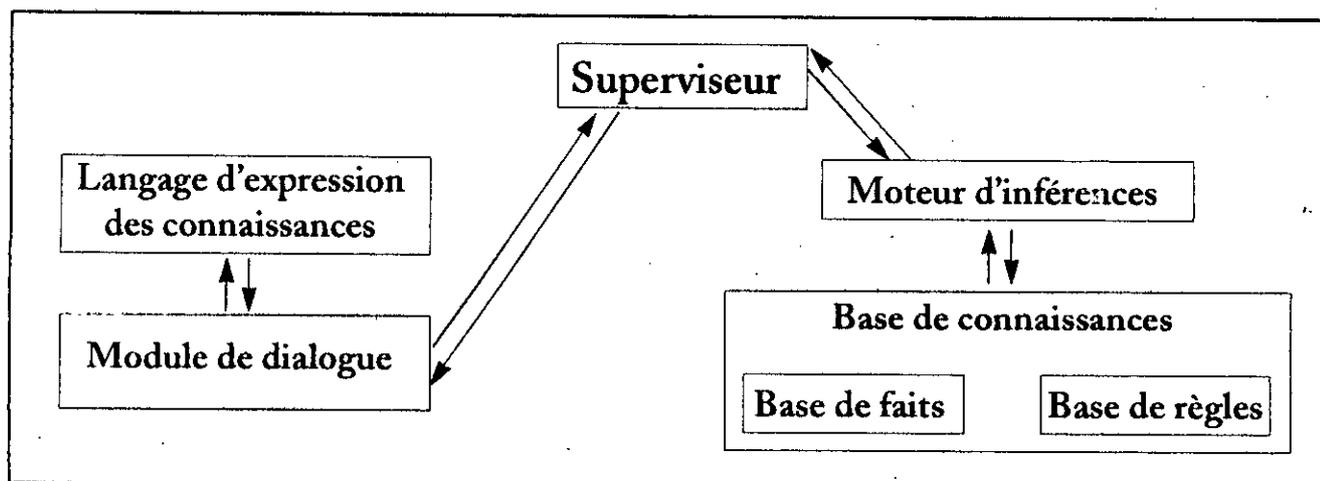


Figure I : Organisation de principe d'un système expert <sup>[18]</sup>

## II.5 - Formulation des règles <sup>[15]</sup>

Quel que soit le langage d'expression des connaissances associé au système expert, les règles sont toujours formulées selon un principe commun: la rédaction associative.

Dans cette formulation, une règle n'est pas invoquée par son nom externe mais par un fragment de l'information contenu dans la règle (accès par filtrage).

Chaque règle est rédigée de façon à permettre par sa représentation, de définir ses conditions de déclenchement et ses effets. Une règle s'écrit alors sous la forme:

**Si Prémisse alors Conclusion**

et est composée de deux parties:

**Règle = Déclencheur + Corps**

Pour invoquer une règle, on fournit un groupe de faits compatibles avec le déclencheur ou filtre de la règle.

- Dans le cas où on raisonne des données vers les buts, c'est à dire lorsque l'on cherche à déduire une conclusion de certains faits, le déclencheur de la règle est sa prémisse. On parle alors de règles en avant.

**Filtre = Prémisse: Règle en avant**

- Dans le cas où on raisonne des buts vers les données, c'est à dire lorsque l'on cherche à vérifier certaines données à partir des buts, le déclencheur de la règle est sa conclusion. On parle de règle en arrière.

**Filtre = Conclusion: Règle en arrière**

## II.6 - Fonctionnement des moteurs d'inférences

Pour pouvoir raisonner sur les connaissances, le moteur d'inférences doit effectuer au vu de la base des faits, le choix des règles à utiliser pour résoudre un problème.

Au démarrage du moteur, la base de faits contient l'énoncé du problème sous forme de données et de buts.

Grâce à la rédaction associative, le moteur sélectionne dans la base des règles, les règles susceptibles d'être déclenchées. Cette phase est appelée phase d'évaluation.

En phase d'exécution, le moteur déclenche, selon différentes stratégies, les règles retenues et infère ainsi de nouveaux faits. Le moteur s'arrête avec succès lorsque le but est atteint, ou avec échec lorsqu'aucune règle n'a pu être déclenchée<sup>[17]</sup> (voir figure II).

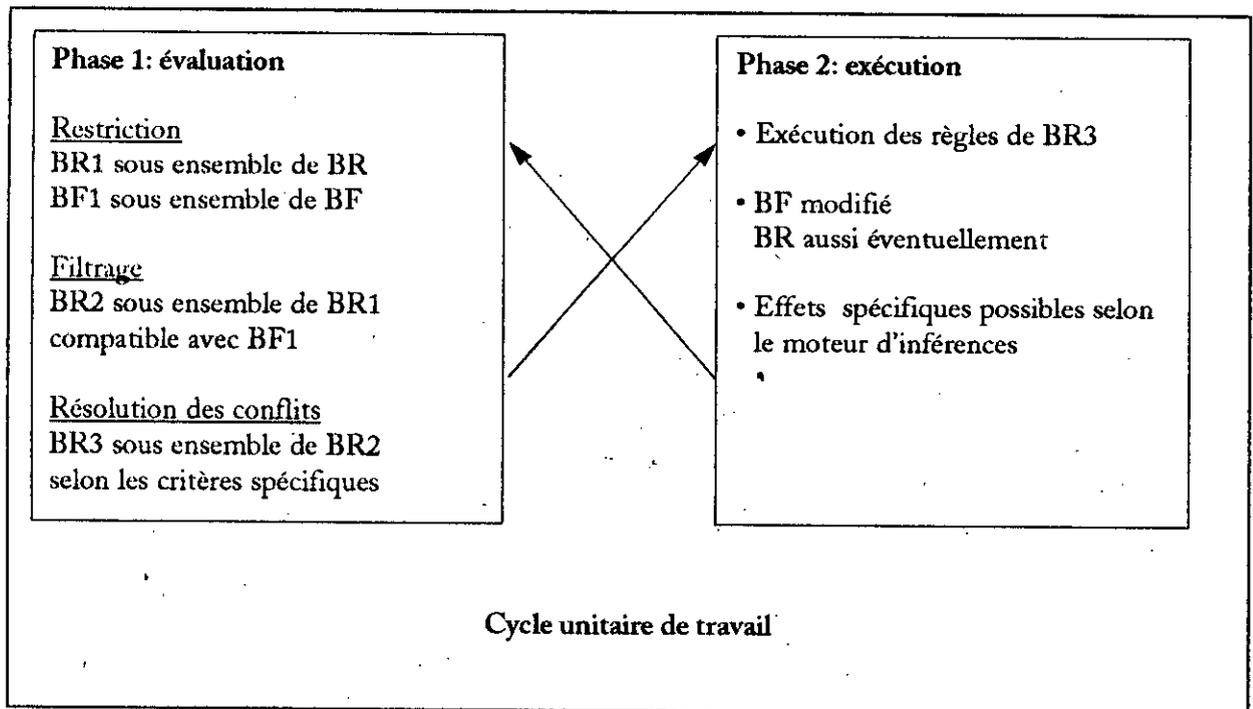


Figure II : Représentation schématique du fonctionnement d'un moteur d'inférences<sup>[17]</sup>

### A - Phase d'évaluation

Elle comporte en fait trois étapes:

#### a. Sélection

Dans cette étape, le moteur détermine à partir de l'état de la base de connaissances, un sous-ensemble de faits BF1 et un sous ensemble de règles BR1 qui nécessitent d'être comparés. Cela peut consister à traiter en priorité pour une certaine durée, un groupe de connaissances.

#### b. Filtrage

Dans cette étape, le moteur compare les déclencheurs de chaque règle retenue dans BR1 avec les faits de BF1, et détermine l'ensemble des règles compatibles avec les faits considérés. Il construit ainsi l'ensemble des conflits BR2 composé des règles déclenchables.

#### c. Résolution des conflits

Dans cette étape, le moteur détermine à partir de BR2 un sous-ensemble de règles BR3 qui doivent être effectivement déclenchées. Le choix de ces règles dépend de la stratégie du moteur. On peut choisir les règles les moins coûteuses, les plus prometteuses (si on possède une heuristique), etc.

### B - Phase d'exécution

Dans cette phase, le moteur met en œuvre, en les déclenchant, les actions contenues dans les corps des règles de BR3. Ces actions vont provoquer un changement d'état de la base des faits. Si cette base ne contient pas encore le but recherché, le moteur retourne en phase d'évaluation et reconsidère la nouvelle base de faits.

### C - Régime «irrévocable» ou «par tentatives»

Lorsque l'ensemble BR3 est vide, si le moteur fonctionne en régime «irrévocable», il s'arrête en échouant. S'il fonctionne en régime «par tentatives», il reconsidère l'ensemble des conflits BR2 du cycle antérieur et redéfinit un nouvel ensemble BR3. On dit qu'il effectue un retour-arrière.

## III - REPRESENTATION DES CONNAISSANCES

### III.1 - Introduction

Les systèmes experts se différencient selon l'aptitude du langage d'expression des connaissances qui leur est associé.

Ces langages, pour plus de performance, offrent souvent en dehors des faits et des règles d'autres classes de connaissances.

### III.2 - Faits<sup>(22)</sup>

On peut classer les faits selon plusieurs catégories faisant appel à des moyens de manipulation spécifiques.

### III.3 - Règles<sup>(22)</sup>

Ce sont des «connaissances opératoires» référant aux faits par rédaction associative.

Selon les systèmes, on dispose de possibilités plus ou moins riches pour décrire les déclencheurs et corps de règles.

On distingue souvent un sous-langage pour déclencheurs et un sous-langage pour corps.

#### III.3.1 - Langages pour déclencheurs

Ils jouent un rôle essentiel dans la rédaction associative, chaque langage offre un modèle de compatibilité entre les faits et les déclencheurs de règles.

##### a. Identité

Dans ce cas, un filtre de règle n'est compatible qu'avec des faits parfaitement identiques.

##### b. Semi-unification

Dans ce cas, les règles peuvent comporter des variables. Les identificateurs de variables étant compatibles avec n'importe quelle constante présente dans les faits (spécialisation universelle):  $F(x)$  compatible avec  $F(a)$ .

##### c. Unification

Dans ce cas, les faits et les règles peuvent comporter des variables. Le filtre  $P(X, f(a))$  est compatible avec le fait  $P(h(g), b)$ .

#### d. Semi-unification avec théorème de filtrage

Le langage est parfois doté de théorèmes qui étendent virtuellement la base de faits. Ces théorèmes permettent de remplacer un filtre non compatible par un filtre équivalent.

#### **III.3.2 - Langages pour corps de règles**

Les effets des corps des règles se traduisent en général par des transformations de la base des faits. En fait, on distingue en plus de ces effets:

- Des instructions de contrôle: effets sur le régime du moteur.
- Des instruction d'entrée/sortie: communication externe.

#### **III.4 - Métarègles<sup>[21]</sup>**

Ce sont des règles susceptibles de préciser la manière d'utiliser les règles.

Ces règles interviennent à un niveau supérieur, pour guider le moteur dans le choix des règles à considérer.

#### **III.5 - Procédures<sup>[21]</sup>**

Certaines règles peuvent contenir dans leur corps des appels de procédure désignées par leur nom et effectuant un traitement particulier.

#### **III.6 - Démons<sup>[21]</sup>**

Ce sont des connaissances opératoires rédigées généralement en rédaction associative et qui exécutent des tâches spécifiques.

Les démons sont immédiatement exploités dès que surviennent certains faits. Ils ne sont pas sujets aux mêmes cycles que les règles.

#### **III.7 - Usage de variables<sup>[24]</sup>**

L'usage des variables permet la factorisation des connaissances. Les moteurs d'inférences sont qualifiés de moteurs d'ordre 1 lorsqu'ils admettent un langage avec variable, ils sont qualifiés de moteurs d'ordre 0 lorsqu'ils admettent un langage sans variables.

### **IV - CRITERES DE DIFFERENCIATION DES MOTEURS D'INFERENCEES**

#### **IV.1 - Introduction**

Les moteurs d'inférences peuvent être différenciés par certains critères qui déterminent leur fonctionnement et leurs aptitudes.

#### **IV.2 - Mode d'invocation des règles<sup>[16]</sup>**

##### **a - Chainage avant**

Un moteur fonctionne en chainage avant lorsque les déclencheurs des règles portent sur les données (faits de valeur de vérité établie): raisonnement des données vers les buts.

##### **b - Chainage arrière**

Un moteur fonctionne en chainage arrière lorsque les déclencheurs de règles portent sur des buts (faits de valeur de vérité à établir). Les buts sont des faits pouvant représenter des hypothèses à vérifier ou des problèmes à résoudre: raisonnement des buts vers les données.

c - **Chaînage mixte**

Un moteur fonctionne en chaînage mixte lorsque les déclencheurs de règles peuvent porter aussi bien sur les données que sur les buts.

d - **Chaînage bidirectionnel**

Un moteur fonctionne en chaînage bidirectionnel lorsque pour un même problème, il utilise tantôt des règles en avant, tantôt des règles en arrière.

IV.3 - Modes d'inférences<sup>[34]</sup>

a - «Moteurs 0»

Les «Moteurs 0» qui n'utilisent pas de variables invoquent des règles de la forme:

$$\text{Si } P \text{ alors } Q$$

Cette règle a pour équivalent en logique la proposition:

$$P \Rightarrow Q$$

Interprétée par: Si  $P \Rightarrow Q$  est vrai et  $P$  vrai alors  $Q$  est vrai (chaînage avant) ou bien par: Si on veut résoudre  $Q$  et que  $P \Rightarrow Q$  est vrai alors il suffit de résoudre  $P$  (chaînage arrière). Ce mode d'inférence est connu en logique sous le nom de «Modus Ponens». Les «Moteurs 0» utilisent le mode d'inférence «Modus Ponens».

b - «Moteurs 1»

Les «Moteurs 1» admettent des variables et utilisent des règles de la forme:  
Si  $P(x)$  alors  $Q(x)$  dont l'équivalent en logique est:

$$P(x) \Rightarrow Q(x), \text{ quelque soit } x.$$

Cette règle est interprétée par:

Si  $P(x) \Rightarrow Q(x)$  est vrai quelque soit  $x$ , alors:

$P(a) \Rightarrow Q(a)$  est vrai.

Ce mode d'inférence est connu en logique sous le nom de «spécialisation universelle».

Les «Moteurs 1» fonctionnent selon la combinaison des modes d'inférences «spécialisation universelle» et «Modus Ponens».

IV.4 - Stratégie de résolution<sup>[28]</sup>

IV.4.1 - Représentation par espace d'états

Les règles étant considérées comme des opérateurs de changement d'état (voir chapitre II, I.3), l'espace des états est obtenu par l'application de toutes les règles possibles depuis l'état initial jusqu'à l'état objectif (but du problème). L'état objectif étant considéré comme une base de faits:

- Contenant le but recherché (chaînage avant);
- Ne contenant plus de problèmes à résoudre (chaînage arrière).

#### **IV.4.2 - Représentation par espace de sous-problèmes**

Dans le cas d'un fonctionnement en chaînage arrière, une règle peut être considérée comme un opérateur de décomposition de problème (voir chapitre II, I.3). L'espace des sous-problèmes est alors représenté par un graphe ET-OU.

#### **IV.4.3 - Rôle de l'étape de sélection dans la stratégie**

L'étape de sélection peut consister à choisir de traiter en priorité les faits les plus récemment engendrés (fonctionnement en profondeur d'abord), ou les plus anciennement engendrés (largeur d'abord), ou bien de choisir les faits selon les critères dictés par une stratégie spécifique.

On voit que l'étape de sélection est étroitement liée à la stratégie de recherche utilisée.

#### **IV.4.4 - Rôle de l'étape de «résolution des conflits»**

Cette étape permet de choisir une méthode de recherche ordonnée en sélectionnant les règles les moins coûteuses ou les plus prometteuses (introduction d'heuristiques) ou les moins complexes. Si la base de règles est ordonnée, cette étape peut consister uniquement à choisir les premières règles de la base.

On peut aussi guider la résolution de conflit en utilisant des métarègles.

#### **IV.4.5 - Régime irrévocable, régime par tentatives**

Un moteur fonctionne en régime irrévocable si les règles choisies lors de la résolution des conflits ne sont pas remises en cause en cas d'échec.

Un régime est dit par tentatives, lorsqu'il admet des retours-arrières pour reconsidérer l'ensemble des conflits en cas d'échec de la recherche.

#### **IV.5 - Non monotonie<sup>[18]</sup>**

Un moteur est dit non monotone lorsqu'il permet la suppression de connaissances (faits ou règles). Ceci présente un intérêt dans les situations où:

- La valeur de vérité d'un fait peut évoluer (cas des traitements interactifs);
- On introduit des hypothèses de travail qu'on doit pouvoir remettre en cause en cas d'impasse.

#### **IV.6 - Intérêt de la structuration de la base de connaissance<sup>[14]</sup>**

Pour améliorer les performances des systèmes experts, on peut structurer la base de connaissances.

##### **IV.6.1 - Structuration en classe de connaissances**

Cette méthode consiste à distinguer plusieurs classes de données ou de règles.

Certaines règles n'étant valides que dans certaines conditions, on peut classer par exemple les règles par contexte. On ne se réfère alors, en fonction du contexte, qu'à une partie de la base de règles.

Des techniques sophistiquées, telles que les réseaux, peuvent aussi être utilisées.

##### **IV.6.2 - Règles en vrac**

Un moteur ne peut tolérer des règles en vrac que si, quelque soit l'ordre des règles, la solution à un problème est toujours identique.

En général, l'ordre des règles contribue efficacement au bon fonctionnement du moteur.

## **V - CONCLUSION**

Nous avons présenté les principales techniques d'intelligence artificielle et défini les composantes et les principes de fonctionnement des systèmes experts. Les concepts étudiés nous permettent d'explicitier la terminologie employée dans la partie consacrée à la conception et à la réalisation de notre système expert de reconnaissance de la parole.

---

# **DEUXIEME PARTIE**

## **CONCEPTION ET REALISATION DU SYSTEME**

**Chapitre I**  
**Position du problème**

**Chapitre II**  
**Représentation interne du signal**

**Chapitre III**  
**Module de prétraitement**

**Chapitre IV**  
**Module d'étiquetage**

**Chapitre V**  
**Module de segmentation**

**Chapitre VI**  
**Module d'identification phonétique**

**Chapitre VII**  
**Module d'analyse lexical**

**Chapitre VIII**  
**Organisation et structure des logiciels**

## CHAPITRE I

### POSITION DU PROBLEME

#### I - OBJECTIFS VISES

Le but assigné à notre travail est la conception d'un système expert d'aide à la reconnaissance automatique de la parole continue multilocuteur.

Un système de reconnaissance automatique de la parole naturelle, au sens le plus large, possédant un vocabulaire illimité, paraît tout à fait utopique eu égard aux difficultés inhérentes au domaine.

Nous situerons donc notre contribution à la réalisation d'un système limité de reconnaissance de la parole continue, dans le sens où le langage admis par le système est un sous-ensemble du langage parlé, de structure simple.

Il s'agit, à partir d'un signal vocal paramétrisé, de reconnaître par des techniques d'intelligence artificielle, les mots qui en constituent le message.

#### II - SYSTEME ENVISAGE

Notre approche est de tenter d'utiliser, dans notre système, les connaissances que possèdent les experts humains qui ont étudié la parole et ce à plusieurs niveaux: acoustique, phonétique et lexical.

Un tel système est dit «fondé sur la connaissance» par opposition aux systèmes dits «auto-organiseurs».

En effet, les systèmes auto-organiseurs n'utilisent pas explicitement les connaissances expertes liées au domaine, mais des échantillons d'apprentissage mémorisés auxquels sont comparés les échantillons candidats à la reconnaissance, par des méthodes mathématiques (voir 1<sup>ère</sup> partie, chapitre I)<sup>(10)</sup>.

Notre choix est conforté par les arguments suivants:

1. Les systèmes «fondés sur la connaissance» sont naturellement amenés à mettre en œuvre des techniques de type intelligence artificielle, objet de notre étude.
2. Un tel système, s'appuyant sur les connaissances des experts, ne nécessite pas d'apprentissage.
3. L'utilisation par le système de connaissances expertes aux niveaux phonétique et acoustique permet une reconnaissance analytique; méthode la plus appropriée à la parole continue.
4. Les connaissances qu'ont les experts s'appuient principalement sur les "invariants" présents dans les spectrogrammes des sons qu'ils étudient. Un système fondé sur de telles connaissances est alors indépendant du locuteur, donc le mieux adapté à la reconnaissance multilocuteur, exigence de notre système.

#### III - VUE GENERALE

L'élaboration d'un système expert fait appel à 3 composantes principales:

- Un moteur d'inférences qui met en œuvre des mécanismes généraux de raisonnement;
- Une base de connaissances comportant l'ensemble des connaissances relatives au domaine d'application, notamment une base de faits et une base de règles;
- Un langage d'expression qui permet une représentation symbolique des connaissances du domaine.

En fait, la spécificité du domaine qui nous intéresse nous conduit à mettre en œuvre une architecture plus complexe pour notre système.

Dans la majorité des domaines d'application, les systèmes experts opèrent sur des problèmes dont les données d'entrée (description de l'état initial ou faits initiaux) sont définies et constantes tout au long d'une même session de travail.

Le signal de parole caractérisé par sa variabilité sur l'axe temporel implique une évolution à chaque instant des données d'entrée. Cet aspect du problème nous impose de prévoir des logiciels spécifiques à la parole pour gérer l'évolution temporelle des faits initiaux.

D'autre part, la reconnaissance automatique de la parole, faisant appel à plusieurs étapes successives de traitement tels que la segmentation, l'étiquetage, l'identification phonétique et la reconnaissance de mots nous conduit à prévoir une architecture modulaire des logiciels où chaque module assure une étape particulière du traitement.

Ces différents modules ne font pas tous appel aux systèmes experts, ce qui implique l'implémentation de logiciels spécifiques à ce domaine d'application. Ces logiciels sont néanmoins écrits avec des méthodes de type intelligence artificielle.

#### IV - SOURCES DE CONNAISSANCES

A défaut de la collaboration d'experts en lecture de spectrogrammes, nous avons été amenés, dans le cadre de cette application, à recueillir les connaissances du domaine à partir de deux sources principales:

- Les descriptions acoustiques et phonétiques des sons et de leurs caractéristiques détaillées publiées par les chercheurs du domaine;
- L'observation et l'étude des sonagrammes et spectrogrammes trouvés dans la littérature ou glanés dans l'en-tourage.

Nous avons essayé, à partir de ces sources, de recueillir le maximum "d'invariants" et d'indices pertinents pour l'ensemble des sons étudiés afin d'implémenter les bases de règles de notre système expert.

#### V - DEMARCHE GLOBALE (voir figure III)

Nous voulons transformer un continuum vocal, paramétrisé et mémorisé dans notre fichier d'entrée, en unités phonétiques identifiées en procédant à l'analyse phonétique du signal.

La difficulté de ce traitement tient au fait qu'il n'existe pas de portion de signal, facilement localisable, qui corresponde de façon biunivoque à un phonème.

Il faut procéder à la segmentation du signal avant de pouvoir effectuer le décodage acoustico-phonétique des segments.

Notre système a donc pour fonctions principales:

- a. L'extraction des paramètres pertinents.
- b. La segmentation du signal en unités bien définies.
- c. L'identification de ces unités.
- d. La reconstruction des mots contenus dans le message d'entrée.

##### V.1 - Extraction des paramètres

Pour des raisons de non stationnarité du signal vocal, nous travaillons sur des échantillons centiseconde de parole, c'est à dire sur des tranches de signal de durées égales à 10 ms, intervalle où le signal peut être considéré comme stationnaire.

Chaque échantillon centiseconde est représenté dans notre fichier par une liste de paramètres appelée «vecteur acoustique».

L'ensemble des vecteurs acoustiques contenus dans le fichier représente la phrase candidate à la reconnaissance. Ce fichier est introduit manuellement dans le système mais issu de signaux réels.

### V.2 - Segmentation du signal

Nous procédons à la segmentation du signal en pseudo-phonèmes, c'est à dire en unités proches du phonème. Nous introduisons à cet effet plusieurs étapes pour apporter des solutions aux problèmes liés à la segmentation.

### V.3 - Identification des unités

Nous procédons à l'identification des unités sus-citées, en utilisant les connaissances que nous avons sur les caractéristiques et les traits distinctifs des phonèmes et ce, en tenant compte des variations allophoniques dues aux altérations provoquées par le contexte de réalisation de chaque phonème.

### V.4 - Reconstruction des mots

Grâce à un lexique contenant le vocabulaire admissible par notre système, nous procédons à l'identification des mots par leur comparaison à un dictionnaire.

## VI - RESTRICTIONS IMPOSEES

L'incertitude sur certaines connaissances que nous avons du domaine et les objectifs visés par notre système, nous ont conduits à imposer certaines restrictions.

### VI.1 - Langage utilisé

- Le langage utilisé est un sous-ensemble du langage parlé de construction simple. Seules certaines structures grammaticales sont admises.
- Le dictionnaire des mots admissibles ne dépasse pas cinquante mots.
- Le manque de connaissances précises sur les semi-voyelles nous impose de ne pas traiter les mots qui en contiennent.
- Les suites voyelles-voyelles ne sont pas admises.

### VI.2 - Contraintes imposées au locuteur

Le locuteur doit respecter les contraintes suivantes pour être coopératif:

- Le débit d'élocution doit être un débit de lecture.
- Les informations extra-linguistiques telles que les hésitations et les interjections doivent être éliminées.

### VI.3 - Conditions d'enregistrement du signal

Notre système tient compte du bruit ambiant pourvu qu'il ne soit pas plus important que le signal lui-même. Nos fichiers d'entrée représentent donc les paramètres d'un signal non dépourvu de bruit.

## CHAPITRE II

### REPRESENTATION INTERNE DU SIGNAL

#### I - INTRODUCTION

En acoustique, un son se définit classiquement au moyen de son amplitude, de sa durée et de son timbre. Le traitement du signal a pour but précisément de quantifier ces trois grandeurs pour faire correspondre à l'onde sonore temporelle une description multidimensionnelle<sup>[10]</sup>.

La reconnaissance automatique de la parole opère sur une description du signal, généralement non temporelle, mais spectrale. De plus, le signal doit être traité pour réduire la redondance et trouver des paramètres pertinents pour la reconnaissance.

En effet, le message utile est noyé parmi d'autres informations non pertinentes pour la reconnaissance. Le problème fondamental qui se pose alors, est de définir les critères précisant la pertinence des paramètres à extraire dans le signal.

#### II - PARAMETRES RETENUS

Les objectifs visés par notre système et les indices significatifs de distinction entre sons déterminent les paramètres pertinents nécessaires à la reconnaissance.

Pour chaque échantillon centisecondes de signal, nous avons retenu les paramètres suivants:

##### II.1 - Energies par bandes de fréquences

Il s'agit des énergies données par un banc de filtres contigus et dont les bandes de fréquences sont déterminées par les indices nécessaires à la reconnaissance. Nous avons retenu les bandes de fréquences suivantes:

$$B1 = [250 - 450] \text{ Hz}$$

$$B2 = [450 - 650] \text{ Hz}$$

$$B3 = [650 - 850] \text{ Hz}$$

$$B4 = [850 - 1050] \text{ Hz}$$

$$B5 = [1050 - 1300] \text{ Hz}$$

$$B6 = [1300 - 1600] \text{ Hz}$$

$$B7 = [1600 - 1900] \text{ Hz}$$

$$B8 = [1900 - 2500] \text{ Hz}$$

$$B9 = [2500 - 3400] \text{ Hz}$$

$$B10 = [3400 - 4300] \text{ Hz}$$

$$B11 = [4300 - 5000] \text{ Hz}$$

$$B12 = [5000 - 6000] \text{ Hz}$$

$$B13 = [6000 - 7000] \text{ Hz}$$

$$B14 = [7000 - 8000] \text{ Hz}$$

##### II.2 - Taux de passage par zéro du signal

Ce paramètre essentiel pour la distinction des sons fricatifs a été rajouté pour la représentation paramétrique du signal. Il est calculé sur le signal temporel échantillonné et numérisé par un algorithme qui comptabilise le nombre de changement de signes des valeurs des échantillons.

### II.3 - Intensité du fondamental

Ce paramètre, indispensable pour la détection des sons voisés, est noté IF0 (intensité du pitch). Sa valeur exacte n'est pas nécessaire, seule sa visibilité sur le spectrogramme détectée par la présence d'énergie en basses fréquences est prise en compte.

### II.4 - Fréquences et intensités des 2 premiers formants

Pour la distinction entre voyelles et la détection de la structure formantique des consonnes sonnantes, nous avons besoin des fréquences et des intensités des 2 premiers formants F1, F2 et IF1, IF2. ces valeurs sont données par l'analyse par sonagraphe ou par l'algorithme de Levinson (méthode de corrélation) avec la modélisation AR.

### II.5 - Centre de gravité spectral du signal

Le centre de gravité spectral est un indice pour les voyelles et les fricatives et est calculé par notre système en numéro de canal sur un vocodeur de 14 canaux. Il est donné par la relation:

$$CDG = \frac{\sum_{i=1}^{14} i E_i}{\sum_{i=1}^{14} E_i}$$

## III - REPRESENTATION INTERNE

Comme nous travaillons en Lisp, langage d'intelligence artificielle qui manipule très aisément les listes, nous avons opté pour la représentation du vecteur acoustique pour une représentation en liste de paramètres. Chaque échantillon centiseconde sera alors représenté par la liste:

ECSi = (E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12 E13 E14 IF0 F1 F2 IF1 IF2 TPZ)

Chaque échantillon de signal est donc représenté par 20 paramètres correspondants à:

E1 Energie dans la bande de fréquence B1

E2 Energie dans la bande de fréquence B2

.....

Ej Energie dans la bande de fréquence Bj

.....

E14 Energie dans la bande de fréquence B14

IF0 Intensité du fondamental

F1, F2 Fréquences des 2 premiers formants

IF1, IF2 Intensités des 2 premiers formants

TPZ Taux de passage par zéro du signal

# ORGANISATION GENERALE DU SYSTEME

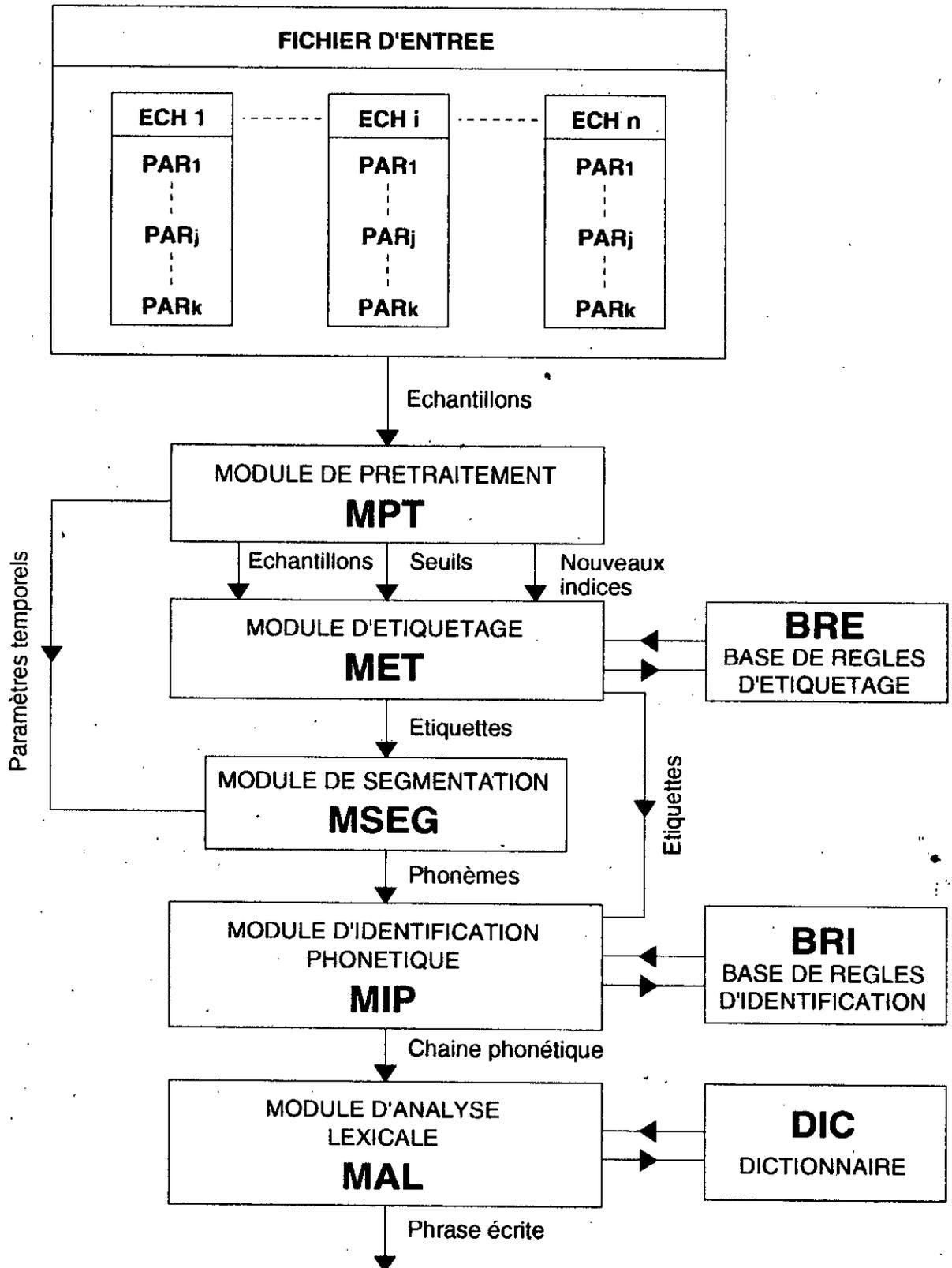


Figure III

## CHAPITRE III

### MODULE DE PRETRAITEMENT

#### I - INTRODUCTION

Pour procéder à l'étiquetage des échantillons centisecondes, notre système expert utilise un ensemble de règles qui permettent de faire correspondre à des indices acoustiques un trait phonétique particulier.

Ces indices acoustiques sont extraits des paramètres du signal représentant l'échantillon et sont comparés à des seuils de référence.

Dans le cas de notre système, qui se veut le plus général possible, ces seuils ne peuvent être connus *a priori*, car dépendant fortement du locuteur et des conditions d'enregistrement.

Nous proposons alors de calculer ces seuils pour Chaque phrase analysée, car nous les savons fortement corrélés à la structure générale du signal candidat à la reconnaissance.

Il semble que l'expert humain opère de la même manière en ayant tour à tour une vision «globale» et «détaillée» de la matière phonique.

#### II - FONCTIONS DU MODULE DE PRETRAITEMENT (voir organigramme A)

##### II.1 - Gestion du fichier d'entrée (voir organigramme A1, organigramme A2)

Le module de prétraitement gère le fichier d'entrée en assurant:

- 1 - La sortie des paramètres des échantillons centisecondes successifs qu'il organise sous forme de listes ordonnées, facilement manipulables.
- 2 - La création de courbes d'évolution temporelle de certains paramètres tels que:
  - Les courbes d'évolution temporelle de l'énergie totale,
  - La courbe d'évolution temporelle de l'énergie basses fréquences: BBF = [250 - 650] Hz,
  - La courbe d'évolution temporelle de l'énergie dans les bandes [250 - 2500] Hz,
  - La courbe d'évolution temporelle des centres de gravité;
- 3 - L'interface avec le module d'étiquetage des échantillons centisecondes;

##### II.2 - Calcul de seuils (voir organigramme A3)

###### II.2.1 - Détection du signal de parole;

Le signal analysé étant supposé entaché de bruits parasites, le module de prétraitement calcule l'énergie maximale de ce bruit de fond.

Ce calcul se fait sur une portion de signal correspondant à un silence, soit sur les 100 premières millisecondes de l'enregistrement où le signal de parole n'a pas encore débuté.

Il s'agit alors de calculer l'énergie totale maximale sur les 10 premiers échantillons, le seuil de parole S-PAR est alors donné par la relation:

$$S-PAR = \text{MAX} [ ET1 \ ET2 \ \dots \ ET10 ]$$

où ET représente l'énergie totale.

- La présence de parole est alors détectée lorsque:

\*  $ET > S\text{-PAR}$

- Un silence est détecté lorsque:

\*  $ET < S\text{-PAR}$

### II.2.2. Détection du début et fin de phrase (voir organigramme A4)

Le début de phrase correspond à l'échantillon qui vérifie:

$$ET (\text{ech}) > S\text{-PAR}$$

La détection de la fin de phrase est un peu plus délicate, car pour ne pas confondre un silence effectif avec le silence d'occlusion des plosives, il est nécessaire que la durée du silence soit supérieure à 200 ms.

La fin de phrase est alors détectée par la relation:

$$ET < S\text{-PAR} \text{ sur plus de 20 échantillons successifs.}$$

### II.2.3. Détection du seuil de voisement (S-VOIS) (voir organigramme A3)

Pour décider si un segment est voisé, on détecte la présence d'énergie dans la bande de fréquences [250 - 650] Hz.

Le bruit ambiant pouvant contenir de l'énergie dans cette bande, on détermine un seuil de voisement en calculant l'énergie maximale du bruit parasite dans cette bande, sur les 100 premières millisecondes de signal qui correspondent à un silence effectif et ce en procédant comme suit:

$$EBF = E1 + E2: \text{Energie basses fréquences}$$

où  $EBF = E [250 - 650] \text{ Hz}$

$E1 = E [250 - 450] \text{ Hz}$

$E2 = E [450 - 650] \text{ Hz}$

$$\text{Et } S\text{-VOIS} = \text{MAX} [EBF1, EBF2 \dots EBF10]$$

Le silence d'une plosive voisée sera détecté par la relation:

$$EBF > S\text{-VOIS}$$

$$ET - EBF < S\text{-PAR}$$

### II.2.4. Détection du seuil de friction (S-FRIC) (voir organigramme A3)

Les fricatives correspondant à un taux de passage par zéro du signal élevé. Pour les détecter, on compare ce taux au taux maximal contenu dans le signal.

Le TPZ maximal est donné par la relation:

$$TPZ \text{ MAX} = \text{MAX} [TPZ1 \dots TPZi \dots TPZn]$$

et le seuil de friction que doit dépasser une fricative est donné (d'après les expériences) par:

$$S\text{-FRIC} = 1/2 \text{ TPZ MAX}$$

II.2.5. Calcul de l'énergie totale maximale contenue dans le signal (ET MAX) (voir organigramme A3)

Nous avons observé, après étude des sonagrammes de plusieurs signaux de parole, que les voyelles sont localisées sur les maxima locaux de la courbe d'évolution temporelle de l'énergie totale, et que les énergie totales des voyelles d'un même continuum vocal ne présentent jamais un rapport inférieur à 1/2.

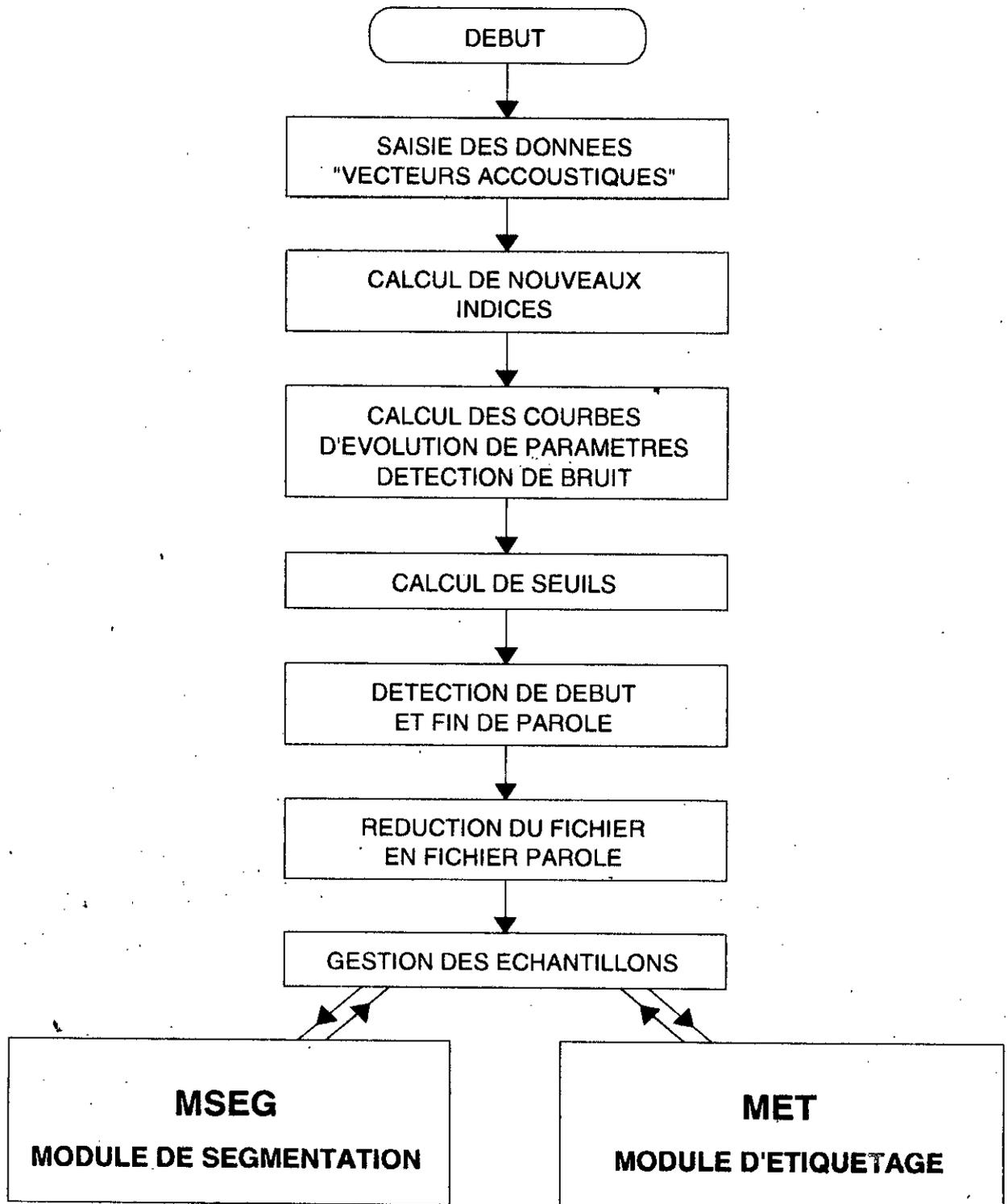
Par conséquent, l'énergie d'une voyelle doit vérifier la relation:

$$ET > 1/2 \text{ ETMAX}$$

où ETMAX, énergie maximale contenue dans le signal, est donné par:

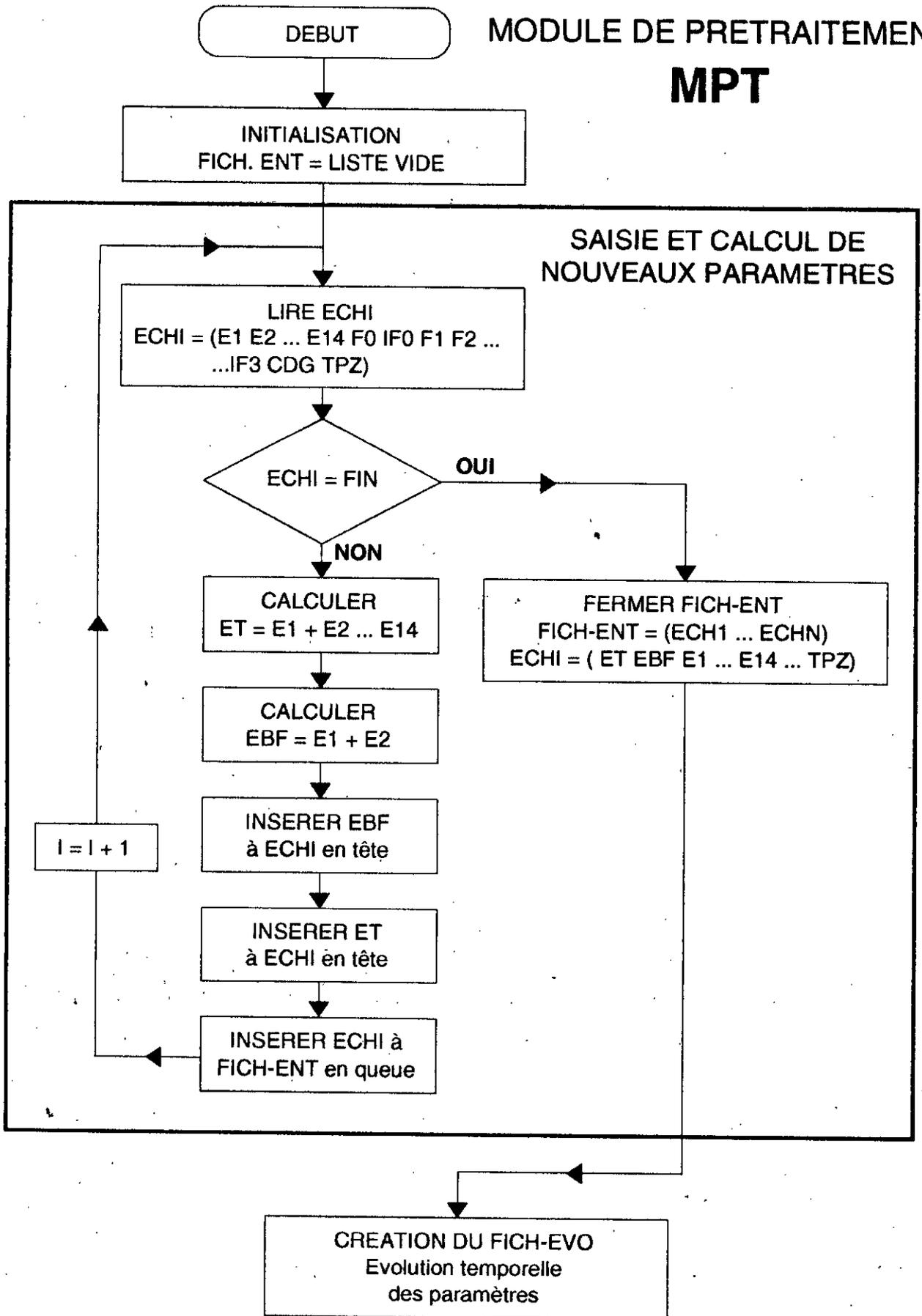
$$\text{ETMAX} = \text{MAX} [\text{ET}_1, \text{ET}_2, \dots \text{ET}_N]$$

# ORGANISATION GENERALE DU MODULE DE PRETRAITEMENT **MPT**



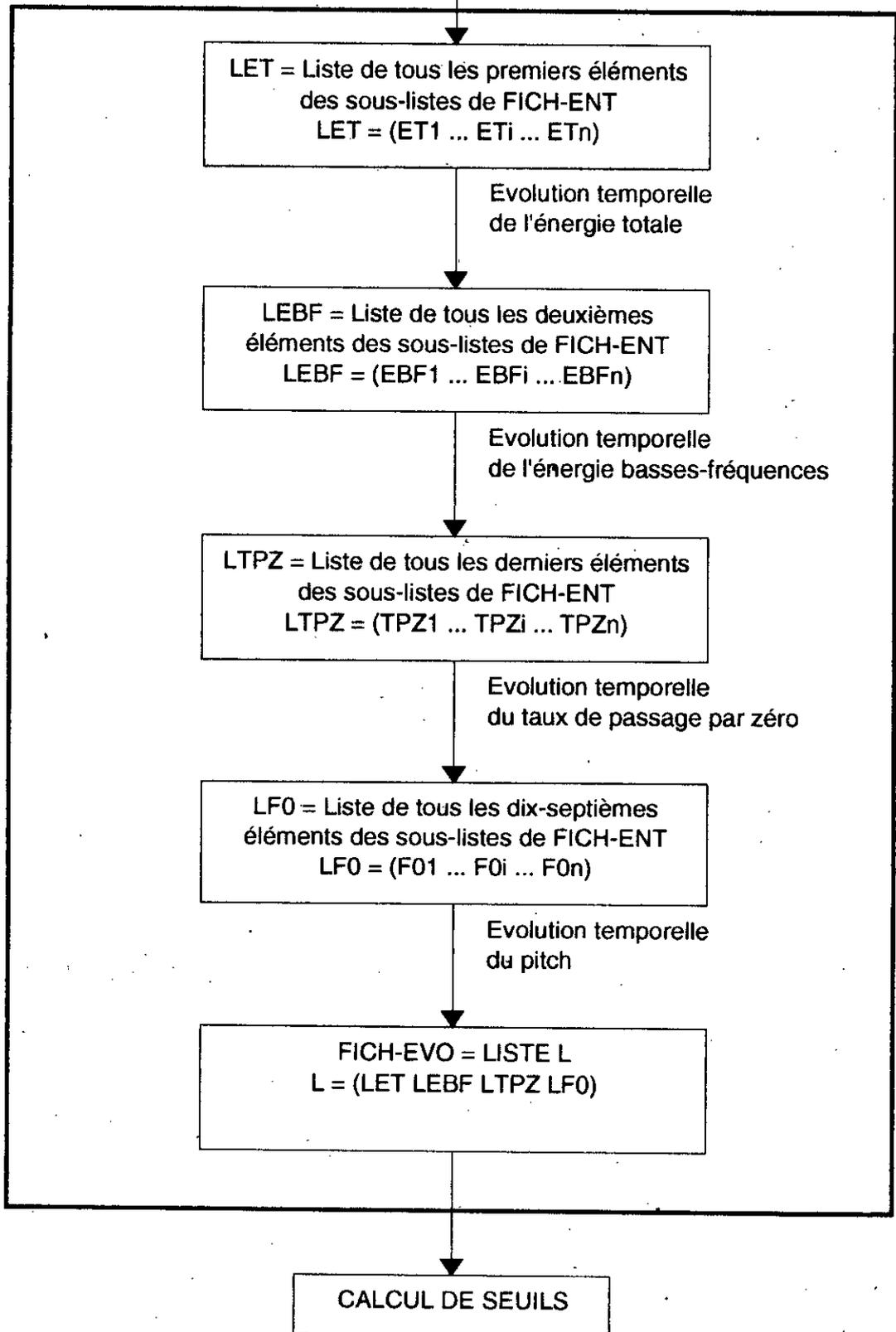
Organigramme. A

# FONCTIONNEMENT DU MODULE DE PRETRAITEMENT MPT

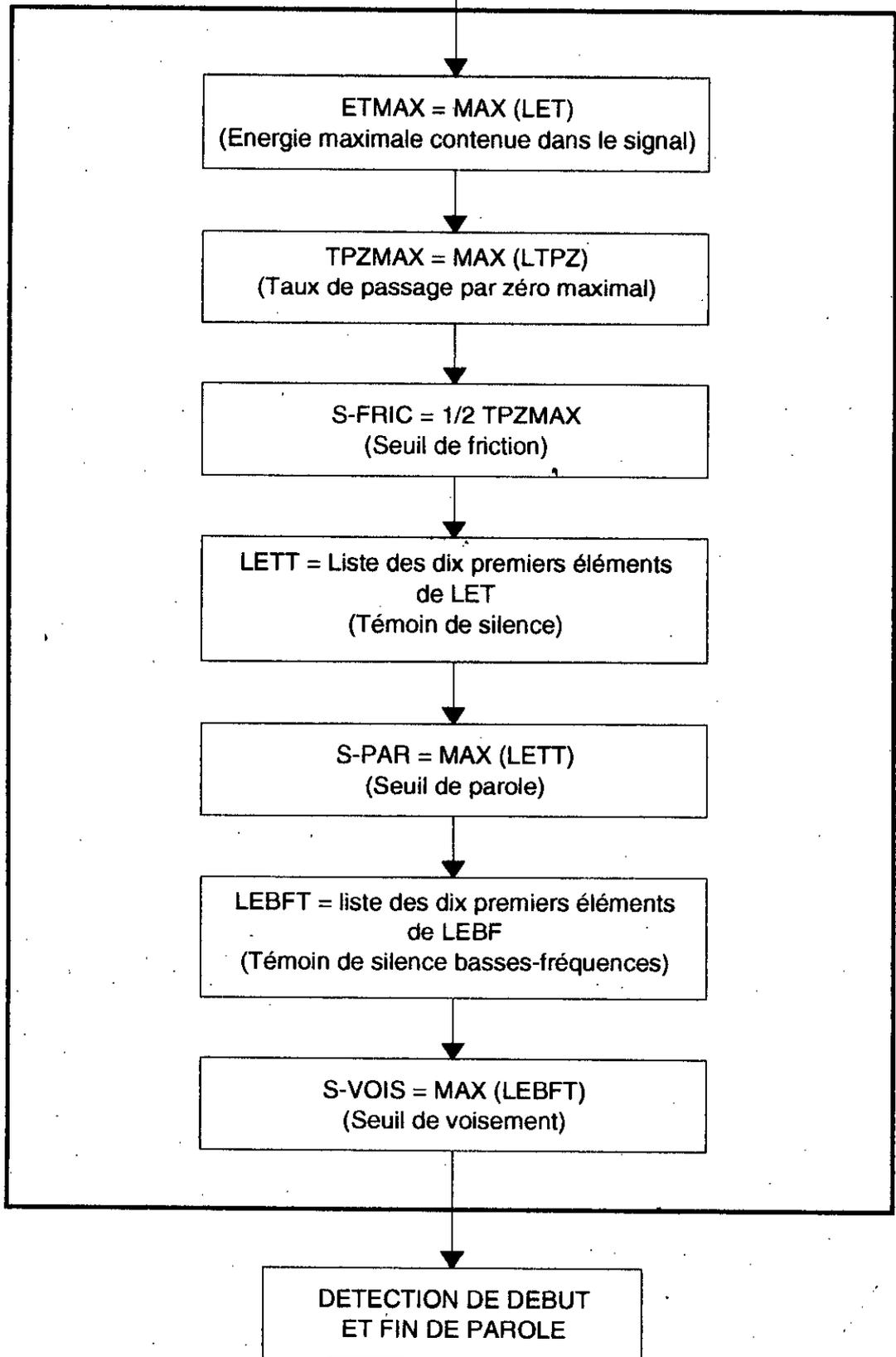


Organigramme. A1

# CREATION DU FICH-EVO



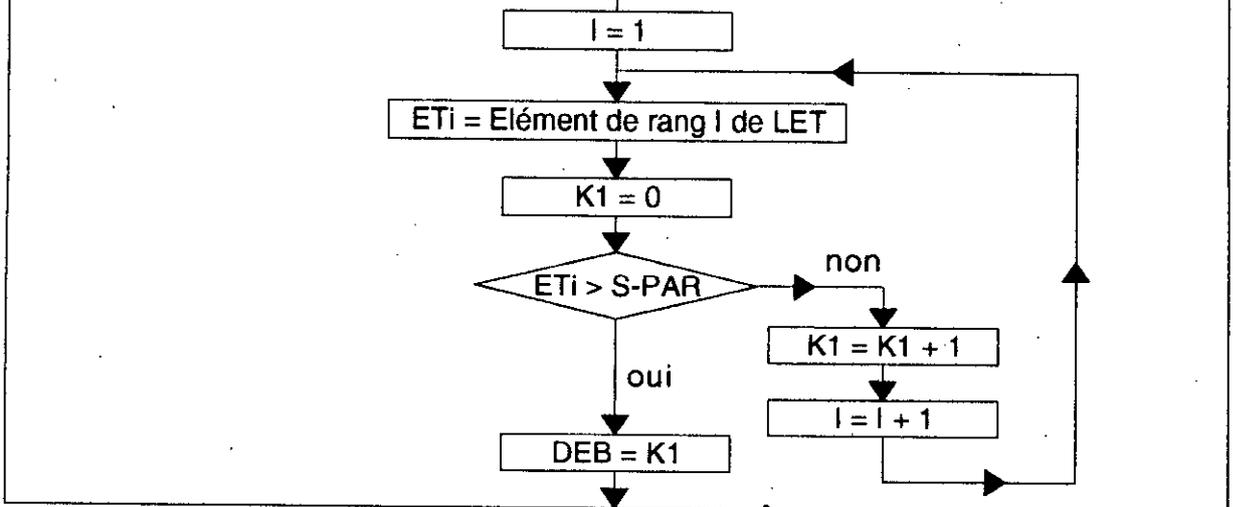
# CALCUL DE SEUILS



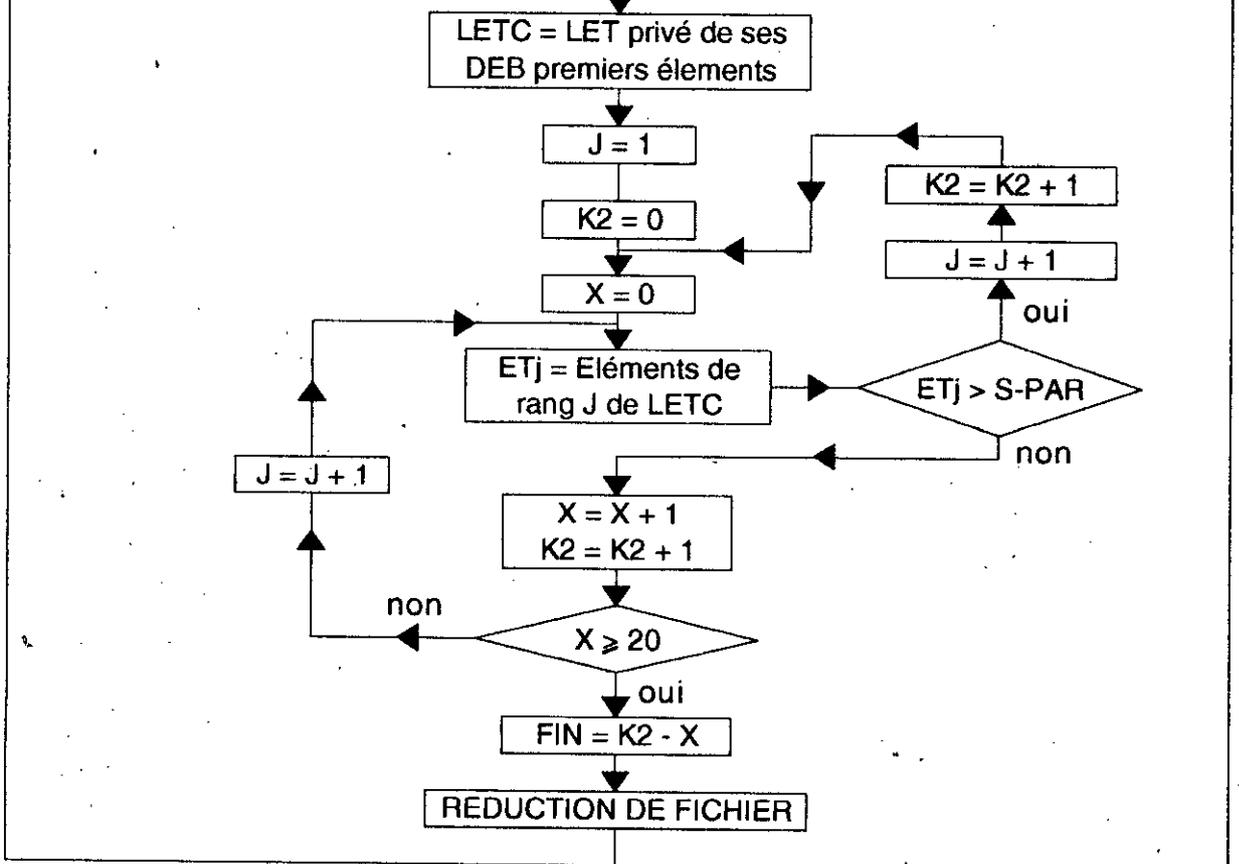
Organigramme. A3

# DETECTION DE DEBUT ET FIN DE PAROLE

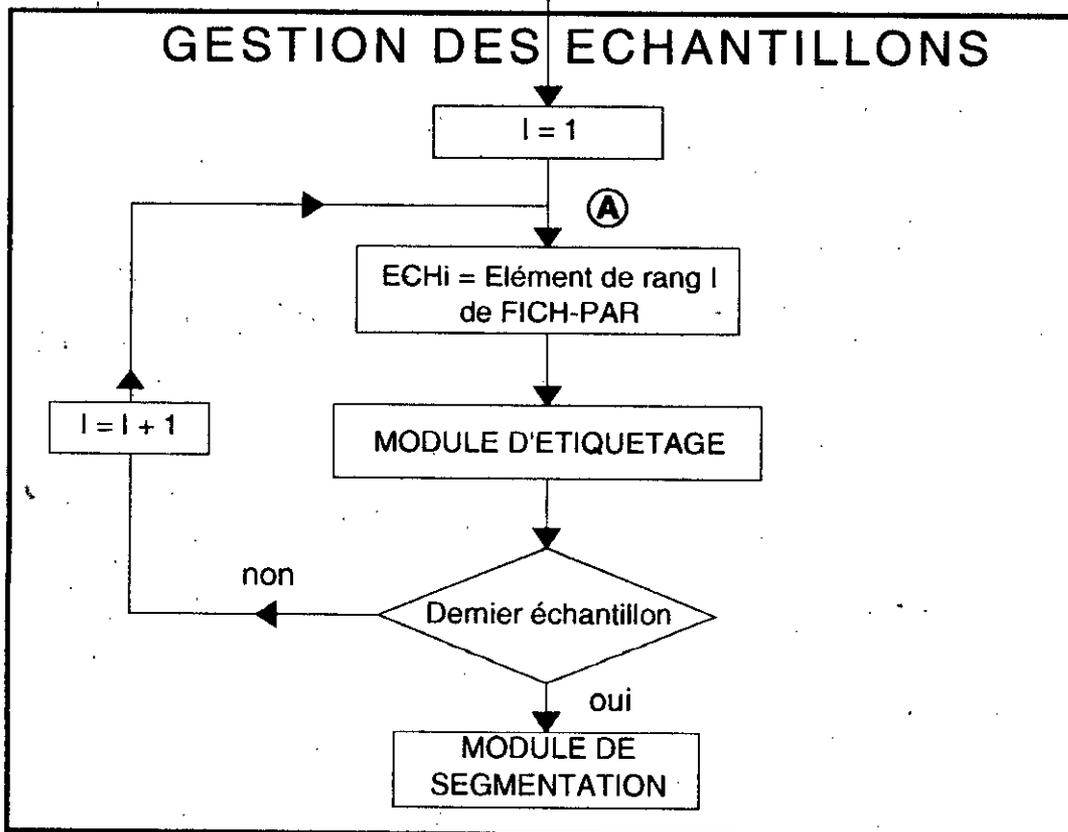
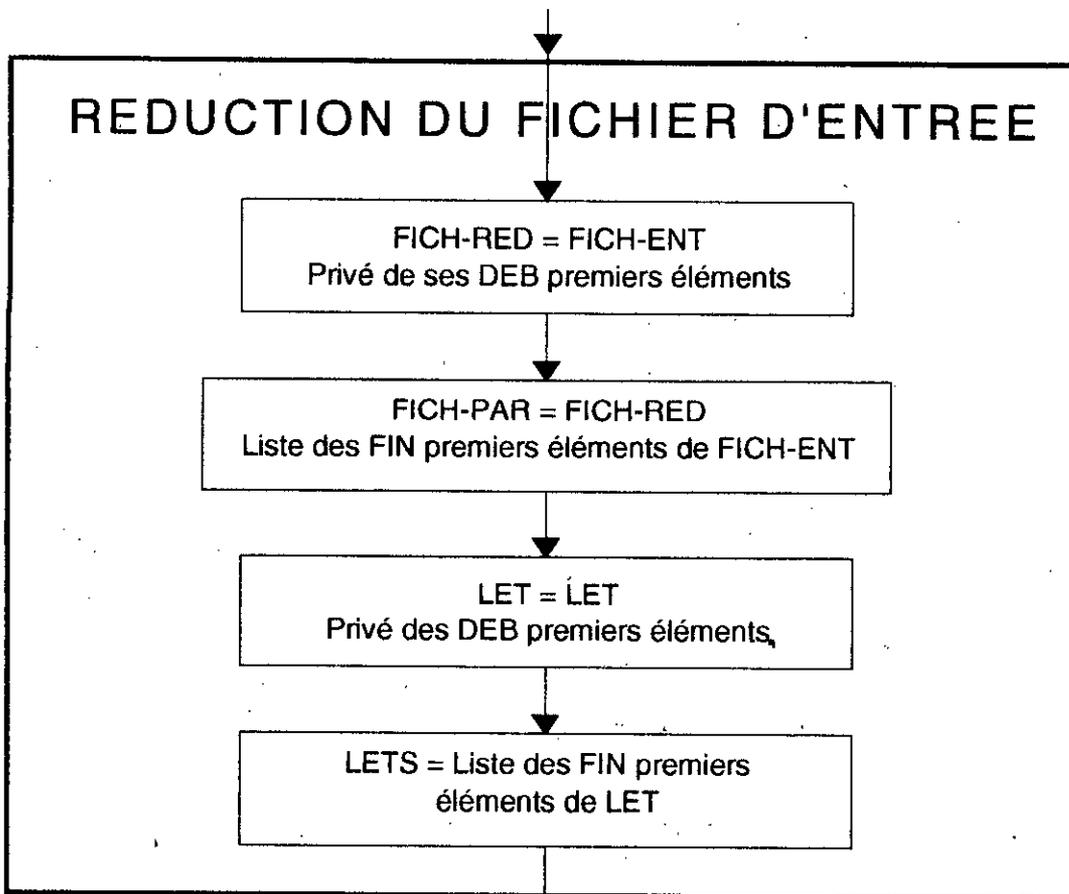
## DETECTION DE DEBUT DE PAROLE



## DETECTION DE FIN DE PAROLE



Organigramme A4



Organigramme A5

## CHAPITRE IV

### MODULE D'ETIQUETAGE

#### I - INTRODUCTION

On distingue l'étiquetage de l'identification dans le sens où l'étiquetage est une identification partielle et locale<sup>[19]</sup>.

Dans un premier temps, nous tentons d'attribuer aux échantillons centisecondes des étiquettes nous permettant de savoir à quelles macro-classes ou classes ils appartiennent.

Ces étiquettes sont inférées par un ensemble de règles contenues dans la «base de règles d'étiquetage».

Les règles opèrent sur des indices acoustiques déterminés par les paramètres de l'échantillon pour identifier des traits phonétiques distinctifs.

L'ensemble des phonèmes d'une langue peut être décrit par un arbre établissant une hiérarchie sur les traits.

#### II - CLASSIFICATION DES PHONEMES<sup>[26]</sup>

On distingue deux macro-classes: les voyelles et les consonnes.

##### II.1 - Les voyelles:

Les voyelles se divisent en deux classes:

- Les voyelles orales,
- Les voyelles nasales.

Nous n'étudions, pour notre part, que les voyelles orales.

##### II.2 - Les consonnes:

Les consonnes peuvent être décomposées en trois ensembles ayant des caractéristiques distinctes:

- Les occlusives,
- Les fricatives,
- Les sonantes.

Les consonnes sonantes caractérisées par une structure de formants se divisent en trois sous-classes:

- Les nasales,
- Les liquides,
- Les semi-consonnes.

Le manque de connaissances précises sur les semi-consonnes et leur trop grande ressemblance avec les voyelles, ne nous permettent pas de les identifier (dans l'état actuel de l'art) avec une bonne performance. C'est pour cette raison que nous ne les avons pas intégrées au système de reconnaissance.

#### III.- TRAITS PHONETIQUES ET INDICES ACOUSTIQUES RETENUS<sup>[10]</sup>

##### III.1 - Détection de la macro-classe voyelle

En première approximation, nous attribuons l'étiquette «voyelle» à tout échantillon vérifiant les indices suivants:

- Structure formantique nette,
- ET > 1/2 ET MAX,

- E [250 - 2500] >> E [2500 - 8000],
- Voisement,
- CDG > 7 et CDG < 8,9.

### III.2 - Détection de la macro-classe consonne

Cette détection est évidemment liée à la détection de la macro-classe voyelle.

### III.3 - Voyelle postérieure

Cette étiquette repose sur les indices formantiques suivants:

- F2 < 1300 Hz

### III.4 - Voyelle antérieure

- F2 > 1900 Hz

### III.5 - Voyelle centrale

- 1300 Hz < F2 < 1900 Hz

### III.6 - Trait d'ouverture des voyelles

L'indice principal est le premier formant F1

- 250 Hz < F1 < 350 Hz: Fermé
- 350 Hz < F1 < 450 Hz: Mi-fermé
- 450 Hz < F1 < 650 Hz: Mi-ouvert
- F1 > 650 Hz: Ouvert

### III.7 - Distinction consonne voisée-consonne sourde

Cette distinction est basée sur les indices suivants:

- Présence du fondamental: I(F0) > 0,
- E [250 - 650] Hz > S-VOIS,

### III.8 - Trait de structure formantique

Un échantillon est considéré à structure formantique nette si les indices suivants sont présents:

- I(F1) > 0,
- I(F2) > 0.

### III.9 - Trait de friction

Un échantillon est étiqueté fricatif s'il présente les indices suivants:

- CDG > 9,
- TPZ > S-FRIC,
- Pas de structure formantique.

### III.10 - Trait de nasalité

Un échantillon est étiqueté nasal si:

- Structure formantique nette,
- E [250 - 450]Hz (où se situe le murmure nasal) est supérieure à l'énergie dans certaines autres bandes de fréquence.

### III.11 - Détection des consonnes liquides

L'étiquette «liquide» est attribuée à un échantillon vérifiant les indices suivants:

- E [1600 - 4300] > E [250 - 1600],
- E [250 - 1600] > E [4300 - 8000].

### III.12 - Détection du trait «Stop»: Silence

Un échantillon correspond à un silence si:

- ET < S-PAR

### III.13 - Détection du trait «Buzz-Bar»

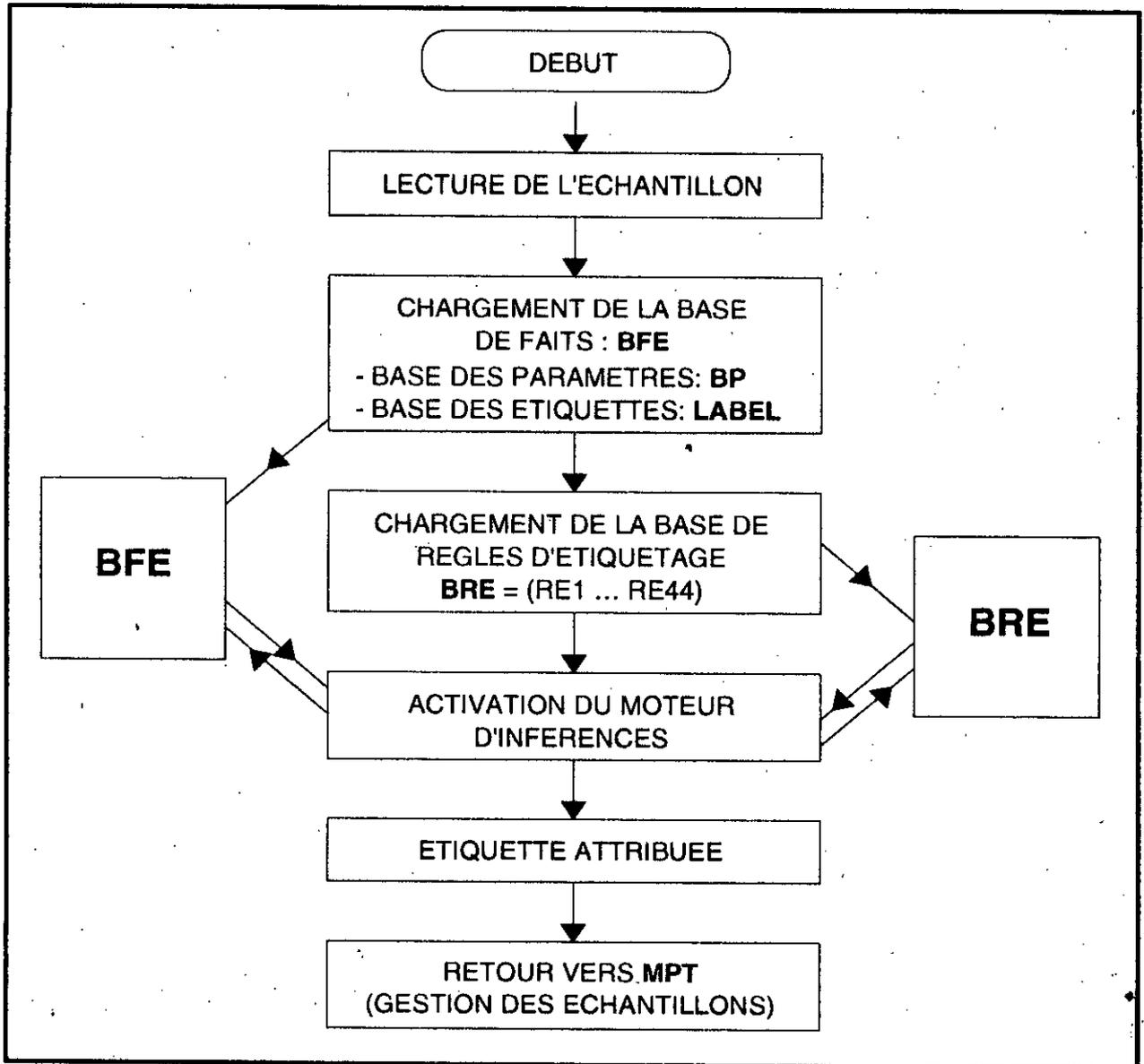
Un échantillon correspond à un silence voisé et est étiqueté «Buzz-Bar» si:

- E [650 - 8000] Hz < S-PAR,
- E [250 - 650] Hz > S-VOIS.

## IV - FONCTIONNEMENT (voir organigramme B)

Le module d'étiquetage est conçu sous la forme d'un système expert. Pour chaque échantillon centiseconde, le moteur d'inférences est activé pour opérer sur une nouvelle base de "faits initiaux" (paramètres de l'échantillon en question). Les échantillons sont gérés par le module de prétraitement qui fait aussi fonction de superviseur. La reconnaissance d'une phrase peut nécessiter des centaines de sessions de travail du système expert, chaque session comprenant plusieurs cycles unitaires d'inférences. Ce système possède 44 règles d'étiquetage et fonctionne selon la stratégie et le mode décrits au chapitre VIII.

# FONCTIONNEMENT DU MODULE D'ETIQUETAGE MET



LABEL = (VOA, VOC, VOP, VFA,  
VFC, VFP, VMOA, VMOC,  
VMOP, VMFA, VMFC, VMFP,  
CN, CL, STOP, BUZZ, CFV,  
CF, IV, IS)

Liste des étiquettes attribuables  
aux échantillons centisecondes

BP = (ET, E1, ... E14, F0, IF0, F1,  
F2, F3, IF1, IF2, IF3, CDG,  
TPZ, S-PAR, S-VOIS,  
S-FRIC, EMAX)

Liste des paramètres utiles à  
l'étiquetage des ECS

Organigramme B

## V - BASE DE REGLES D'ETIQUETAGE

- RE1: SI  $I(F_0) \neq 0$   
 ET  $E [250 - 650] > S - VOIS$   
 ALORS T-VOIS = VOISE
- RE2: SI T-VOIS  $\neq$  VOISE  
 ALORS T-VOIS = NON-VOISE
- RE3: SI T-VOIS = VOISE  
 ET  $I(F_1) \neq 0$   
 ET  $I(F_2) \neq 0$   
 ALORS T-STFOR = FORMANTIQUE
- RE4: SI T-STFOR  $\neq$  FORMANTIQUE  
 ALORS T-STFOR = N.FORM
- RE5: SI T-STFOR = FORMANTIQUE  
 ET  $E.T > 1/2 ETMAX$   
 ET  $E [250 - 2500] > 1/4 ET$   
 ET  $CDG > 7$   
 ET  $CDG < 8,9$   
 ALORS MACRO-CLASSE = VOYELLE
- RE6: SI MACRO-CLASSE  $\neq$  VOYELLE  
 ALORS MACRO-CLASSE = CONSONNE
- RE7: SI MACRO-CLASSE = VOYELLE  
 ET  $F_2 < 1300 \text{ Hz}$   
 ALORS TRAIT-ANT = POST
- RE8: SI MACRO-CLASSE = VOYELLE  
 ET  $F_2 > 1900 \text{ Hz}$   
 ALORS TRAIT-ANT = ANT
- RE9: SI MACRO-CLASSE = VOYELLE  
 ET  $F_2 < 1900 \text{ Hz}$   
 ET  $F_2 > 1300 \text{ Hz}$   
 ALORS TRAIT-ANT = CENT
- RE10: SI MACRO-CLASSE = VOYELLE  
 ET  $F_1 > 250 \text{ Hz}$   
 ET  $F_1 < 350 \text{ Hz}$   
 ALORS TRAIT-OUV = FERM

- RE11: SI MACRO-CLASSE = VOYELLE  
 ET F1 > 350 Hz  
 ET F1 < 450 Hz  
 ALORS TRAIT-OUV = MI-FERM
- RE12: SI MACRO-CLASSE = VOYELLE  
 ET F1 > 450 Hz  
 ET F1 < 650 Hz  
 ALORS TRAIT-OUV = MI-OUV
- RE13: SI MACRO-CLASSE = VOYELLE  
 ET F1 > 650 Hz  
 ALORS TRAIT-OUV = OUV
- RE14: SI MACRO-CLASSE = CONSONNE  
 ET CDG > 9  
 ET TPZ > S-FRIC  
 ET T-STFOR = N.FORM  
 ALORS CLASSE = FRIC
- RE15: SI MACRO-CLASSE = CONSONNE  
 ET ET < S-PAR  
 ALORS CLASSE = STOP
- RE16: SI MACRO-CLASSE = CONSONNE  
 ET E [650 - 8000] < S-PAR  
 ET T-VOIS = VOISE  
 ALORS CLASSE = BUZZ
- RE17: SI MACRO-CLASSE = CONSONNE  
 ET T-STFOR = FORMANTIQUE  
 ET E [250 - 450] > MAXI [450 - 1600] + 4dB  
 ET E [250 - 450] > MAXI [1600 - 4300] + 8dB  
 ET E [250 - 450] > E [650 - 850] + 12dB  
 ET E [450 - 650] > E [650 - 850] + 4dB  
 ET E [250 - 450] > E [850 - 1050] + 8dB  
 ALORS CLASSE = NASALE
- RE18: SI MACRO-CLASSE = CONSONNE  
 ET T-STFOR = FORMANTIQUE  
 ET E [250 - 450] > MAXI [450 - 1600] + 4dB  
 ET E [250 - 450] > MAXI [1600 - 4300] + 8dB  
 ET E [250 - 450] > E [650 - 850] + 12dB  
 ET E [450 - 650] > E [650 - 850] + 4dB  
 ET E [250 - 450] > E [1050 - 1300] + 8dB  
 ALORS CLASSE = NASALE

- RE19: SI MACRO-CLASSE = CONSONNE  
 ET T-STFOR = FORMANTIQUE  
 ET E [1600 - 4300] > E [250 - 1600]  
 ET E [250 - 1600] > E [4300 - 8000]  
 ALORS CLASSE = LIQUIDE
- RE20: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = ( )  
 ALORS CLASSE = INC
- RE21: SI MACRO-CLASSE = VOYELLE  
 ET T-OUV = OUV  
 ALORS ATTV = VO
- RE22: SI MACRO-CLASSE = VOYELLE  
 ET T-OUV = MI-OUV  
 ALORS ATTV = VMO
- RE23: SI MACRO-CLASSE = VOYELLE  
 ET T-OUV = MI-FER  
 ALORS ATTV = VMF
- RE24: SI MACRO-CLASSE = VOYELLE  
 ET T-OUV = FER  
 ALORS ATTV = VF
- RE25: SI ATTV = VO  
 ET T-ANT = ANT  
 ALORS ETIQ = VOA
- RE26: SI ATTV = VO  
 ET T-ANT = CENT  
 ALORS ETIQ = VOC
- RE27: SI ATTV = VO  
 ET T-ANT = POST  
 ALORS ETIQ = VOP
- RE28: SI ATTV = VMO  
 ET T-ANT = ANT  
 ALORS ETIQ = VMOA
- RE29: SI ATTV = VMO  
 ET T-ANT = CENT  
 ALORS ETIQ = VMOC

- RE30: SI ATTV = VMO  
 ET T-ANT = POST  
 ALORS ETIQ = VMOP
- RE31: SI ATTV = VMF  
 ET T-ANT = ANT  
 ALORS ETIQ = VMFA
- RE32: SI ATTV = VMF  
 ET T-ANT = CENT  
 ALORS ETIQ = VMFC
- RE33: SI ATTV = VMF  
 ET T-ANT = POST  
 ALORS ETIQ = VMFP
- RE34: SI ATTV = VF  
 ET T-ANT = ANT  
 ALORS ETIQ = VFA
- RE35: SI ATTV = VF  
 ET T-ANT = CENT  
 ALORS ETIQ = VFC
- RE36: SI ATTV = VF  
 ET T-ANT = POST  
 ALORS ETIQ = VFP
- RE37: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = NASALE  
 ALORS ETIQ = CN
- RE38: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = LIQUIDE  
 ALORS ETIQ = CL
- RE39: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = FRIC  
 ET T-VOIS = VOISE  
 ALORS ETIQ = CFV
- RE40: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = FRIC  
 ET T-VOIS = N.VOIS  
 ALORS ETIQ = CF

RE41: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = STOP  
 ALORS ETIQ = STOP

RE42: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = BUZZ  
 ALORS ETIQ = BUZZ

RE43: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = INC  
 ET T-VOIS = VOIS  
 ALORS ETIQ = IV

RE44: SI MACRO-CLASSE = CONSONNE  
 ET CLASSE = INC  
 ET T-VOIS = N.VOIS  
 ALORS ETIQ = IS

## CHAPITRE V

### MODULE DE SEGMENTATION

#### I - INTRODUCTION

L'unité de décision choisie dans notre système étant le phonème, le module de segmentation doit découper le signal en unités proches du phonème «les pseudo-phonèmes».

Le module de segmentation construit des unités potentiellement proches du phonème quant à sa durée, sur des critères acoustiques.

Dans un second temps, le module recompose ces segments en procédant à des fusions ou des dissociations pour former les phonèmes.

Le module d'étiquetage ayant fourni l'essentiel de l'information, il serait aisé de recomposer les phonèmes en regroupant dans un même segment tous les échantillons de même étiquette.

Néanmoins, cette méthode s'avère insuffisante car une erreur d'étiquetage d'un échantillon pénaliserait tout le reste du traitement.

D'autre part, nous savons qu'un phonème d'une classe donnée peut présenter sur certaines phases de sa réalisation, les caractéristiques d'un phonème d'une autre classe.

Un segment phonétique de la classe des liquides peut être, par exemple, formé d'un certain nombre d'échantillons présentant les caractéristiques d'une fricative.

De plus, les règles de détection des voyelles, ne sont pas toujours précises et peuvent donner dans certains contextes des étiquettes erronées.

Pour une segmentation plus fiable, nous proposons la démarche suivante:

#### II - STRATEGIE ADOPTÉE (voir organigramme C)

##### 1<sup>ère</sup> étape: segmentation en pseudo-syllabes (voir organigramme C1)

Sachant que la syllabe est une unité facilement localisable, car généralement comprise entre deux minima d'énergie totale, nous construisons dans une première étape des unités pseudo-syllabiques, en formant des segments vérifiant les conditions suivantes:

- a - Borne inférieure du segment égale à un minimum local sur la courbe d'énergie totale.
- b - Borne supérieure du segment égale au prochain minimum local.

Chaque segment ainsi constitué contient un maximum local d'énergie totale correspondant généralement à un noyau vocalique.

##### 2<sup>ème</sup> étape: réduction des segments excédentaires (voir organigramme C2)

Les segments «pseudo-syllabiques» construits par la première étape sont analysés pour vérifier s'ils présentent bien des caractéristiques compatibles avec une syllabe. Ils doivent notamment présenter les caractéristiques suivantes:

- La durée du segment doit être supérieure à 120 ms (durée minimale d'une syllabe);
- Le segment doit contenir au moins 6 échantillons étiquetés «voyelle»;

- Le maximum contenu dans le segment doit vérifier la relation  $ET > 1/2 ETMAX$ .

Les segments ne satisfaisant pas à ces trois conditions seront concatenés aux segments gauche et droit pour former la nouvelle pseudo-syllabe.

**3<sup>ème</sup> étape (voir organigramme C3, organigramme C4)**

La voyelle contenue dans la syllabe est délimitée dans la zone stationnaire autour du maximum et doit satisfaire la condition: l'énergie totale est très proche de l'énergie du maximum.

Le reste du segment correspond à une réalisation consonantique.

**4<sup>ème</sup> étape (voir organigramme C5)**

A cette étape, les phonèmes peuvent être localisés et déterminés en regroupant pour Chaque segment de l'étape 3, les échantillons qui portent la même étiquette. Avec la condition supplémentaire concernant les plosives:

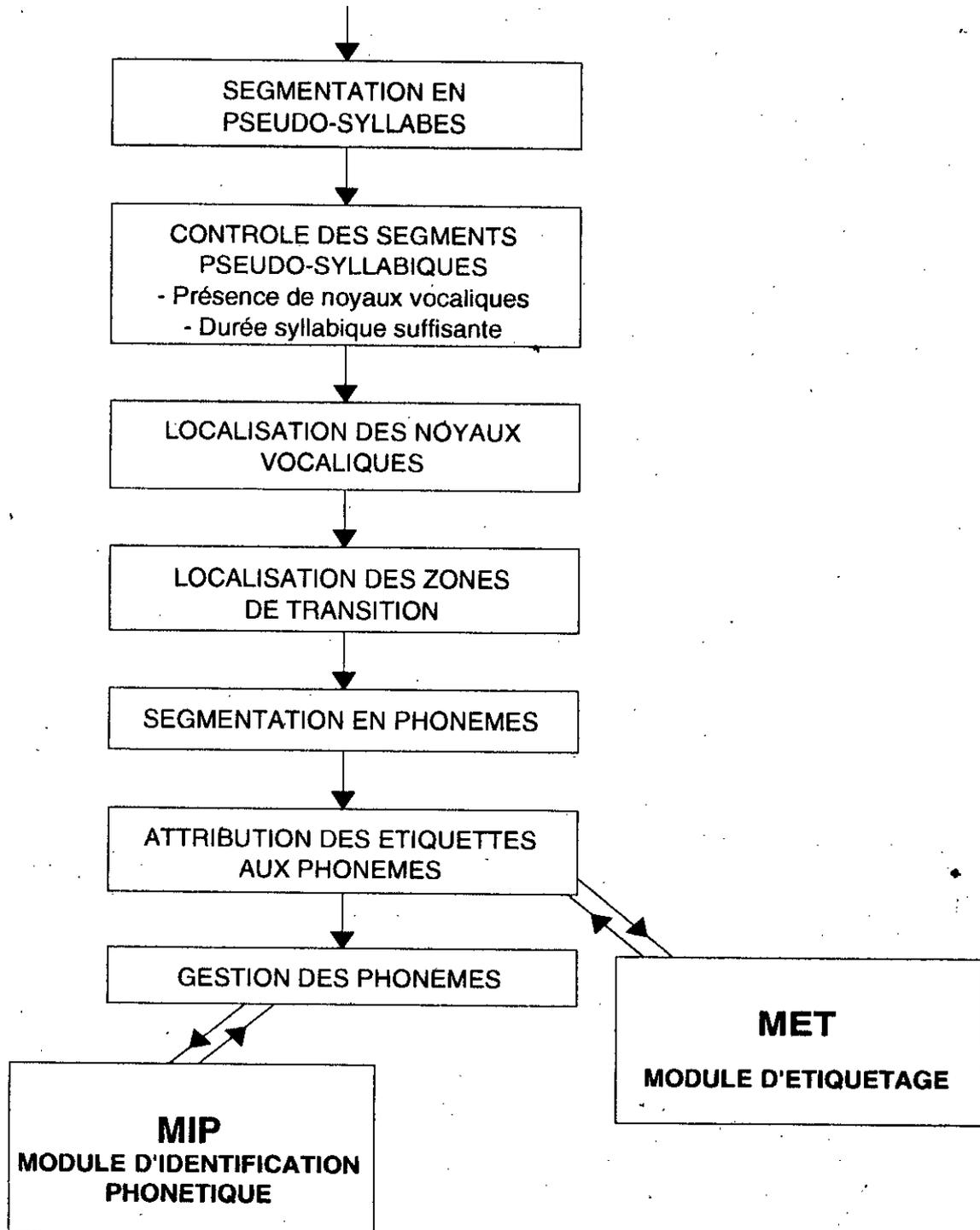
\* Les segments étiquetés STOP sont concatenés aux segments qui les suivent si ceux-ci sont étiquetés FRICATIF ou INCONNU.

\* De même pour les segments étiquetés BUZZ.

Et une nouvelle étiquette PLOS leur est attribuée.

Les segments portant la même étiquette (à l'exception de STOP et BUZZ) et de durée inférieure à 60 ms sont concatenés aux segments adjacents.

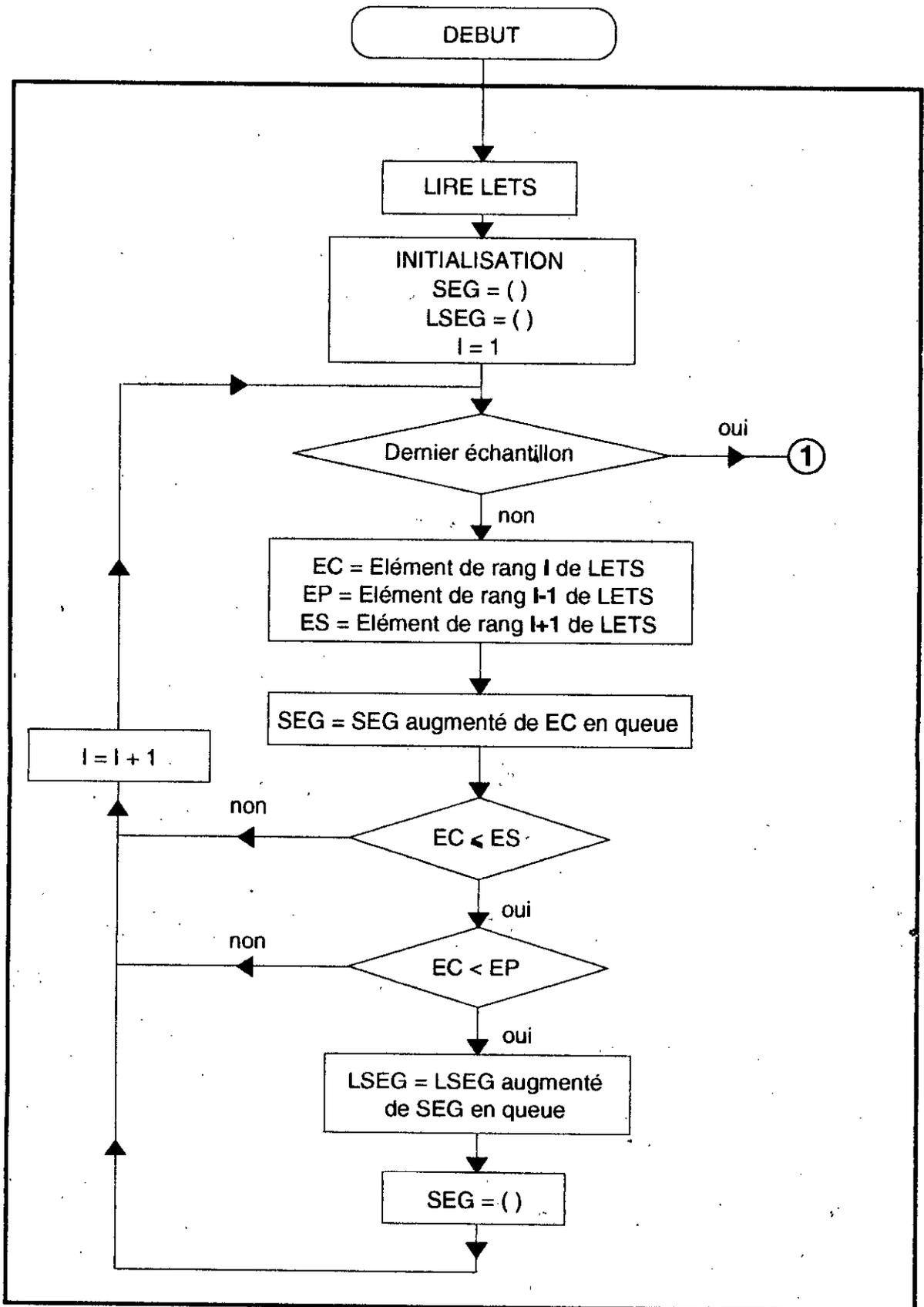
# ORGANISATION GENERALE DU MODULE DE SEGMENTATION **MSEG**



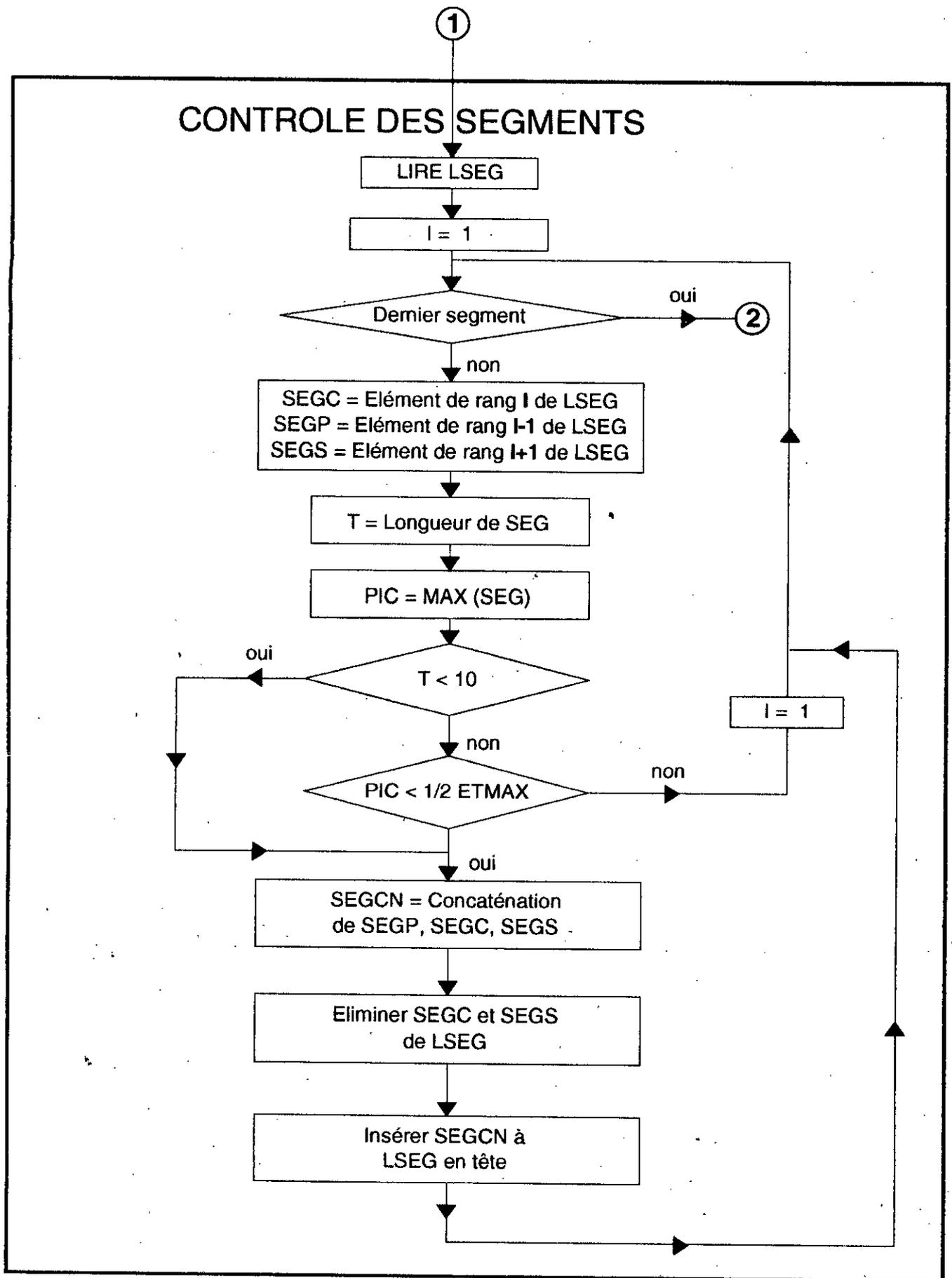
Organigramme C

# FONCTIONNEMENT DU MODULE DE SEGMENTATION

## MSEG

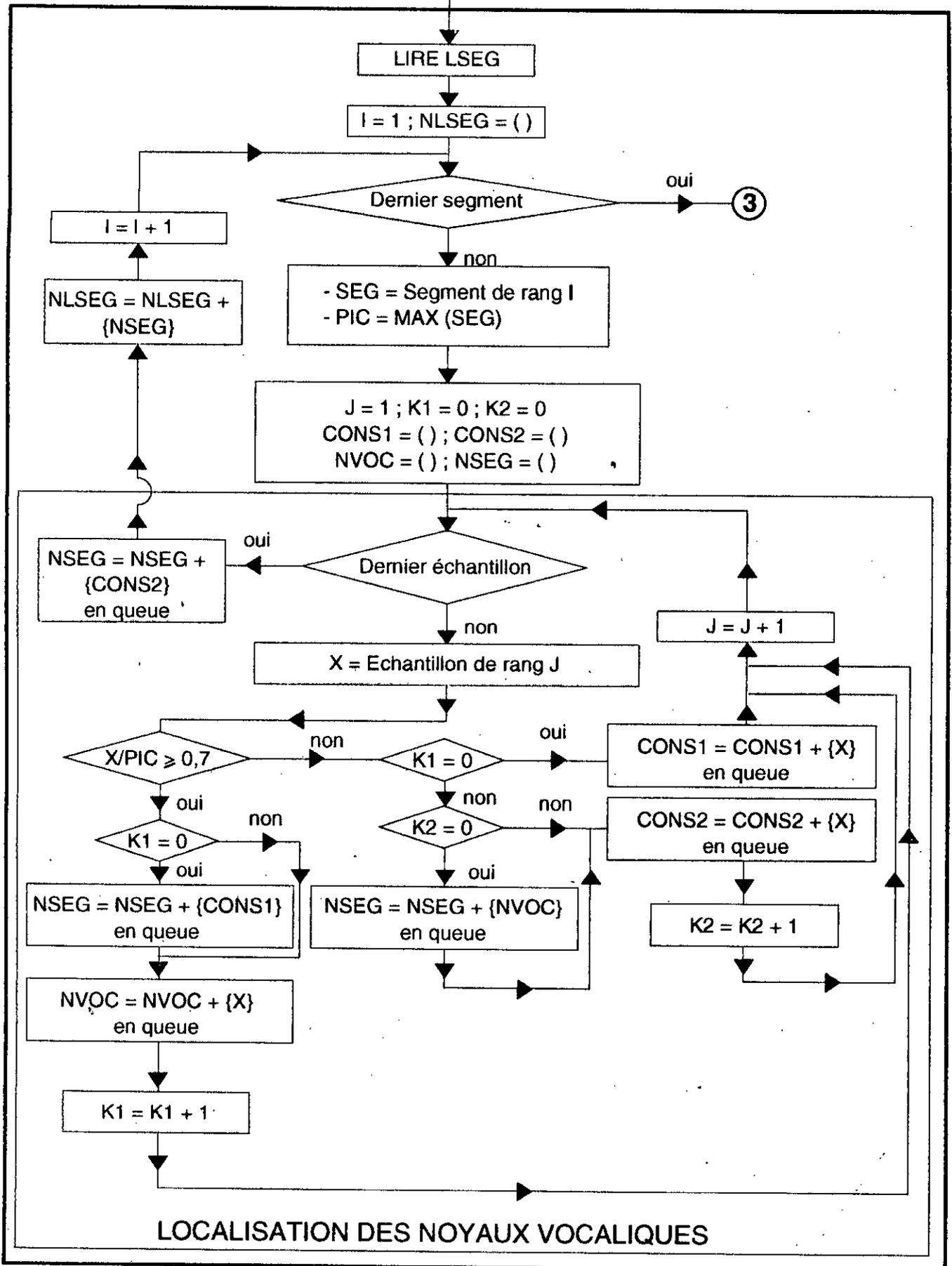


Organigramme C1



Organigramme C2

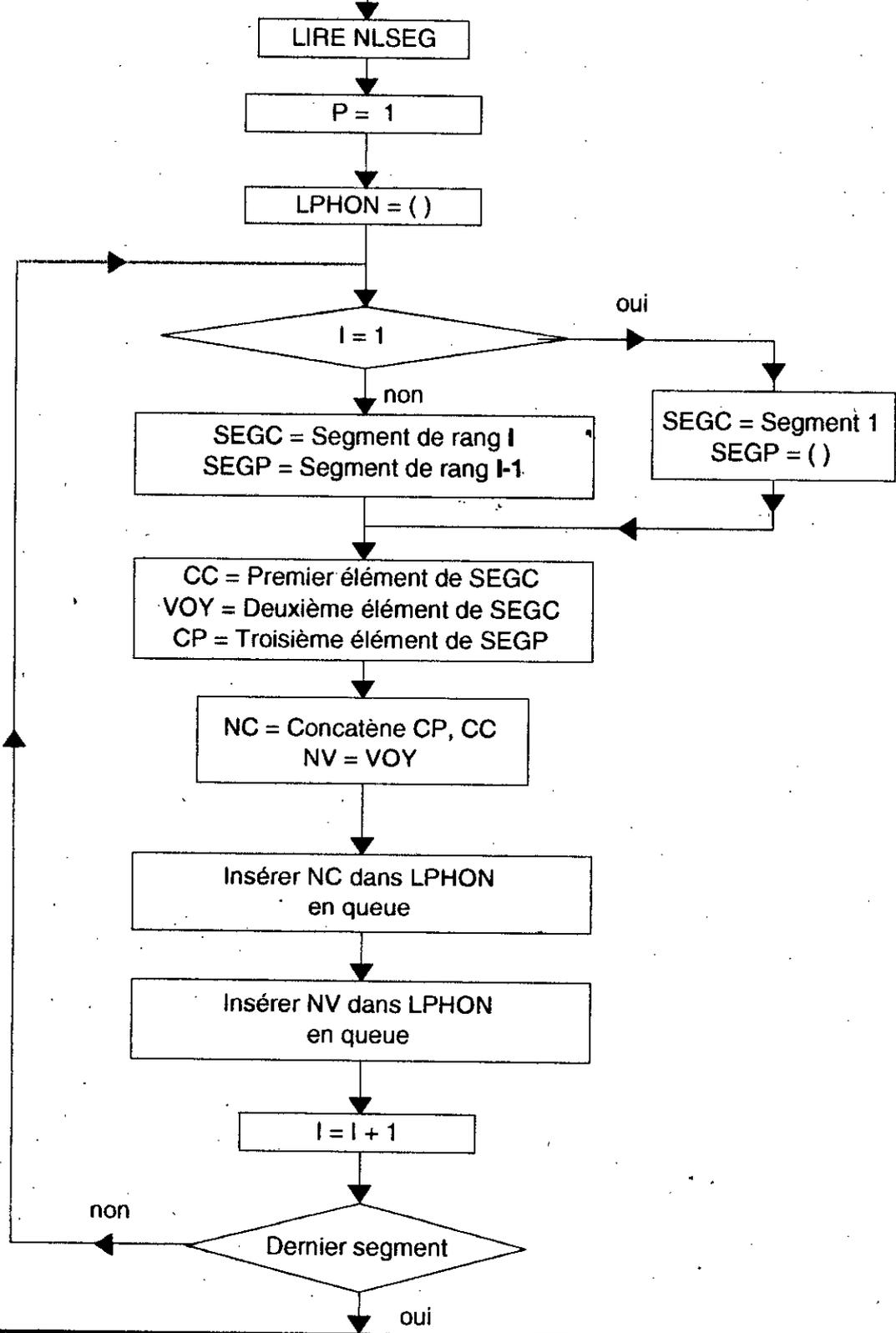
## ② SEGMENTATION EN PHONEMES



Organigramme C3

3

# COMPRESSION DES ZONES DE TRANSITION

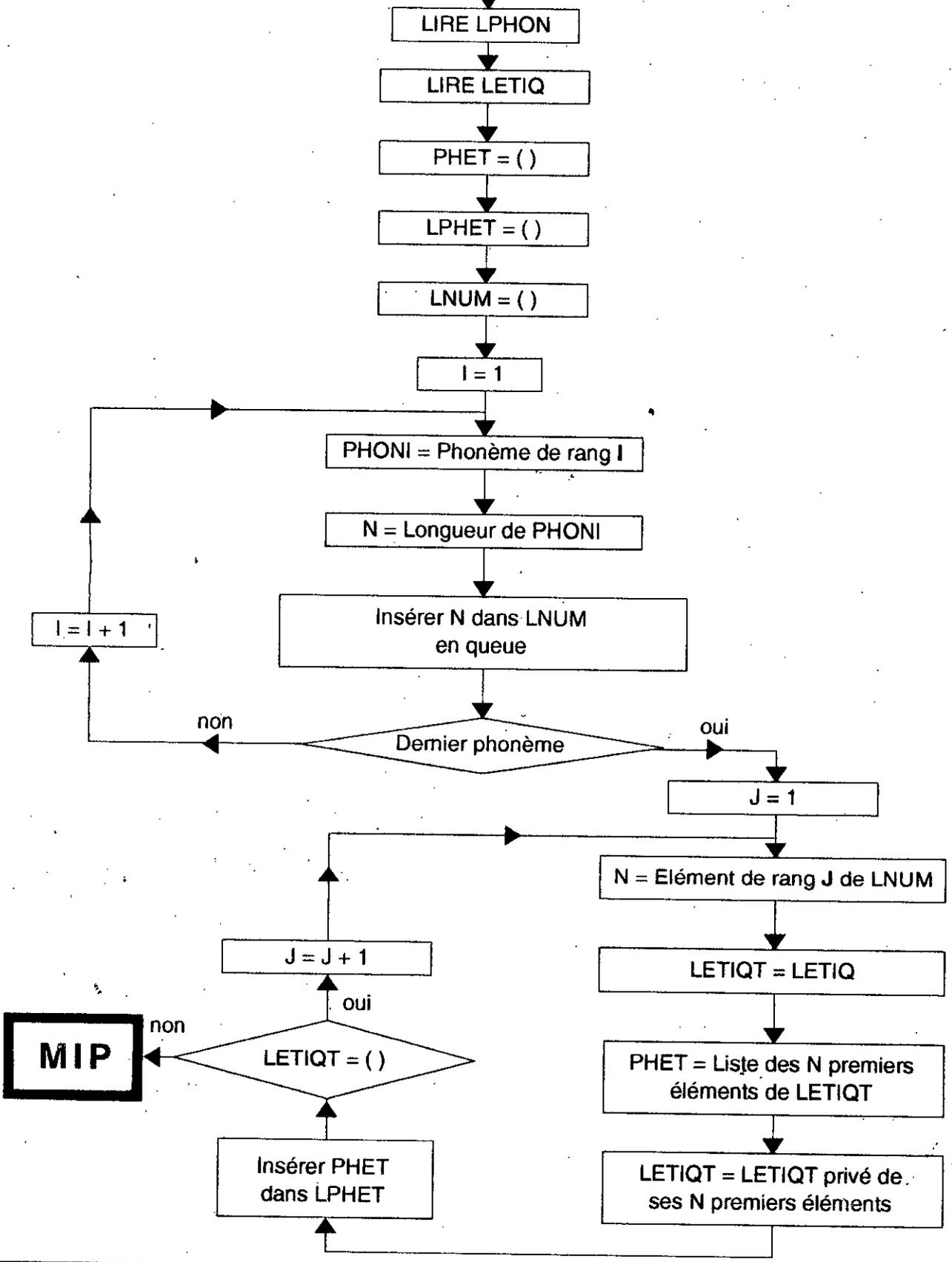


4

Organigramme C4

④

# ATTRIBUTION DES ETIQUETTES AUX PHONEMES



Organigramme C5

## CHAPITRE VI

### MODULE D'IDENTIFICATION PHONETIQUE

#### I - INTRODUCTION

Le module d'identification phonétique a pour fonction de reconnaître le phonème qui lui est présenté à son entrée.

Dans notre cas, le signal est segmenté en phonèmes et Chaque phonème est donné avec la liste des étiquettes des échantillons qui le composent.

Les étiquettes ne sont pas toutes identiques pour un même phonème, car nous avons pris le soin de ne pas segmenter le signal par regroupement des étiquettes.

Cette précaution a été prise pour pouvoir tenir compte des erreurs d'étiquetage éventuelles, des erreurs d'extractions des paramètres et des parasites acoustiques présents dans le signal, qui peuvent altérer localement un son.

L'étiquette du phonème à identifier sera dans notre cas déduite d'une statistique sur les étiquettes qui le composent.

De plus, cette stratégie adoptée permet de prendre en compte les phonèmes dont les réalisations acoustiques comportent des phases discontinues tels que les plosives.

#### II - METHODOLOGIE D'IDENTIFICATION (voir organigramme D)

##### II.1 - Calcul du taux de présence d'un trait phonétique donné (voir organigramme D1)

Pour chaque phonème, et pour chaque trait phonétique donné, le degré de confiance du trait est calculé par une statistique sur sa présence dans le phonème.

- Soit ETIQ, l'étiquette du trait que l'on veut attribuer au phonème.
- Soit N, le nombre d'échantillons d'étiquettes ETIQ contenu dans le phonème.
- Soit L, la longueur du phonème (nombre d'échantillons qui le composent).

Le degré de confiance du trait phonétique est donné par:

$$C = N/L \text{ donné en \%}$$

##### II.2 - Cas de plosives (voir organigramme D2)

- L'étiquette plosive voisée (CPV) est attribuée à un phonème lorsque l'on distingue une partie étiquetée BUZZ d'une partie étiquetée CFV ou IV.
- L'étiquette plosive sourde (CP) est attribuée à un phonème lorsque l'on distingue une partie étiquetée STOP et une partie étiquetée CF ou IS.

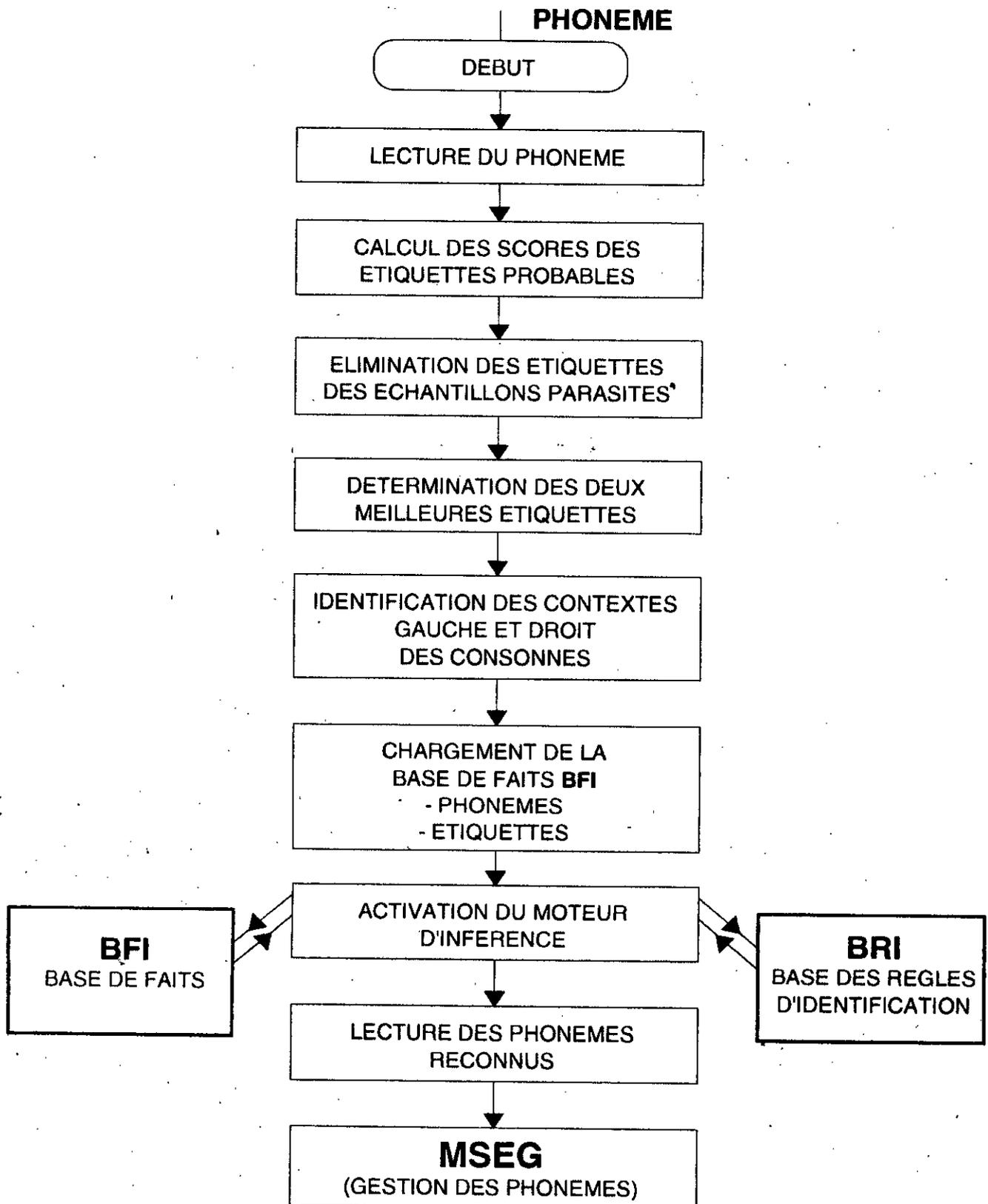
**II.3 - Détermination du lieu d'articulation des consonnes (voir organigramme D)**

La détermination du lieu d'articulation des consonnes se base sur un indice important lié aux transitions formantiques. Nous avons retenu, pour notre part, des indices sur la distribution des énergies. Sachant que, selon le lieu d'articulation, cette distribution varie. Par exemple, les plosives labiales sont caractérisées par une plus grande intensité en basses fréquences.

Le lieu d'articulation est déterminé par un ensemble de 29 règles contenues dans la base de règles d'identification phonétique (BRI) manipulée par un moteur d'inférences. Ces règles permettent, à partir de la détermination du lieu d'articulation et de l'étiquette attribuée, d'identifier le phonème candidat. Cette identification fait appel à un logiciel de type système expert dont les caractéristiques sont données au chapitre VIII.

# ORGANISATION GENERALE DU MODULE D'IDENTIFICATION

## MIP

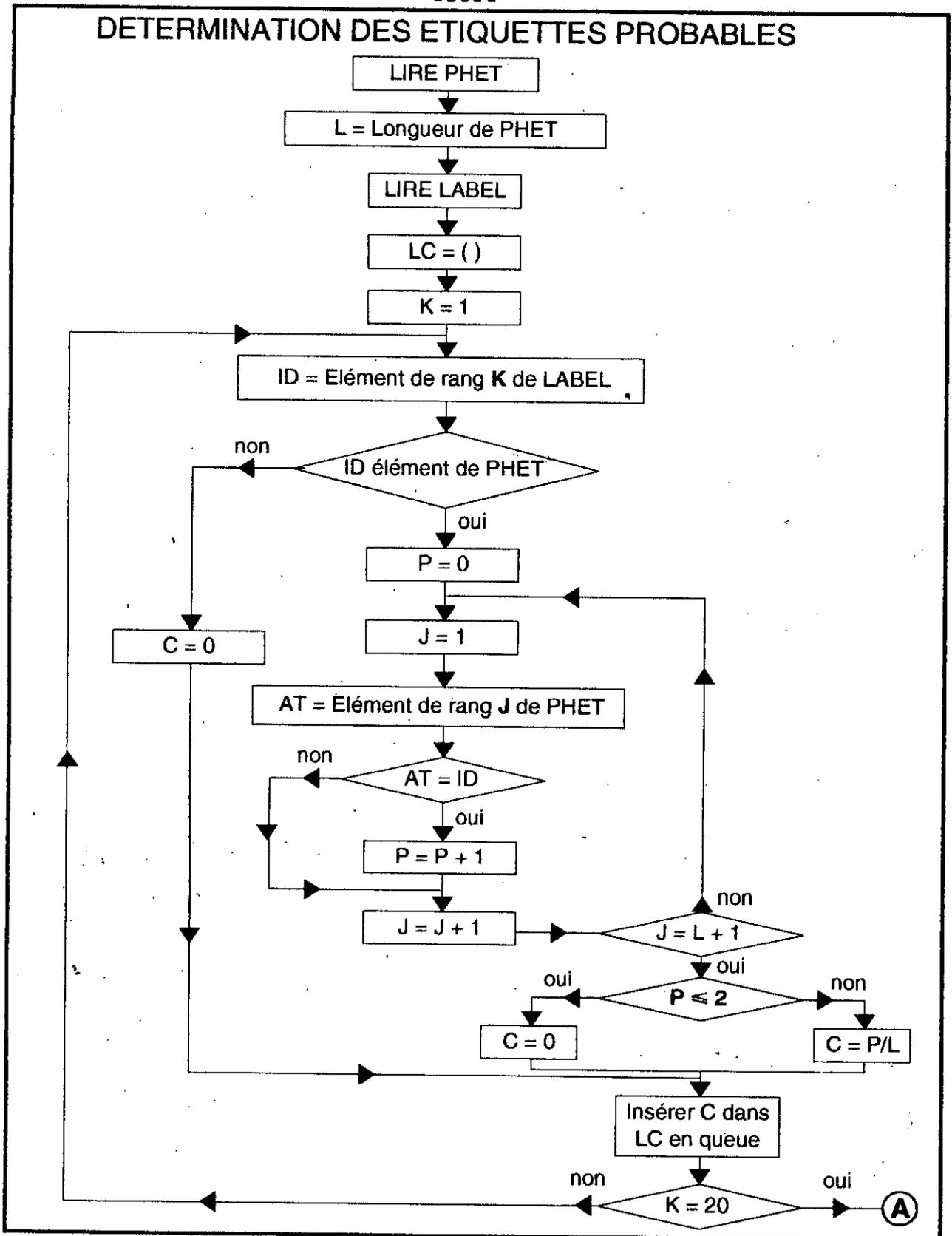


Organigramme D

# FONCTIONNEMENT DU MODULE D'IDENTIFICATION

## MIP

### DETERMINATION DES ETIQUETTES PROBABLES



Organigramme D1

(A)

# RECHERCHE DU TREILLIS PHONETIQUE

LIRE LC

TMAX1 = MAX(LC)

R1 = Rang de TMAX1

IDF1 = Elément de rang R1 de LABEL

PHON1 = (IDF1 TMAX1)

Substituer 0 à TMAX1 dans LC

TMAX2 = MAX(LC)

R2 = Rang de TMAX2

IDF2 = Elément de rang R2 de LABEL

PHON2 = (IDF2 TMAX2)

TREIL = Liste (PHON1 PHON2)

**BRI**  
BASE DE REGLES  
D'IDENTIFICATION

## III - BASE DE REGLES D'IDENTIFICATION

- RI1 : SI PHON1 = VOC  
ALORS PHONEME = / a /
- RI2 : SI PHON1 = VMOC  
ALORS PHONEME = / œ /
- RI3 : SI PHON1 = VMOP  
ALORS PHONEME = / ɔ /
- RI4 : SI PHON1 = VMFA  
ALORS PHONEME = / c /
- RI5 : SI PHON1 = VMFC  
ALORS PHONEME = / Ø /
- RI6 : SI PHON1 = VMFP  
ALORS PHONEME = / o /
- RI7 : SI PHON1 = VFA  
ALORS PHONEME = / i /
- RI8 : SI PHON1 = VFC  
ALORS PHONEME = / y /
- RI9 : SI PHON1 = VFP  
ALORS PHONEME = / u /
- RI10 : SI PHON1 = BUZZ  
ET PHON2 = IV  
ALORS PHONN = CPV
- RI11 : SI PHON1 = BUZZ  
ET PHON2 = CFV  
ALORS PHONN = CPV
- RI12 : SI PHON1 = STOP  
ET PHON2 = CF  
ALORS PHONN = CP
- RI13 : SI PHON1 = STOP  
ET PHON2 = IS  
ALORS PHONN = CP

- RI14 : SI PHON = CN  
ALORS PHONEME = (m n)
- RI15 : SI PHON = CL  
ALORS PHONEME = / l /
- RI16 : SI PHON1 = CFV  
ET EM [250-1600] / EM [1600-8000] = 0  
ALORS PHONEME = / v /
- RI17 : SI PHON = CFNV  
ET EM [250-1600] / EM [1600-8000] = 0  
ALORS PHONEME = / f /
- RI18 : SI PHON = CFV  
ET EM [250-3400] / EM [3400-8000] = 0  
ALORS PHONEME = / z /
- RI19 : SI PHON = CFNV  
ET EM [250-3400] / EM [3400-8000] = 0  
ALORS PHONEME = / s /
- RI20 : SI PHON = CF  
ET EM [250-2500] / EM [2500-5000] = 0  
ALORS PHONEME = / j /
- RI21 : SI PHON = CFV  
ET EM [250-2500] / EM [2500-5000] = 0  
ALORS PHONEME = / ʒ /
- RI22 : SI PHONN = CPV  
ET CDG < 7  
ALORS PHONEME = / b /
- RI23 : SI PHONN = CP  
ET CDG < 7  
ALORS PHONEME = / p /
- RI24 : SI PHONN = CPV  
ET CDG > 7  
ALORS PHONEME = / d /
- RI25 : SI PHONN = CP  
ET CDG > 7  
ALORS PHONEME = / t /

RI 26 : SI PHONN = CPV  
ET EM [250-2500] < EM [2500-5000]  
ET EM [2500-5000] > EM [5000-8000]  
ALORS PHONEME = / g /

RI 27 : SI PHONN = CP  
ET EM [250-2500] < EM [2500-5000]  
ET EM [2500-5000] > EM [5000-8000]  
ALORS PHONEME = / k /

RI28 : SI PHON = IV  
ALORS PHONEME = IV

RI29 : SI PHON = IS  
ALORS PHONEME = IS

## CHAPITRE VII

### MODULE D'ANALYSE LEXICALE

#### I - INTRODUCTION

Le rôle de ce module est de reconstituer la phrase candidate à la reconnaissance à partir de la chaîne de phonèmes (ou treillis de phonèmes) donnée à la sortie du module d'identification phonétique.

Il s'agit de passer d'une suite de phonèmes probablement entachée d'erreurs, à une phrase éventuellement incomplète sous la forme écrite, c'est-à-dire en lui faisant correspondre une orthographe.

Ceci nécessite de traiter le passage de la forme orale à la forme écrite.

Notre système étant un système limité de reconnaissance de la parole, le vocabulaire admis est un sous-ensemble du français, que nous avons codé dans un dictionnaire.

Le dictionnaire comporte les mots du vocabulaire dans leurs deux formes: orale et écrite.

Le problème-clé à résoudre est de trouver la procédure d'accès à ces représentations.

#### II - MODELE D'ACCES AU LEXIQUE

Nous avons opté pour un modèle d'accès de type recherche active dans le dictionnaire, selon une méthode inspirée du modèle de la "cohorte" de Marslen-Wilson (voir chapitre I, 1<sup>re</sup> partie).

Les mots du lexique sont représentés sous leur forme complète. Ce choix est dicté par la petite taille de notre dictionnaire; car notre objectif est expérimental.

#### III - STRUCTURE DU DICTIONNAIRE

Le dictionnaire comporte tous les mots du lexique (une vingtaine dans le cas des tests que nous avons effectués). Les mots sont représentés par leurs phonèmes dans le domaine oral et par leur graphème dans le domaine écrit. Par exemple: le mot «chat» est représenté dans notre dictionnaire par : «ʃa» et «chat».

Le dictionnaire n'est pas figé et peut être redéfini pour chaque nouvelle application.

#### IV - STRATEGIE DE RECONNAISSANCE DES MOTS (voir organigramme E)

Nous faisons une reconnaissance de "gauche à droite", en utilisant comme première clef d'accès le premier phonème issu du module d'identification.

La première clef d'accès va sélectionner dans le dictionnaire la cohorte de mots qui débutent par ce phonème, cette cohorte est progressivement réduite par le phonème suivant jusqu'à identification du mot.

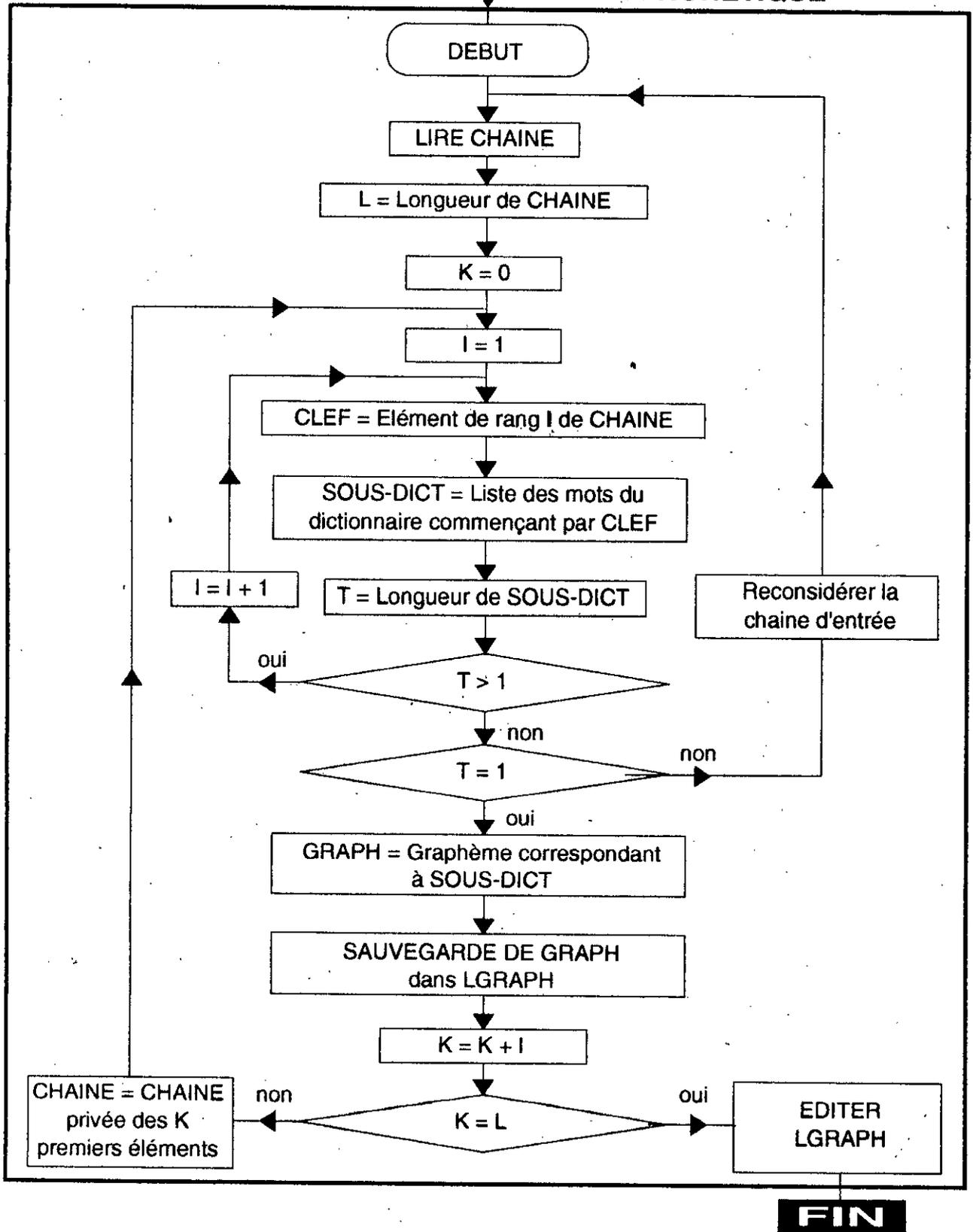
- Si aucun mot n'est sélectionné, nous effectuons un retour arrière en remplaçant la clef d'accès par le phonème de poids inférieur présent dans le treillis.

- Lorsqu'un mot a été sélectionné, nous le sauvegardons et recommençons l'opération avec le mot suivant.

# FONCTIONNEMENT DU MODULE D'ANALYSE LEXICALE

## MAL

CHAINE PHONETIQUE



Organigramme E

## CHAPITRE VIII

### ORGANISATION ET STRUCTURE DES LOGICIELS

#### I - LANGAGES UTILISES

Les langages mis en œuvre pour le développement de notre système sont les langages Lisp et Prolog. Ces deux langages spécifiques au domaine de l'intelligence artificielle se distinguent des langages de programmation traditionnels par les possibilités de programmation dynamique qu'ils offrent<sup>(15)</sup>.

##### I.1 - Langage Lisp

Les modules de prétraitement MPT, de segmentation MSEG et le module d'analyse lexicale MAL sont écrits en Lisp. La récursivité offerte par Lisp nous permet une programmation dynamique et évolutive nécessaire au traitement de données dont on ne connaît pas la taille a priori. En effet, la longueur de la phrase à analyser ne peut être connue à l'avance. Lisp nous offre l'avantage de séparer nettement le programme des données à traiter. Sa structure arborescente permet une représentation des données très souple et une manipulation aisée d'objets symboliques.

Les programmes sont écrits sous forme procédurale, Chaque procédure représentant une fonction pour Lisp. Une procédure peut être appelée par n'importe quelle autre procédure ou par elle même (récursivité).

Les programmes n'ont pas une structure séquentielle. Les organigrammes donnés par souci de clarté et de lisibilité rendent assez mal le fonctionnement des procédures.

##### \* Exemple :

La procédure qui permet de détecter la fin de parole dans le module de prétraitement est écrite en Lisp de manière déclarative comme suit:

<u>Procédure</u>	Detect-sil (LET, K, X) 1. <u>Début</u> 2. ET=premier élément de LET 3. Si ET > S-PAR alors 3.1. <u>Début</u> 3.2. LET=LET privé de son premier élément 3.3. X=0 3.4. K=K+1 3.5. Detect-sil (LET, K, X) 3.6. <u>Fin</u> 4. K=K+1 5. X=X+1 6. Si X≥20 alors renvoyer K-X 7. LET=LET privé de son premier élément 8. Detect-sil (LET, K, X) 9. <u>Fin</u>
------------------	---

Pour des raisons d'intelligibilité, nous avons traduit en français, les fonctions Lisp utilisées.

### **I.2 - Langage Prolog**

Les modules d'étiquetage MET et d'identification phonétique MIP sont écrits en Prolog. Prolog est un moteur d'inférences.

C'est un moteur monotone, d'ordre 1 (admettant des variables), fonctionnant en régime par tentatives avec retour arrière, avec une stratégie en profondeur d'abord et un mode d'invocation des règles en chaînage arrière.

En plus du moteur d'inférences, Prolog offre plusieurs autres fonctions de calcul et de manipulation de symboles.

Le fonctionnement en chaînage arrière nous conduit à raisonner des buts vers les données.

Pour les besoins de notre système, nous avons modifié la stratégie de raisonnement de Prolog en lui imposant une stratégie en "largeur d'abord".

## **II - STRUCTURATION DES BASES DE REGLES**

Pour des raisons d'homogénéité, nous avons adopté la même structure pour le module d'étiquetage et le module d'identification phonétique.

Le chaînage arrière impose de rédiger les règles sous la forme:

Conclusion : Propositions

Interprétée par: pour résoudre le problème contenu dans "conclusion", il suffit de résoudre les sous-problèmes contenus dans "propositions". La partie déclencheur d'une règle est sa partie conclusion.

Pour améliorer les performances des modules MET et MIP, nous avons structuré la base de connaissances en ordonnant les règles de la plus robuste à la moins sûre.

## **III - INTERACTION ENTRE LISP ET PROLOG**

Le programme principal qui contrôle le système est écrit en Lisp. Prolog est invoqué sous la direction de Lisp pour traiter les modules MET et MIP.

---

# **TROISIEME PARTIE**

## **TESTS ET RESULTATS**

## TESTS ET RESULTATS

### I - TESTS

Nous avons testé notre système avec les phrases "ce chat subit" et "ce petit phoque" dont les écritures phonémiques correspondent respectivement à "sø ʃa sybi" et "sø pøti fokø".

Ces phrases, incomplètes du point de vue sémantique, ont été choisies car elles regroupent:

- Pour la phrase "ce chat subit":
  - quatre voyelles différentes,
  - deux fricatives sourdes,
  - une plosive voisée.
- Pour la phrase "ce petit phoque":
  - deux voyelles différentes,
  - deux plosives sourdes,
  - une fricative sourde.

Les paramètres de ces phrases ont été calculés à partir de sonagrammes des sons qui les composent.

La composition de ces phrases nous permet de tester les indices les plus robustes contenus dans notre système.

Nous avons introduit volontairement des phases de silence non dépourvu de bruit pour tester la validité du logiciel déterminant les début et fin de parole.

Nous avons également inséré des pics d'énergie sans signification acoustique pour tester le logiciel qui contrôle la segmentation.

Pour un meilleur suivi des différentes étapes de reconnaissance, le système imprime les résultats partiels issus de chaque module.

#### I.1 - Fichier d'entrée

Les paramètres de chaque échantillon centiseconde composant la phrase candidate ont été introduits manuellement dans un fichier que l'on a appelé "FICH-ENT" (voir listings).

#### I.2 - Module de prétraitement

Ce module calcule pour chaque échantillon centiseconde:

- le centre de gravité spectral;
- l'énergie totale;
- l'énergie dans la bande [250 - 650] Hz.

L'évolution temporelle de ces paramètres est donnée respectivement par:

LCDG, LET, LEBF (voir listings).

Le module de prétraitement calcule aussi:

- l'énergie totale maximale contenue dans le signal ETMAX;
- le seuil de parole S-PAR;
- le début de parole DEB;
- la fin de parole FIN;

- le seuil de voisement S-VOIS;
- le seuil de friction S-FRIC.

Il réduit le fichier d'entrée en un fichier ne contenant plus que le signal de parole effective FICH-PAR.

### **1.3 - Module d'étiquetage**

Le module d'étiquetage, en utilisant des règles contenues dans la base de règles BRE, attribue à chaque échantillon centiseconde une étiquette. La liste des étiquettes attribuées est contenue dans LETIQ.

### **1.4 - Module de segmentation**

Dans une première étape, il segmente le signal en pseudo-syllabes en opérant une découpe sur la courbe d'évolution temporelle de l'énergie totale. Les segments pseudo-syllabiques sont donnés par LSEG.

Dans une seconde étape, le module de segmentation, tout en effectuant un contrôle sur d'éventuels segments erronés, découpe le signal en unités phonétiques données par LPHON (voir listings).

Enfin, il construit la liste des phonèmes étiquetés LPHET en utilisant la liste des étiquettes LETIQ.

### **1.5 - Module d'identification phonétique**

Le module d'identification phonétique construit la liste LTREIL des treillis phonétiques en attribuant à chaque phonème identifié un coefficient de confiance. Il détermine, grâce à la base de règles BRI, la chaîne de phonèmes CHAINE composant la phrase test.

### **1.6 - Module d'analyse lexicale**

A partir d'un dictionnaire contenant une vingtaine de mots, le module d'analyse lexicale effectue la reconnaissance des mots et détermine l'orthographe de la chaîne phonétique donnée par LGRAPH (voir listings).

## **II - INTERPRETATION DES RESULTATS**

Ces tests ont donné un taux de reconnaissance de 100%, les phrases candidates ayant été entièrement reconnues. Cette apparente perfection est due, en partie, à la nature des phrases test qui ne comportent que des plosives et des fricatives.

Les indices retenus pour l'identification de ces consonnes et des voyelles en général semblent donc robustes.

Ces tests nous permettent d'affirmer que:

- le voisement est bien détecté par le système;
- les plosives et les fricatives sont bien détectées et bien discriminées.

La phrase "le canard lit" testée par le système a donné lieu à une bonne segmentation par étiquettes mais la recherche des noyaux vocaliques a donné des résultats erronés en raison des transitions "liquide-voyelle" qui ne sont pas aussi nettes que les transitions "fricative-voyelle" ou "plosive-voyelle". De plus, les indices retenus pour les liquides ne semblent pas robustes.

## **III - CONCLUSIONS**

Ces tests nous ont permis de vérifier la validité de nos logiciels et la robustesse des connaissances implémentées pour la reconnaissance des voyelles, des fricatives et des plosives.

Le test sur la phrase "le canard lit", qui comporte 3 consonnes liquides et une consonne nasale, choisie sciemment, a permis de poser le problème du manque de connaissances performantes sur ce type de consonnes. En effet, la disponibilité d'autres sources de connaissances est souhaitable pour étendre le système à un langage moins contraint et traiter ainsi des vocabulaires plus riches et des structures linguistiques plus élaborées.

Test 1

Phrase candidate : "ce chat subit"

FICH-ENT= ( 0 )  
 ( 0 )  
 ( 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 )  
 ( 0 0 0 0 0 0 0 0 0 4 6 4 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 6 6 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 7 6 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 9 6 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 0 80 )  
 ( 0 0 0 0 0 0 0 0 0 4 19 6 2 4 0 0 0 0 0 0 80 )  
 ( 9 0 0 0 0 0 9 9 0 18 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 9 0 0 0 0 0 9 9 0 18 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 10 0 0 0 0 0 10 10 0 20 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 11 0 0 0 0 0 11 11 0 22 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 12 0 0 0 0 0 12 12 0 24 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 8 0 0 0 0 0 8 8 0 16 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 8 0 0 0 0 0 5 6 0 16 0 0 0 0 0 V 0,4 1,6 V V 40 )  
 ( 0 0 0 0 0 0 0 8 4 2 4 2 0 0 0 0 0 0 60 )

Fichier d'entrée : Liste des échantillons centiseconde ECSi où ECSi est représenté par une liste de 20 paramètres.  
 ECSi = (E1 E2 E3 E4 ... E14 IF0 F1 F2 IF1 IF2 TPZ)

(0 0 0 0 0 0 10 5 2,5 5 2,5 0 0 0 0 0 0 60)  
 (0 0 0 0 0 0 10 5 2,5 5 2,5 0 0 0 0 0 0 60)  
 (0 0 0 0 0 0 12 6 3 6 3 0 0 0 0 0 0 60)  
 (0 0 0 0 0 0 12 6 3 6 3 0 0 0 0 0 0 60)  
 (0 0 0 0 0 0 16 8 4 8 4 0 0 0 0 0 0 60)  
 (0 0 0 0 0 0 16 8 4 8 4 0 0 0 0 0 0 60)  
 (0 0 0 0 0 0 16 8 4 8 4 0 0 0 0 0 0 60)  
 (0 5 9 0 0 9 0 0 0 11 11 0 0 0 V 7,5 1,4 V V 36)  
 (0 5 9 0 0 9 0 0 0 11 11 0 0 0 V 7,5 1,4 V V 36)  
 (0 5 11 0 0 11 0 0 0 14 14 0 0 0 V 7,5 1,4 V V 36)  
 (0 7 13 0 0 13 0 0 0 16 16 0 0 0 V 7,5 1,4 V V 36)  
 (0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
 (0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
 (0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
 (0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
 (0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
 (0 6 12 0 0 12 0 0 0 15 15 0 0 0 V 7,5 1,4 V V 36)  
 (0 5 9 0 0 9 0 0 0 11 11 0 0 0 V 7,5 1,4 V V 36)  
 (0 0 0 0 0 0 0 0 4 6 4 2 4 0 0 0 0 80)  
 (0 0 0 0 0 0 0 0 4 6 6 2 4 0 0 0 0 80)  
 (0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 80)  
 (0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 80)  
 (0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 80)  
 (0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 80)  
 (0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 80)  
 (0 0 0 0 0 0 0 0 4 6 4 2 4 0 0 0 0 80)  
 (10 0 0 0 0 10 5 15 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (13 0 0 0 0 12 6 18 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (15 0 0 0 0 15 7 23 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (15 0 0 0 0 15 7 23 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (20 0 0 0 0 20 10 30 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (20 0 0 0 0 20 10 30 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (20 0 0 0 0 20 10 30 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (20 0 0 0 0 20 10 30 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (15 0 0 0 0 15 7 23 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (6 1 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0)  
 (4 1 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0)  
 (2 1 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0)  
 (3 0 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0)  
 (4 1 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0)  
 (2,5 2,5 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)



FICH-PAR= ( 20 0 0 0 0 0 0 0 0 0 0 4 6 4 2 4 0 0 0 0 0 80)  
(22 0 0 0 0 0 0 0 0 0 0 4 6 6 2 4 0 0 0 0 0 80)  
(23 0 0 0 0 0 0 0 0 0 0 4 7 6 2 4 0 0 0 0 0 80)  
(25 0 0 0 0 0 0 0 0 0 0 4 9 6 2 4 0 0 0 0 0 80)  
(30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
(30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
(30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
(30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
(30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
(35 0 0 0 0 0 0 0 0 0 0 4 19 6 2 4 0 0 0 0 0 80)  
(45 9 9 0 0 0 0 0 9 9 0 18 0 0 0 0 V 0,4 1,6 V V 40)  
(45 9 9 0 0 0 0 0 9 9 0 18 0 0 0 0 V 0,4 1,6 V V 40)  
(50 10 10 0 0 0 0 0 10 10 0 20 0 0 0 0 V 0,4 1,6 V V 40)  
(55 11 11 0 0 0 0 0 11 11 0 22 0 0 0 0 V 0,4 1,6 V V 40)  
(60 12 12 0 0 0 0 0 12 12 0 24 0 0 0 0 V 0,4 1,6 V V 40)  
(80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
(80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
(80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
(40 8 8 0 0 0 0 0 8 8 0 16 0 0 0 0 V 0,4 1,6 V V 40)  
(35 8 8 0 0 0 0 0 5 6 0 16 0 0 0 0 V 0,4 1,6 V V 40)  
(20 0 0 0 0 0 0 0 8 4 2 4 2 0 0 0 0 0 0 0 60)  
(25 0 0 0 0 0 0 0 10 5 2,5 5 2,5 0 0 0 0 0 0 0 60)  
(25 0 0 0 0 0 0 0 10 5 2,5 5 2,5 0 0 0 0 0 0 0 60)  
(30 0 0 0 0 0 0 0 12 6 3 6 3 0 0 0 0 0 0 0 60)  
(30 0 0 0 0 0 0 0 12 6 3 6 3 0 0 0 0 0 0 0 60)  
(40 0 0 0 0 0 0 0 16 8 4 8 4 0 0 0 0 0 0 0 60)  
(40 0 0 0 0 0 0 0 16 8 4 8 4 0 0 0 0 0 0 0 60)  
(40 0 0 0 0 0 0 0 16 8 4 8 4 0 0 0 0 0 0 0 60)  
(45 5 0 5 9 0 0 9 0 0 0 11 11 0 0 0 V 7,5 1,4 V V 36)  
(45 0 0 5 9 0 0 9 0 0 0 11 11 0 0 0 V 7,5 1,4 V V 36)  
(55 5 0 5 11 0 0 11 0 0 0 14 14 0 0 0 V 7,5 1,4 V V 36)  
(65 7 0 7 13 0 0 13 0 0 0 16 16 0 0 0 V 7,5 1,4 V V 36)  
(80 8 0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
(80 8 0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
(80 8 0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
(80 8 0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
(80 8 0 8 16 0 0 16 0 0 0 20 20 0 0 0 V 7,5 1,4 V V 36)  
(60 6 0 6 12 0 0 12 0 0 0 15 15 0 0 0 V 7,5 1,4 V V 36)

Fichier parole : Fichier ne contenant que le échantillon représentant de la parole effect.

(45 5 0 5 9 0 0 9 0 0 0 11 11 0 0 0 0 V 7,5 1,4 V V 36)  
 (20 0 0 0 0 0 0 0 0 0 0 4 6 4 2 4 0 0 0 0 0 80)  
 (22 0 0 0 0 0 0 0 0 0 0 4 6 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 4 6 4 2 4 0 0 0 0 0 80)  
 (40 10 10 0 0 0 0 0 10 5 15 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (50 13 13 0 0 0 0 0 12 6 18 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (60 15 15 0 0 0 0 0 15 7 23 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (60 15 15 0 0 0 0 0 15 7 23 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (80 20 20 0 0 0 0 0 20 10 30 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (80 20 20 0 0 0 0 0 20 10 30 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (80 20 20 0 0 0 0 0 20 10 30 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (80 20 20 0 0 0 0 0 20 10 30 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (60 15 15 0 0 0 0 0 15 7 23 0 0 0 0 0 0 V 0,3 1,8 V V 35)  
 (7 7 6 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0 0)  
 (5 5 4 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0 0)  
 (3 3 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0 0)  
 (3 3 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0 0)  
 (5 5 4 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 V 0 0 0 0 0)  
 (25 5 2,5 2,5 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (25 5 2,5 2,5 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (25 5 2,5 2,5 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (30 10 5 5 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (50 12,5 12,5 0 0 0 0 0 12,5 12,5 12,5 0 0 0 0 0 0 V 0,25 2,2 V V 40)  
 (60 15 15 0 0 0 0 0 15 15 15 0 0 0 0 0 0 V 0,25 2,2 V V 40)  
 (80 20 20 0 0 0 0 0 20 20 20 0 0 0 0 0 0 V 0,25 2,2 V V 40)  
 (80 20 20 0 0 0 0 0 20 20 20 0 0 0 0 0 0 V 0,25 2,2 V V 40)  
 (80 20 20 0 0 0 0 0 20 20 20 0 0 0 0 0 0 V 0,25 2,2 V V 40)  
 (60 15 15 0 0 0 0 0 15 15 15 0 0 0 0 0 0 V 0,25 2,2 V V 40)  
 (60 15 15 0 0 0 0 0 15 15 15 0 0 0 0 0 0 V 0,25 2,2 V V 40)  
 (60 15 15 0 0 0 0 0 15 15 15 0 0 0 0 0 0 V 0,25 2,2 V V 40)  
 (5 2 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0))

LET=(0 0 0 7 5 5 5 5 5 5 5 7 7 7 7 7 7 7 7 7 20 22 23 25 30 30 30 30 30 35  
 45 45 50 55 60 80 80 80 40 35 20 25 25 30 30 40 40 40 45 45 55 65 80 80  
 80 80 80 60 45 20 22 30 30 30 30 30 30 40 50 60 60 80 80 80 80 60 7 5  
 3 3 5 25 25 25 30 50 60 80 80 80 60 60 60 5 5 5 0 0 0 0 0 0 0 0 0 0  
 0 0 0 0 0 0 0 0 0)

Evolution temporelle de l'énergie totale

LEBF=(0 0 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 0 0 0 0 0 0 0 0 0 9 9 10 11  
 12 16 16 16 8 8 0 0 0 0 0 0 0 0 5 5 5 7 8 8 8 8 8 6 5 0 0 0 0 0 0  
 0 10 13 15 15 20 20 20 20 15 7 5 3 3 5 5 5 5 10 12,5 15 20 20 20 15 15  
 15 2 2 2 0)

Evolution temporelle de l'énergie en basses fréquences [250-650] Hz

S-PAR=7      Seuil de parole

ETMAX=80      Energie totale maximale contenue dans le signal

S-VOIS=2      Seuil de voisement

DEB=20      Début de parole

FIN=74      Fin de parole

LETS=(20 22 23 25 30 30 30 30 30 35 45 45 50 55 60 80 80 80 40 35 20 25 25  
 30 30 40 40 40 45 45 55 65 80 80 80 80 80 60 45 20 22 30 30 30 30 30  
 30 40 50 60 60 80 80 80 80 60 7 5 3 3 5 25 25 25 30 50 60 80 80 80 60  
 60 60 5)

LCDG=(11,8 11,8 11,8 11,7 11,6 11,6 11,6 11,6 11,6 11,5 7,2 7,2 7,2 7,2 7,2  
 7,2 7,2 7,2 7,2 7,2 9,4 9,4 9,4 9,4 9,4 9,4 9,4 9,4 7,2 7,2 7,2 7,2 7,2  
 7,2 7,2 7,2 7,2 7,2 7,2 11,6 11,6 11,6 11,6 11,6 11,6 11,6 11,6 7,3 7,3  
 7,3 7,3 7,3 7,3 7,3 7,3 7,3 0 0 0 0 0 6,2 6,2 6,2 5,7 7,1 7,1 7,1 7,1  
 7,1 7,1 7,1 7,1 3)

Evolution temporelle du centre de gravité spectral

S-FRIC=40      Seuil de friction

LABEL=(VOA VOC VOP VFA VFC VFP VMOA VMOC VMOP VMFA VMFC VMFP CN CL STOP BUZZ  
CFV CF IV IS)

LETIQ=(CF CF CF CF CF CF CF CF CF CF VMFC VMFC VMFC VMFC VMFC VMFC VMFC VMFC  
VMFC VMFC CF VOC VOC VOC VOC VOC VOC VOC VOC  
VOC VOC CF CF CF CF CF CF CF CF VFC VFC VFC VFC VFC VFC VFC VFC VFC  
BUZZ BUZZ BUZZ BUZZ BUZZ IV IV IV IV VFA VFA VFA VFA VFA VFA VFA VFA  
STOP) Liste des étiquettes attribuées

LSEG=((20 22 23 25 30 30 30 30 30 35) (45 45 50 55 60 80 80 80 40 35 20) (25  
25 30 30 40 40 40) (45 45 55 65 80 80 80 80 80 60 45 20) (22 30 30 30  
30 30 30 40) (50 60 60 80 80 80 80 60) (7 5 3 3 5 25 25 25 30) (50 60  
80 80 80 60 60 60 5))

LPHON=((20 22 23 25 30 30 30 30 30 35 45 45 50 55 60 80 80 80 40 35 20) (25  
25 30 30 40 40 40 45 45 55 65 80 80 80 80 80 60 45 20) (22 30 30 30 30  
30 30 40 50 60 60 80 80 80 80 60 7 5 3) (3 5 25 25 25 30 50 60 80 80  
80 60 60 60 5))

LPHET=((CF CF CF CF CF CF CF CF CF CF) (VMFC VMFC VMFC VMFC VMFC VMFC VMFC  
VMFC VMFC VMFC CF) (CF CF CF CF CF CF CF) (VOC VOC VOC VOC VOC VOC VOC  
VOC VOC VOC VOC CF) (CF CF CF CF CF CF CF VFC) (VFC VFC VFC VFC VFC VFC  
VFC VFC) (BUZZ BUZZ BUZZ BUZZ BUZZ IV IV IV IV) (VFA VFA VFA VFA VFA  
VFA VFA VFA STOP)) Liste des phonèmes étiquetés

LTREIL=((CF 1) (VMFC 0,9) (CF 1) (VOC 0,9) (CF 0,8) (VFC 1) ((BUZZ 0,55) (IV  
0,44)) (VFA 1)) Liste des treillis phonétiques avec calcul du score

CHAINE=(s 0 \$ a s y b i) Chaîne phonétique identifiée

DICTIONNAIRE=((abime abimer) (abone abonner) (abor abord) (ami ami)  
(bak bac) (dute douter) (fisel0 ficelle) (gamin0 gamine)  
(ku coup) (pase passer) (s0 ce) (so sot) (sybi subit)  
(\$a chat) (\$0mine cheminée) (\$ipi chipie) (\$y\$ote chuchoter)  
(tape taper))

LGRAPH=((ce) (chat) (subit))

SOLUTION: "ce chat subit"

Test 2

Phrase candidate : "ce petit phoque"



(0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 0 0 0 0 0 0 0 0 0 0 40)  
 (9 0 0 0 0 0 9 9 0 18 0 0 0 0 V 0,4 1,6 V V 40)  
 (9 0 0 0 0 0 9 9 0 18 0 0 0 0 V 0,4 1,6 V V 40)  
 (10 0 0 0 0 0 10 10 0 20 0 0 0 0 V 0,4 1,6 V V 40)  
 (11 0 0 0 0 0 11 11 0 22 0 0 0 0 V 0,4 1,6 V V 40)  
 (12 0 0 0 0 0 12 12 0 24 0 0 0 0 V 0,4 1,6 V V 40)  
 (16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
 (16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
 (16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
 (8 0 0 0 0 0 8 8 0 16 0 0 0 0 V 0,4 1,6 V V 40)  
 (8 0 0 0 0 0 5 6 0 16 0 0 0 0 V 0,4 1,6 V V 40)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (12,5 0 0 0 0 0 0 12,5 12,5 12,5 0 0 0 0 V 0,25 2,2 V V 40)  
 (15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (20 0 0 0 0 0 0 20 20 20 0 0 0 0 V 0,25 2,2 V V 40)  
 (20 0 0 0 0 0 0 20 20 20 0 0 0 0 V 0,25 2,2 V V 40)  
 (20 0 0 0 0 0 0 20 20 20 0 0 0 0 V 0,25 2,2 V V 40)  
 (15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (0 0 0 0 0 0 4 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (0 0 0 0 0 0 4 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (0 0 0 0 0 0 4 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (0 0 0 0 0 0 4 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (0 0 0 0 0 0 4 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (0 0 0 0 0 0 4 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (0 0 0 0 0 0 4 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (10 0 10 0 0 0 0 0 15 15 10 0 0 0 V 0,4 0,7 V V 25)  
 (10 0 10 0 0 0 0 0 15 15 10 0 0 0 V 0,4 0,7 V V 25)  
 (10 0 10 0 0 0 0 0 15 15 10 0 0 0 V 0,4 0,7 V V 25)  
 (10 0 10 0 0 0 0 0 15 15 10 0 0 0 V 0,4 0,7 V V 25)  
 (10 0 10 0 0 0 0 0 15 15 10 0 0 0 V 0,4 0,7 V V 25)  
 (10 0 10 0 0 0 0 0 15 15 10 0 0 0 V 0,4 0,7 V V 25)



FICH-PAR= (20 0 0 0 0 0 0 0 0 0 0 0 4 6 4 2 4 0 0 0 0 0 80)  
 (22 0 0 0 0 0 0 0 0 0 0 0 4 6 6 2 4 0 0 0 0 0 80)  
 (23 0 0 0 0 0 0 0 0 0 0 0 4 7 6 2 4 0 0 0 0 0 80)  
 (25 0 0 0 0 0 0 0 0 0 0 0 4 9 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 0 4 14 6 2 4 0 0 0 0 0 80)  
 (30 0 0 0 0 0 0 0 0 0 0 0 4 19 6 2 4 0 0 0 0 0 80)  
 (45 9 9 0 0 0 0 0 9 9 0 18 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (45 9 9 0 0 0 0 0 9 9 0 18 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (50 10 10 0 0 0 0 0 10 10 0 20 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (55 11 11 0 0 0 0 0 11 11 0 22 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (60 12 12 0 0 0 0 0 12 12 0 24 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (40 8 8 0 0 0 0 0 8 8 0 16 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (35 8 8 0 0 0 0 0 5 6 0 16 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (14 0 0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 0 0 0 0 0 0 0 0 0 0 40)  
 (14 0 0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 0 0 0 0 0 0 0 0 0 0 40)  
 (45 9 9 0 0 0 0 0 9 9 0 18 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (45 9 9 0 0 0 0 0 9 9 0 18 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (50 10 10 0 0 0 0 0 10 10 0 20 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (55 11 11 0 0 0 0 0 11 11 0 22 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (60 12 12 0 0 0 0 0 12 12 0 24 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (40 8 8 0 0 0 0 0 8 8 0 16 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (40 8 8 0 0 0 0 0 5 6 0 16 0 0 0 0 0 V 0,4 1,6 V V 40)  
 (0 0)  
 (0 0)  
 (0 0)  
 (0 0)  
 (0 0)  
 (20 0 0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (20 0 0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (20 0 0 0 2,5 2,5 2,5 2,5 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 1,2 V 0 0 0 0 40)  
 (50 12,5 12,5 0 0 0 0 0 12,5 12,5 12,5 0 0 0 0 0 V 0,25 2,2 V V 40)

(60 15 15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (80 20 20 0 0 0 0 0 0 20 20 20 0 0 0 0 V 0,25 2,2 V V 40)  
 (80 20 20 0 0 0 0 0 0 20 20 20 0 0 0 0 V 0,25 2,2 V V 40)  
 (80 20 20 0 0 0 0 0 0 20 20 20 0 0 0 0 V 0,25 2,2 V V 40)  
 (60 15 15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (60 15 15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (60 15 15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (60 15 15 0 0 0 0 0 0 15 15 15 0 0 0 0 V 0,25 2,2 V V 40)  
 (28 0 0 0 0 0 0 0 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (28 0 0 0 0 0 0 0 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (28 0 0 0 0 0 0 0 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (28 0 0 0 0 0 0 0 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (28 0 0 0 0 0 0 0 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (28 0 0 0 0 0 0 0 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (28 0 0 0 0 0 0 0 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (28 0 0 0 0 0 0 0 4 4 4 4 4 4 0 0 0 0 0 0 40)  
 (60 10 10 0 10 0 0 0 0 15 15 10 0 0 0 0 V 0,4 0,7 V V 25)  
 (60 10 10 0 10 0 0 0 0 15 15 10 0 0 0 0 V 0,4 0,7 V V 25)  
 (60 10 10 0 10 0 0 0 0 15 15 10 0 0 0 0 V 0,4 0,7 V V 25)  
 (60 10 10 0 10 0 0 0 0 15 15 10 0 0 0 0 V 0,4 0,7 V V 25)  
 (60 10 10 0 10 0 0 0 0 15 15 10 0 0 0 0 V 0,4 0,7 V V 25)  
 (60 10 10 0 10 0 0 0 0 15 15 10 0 0 0 0 V 0,4 0,7 V V 25)  
 (60 10 10 0 10 0 0 0 0 15 15 10 0 0 0 0 V 0,4 0,7 V V 25)  
 (60 10 10 0 10 0 0 0 0 15 15 10 0 0 0 0 V 0,4 0,7 V V 25)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)  
 (30 0 0 0 0 0 0 0 0 10 10 10 0 0 0 0 0 0 0 0 50)  
 (30 0 0 0 0 0 0 0 0 10 10 10 0 0 0 0 0 0 0 0 50)  
 (30 0 0 0 0 0 0 0 0 10 10 10 0 0 0 0 0 0 0 0 50)  
 (30 0 0 0 0 0 0 0 0 10 10 10 0 0 0 0 0 0 0 0 50)  
 (45 9 9 0 0 0 0 0 9 9 0 18 0 0 0 0 V 0,4 1,6 V V 40)  
 (45 9 9 0 0 0 0 0 9 9 0 18 0 0 0 0 V 0,4 1,6 V V 40)  
 (50 10 10 0 0 0 0 0 10 10 0 20 0 0 0 0 V 0,4 1,6 V V 40)  
 (55 11 11 0 0 0 0 0 11 11 0 22 0 0 0 0 V 0,4 1,6 V V 40)  
 (60 12 12 0 0 0 0 0 12 12 0 24 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
 (80 16 16 0 0 0 0 0 16 16 0 32 0 0 0 0 V 0,4 1,6 V V 40)  
 (40 8 8 0 0 0 0 0 8 8 0 16 0 0 0 0 V 0,4 1,6 V V 40)  
 (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0))



LPHET=((CF CF CF CF CF CF CF CF CF CF)(VMFC VMFC VMFC VMFC VMFC VMFC VMFC VMFC  
 VMFC VMFC)(STOP STOP STOP STOP STOP IS IS)(VMFC VMFC VMFC VMFC VMFC  
 VMFC VMFC VMFC VMFC VMFC)(STOP STOP STOP STOP STOP CF CF CF)(VFA VFA  
 VFA VFA VFA VFA VFA VFA VFA)(CF CF CF CF CF CF CF)(VMFP VMFP VMFP VMFP  
 VMFP VMFP VMFP VMFP)(STOP STOP STOP STOP STOP CF CF CF CF)(VMFC VMFC  
 VMFC VMFC VMFC VMFC VMFC VMFC STOP))

LTREIL=((CF 1) (VMFC 1) ((STOP 0,7) (IS 0,3)) (VMFC 1) ((STOP 0,6) (CF 0,4))  
 (VFA 1) (CF 1) (VMFP 1) ((STOP 0,5) (CF 0,4)) (VMFC 1))

CHAINE=(s 0 p 0 t i f o k)

DICTIONNAIRE=((abime abimer) (abone abonner) (abor abord) (ami ami)  
 (bak bac) (dute douter) (fisel0 ficelle) (fok0 phoque) (gamin0  
 gamine) (ku coup) (pase passer) (p0ti petit)  
 (s0 ce) (so sot) (sybi subit) (\$a chat) (\$0mine cheminée) (\$ipi  
 chipie) (\$y\$ote chuchoter) (tape taper))

LGRAPH=((ce) (petit) (phoque))

SOLUTION: "ce petit phoque"

## CONCLUSION ET PERSPECTIVES

### I - CONCLUSION

Nous avons tenté, à travers l'élaboration de notre système, d'apporter une contribution au délicat problème de la reconnaissance automatique de la parole continue multilocuteur.

Les résultats obtenus, variables selon les phrases testées, montrent que pour des domaines d'application limités, le système s'avère performant. Mais nous sommes encore très loin du vieux rêve de la communication homme-machine en langage naturel.

En effet, les tests ont montré que la faiblesse du système réside dans le manque de robustesse de certaines connaissances implémentées. De plus, il semble que les 73 règles utilisées par le système ne suffisent pas à décrire toutes les variations allophoniques des sons étudiés.

L'architecture modulaire, adoptée pour la conception du système permet d'appréhender le problème général de reconnaissance en décomposant la résolution en différentes étapes. Cette architecture confère à chacun des modules réalisés une autonomie qui lui permet de subir d'éventuelles modifications sans nécessiter d'interventions sur les autres modules.

De plus, l'approche "Intelligence artificielle" permet des possibilités d'évolution et d'extension des bases de connaissances sans modifier les logiciels implémentés.

Les perspectives restent donc ouvertes et les possibilités d'amélioration s'avèrent prometteuses.

Les performances du système peuvent être considérablement augmentées par l'enrichissement des bases de connaissances grâce à la collaboration d'experts du domaine.

### II - PERSPECTIVES

Une chaîne d'acquisition et de traitement du signal de parole permettrait d'évaluer avec précision les performances du système à partir de tests effectués sur des phrases réelles.

De tels tests permettraient sûrement, par l'observation d'un grand nombre de phrases réelles d'extraire de nouvelles connaissances pour enrichir ou même modifier les règles implémentées.

Le système peut être étendu à un langage moins contraint et moins limité, par la réalisation d'un module d'analyse lexicale plus élaboré.

La taille du dictionnaire et la complexité du langage peuvent être considérablement augmentées par l'introduction de contraintes contextuelles de niveaux supérieurs.

Il est possible d'étendre notre système à un système de compréhension de la parole par l'implémentation de modules de plus hauts niveaux tels que : l'analyse prosodique, l'analyse syntaxique, sémantique et pragmatique.

Par ailleurs, les modules d'étiquetage et d'indentification phonétique, ayant des structures indépendantes, peuvent être isolés et servir de bases à l'implémentation de systèmes experts opérant dans d'autres domaines d'application.

## BIBLIOGRAPHIE

# BIBLIOGRAPHIE

- [ 1 ] A. AGAPEYEFF : "Expert systems, Fifth generation and UK suppliers", NC Publications, Manchester, (1983).
- [ 2 ] J.S. BENETT : "An expert system for computer fault diagnosis", Voth IJCAI, pp.843-845, Vancouver, août 1981.
- [ 3 ] A. BONNEAU : Indices de reconnaissance des consonnes occlusives sourdes du français". Thèse de 3<sup>e</sup> cycle, Université d'Aix Marseille I 1984.
- [ 4 ] A. BONNEAU, M. ROSSI, R.G. MERCIER : "Hierarchical recognition of french vowels by SERAC IROISE expert-system", Séminaire franco-suédois, Grenoble 1985.
- [ 5 ] A. BONNET : "Applications de l'intelligence artificielle : Les systèmes experts". Rairo Informatique Vo 1.15, n°4, pp.325-341, 1981.
- [ 6 ] A. BONNET : L'intelligence artificielle, promesses et réalités. Inter-éditions, Paris, 1984.
- [ 7 ] B. BUCHANAN : "Experts systems", Journal of Automated reasoning, Vol. 1, n°1, pp.28-35 (1985).
- [ 8 ] B.G. BUCHANAN : "New research on expert systems. Machine Intelligence" 10, Ed. by JE. Hayes-Roth pp. 269-299, 1982.
- [ 9 ] H. CALLAIRE : "Les systèmes-experts. Une percée de l'intelligence artificielle". La recherche, Vol 13, n° 133, pp. 644-645 (1982).
- [ 10 ] CALLIOPE : La parole et son traitement automatique» Edition Masson, 1989.
- [ 11 ] N. CARBONNEL : "APHODEX : Design and implementation of an acoustic phonetic decoding expert system" IEEE ICASSP pp. 1201-1204, Tokyo, 1986.
- [ 12 ] N. CARBONNEL : "Techniques d'intelligence artificielle en décodage acoustico-phonétique" 14<sup>èmes</sup> journées d'études sur la parole Paris pp. 299-303, 1985.
- [ 13 ] M. CAYROL, H. FARRENY : "ARGOS II Un système de production pour écrire des résolveurs experts". Cong.AFCET.Info, pp. 351-361, Nancy, nov.80.
- [ 14 ] D. CHANIAT : "Systèmes-experts : Les applications, Micro et Robots", n° 16, pp.61-71, Mars 1985.
- [ 15 ] E. CHOURAQUI, H. FARRENY : "Modélisation du raisonnement et de la connaissance", Journée CNRS , Intelligence artificielle, GIF sur Yvette, Mars 1985.
- [ 16 ] E. CHOURAQUI : Méthodes de l'intelligence artificielle Hermès Pub. 1985.
- [ 17 ] M.O. CORDIER : "Les systèmes-experts" La recherche, n° 151 pp. 60-70, Janv.1984.
- [ 18 ] H. FARRENY : Les systèmes-experts : Principes et exemples Cepaduès Editions (1985).
- [ 19 ] J.M. HATON, D. FOHR : "Paramétrisation acoustique et décodage phonétique fondé sur des connaissances pour la parole continue multilocuteurs". Actes du séminaire «Gréco-PRC» Communication Homme Machine - 1988.
- [ 20 ] J.P. HATON, J.M. PIERREL, G. RERENOU, J. CAELEN, J.L. GAUVAIN : Reconnaissance automatique de la parole, Edition Dunod Informatique 1991.
- [ 21 ] J.L. LAURRIRE : "Représentation des connaissances". Université Paris VI, 1984..
- [ 22 ] J.L. LAURRIRE : "Représentation et utilisation des connaissances.: Les systèmes experts", Université Paris VI, 1984.
- [ 23 ] J.M. LIENAR, J.J. MARIANI : "Système de reconnaissance acoustique de mots isolés. MOISE", Dossier ANVAR n° 50312, 1982.

- [ 24 ] H. MELONI, J. GISPERT J. GUIZOL : "An expert-system for analytic word identification in continuous speech". 5ème J.I.: Les systèmes experts et leurs applications, Avignon, Mai 1985.
- [ 25 ] C. MERCIER : "Expert-systems, Approach to acoustic-phonetic decoding and word recognition", CNET, Centre Lannion A. pp. 201-221, 1990.
- [ 26 ] C. MERCIER : Analyse acoustique et transcription phonétique du signal de parole, Nancy, 1977.
- [ 27 ] N. NILSON : "Principles of artificial intelligence", Troga Publishing Company (1980).
- [ 28 ] N. NILSON : Problem-solving method in artificial intelligence. Mc Graw-Hill Company (1971)
- [ 29 ] PASTED : Un système de démonstrations automatique de théorèmes utilisant connaissances et métaconnaissances en mathématiques. Thèse d'état Univ. de Paris VI (1984).
- [ 30 ] D. RANDALL : "Diagnostic reasoning based on structure and behaviour", Massachusetts Institute of Technology MIT Press, Cambridge, USA, 1983.
- [ 31 ] R. REITER : A logic for default reasoning, University of British Columbia, Canada 1979.
- [ 32 ] R. REITER : A theory of diagnosis from first principles, University of Toronto, 1986.
- [ 33 ] E. RICH : Artificial intelligence. Mc Graw Hill, 1983.
- [ 34 ] E. SEYDEN : Systèmes-experts : Simuler l'intelligence, Sciences et techniques, n°4, pp.46-63, Mai 1984.
- [ 35 ] P.E. STERNE, M. ESKENASI, D. MEMMI : "An expert system for speech spectrogram reading". IEEE ICASSP, Tokyo, 1986.