UNIVERSITE DE NANCY 1
N° d'enregistrement au CNRS : A.0.10.930

THÈSE

présentée à

L'UNIVERSITE DE NANCY 1

pour obtenir le grade de
DOCTEUR ES-SCIENCES PHYSIQUES

par

Jacques BREMONT

Licencié ès-sciences physiques
Docteur en spécialité

SUJET : **CONTRIBUTION A LA RECONNAISSANCE AUTOMATIQUE DE LA PAROLE PAR LES SOUS-ENSEMBLES FLOUS.**

---

soutenue publiquement le 11 Avril 1975
devant la Commission d'Examen

Membres du Jury

président : **M.Y. BERNARD**

examinateurs : **M.M. E. LEIPP**
R. MAINARD
R. CARRE
A. FRUHLING
M. LAMOTTE

1975

UNIVERSITE DE NANCY 1
N° d'enregistrement au CNRS : A.0.10.930

THÈSE

présentée à

L'UNIVERSITE DE NANCY 1

pour obtenir le grade de
DOCTEUR ES-SCIENCES PHYSIQUES

par

Jacques BREMONT

Licencié ès-sciences physiques
Docteur en spécialité

SUJET : **CONTRIBUTION A LA RECONNAISSANCE AUTOMATIQUE DE LA PAROLE PAR LES SOUS-ENSEMBLES FLOUS.**

---

soutenue publiquement le 11 Avril 1975
devant la Commission d'Examen

Membres du Jury

président : **M.Y. BERNARD**

examinateurs : **M.M. E. LEIPP**
R. MAINARD
R. CARRE
A. FRUHLING
M. LAMOTTE

1975

IMP~CUCES

A mes Parents.

A ma Femme.

A Xavier et Virginie.

AVANT - PROPOS

Ce travail a été effectué au Laboratoire d'Electricité et d'Automatique de l'Université de Nancy I dans le cadre des activités du groupe d'Etude sur la Parole.

Nous exprimons toute notre gratitude à Monsieur le Professeur A. FRUHLING, Directeur du L.E.A., pour le soutien qu'il n'a cessé de nous prodiguer en toutes circonstances et pour l'intérêt qu'il a porté à ce travail.

Nous remercions également Monsieur M.-Y. BERNARD, Professeur au CNAM, qui, malgré ses nombreuses activités, a accepté de s'intéresser à ce mémoire et de nous faire l'honneur de présider le Jury d'examen.

Nous sommes reconnaissants à Monsieur E. LEIPP, Directeur du Laboratoire d'Acoustique de l'Université de Paris VI, de la confiance et de l'encouragement qu'il nous a apportés quant à l'application de notre nouvelle méthode dans le domaine de la Reconnaissance de la Parole.

Nous remercions également Monsieur R. CARRE, Chargé de Recherche au CNRS, d'avoir accepté de prendre connaissance de notre travail et de faire partie du Jury.

Nous remercions
Monsieur R. MAINARD, Directeur de l'I.U.T. de Nancy, de la confiance qu'il nous a toujours manifestée et de sa présence dans la Commission d'Examen.

Nous tenons à remercier vivement Monsieur M. LAMOTTE, Chargé de Recherche au CNRS, qui, par sa critique constructive, a permis l'éclaircissement des parties délicates de ce travail ; qu'il trouve ici le témoignage d'une amitié de longue date.

Nous remercions tous nos camarades du L.E.A. qui nous ont indirectement aidé par leur présence chaleureuse et également les personnes qui ont contribué à la présentation de ce mémoire.

TABLE DES MATIERES

INTRODUCTION

Dans le domaine de l'intelligence artificielle, la reconnaissance de la parole est un problème particulier de reconnaissance de formes.

Consistant à affecter une catégorie à une forme vocale présentée (phonème, mot ou phrase), elle intéresse plusieurs disciplines telles que l'acoustique, l'automatique, l'électronique, la phonétique, la linguistique, etc...

Si des efforts importants en reconnaissance de la parole concernent la construction des "images sonores", il n'en demeure pas moins vrai que la reconnaissance de ces formes, dont la variabilité est grande, doit faire intervenir la redondance des informations autres qu'acoustiques.

Avec un matériel important et de longues durées de calcul, l'exploitation optimale de l'information intrinsèque, permet de traiter des lexiques de quelques centaines de mots isolés. L'exploitation de l'information extrinsèque, linguistique ou sémantique, n'est possible que par un système de reconnaissance adjoint au précédent qui gonfle encore la taille et l'inertie du système global de reconnaissance.

Dans ce contexte général, il importe, pour devenir opérationnel, de simplifier le plus possible chacun de ces systèmes. C'est par une "certaine façon de penser" originalement appliquée dans ces préoccupations que nous estimons avoir apporté un progrès.

L'étude qui suit comprend 5 parties :

Dans la première partie, nous présentons la reconnaissance de la parole comme un cas particulier de la reconnaissance des formes. Puis, partant d'une hypothèse concernant le mécanisme de la reconnaissance naturelle, nous introduisons la notion de flou dans la description d'une image.

La deuxième partie est consacrée à la formulation élémentaire de la théorie du flou telle qu'elle a été conçue en 1965 par L.-A. ZADEH [Z 1] et introduite en France par A. KAUFMANN [K 1].

La troisième partie concerne la mise en œuvre de cet algorithme par l'identification à des sous-ensembles flous des formes acoustiques fournies par l'analyseur vocal. L'élaboration d'un indice de similarité permet d'effectuer la reconnaissance de la parole en temps réel grâce au calculateur T 2000 du Laboratoire d'Electricité et d'Automatique. Nous présentons alors quelques résultats.

Dans une quatrième partie, afin d'augmenter le taux de reconnaissance pour un seul locuteur, on agit sur plusieurs paramètres importants comme la segmentation, la modulation de l'indice, etc... , pour étudier leur effet sur le taux de reconnaissance, le temps de réponse et le volume de stockage d'un lexique. On compare alors les résultats obtenus à ceux d'une équipe qui travaille par programmation dynamique avec le même matériel.

Dans la cinquième partie, nous considérons le problème très important de l'adaptation de la méthode à plusieurs locuteurs.

CHAPITRE I

RECONNAISSANCE DES FORMES ET RECONNAISSANCE AUTOMATIQUE DE LA PAROLE

1) *INTRODUCTION.*

Le problème de la reconnaissance de formes se présente continuellement dans la vie courante.

C'est ainsi que dès son plus jeune âge, l'enfant distingue, mémorise et reconnaît un visage familier parmi l'ensemble des visages qu'il rencontre. Il apprend à parler, c'est-à-dire à reproduire une suite de sons, d'images sonores qu'il a perçues et analysées inconsciemment pour les reconnaître. La lecture n'est pour lui que l'identification d'une suite de caractères ou de mots qu'il analyse pour les mémoriser.

Il suffit qu'un enfant ait quelque retard ou handicap dans le développement sensoriel, nerveux ou psychique pour que son entourage prenne conscience de l'importance des problèmes de la reconnaissance des formes dans l'apprentissage du langage et de la lecture par exemple. Les méthodes audio-visuelles ou gestuelles d'éducation se ramènent pour la plupart à des identifications de formes avec apprentissage pour la reconnaissance de celles-ci.

Non seulement les éducateurs, linguistiques et phychologues, mais aussi les économistes, les biologistes, les médecins et plus généralement les personnes qui s'occupent de décisions ou de classifications sont conduits à prendre conscience et à résoudre des problèmes de reconnaissance de formes.

Dans chacune de ces situations, la démarche reste à peu près la même : on analyse d'abord le plus objectivement possible une situation (une forme) en extrayant les paramètres les plus pertinents, mesurés avec les capteurs convenables. Ces paramètres sont mémorisés et servent de référence dans la phase de reconnaissance où une situation nouvelle est associée à une situation déjà recontrée, et identifiée.

S'il faut admettre qu'actuellement l'intelligence humaine n'a pas la possibilité d'expliquer son propre fonctionnement, il n'en reste pas moins que rien n'interdit d'envisager et de mettre en œuvre des processus artificiels dont le but est le même. C'est ainsi que depuis une vingtaine d'années, l'apparition de l'ordinateur a permis de procéder à la résolution automatique de problèmes posés par le fonctionnement de l'intelligence ; c'est le domaine de l'intelligence artificielle qui inclut la reconnaissance de formes.

Avant d'exposer notre contribution à ces préoccupations, nous nous proposons de faire un rapide tour d'horizon de la reconnaissance de formes pour aborder ensuite le cas particulier de la reconnaissance de la parole.

2) *RECONNAISSANCE DES FORMES.*

La reconnaissance de formes telles que les lettres de l'alphabet, les éléments de photographies aériennes, les phonèmes, etc... consiste à attribuer à chacune de ces formes une catégorie de classement.

Des tâches de classement, susceptibles d'être réalisées par une machine, sont données par le tableau non exhaustif suivant.

Tâches	Données	Réponses
Diagnostic médical	- symptômes - électro-encéphalogrammes, etc...	état physiologique
Prévisions du temps	mesures météorologiques	prévisions
Reconnaissance de l'écriture	signal optique	identification de caractères
Reconnaissance de la parole	signal acoustique	identification du phonème ou du mot
Triage de photographies	signal optique	catégories de photographies

TABLEAU 1

Une des difficultés principales de ce problème résulte du manque de définition précise des informations caractéristiques d'une forme réelle, qui sont généralement entachées d'un bruit aléatoire comme le grain pour les photographies ou les bruits parasites pour un son.

A - Système de reconnaissance

Un système de reconnaissance est un dispositif constitué schématiquement de trois étages.

Au premier étage, un ou plusieurs capteurs saisissent les caractères descriptifs de la forme et fournissent l'information d'entrée.

Le choix du capteur est important car il conditionnera le fonctionnement du système de reconnaissance qui sera plus ou moins complexe, rapide ou efficace suivant le capteur utilisé. Ce dernier relèvera les paramètres considérés comme étant les plus pertinents pour la reconnaissance. Malheureusement, les paramètres ne sont pas toujours connus exactement et présentent en outre entre eux certaines corrélations et parfois des redondances. Un des problèmes les plus importants en reconnaissance de formes est justement le choix de ces paramètres.

Le rôle du deuxième étage est de transformer les informations brutes des capteurs, entachées de bruit, en données directement exploitables par l'étage de décision. Ce prétraitement extrait les caractères utiles par un filtrage convenable en éliminant les paramètres redondants ou superflus, et les présente sous un aspect plus simple, souvent en codage binaire.

La décision de classement est prise, au troisième étage, grâce à un algorithme de comparaison à des formes de référence.

Les méthodes de reconnaissance sont très nombreuses [F 1]. Elles sont classées en méthodes analytiques où chaque forme est décrite en fonction de ses composants et des relations entre ces composants, et en méthodes globales où les formes sont présentées ou reconnues dans leur ensemble.

Suivant la description des formes, on distingue encore les méthodes de reconnaissance déterministes et probabilistes.

Pour chacune de ces méthodes le classement s'effectue par l'utilisation d'une ou plusieurs fonctions de discrimination établies à l'aide de paradigmes.

Dans le cas où les formes sont complexes ou mal définies, un processus d'apprentissage permet de construire ces fonctions de discrimination par approches successives.

B - Automatisation de la reconnaissance des formes

Si la reconnaissance des formes à l'échelle humaine est en général intuitive, sinon inconsciente, donc simple en apparence, il n'en est pas de même en intelligence artificielle où le processus passe par les phases distinctes d'analyse et de décision accompagnée éventuellement de l'apprentissage. Le problème se complique encore par l'obligation d'accumuler des données importantes, souvent redondantes mais nécessaires, que le calculateur est incapable d'explorer en simultanéité. Pour réduire le temps de réponse, il est donc indispensable de comprimer les données afin de réaliser un stockage important de formes de référence, tout en gardant la possibilité d'exploiter rapidement ces données.

3) CAS DE LA PAROLE.

Le problème de la reconnaissance de la parole entre parfaitement dans le cadre précédent.

A - Reconnaissance naturelle de la parole

a) Position du problème :

La reconnaissance naturelle de la parole s'effectue quand, prêtant attention à un locuteur, nous captons et nous interprétons l'information transmise sous forme sonore.

Dans la reconnaissance automatique, on se propose de remplacer par un système physique autonome et aussi simple que possible les fonctions naturelles de perception et de compréhension dont il nous paraît nécessaire de définir un modèle.

b) Modèle possible pour la reconnaissance naturelle :

L'audition est organisée pour percevoir et reconnaître les phénomènes acoustiques.

L'information acoustique est d'une richesse inouïe. Or le cerveau humain n'a pas une capacité infinie et il est impensable qu'il puisse stocker intégralement les informations de tous ordres accumulées durant des années. Il serait donc capable de ne conserver que le squelette le plus informatif des formes perçues. Ce filtrage de l'information pourraît résulter de l'aspect autocorrélatif du signal par comparaison entre elles de brèves tranches succesives, ce qui permettrait de déceler des changements éventuels entre tranches, par comparaison de proche en proche. Si, instantanément, toute l'information est disponible, ce processus autocorrélatif supprimerait tout ce qui ne change pas dans un délai donné et qui représente une redondance considérable dans le message perçu.

La redondance, dans la perception, quasiment inutile pour la reconnaissance, serait donc réduite selon le schéma précédent. La difficulté est de connaître la limite de ce dépouillement.

Pour la phase de décision, les images stockées après apprentissage éventuel ne seraient ainsi plus que des squelettes informatifs servant de référence aux signaux à reconnaître à chaque instant. Pour la reconnaissance d'une forme, son image est comparée aux images stockées. Un simple mécanisme de superposition entre deux images suffit pour déterminer les points en coïncidence. Si tous les points sont en coïncidence ou presque, donc les images en corrélation, il y a reconnaissance [L 1].

B - Reconnaissance automatique de la parole

Le point sur l'ensemble des travaux et leur tendance en reconnaissance de la parole a été effectué par J.-P. HATON [H 1] et les expériences les plus significatives en France sont les suivantes :

Les paramètres analytiques du signal sont souvent les fréquences de filtres répartis sur le spectre sonore.

Dans certains cas ce sont la position et l'amplitude des formants ou encore les informations de passage par zéro, avec cependant pour ces cas, de longues durées d'exploitation et de grands besoins en mémoires. Parfois, on ajoute encore des données syntaxiques ou sémantiques.

Les décisions de classement sont prises par des méthodes diverses : perceptron ou matrice d'apprentissage comme dans nos premiers travaux [B 1], [L 2], [V 1], [H 2], programmation dynamique dans des travaux plus récents [H 3].

C'est ainsi que pour la programmation dynamique les chercheurs du C.N.E.T. atteignent un taux de 73 % pour 1000 mots [B 2] et M. MLOUKA et J.-S. LIENARD de 90 % pour 100 mots [M 1]. De son côté, J.-P. HATON, dans notre groupe, reconnaît en temps réel une quarantaine de mots avec un taux voisin de 100 %.

C - Objectifs poursuivis

Les meilleures méthodes donnent ainsi des résultats équivalents en ce qui concerne l'étage de décision dans la mesure où l'on ne se préoccupe ni de l'encombrement en mémoire, ni du temps de calcul. S'il est sûr que la mise en oeuvre de ressources plus importantes ferait progresser le problème à plus ou moins long terme, en ce qui nous concerne, notre objectif est bien davantage de concevoir un système simple et opérationnel en temps réel, sans recourir à des moyens énormes en "matériel et logiciel".

Pour réaliser cet objectif, nous formaliserons la modélisation précédente de la reconnaissance naturelle, dans le cadre de la théorie des sous-ensembles flous.

4) *INTRODUCTION DES SOUS-ENSEMBLES FLOUS.*

Dans les domaines de la théorie de l'information, de la linguistique, dans les problèmes d'apprentissage, en classification et en contrôle automatique, en théorie de la décision, dans les problèmes de reconnaissance de formes et dans beaucoup d'autres domaines d'investigations, les données et les contraintes des problèmes et leurs solutions ne sont pas toujours parfaitement définies.

Les problèmes de reconnaissance des formes, ou plus généralement de décision, souffrent sans doute de l'imperfection des capteurs mais bien davantage de l'imprécision des "contours" (de la définition) de ces formes.

Pour illustrer cette imprécision, considérons les objets de couleur verte dans un ensemble d'objets de couleurs quelconques.

S'il est évident que la couleur verte est distincte du rouge ou du bleu, la notion de vert reste cependant vague ou "floue", pour caractériser la nuance de la couleur d'un objet considéré. On dira que par rapport à un vert bien déterminé de référence, d'autres verts seront comparés grâce à un certain degré d'appartenance dont l'évaluation reste d'ailleurs difficile et subjective. Les objets verts, considérés avec leur degré d'appartenance (éventuellement nul) constituent le sous-ensemble flou des objets verts.

De manière générale, des situations quelconques appréciées par des paramètres, souvent arbitraires et mal évaluables, pourront être considérées comme des sous-ensembles flous et traitées mathématiquement comme tels pour résoudre les problèmes qu'elles posent. Nous verrons que ce traitement est essentiellement caractérisé par l'importance plus grande accordée à la combinaison des taches floues qu'à ces taches elles-mêmes.

C'est ainsi que nous considérerons le problème de la reconnaissance de la parole, en assimilant les formes données par le prétraitement à des sous-ensembles flous auxquels nous appliquerons un traitement de corrélation selon le schéma proposé pour la reconnaissance naturelle.

==========

CHAPITRE II

SOUS-ENSEMBLES FLOUS

1) *INTRODUCTION.*

La théorie des sous-ensembles flous, proche des algèbres multivalentes, constitue une extension de la logique formelle.

Ses concepts de base ont été énoncés par L.-A. ZADEH à partir de 1965 [Z1 à Z4]. Introduits en France en 1969, ils ont été rassemblés par A. KAUFMANN [K 1].

Cette théorie permet de formuler des processus extrêmement variés dont la complexité est incompatible avec une décomposition en éléments simples et indépendants dans les domaines les plus divers : théorie de la décision, taxonomie, reconnaissance de formes, etc...

2) *ENSEMBLE DES PARTIES D'UN ENSEMBLE.*

Considérons deux ensembles A et E, où A est inclus dans E : $A \subset E$.

A tout ensemble E, nous pouvons associer l'ensemble P(E) de toutes ses parties ; nous aurons :

$$P(E) = \{A / A \subset E\}$$

avec en particulier $\begin{cases} \phi \in P(E) \\ E \in P(E) \end{cases}$ où ϕ désigne l'ensemble vide

Si $\overline{A}$ est le complémentaire de A par rapport à E, A U B et $A \cap B$ respectivement la réunion et l'intersection des sous-ensembles A et B ($A \subset E$, $B \subset E$), alors les parties A, B et C de l'ensemble E jouissent des propriétés duales suivantes :

$\forall$ A, B et C $\in$ P(E)

$$\begin{cases} A \cup B = B \cup A \\ A \cap B = B \cap B \end{cases} \quad (1)$$

$$\begin{cases} A \cup (B \cup C) = (A \cup B) \cup C \\ A \cap (B \cap C) = (A \cap B) \cap C \end{cases} \quad (2)$$

$$\begin{cases} A \cup A = A \\ A \cap A = A \end{cases} \quad (3)$$

$$\begin{cases} A \cup (B \cap C) = (A \cup B) \cap (A \cup B) \\ A \cap (B \cup C) = (A \cap B) \cup (A \cap B) \end{cases} \quad (4)$$

$$\begin{cases} \overline{A \cup B} = \overline{A} \cap \overline{B} \\ \overline{A \cap B} = \overline{A} \cup B \end{cases} \quad (5)$$

$$\begin{cases} A \cup \phi = A \\ A \cap E = A \end{cases} \quad (6)$$

$$\begin{cases} A \cap \phi = \phi \\ A \cup E = E \end{cases} \quad (7)$$

$$\{ \overline{\overline{A}} = A \quad (8)$$

$$\begin{cases} A \cup \overline{A} = E \\ A \cap \overline{A} = \phi \end{cases} \quad (9)$$

Ces propriétés caractérisent une Algèbre de Boole.

3) *NOTIONS DE LA THEORIE DES TREILLIS.*

A - Ensembles partiellement ordonnés

Un ensemble P partiellement ordonné est un ensemble sur lequel est défini un ordre, c'est-à-dire une relation binaire, réflexive, antisymétrique et transitive. En notant par $\geqslant$ la relation d'ordre nous avons :

$$x \leqslant y \quad \text{ssi} \quad y \geqslant x \qquad \forall\, x, y \in P \tag{10}$$

$$x < y \quad \text{ssi} \quad x \leqslant y \text{ et } x \neq y \qquad \forall\, x, y \in P \tag{11}$$

$$x > y \quad \text{ssi} \quad x \geqslant y \text{ et } x \neq y \qquad \forall\, x, y \in P \tag{12}$$

Si x et y sont les éléments de P, nous dirons que x couvre y si $x > y$, et s'il n'existe pas d'éléments z tel que : $x > z > y$.

En représentant les éléments de P par des points dans le plan, de telle sorte que x soit plus haut que y lorsque $x > y$, et en dessinant le segment $[x , y]$ lorsque x couvre y, nous obtenons le diagramme de Hasse de P.

Deux ensembles partiellement ordonnés sont isomorphes s'ils sont représentés par le même diagramme de Hasse.

S'il existe, dans P, un élément O tel que $x \geqslant O$, $\forall\, x \in P$, cet élément est l'élément nul de P.

On déduit de l'antisymétrie de la relation $\geqslant$ qu'un ensemble partiellement ordonné n'a qu'un élément nul au plus.

S'il existe dans P un élément I tel que $I \geqslant x$, $\forall x \in P$, cet élément est l'élément universel de P et il est unique s'il existe.

Un élément x de P est majorant des sous-ensembles Q de P si $x \geqslant y$, $\forall y \in Q$.

Un élément x de P est un minorant de Q si $y \geqslant x$, $\forall y \in Q$.

La borne supérieure de Q est le plus petit majorant de Q ; elle est unique si elle existe.

La borne inférieure de Q est le plus grand minorant de Q ; elle est unique si elle existe.

B - Treillis

a) Définition ; propriétés fondamentales :

Un treillis L est un ensemble partiellement ordonné tel que toute paire d'éléments {x, y} admet une borne supérieure et une borne inférieure notées respectivement par $x \nabla y$ et $x \Delta y$.

Tout treillis L jouit des six propriétés suivantes :

$\forall x, y$ et $z \in L$

$$
\begin{cases}
1) & x \nabla y = y \nabla x & (13) \\
2) & x \Delta y = y \Delta x & (14) \\
3) & x \nabla (y \nabla z) = (x \nabla y) \nabla z & (15) \\
4) & x \Delta (y \Delta z) = (x \Delta y) \Delta z & (16) \\
5) & x \nabla (x \Delta y) = x & (17) \\
6) & x \Delta (x \nabla y) = x & (18)
\end{cases}
$$

b) Treillis modulaire, distributif, complémenté :

Définition 1. Le treillis L est un treillis modulaire s'il vérifie l'implication :

$$x > z \Rightarrow x \Delta (y \nabla z) = (x \Delta y) \nabla z, \forall x,y,z \in L .$$

Définition 2. Le treillis L est un treillis distributif si les lois ∇ et Δ se distribuent mutuellement :

$$x \nabla (y \Delta z) = (x \nabla y) \Delta (x \nabla z) , \forall x,y,z \in L \quad (19)$$

$$x \Delta (y \nabla z) = (x \Delta y) \nabla (x \Delta z) , \forall x,y,z \in L \quad (20)$$

Définition 3. Si L est un treillis avec un élément nul O et un élément universel I, et si x et y sont deux éléments de L tels que $x \Delta y = 0$ et $x \nabla y = I$, alors y est un complément de x et y est un complément de x.

Théorème : Dans un treillis distributif, si un élément a un complément, celui-ci est unique : $\bar{x}$.

c) Algèbre de Boole :

Une algèbre de Boole est un treillis distributif et complémenté.

Ses propriétés duales sont :

$\forall x,y$ et $z \in L$

$$\begin{cases} x \nabla y = y \nabla x & (21) \\ x \Delta y = y \Delta x & (22) \end{cases}$$

$$\begin{cases} x \nabla (y \nabla z) = (x \nabla y) \nabla z & (23) \\ x \Delta (y \Delta z) = (x \Delta y) \Delta z & (24) \end{cases}$$

$$\begin{cases} x \nabla x = x & (25) \\ x \Delta x = x & (26) \end{cases}$$

$$\begin{cases} x \nabla (y \Delta z) = (x \nabla y) \Delta (x \nabla z) & (27) \\ x \Delta (y \nabla z) = (x \Delta y) \nabla (x \Delta z) & (28) \end{cases}$$

$$\begin{cases} \overline{x \nabla y} = \overline{x} \Delta \overline{y} & (29) \\ \overline{x \Delta y} = \overline{x} \nabla \overline{y} & (30) \end{cases}$$

$$\begin{cases} x \nabla 0 = x & (31) \\ x \Delta I = x & (32) \end{cases}$$

$$\begin{cases} x \nabla I = I & (33) \\ x \Delta 0 = 0 & (34) \end{cases}$$

$$\begin{cases} x \nabla \overline{x} = I & (35) \\ x \wedge \overline{x} = 0 & (36) \end{cases}$$

$$\{ \overline{(\overline{x})} = x \quad (37)$$

Exemple : La comparaison des propriétés (1 à 9) et (21 à 37) montre que l'ensemble des parties de E, P(E), ordonné par l'inclusion et où la borne supérieure et la borne inférieure d'une paire {A, B} sont respectivement A U B et A ∩ B, est une algèbre de Boole.

Exemple : L'ensemble des réels que constitue l'intervalle {0,1} ordonné par la relation $\leqslant$, et où la borne supérieure et la borne inférieure d'une paire {x,y} sont respectivement max(x,y) et min(x,y), est un treillis distributif, qui possède un élément nul et un élément universel, mais qui n'est pas complémenté.

En effet :

$\forall x,y \text{ et } z \in [0,1]$

$$\begin{cases} \max(x,y) = \max(y,x) & (38) \\ \min(x,y) = \min(y,x) & (39) \end{cases}$$

$$\begin{cases} \max[x,\max(y,z)] = \max[\max(x,y),z] & (40) \\ \min[x,\min(y,z)] = \min[\min(x,y),z] & (41) \end{cases}$$

$$\begin{cases} \max[x,\min(x,y)] = x & (42) \\ \min[x,\max(x,y)] = x & (43) \end{cases}$$

Il résulte alors des relations (13 à 18) que $[0,1]$, $\leqslant$, max,min est un treillis.

La relation (40) par exemple implique que le treillis considéré est distributif.

Enfin $x \geqslant 0 \quad \forall x \in [0,1]$ et (44)

$x \leqslant 1 \quad \forall x \in [0,1]$ (45)

donne 0 et 1 respectivement pour l'élément nul et l'élément universel de ce treillis.

Pour que le treillis soit complémenté, il faudrait que

$$\forall\, x \in [\underline{0},1] \quad , \quad \exists\, \bar{x} \in [\underline{0},1] : \max(x,\bar{x}) = 1 \quad \text{et} \quad \min(x,\bar{x}) = 0 \ ,$$

ce qui n'est évidemment pas réalisé ici.

C - Valuation ; treillis métrique ; distance

Définition : Une fonction réelle v, définie sur un treillis L, est une variation de $\forall x, y \in L$

$$v(x) + v(y) = v(x \nabla y) + v(x \Delta y) \qquad (46)$$

Elle est positive ssi

$$x > y \implies v(x) > v(y).$$

Définition : Un treillis métrique est un treillis sur lequel est définie une valuation positive.

Définition : Soit E un ensemble et d une application de E x E dans R^+ telle que

$$\begin{aligned} &1)\ \forall x, y \in E : d(x,y) = 0 \ \text{ ssi } x = y \ ; \\ &2)\ \forall x, y \in E : d(x,y) = d(y,x) \qquad (47) \\ &3)\ \forall x,y,z \in E : d(x,z) \leqslant d(x,y) + d(y,z) \end{aligned}$$

d est alors appelée une distance sur E.

4) *SOUS-ENSEMBLES FLOUS*.

A - Définitions

Définition 1. Soit E un ensemble et A un sous-ensemble non vide de E ; alors un sous-ensemble flou $\underset{\sim}{A}$ dans E est un couple $(E, \mu_{\underset{\sim}{A}})$ où $\mu_{\underset{\sim}{A}}$, application de E dans $[0,1]$, est la fonction d'appartenance de $\underset{\sim}{A}$.

Si $e \in E$, alors $\mu_{\underset{\sim}{A}}(e)$ est le degré d'appartenance de l'élément e au sous-ensemble flou $\underset{\sim}{A}$.

Dans le cas particulier où $\mu_{\underset{\sim}{A}} = \{0,1\}$, alors $\mu_{\underset{\sim}{A}}$ est la fonction caractéristique d'un sous-ensemble ordinaire A de E. En effet

$$\mu_{\underset{\sim}{A}}(e) = 0 \quad \Rightarrow \quad e \notin A$$
$$\mu_{\underset{\sim}{A}}(e) = 1 \quad \Rightarrow \quad e \in A$$

Définition 2. Le sous-ensemble flou $\underset{\sim}{A}$ est vide si et seulement si

$$\mu_{\underset{\sim}{A}}(e) = 0 \qquad \forall e \in E.$$

Définition 3. Deux sous-ensembles flous $\underset{\sim}{A}$ et $\underset{\sim}{B}$ dans E sont égaux si et seulement si $\mu_{\underset{\sim}{A}}(e) = \mu_{\underset{\sim}{B}}(e) \quad \forall e \in E$.
On note $\underset{\sim}{A} = \underset{\sim}{B}$.

Définition 4. Un sous-ensemble flou $\underset{\sim}{A}$ dans E est inclus dans le sous-ensemble flou $\underset{\sim}{B}$ dans E si et seulement si

$$\mu_{\underset{\sim}{B}}(e) \geq \mu_{\underset{\sim}{A}}(e) \qquad \forall e \in E \ ;$$

On dira aussi que $\underset{\sim}{A}$ est plus petit que $\underset{\sim}{B}$, ou que $\underset{\sim}{B}$ est plus grand que $\underset{\sim}{A}$.

Définition 5. Le complément de l'ensemble flou $\underset{\sim}{A}$ dans E est l'ensemble flou $\underset{\sim}{\bar{A}}$ dans E défini par :

$$\mu_{\underset{\sim}{\bar{A}}}(e) = 1 - \mu_{\underset{\sim}{A}}(e) \qquad \forall e \in E.$$

Définition 6. La réunion de deux sous-ensembles flous $\underset{\sim}{A}$ et $\underset{\sim}{B}$ dans E est l'ensemble flou $\underset{\sim}{A} \cup \underset{\sim}{B}$ dans E, défini par :

$$\mu_{\underset{\sim}{A} \cup \underset{\sim}{B}}(e) = \max\left[\mu_{\underset{\sim}{A}}(e), \mu_{\underset{\sim}{B}}(e)\right] \quad , \quad \forall e \in E.$$

Définition 7. L'intersection de deux sous-ensembles flous $\underset{\sim}{A}$ et $\underset{\sim}{B}$ dans E est l'ensemble flou $\underset{\sim}{A} \cap \underset{\sim}{B}$ dans E, défini par :

$$\mu_{\underset{\sim}{A} \cap \underset{\sim}{B}}(e) = \min\left[\mu_{\underset{\sim}{A}}(e), \mu_{\underset{\sim}{B}}(e)\right] \quad , \quad \forall e \in E.$$

Définition 8. Si $\underset{\sim}{A}_i$ est un sous-ensemble flou dans E, $\forall i \in I$, où I est un ensemble fini ou infini, alors $\bigcup_{i \in I} \underset{\sim}{A}_i$ est l'ensemble flou dans E défini par :

$$\mu_{\bigcup_i \underset{\sim}{A}_i}(e) = \sup_{i \in I} \mu_{\underset{\sim}{A}_i}(e) \qquad \forall e \in E.$$

De manière analogue, $\bigcap_{i \in I} \underset{\sim}{A}_i$ est l'ensemble flou dans E défini par :

$$\mu_{\bigcap_i \underset{\sim}{A}_i}(e) = \inf_{i \in I} \mu_{\underset{\sim}{A}_i}(e) \qquad \forall e \in E.$$

Définition 9. Si $E = \{e_1, \ldots, e_n\}$ est un ensemble fini et si $\underset{\sim}{A}$ est un sous-ensemble flou dans E, alors le cardinal de $\underset{\sim}{A}$ est :

$$|\underset{\sim}{A}| = \sum_{i=1}^{i=n} \mu_{\underset{\sim}{A}}(e_i)$$

Définition 10. Le produit algébrique dans E de deux sous-ensembles flous $\underset{\sim}{A}$ et $\underset{\sim}{B}$ est défini par :

$$\mu_{\underset{\sim}{A}.\underset{\sim}{B}}(e) = \mu_{\underset{\sim}{A}}(e) \cdot \mu_{\underset{\sim}{B}}(e) \qquad \forall e \in E.$$

Définition 11. La somme algébrique dans E de deux sous-ensembles flous $\underset{\sim}{A}$ et $\underset{\sim}{B}$ est défini par :

$$\mu_{\underset{\sim}{A}+\underset{\sim}{B}}(e) = \mu_{\underset{\sim}{A}}(e) + \mu_{\underset{\sim}{B}}(e) - \mu_{\underset{\sim}{A}}(e) \cdot \mu_{\underset{\sim}{B}}(e) \qquad \forall e \in E$$

B - Propriétés

On a les propositions suivantes :

α) $\underset{\sim}{A} \cup \underset{\sim}{B}$ est le plus petit sous-ensemble flou contenant $\underset{\sim}{A}$ et $\underset{\sim}{B}$.

β) $\underset{\sim}{A} \cap \underset{\sim}{B}$ est le plus grand sous-ensemble flou contenu dans $\underset{\sim}{A}$ et $\underset{\sim}{B}$.

Démontrons α) par exemple.

Soit $\underset{\sim}{C}$ un sous-ensemble flou dans E, contenant $\underset{\sim}{A}$ et $\underset{\sim}{B}$; il résulte de la définition 4 que

$$\mu_{\underset{\sim}{C}}(e) \geq \mu_{\underset{\sim}{A}}(e) \quad , \quad \forall e \in E$$

et

$$\mu_{\underset{\sim}{C}}(e) \geq \mu_{\underset{\sim}{B}}(e) \quad , \quad \forall e \in E$$

d'où $$\mu_{\underset{\sim}{C}}(e) \geq \max\left[\mu_{\underset{\sim}{A}}(e) , \mu_{\underset{\sim}{B}}(e)\right] \quad , \forall e \in E$$

d'où $(\underset{\sim}{A} \cup \underset{\sim}{B}) \subset \underset{\sim}{C}$ selon la définition 6.

C - Exemples : cas de fonctions d'appartenance à éléments discrets

Dans E, considérons les sous-ensembles flous

$$\underset{\sim}{A} = \{(e_1 ; 0,7) , (e_2 ; 0,8) , (e_3 ; 0,9)\}$$

et $$\underset{\sim}{B} = \{(e_1 ; 0,8) , (e_2 ; 0,7) , (e_3 ; 0,9)\}$$

On a pour eux :

$$\underset{\sim}{A} \cup \underset{\sim}{B} = \{(e_1 ; 0,8) , (e_2 ; 0,8) , (e_3 ; 0,9)\}$$

$$\underset{\sim}{A} \cap \underset{\sim}{B} = \{(e_1 ; 0,7) , (e_2 ; 0,7) , (e_3 ; 0,9)\}$$

$$\overline{\underset{\sim}{A}} = \{(e_1 ; 0,3) , (e_2 ; 0,2) , (e_3 ; 0,1)\}$$

Cas de fonctions d'appartenance non discrètes

Un exemple de sous-ensemble flou est le sous-ensemble $\underset{\sim}{A}$ des nombres réels "suffisamment plus grands" que 10 ; on pourra prendre pour fonction d'appartenance

$$\mu_{\underset{\sim}{A}}(e) = \left(1 + \frac{10^3}{(e-10)^2}\right)^{-1} \qquad \forall e \in R$$

D - Structure de l'ensemble des sous-ensembles flous dans E

Si $\underset{\sim}{A}$, $\underset{\sim}{B}$ et $\underset{\sim}{C}$ sont des sous-ensembles flous dans E avec $e \in E$, en notant $x = \mu_{\underset{\sim}{A}}(e)$, $y = \mu_{\underset{\sim}{B}}(e)$ et $z = \mu_{\underset{\sim}{C}}(e)$, on a pour x, y, z $\in [0,1]$:

Commutativité :

$$\begin{cases} \max (x,y) = \max (y,x) \\ \min (x,y) = \min (y,x) \end{cases} \qquad (48)$$

Associativité :

$$\begin{cases} \max\left[x,\max(y,z)\right] = \max\left[\max(x,y),z\right] \\ \min\left[x,\min(y,z)\right] = \min\left[\min(x,y),z\right] \end{cases} \quad 49)$$

Idempotence :

$$\begin{cases} \max\left[x,x\right] = x \\ \min\left[x,x\right] = x \end{cases} \quad (50)$$

Distributivité :

$$\begin{cases} \max\left[x,\min(y,z)\right] = \min\left[\max(x,y),\max(y,z)\right] \\ \min\left[x,\max(y,z)\right] = \max\left[\min(x,y),\min(y,z)\right] \end{cases} \quad (51)$$

Lois de De Morgan :

$$\begin{cases} 1-\max(x,y) = \min\left[(1-x),(1-y)\right] \\ 1-\min(x,y) = \max\left[(1-x),(1-y)\right] \end{cases} \quad (52)$$

Existence d'un élément nul 0 et d'un élément universel 1 :

$$\begin{cases} \max(0,x) = x \\ \min(1,x) = x \end{cases} \quad (53)$$

et

$$\begin{cases} \max(x,1) = 1 \\ \min(x,0) = 0 \end{cases} \quad (54)$$

Involution :

$$\{1 - (1-x) = x \quad (55)$$

et

$$\begin{cases} \max(x,1-x) \neq 1 \\ \min(x,1-x) \neq 0 \end{cases} \quad (56)$$

Un sous-ensemble flou est donc un treillis distributif avec un élément nul 0 et un élément universel 1, mais qui n'est pas complémenté.

Le second exemple de la page 18 constitue un tel sous-ensemble.

E - Distances entre sous-ensembles flous ; ensemble ordinaire le plus proche ; mesure de flou

a) Distance de Hamming :

Le degré d'appartenance $\mu_{\underset{\sim}{A}}(e)$ constitue une valuation pour le sous-ensemble flou $\underset{\sim}{A}$.

Pour deux sous-ensembles flous $\underset{\sim}{A}$ et $\underset{\sim}{B}$

$$\delta(\underset{\sim}{A},\underset{\sim}{B}) = \frac{1}{n} \sum_{i=1}^{n} \left[\mu_{\underset{\sim}{A}}(e_i) - \mu_{\underset{\sim}{B}}(e_i) \right] \qquad (57)$$

est une distance selon le paragraphe D.

En effet, on a :

$$\delta(\underset{\sim}{A},\underset{\sim}{B}) = 0$$

$$\delta(\underset{\sim}{A},\underset{\sim}{B}) = \delta(\underset{\sim}{B},\underset{\sim}{A})$$

$$\delta(\underset{\sim}{A},\underset{\sim}{B}) + \delta(\underset{\sim}{B},\underset{\sim}{C}) = \frac{1}{n} \sum_{i=1}^{n} |\mu_{\underset{\sim}{A}}(e_i) - \mu_{\underset{\sim}{B}}(e_i)| + \sum_{i=1}^{n} |\mu_{\underset{\sim}{B}}(e_i) - \mu_{\underset{\sim}{C}}(e_i)|$$

$$= \frac{1}{n} \sum_{i=1}^{n} |\mu_{\underset{\sim}{A}}(e_i) - \mu_{\underset{\sim}{B}}(e_i)| + |\mu_{\underset{\sim}{B}}(e_i) - \mu_{\underset{\sim}{C}}(e_i)|$$

$$\geq \frac{1}{n} \sum_{i=1}^{n} |\mu_{\underset{\sim}{A}}(e_i) - \mu_{\underset{\sim}{C}}(e_i)| = \delta(\underset{\sim}{A},\underset{\sim}{C})$$

La distance $\delta(\underset{\sim}{A},\underset{\sim}{B})$ est la "distance floue" de Hamming entre les sous-ensembles flous $\underset{\sim}{A}$ et $\underset{\sim}{B}$.

b) Distance euclidienne :

On définit de façon analogue la distance floue euclidienne :

$$e(\underset{\sim}{A},\underset{\sim}{B}) = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^{n} (\mu_{\underset{\sim}{A}}(e_i) - \mu_{\underset{\sim}{B}}(e_i))^2} \qquad (58)$$

c) Ensemble ordinaire le plus proche :

L'ensemble $A_o \subset E$ qui a pour fonction d'appartenance

$$\mu_{A_o}(e) = \begin{cases} 0 & \text{si } \mu_{\underset{\sim}{A}}(e) \leq 0,5 \\ 1 & \text{si } \mu_{\underset{\sim}{A}}(e) > 0,5 \end{cases} \qquad (59)$$

est l'ensemble ordinaire le plus proche du sous-ensemble flou $\underset{\sim}{A}$.

d) Indice de flou :

Un ensemble flou est d'autant plus flou qu'il est éloigné de son ensemble ordinaire le plus proche. On définit l'indice de flou de $\underset{\sim}{A}$ par :

$$\upsilon(\underset{\sim}{A}) = 2\ \delta(\underset{\sim}{A},A_o) \qquad (60)$$

On a :

$$\left.\begin{array}{l} 0 \leq \upsilon(A) \leq 1 \\ \upsilon(\underset{\sim}{A}) = 0 \quad \text{ssi} \quad A = A_o \\ \upsilon(\underset{\sim}{A}) = 1 \quad \text{ssi} \quad \mu_{\underset{\sim}{A}}(e_i) = 0,5 \\ \upsilon(\underset{\sim}{A}) = \upsilon(\overline{\underset{\sim}{A}}) \end{array}\right\} \forall e_i \in E$$

On pourrait tout aussi bien définir le flou par l'indice quadratique

$$\mu(\underset{\sim}{A}) = 1 - 2\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\mu_{\underset{\sim}{A}}(e_i) - 0,5\right)^2} \qquad (61)$$

F - Sous-ensembles flous et probabilités

Si la théorie des sous-ensembles flous et celle des probabilités ont pour objet les phénomènes qui présentent un certain degré d'indétermination, l'analogie s'arrête là, car la formulation des problèmes et leur résolution sont complètement différentes.

En termes de probabilité, les données qui résultent d'un grand nombre de mesures suivent des lois de distribution, d'histogrammes, etc... Un problème est résolu en fonction de ces lois.

Avec les sous-ensembles flous, les données, éléments et degrés d'appartenance correspondants résultent parfois d'un choix subjectif guidé ou non par l'expérience, ou comme nous le ferons, de la mesure d'une seule occurrence. La résolution des problèmes s'effectue par des considérations d'identité, de voisinage, de complémentarité, etc...

L'analogie entre la probabilité et le degré d'appartenance, souvent d'ailleurs fixé par des appréciations statistiques, n'est qu'apparente, les degrés d'appartenance n'ayant pas 1 pour somme, sans compter que leur domaine d'application n'est pas forcément [0,1] (Cf. [G 1]).

Les paramètres mal définis utilisés dans la reconnaissance de la parole seront envisagés dans la suite, d'une manière floue, qui s'apparente à la reconnaissance naturelle dans le modèle proposé au chapitre I. Le chapitre qui s'achève donne les bases nécessaires à la poursuite de ce travail.

==========

CHAPITRE III

FORMES ACOUSTIQUES RECONNUES COMME DES SOUS-ENSEMBLES FLOUS

1) *SIGNAL VOCAL.*

Le signal à traiter provient de l'appareil phonatoire humain constitué par un oscillateur (le larynx) et des résonateurs (cavités buccales, laryngée et nasales), émettant la voix qui devient parole lorsqu'elle est chargée de signification [L 3] .

Les sons vocaux, c'est-à-dire les voyelles et les consonnes sonores, sont obtenus par l'action des cordes vocales, qui transforment en train d'impulsions périodiques le flux d'air provenant des poumons. Les consonnes résultent d'une ouverture brusque d'une cavité ou d'une striction, accompagnée ou non d'un voisement [H 4] , [V 2] .

La parole est formée de phonèmes dont il est parfois intéressant d'étudier des groupements de deux ou trois ; E. LEIPP utilise ainsi les diphonèmes comme éléments de base et souligne le caractère informant des transitions entre phonèmes.

La parole, succession de régimes soutenus et transitoires, se présente comme un signal complexe dont la représentation usuelle est celle de l'amplitude dans l'espace temps-fréquence.

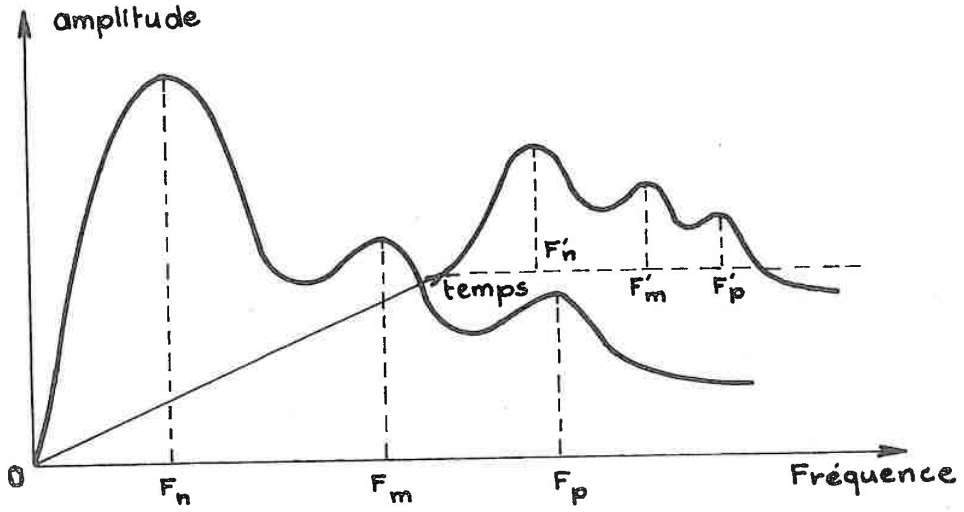


Figure 1.

Sur de telles représentations (figure 1), on note les zones de fréquences, ou formants (F_n, F_m, F_p), où l'amplitude est maximale. Ces formants qui correspondent aux fréquences de résonance des cavités vocales, c'est-à-dire à leur forme à l'instant considéré, se déplacent et se déforment dans le temps.

Si, sur le plan fréquence-temps, on projette l'amplitude, celle-ci correspond à un noircissement plus ou moins important dans le sonagramme ainsi obtenu.

2) *FORMES ACOUSTIQUES ; DESCRIPTION EN TERMES DE FLOU.*

La forme acoustique à reconnaître provient de l'action sur le signal vocal précédent d'un analyseur codeur qui fournit,100 fois par seconde,un échantillon codé sous forme binaire à la sortie de 25 filtres répartis logarithmiquement dans le spectre des fréquences. Le codage donne 1 si la sortie d'un filtre est supérieure à la sortie du filtre précédent,et 0 dans le cas contraire et du niveau inférieur à un certain seuil ; il fonctionne donc en amplitudes relatives, ce qui rend le système indépendant de la dynamique tout en réduisant l'information [D 1] .

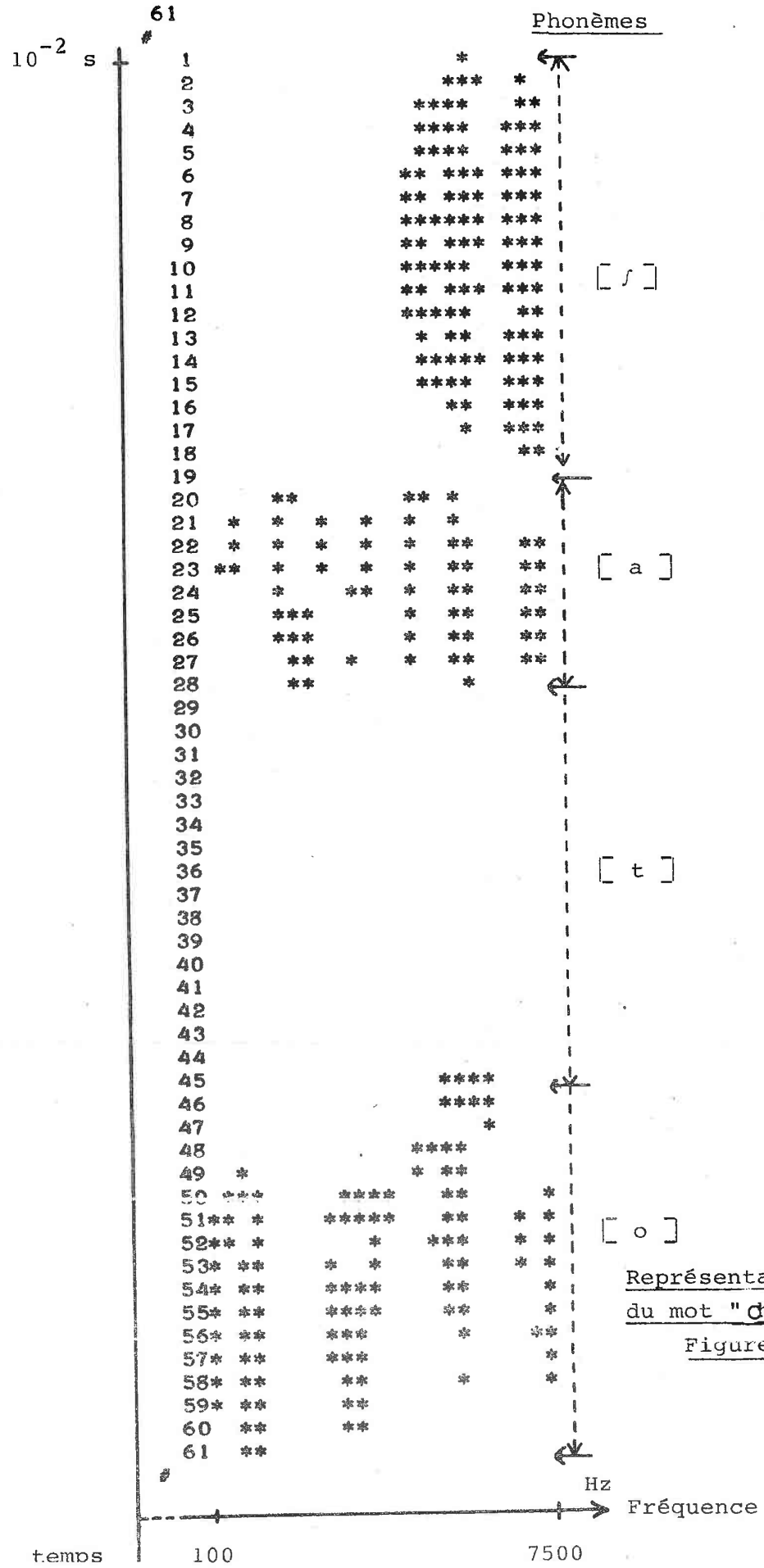
On obtient ainsi, avec un échantillonnage suffisamment rapide, une forme très analogue à un sonagramme (figure 2).

Comme le montre l'expérience, une telle forme n'est définie qu'"à peu près", tant pour sa durée et pour la position temporelle des groupes de 1, que pour la densité des 1 et leur position fréquentielle.

Ce caractère très perceptible déjà pour un même locuteur (bien évidemment toujours "en élocution normale") est particulièrement accentué pour des locuteurs différents, comme le montrent les exemples de la figure 3, relatifs au mot "Assez" prononcé par les locuteurs B et L.

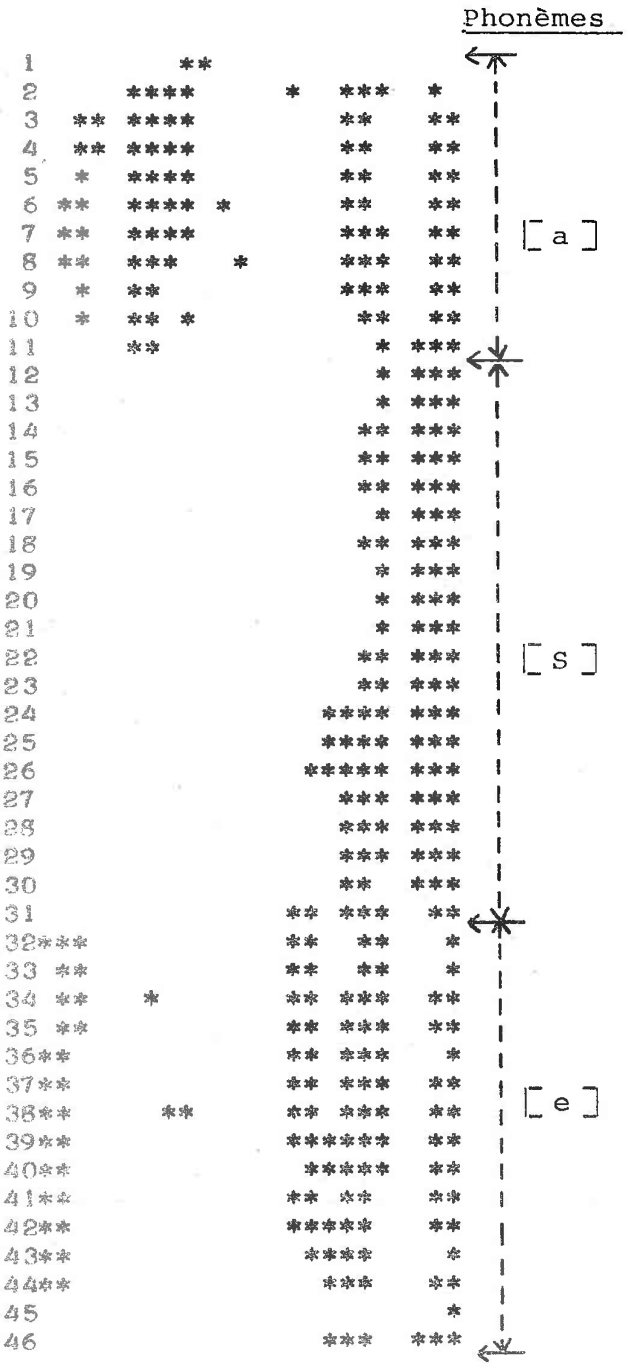
Cette circonstance constitue une des plus grandes difficultés de la reconnaissance automatique.

En ce qui concerne la durée, si elle varie entre de larges limites (de 30 à 50 %), son caractère fondamental, toujours vérifié, est qu'en valeur relative, la position des différentes zones temporelles caractéristiques d'un mot reste invariante "grosso-modo" comme le montre la figure 3.

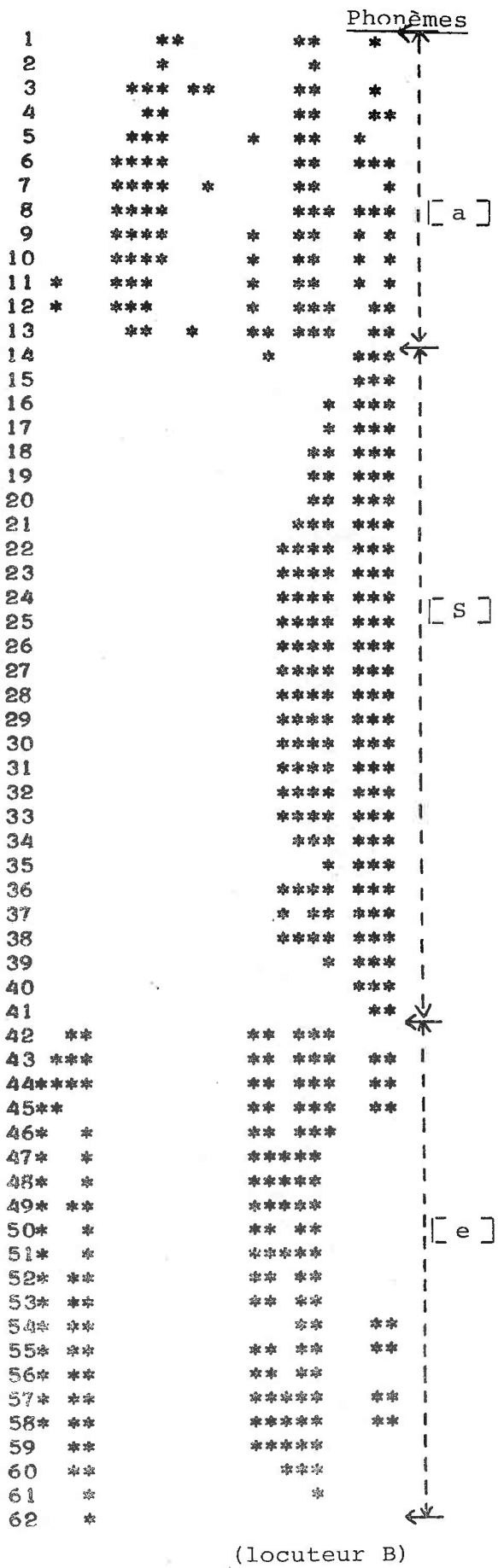


Représentation binaire du mot "chateau"

Figure 2.



(locuteur L)



(locuteur B)

Figure 3.

mot "Assez" ([a] [s] [e]).

Pour la pratique de la reconnaissance qui se fait par référence à des mots mémorisés, le caractère précédent est fondamental pour la normalisation des durées, condition évidemment nécessaire pour simplifier la comparaison.

Bien entendu, comme les limites des zones temporelles ne sont pas parfaitement définies, l'homothétie de normalisation est une "homothétie floue", au sens de peu précise.

En ce qui concerne la position des 1 dans les groupements, on remarque que la considération de bandes de fréquences en nombre inférieur aux 24 fréquences d'analyse, donne une approximation valable pour l'aspect général de ces groupements. Dans l'exemple de la figure 3, il semble que la prise en compte de trois zones fréquentielles (basse, moyenne, haute) soit suffisante, mais les résultats de nos expériences ont montré qu'il est préférable d'en prendre quatre.

On obtient ainsi une diminution importante de l'information à traiter, sans pour autant affecter notablement les formes imprécises de départ.

En reconnaissance automatique, la décomposition précédente en segment, effectuée à vue, doit se faire automatiquement et en temps réel.

3) SEGMENTATION AUTOMATIQUE.

La segmentation automatique résulte de l'usage simultané de deux algorithmes indépendants en parallèle, appliqués à la forme vocale binaire.

Pour le premier, la distinction entre les régimes soutenus et transitoires s'obtient par la mesure de la dissemblance de deux échantillons successifs.

Soit $\vec{X}_t$ le vecteur dont les composantes $X_t(i)$ sont les 24 bits de l'échantillon à l'instant t.

Nous appelons transition $n_t(i)$ le changement d'état de $X_t(i)$ à l'instant t :

$$n_t(i) = X_{t-1}(i) \oplus X_t(i) \qquad (62)$$

La quantité $n_t = \sum_{i=1}^{i=24} n_t(i)$ mesure le nombre de transitions entre deux échantillons successifs.

$n_t = 0$ caractérise donc un régime soutenu

et $n_t \neq 0$ un régime transitoire.

Comme en réalité les deux régimes ne se différencient pas aussi catégoriquement avec des fluctuations d'importance croissante avec le nombre de 1, on s'accorde, avec un seuil S_o, une marge d'erreur sur n_t qu'en outre on lie au nombre N_t des "1" sur un échantillon.

En fait, pour diminuer encore l'incidence des fluctuations et rendre plus franches les transitions, n_t et N_t sont calculés en cumulant les résultats pour les paires successives d'échantillons X_{t-2} et X_{t-1} , X_{t-1} et X_t.

Pour test de décision du transitoire, nous prendrons :

$$n_t \geqslant K.N_t + S_o \qquad (63)$$

Après quelques essais, nous avons retenu les valeurs :

$$\begin{cases} K = 0,3 \\ S_o = 1,2 \end{cases}$$

qui donnent en moyenne la concordance la meilleure avec la segmentation à vue.

En fait, comme le calcul a lieu en valeur entière, le critère s'écrit en pratique avec les remarques précédentes

$$10\ (TR1 + TR2) \geqslant 3\ (N1 + N2 + N3) + 12 \qquad (64)$$

où N3 = nombre de "1" dans l'échantillon pris au temps t
N2 = nombre de "1" dans l'échantillon pris au temps t-1
N1 = nombre de "1" dans l'échantillon pris au temps t-2
TR1 = distance de Hamming entre les échantillons t et t-1
TR2 = distance de Hamming entre les échantillons t-1 et t-2.

Sur la figure 4a et b, relative au mot "Charité", on détaille l'application de ce test à chaque échantillon qui reçoit, selon le cas, l'étiquette soutenu (S) ou transitoire (T).

On prend alors comme point de segmentation l'instant de passage entre un groupe de S et un groupe de T pour ranger dans un tableau TSG1 les numéros des échantillons où il y a segmentation.

Les segmentations parasites dues au mode même de calcul qui apparaissent après un silence, sont clairement

SEGMENTATION DU MOT "Charité" (début)

N_t	$N_1+N_2+N_3$	n_t	$(TR_1+TR_2)\times 10$	$(3\times \Sigma N)+12$			TSG1	TSG2	TSG3
3	3	3	30	21	T	1	S	S	
5	8	2	50	36	T	2			
5	13	0	20	51	S	3			
5	15	0	0	57	S	4			
6	16	1	10	60	S	5			
7	18	1	20	66	S	6			
6	19	1	20	69	S	7			
7	20	1	20	72	S	8			F
6	19	3	40	69	S	9			
7	20	1	40	72	S	10			
9	22	2	30	78	S	11			
8	26	1	30	90	S	12			
9	26	1	20	90	S	13			
9	26	0	10	90	S	14			
9	27	0	0	93	S	15			
6	24	3	30	84	S	16			
7	22	3	60	78	S	17			
4	17	3	60	63	S	18			
3	14	1	40	54	S	19			
1	8	2	30	36	S	20	S		F
3	7	2	40	33	T	21		S	
7	11	6	80	45	T	22		V	
9	19	2	80	69	T	23			
5	21	6	80	75	T	24			
7	21	2	80	75	T	25			
9	21	8	100	75	T	26			
11	27	2	100	93	T	27	S		
10	30	3	50	102	S	28			
12	33	3	60	111	S	29			
9	31	3	60	105	S	30			
10	31	2	50	105	S	31			
4	23	4	60	81	S	32			
4	18	2	60	66	S	33	S	V	F
0	8	4	60	36	T	34		P	
0	0	0	40	12	T	35	S		
0	0	0	0	12	S	36			

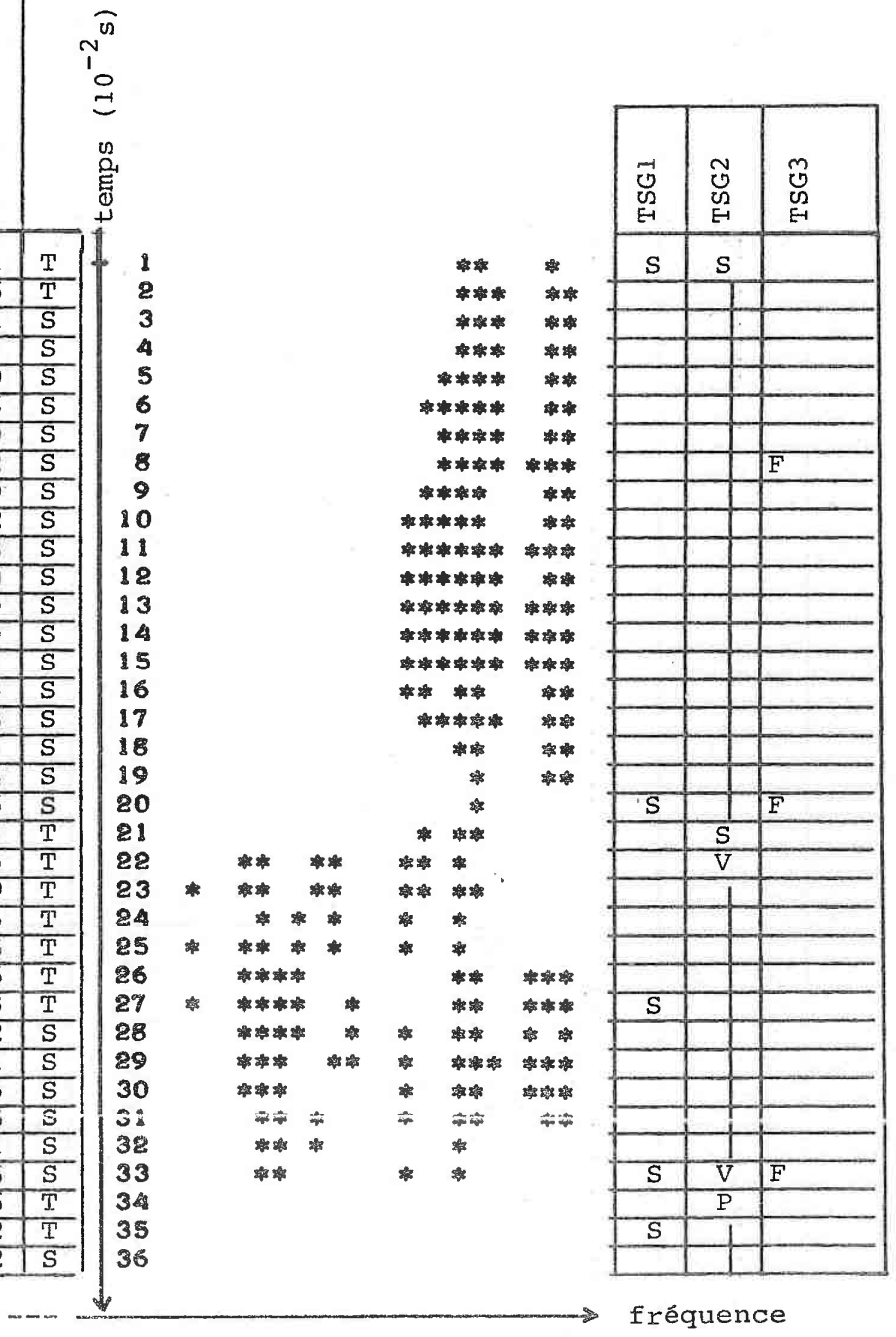


Figure 4 a

N_t	$N_1+N_2+N_3$	n_t	$(TR_1+TR_2)x10$	$(3x\ \Sigma N) + 12$		temps $(10^{-2}s)$		TSG1	TSG2	TSG3
0	0	0	0	12	S	37				
0	0	0	0	12	S	38 S		S	P	F
3	3	3	30	21	T	39	***		S	
4	7	1	40	33	T	40	****	S		
5	12	1	20	48	S	41	*****			
4	13	1	20	51	S	42	****			
5	14	1	20	54	S	43	*****			
5	14	0	10	54	S	44	*****			
4	14	1	10	54	S	45	****			
4	13	0	10	51	S	46	****			
4	12	0	0	48	S	47	****			
2	10	2	20	42	S	48	**	S	S	F
0	6	2	40	30	T	49				
0	2	0	20	18	S	50				
0	0	0	0	12	S	51				
0	0	0	0	12	S	52				
0	0	0	0	12	S	53				
0	0	0	0	12	S	54				
0	0	0	0	12	S	55				
0	0	0	0	12	S	56				
0	0	0	0	12	S	57				
0	0	0	0	12	S	58				
0	0	0	0	12	S	59				
0	0	0	0	12	S	60				
0	0	0	0	12	S	61				
0	0	0	0	12	S	62				
0	0	0	0	12	S	63				
0	0	0	0	12	S	64				
0	0	0	0	12	S	65				
0	0	0	0	12	S	66				
0	0	0	0	12	S	67				
0	0	0	0	12	S	68				
0	0	0	0	12	S	69 S		S	P	F
4	4	4	40	24	T	70	****		V	
3	7	1	50	33	T	71	***	S		
5	12	2	30	48	S	72	*****			
7	15	2	40	57	S	73	***** **			
3	21	2	40	75	S	74*	****** **			
8	24	1	30	84	S	75*	****** *			
9	26	3	40	90	S	76* *	** *** **			
9	26	0	30	90	S	77* *	** *** **			
10	28	1	10	96	S	78* *	****** **			
10	29	2	30	99	S	79* **	***** **			
9	29	1	30	99	S	80 **	***** **			
8	27	1	20	93	S	81 **	**** **			
5	22	3	40	78	S	82 *	*** *			
4	17	1	40	63	S	83 *	***			
2	11	2	30	45	S	84	**		V	F

SEGMENTATION DU MOT " Charité" (fin) Fréquence

Figure 4 b

visibles sur l'exemple exposé. Pour les éliminer de même que pour diminuer d'autres segmentations accidentelles, nous utilisons conjointement un second algorithme.

Selon leur aspect acoustique, les phonèmes sont classés en grandes catégories (voyelles, fricatives, etc...). Dans nos formes binaires, certaines de ces catégories se détectent très facilement :

- les occlusions qui précèdent les plosives et se manifestent par une succession en moyenne de 6 échantillons dépourvus de 1 ;

- les fricatives et les sifflantes dont la zone active est celle des hautes fréquences ;

- les autres phonèmes (voyelles, nasales, liquides, ...) dont la zone active s'étend sur tout le spectre.

Pour le deuxième algorithme, un échantillon étant rangé en mémoire selon :

MOT1 pour les basses fréquences,
MOT2 pous les hautes fréquences,

deux tests sont alors suffisants pour déceler la catégorie :

plosives(PLOS) avec MOT1 = MOT2 = 0

sifflantes(SIF) avec MOT1 = 0 et MOT2 $\neq$ 0

et autres (VOY) avec MOT1 $\neq$ 0 et MOT2 $\neq$ 0.

La prise en compte d'une de ces catégories n'a lieu qu'après 4 échantillons de même type.

Cette procédure permet d'effectuer une décomposition grossière du mot en zones plus ou moins longues dont les bornes sont rangées dans un tableau TSG2 (figure 4a,b).

Remarque 1. La zone étiquettée "Plosive" ne concerne qu'un seul phonème, la plosive en question.

Les zones étiquettées "sifflantes" ou "voyelles" peuvent recouvrir par contre plusieurs phonèmes. Par exemple :

dans "Rutabaga" ([R] [Y] [t] [a] [b] [a] [g] [a]) : tous les [a] sont détectés comme "voyelle" et ne durent qu'un seul phonème puisqu'ils sont séparés les uns des autres par des plosives ;

dans "Zoulou" ([Z] [U] [I] [U]) : la partie soulignée est aussi détectée comme "voyelle" bien que comportant trois phonèmes.

Remarque 2. Le second algorithme présente des insuffisances lorsque les formes ne sont pas très représentatives. Ainsi un [i] dont l'activité dans les basses fréquences est assez faible et qui donne parfois MOT1 = 0,est alors classé comme

une "sifflante" (comme pour "charité" (figure 4a,b)).

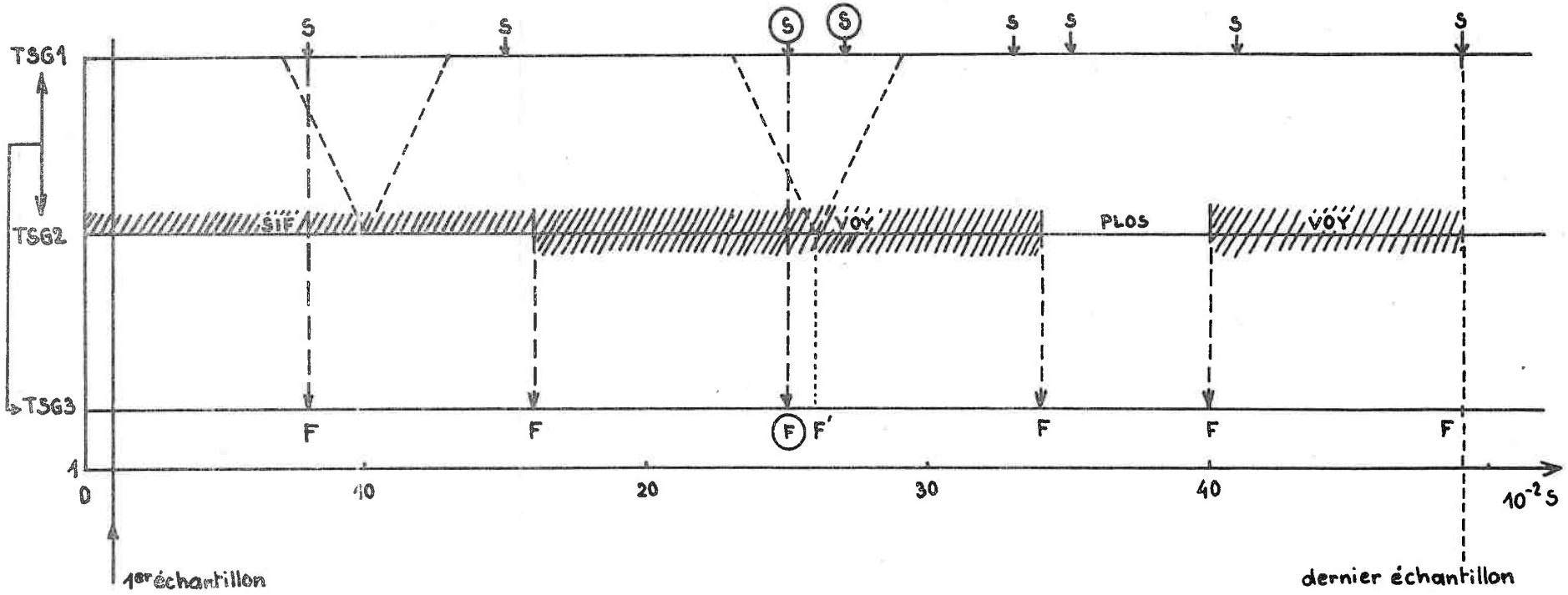
Les résultats définitifs de segmentation, reportés sur le tableau TSG3, se déduisent de la manière suivante de l'application des algorithmes précédents.

Les zones "plosives" de TSG2 qui correspondent à l'occlusion devant un seul phonème sont reportées telles quelles dans TSG3. Pour les zones "voyelles" ou "sifflantes" qui recouvrent un ou plusieurs phonèmes, on prend d'abord,à partir de la borne initiale, un segment de 10 échantillons, longueur moyenne d'un phonème. On cherche alors si la fin F de ce segment coïncide avec un point de segmentation de TSG1 à ±3 échantillons près. En cas de coïncidence, ce point de TSG1 est reporté dans TSG3 et sert de nouvelle borne pour le segment suivant. Dans le cas contraire, rare, la fin F du segment est prise comme point de segmentation.

Lorsqu'enfin une voyelle ou sifflante ne couvre pas un phonème, elle est prise en compte pour un segment.

Ce processus est illustré par la figure 5, fictive pour que tous les cas y apparaissent.

Il est mis en œuvre grâce à un sous-programme convenable de segmentation.



NB : si (S) n'existait pas nous aurions eu F' au lieu de (F)

Illustration de l'algorithme de segmentation

Fig.5

4) *FORME VOCALE COMME SOUS-ENSEMBLE FLOU.*

A - Généralités

Les diverses occurences d'un même mot ont ainsi la même forme générale, mais avec des différences locales sensibles.

Cette variabilité des données conduit à interprêter cette forme en termes de flou, sans qu'on y recherche ses caractéristiques particulières difficiles à isoler, position et amplitude des formants, durée des occlusives, etc... On traite ainsi globalement les différentes parties du mot aux limites mal définies et dont les caractères particuliers sont noyés dans l'aspect général.

Pour développer cette idée, on définit les termes suivants :

Zone temporelle : une zone temporelle est déterminée et repérée en position relative dans le mot.

Zone fréquentielle : une zone fréquentielle correspond à l'un des quatre groupements déjà évoqués des filtres adjacents.

Pavé : un pavé est l'intersection d'une zone temporelle et d'une zone fréquentielle. La forme globale d'un mot se présente alors comme une mosaïque de "pavés".

Critère : le critère est l'activité dans une bande de fréquence et s'évalue par le comptage des "1" dans cette bande.

B - Ensemble de référence

L'ensemble de référence E est l'ensemble, pour tous les mots du lexique enregistré, de tous les pavés x_{ijk} possibles concernant la $i^{ème}$ zone fréquentielle et la $j^{ème}$ zone temporelle du $k^{ème}$ mot où :

$$\begin{aligned} & i \in I \qquad I = \{1, 2, 3, 4\} \\ & j \in J \qquad J = \{1, \ldots, 1, \ldots, NZ\} \\ & k \in K \qquad K = \{1, \ldots, m, \ldots, M\}, \end{aligned} \tag{66}$$

NZ étant le plus grand nombre admissible de zones temporelles pour un mot et M le plus grand nombre possible de mots du lexique.

Le référentiel ne définit pas l'ensemble des mots du lexique, mais l'ensemble des pavés au moins nécessaires à construire ce dictionnaire et qui est l'encombrement mémoire.

Nous définissons alors les sous-ensembles flous sur l'ensemble de référence E.

C - Sous-ensembles flous sur l'ensemble E

Le $m^{ième}$ mot prononcé se compose de 4 bandes fréquentielles et de l_1 zones temporelles. Chaque pavé du référentiel appartient à ce mot avec un certain degré d'appartenance μ dont l'évaluation est un caractère de ce mot. Nous définissons ce degré d'appartenance comme la mesure de l'occupation du pavé, c'est-à-dire la densité des "1" ; sa valeur, comprise dans l'intervalle $[0,1]$ est évidemment nulle pour tous les pavés qui n'entrent pas dans la constitution de ce mot.

Le mot m est le sous-ensemble flou $\underset{\sim}{A}_m$ défini par les couples

$$\{(x_{ijk} , \mu_{\underset{\sim}{A}_m}(x_{ijk})\} \quad , \forall x_{ijk} \in E \tag{67}$$

où

$$\mu_{\underset{\sim}{A}_m}(x_{ijk}) = 0 \text{ pour } k \neq m,\ j \notin \{1, \ldots, l_i\}$$

et

$$\mu_{\underset{\sim}{A}_m}(x_{ijk}) \in [0,1] \quad \text{dans les autres cas.}$$

On pourrait considérer le sous-ensemble flou d'un mot comme le sous-ensemble ordinaire le plus proche avec le degré d'appartenance 0 ou 1. Un mot composé de l zones temporelles définies par quatre critères serait alors caractérisé par une des $2^{4 \cdot l}$ combinaisons. Ce codage très simple serait suffisant pour reconnaître des mots très différents les uns

des autres, c'est-à-dire donnant à coup sûr des combinaisons binaires différentes, mais forcément en nombres très limités. Dans la réalité, il est nécessaire de caractériser plus finement des mots acoustiquement voisins pour lesquels on utilise des degrés d'appartenance dans $[0,1]$.

5) *ALGORITHME DE RECONNAISSANCE*.

A - Généralités

Nous venons de décrire les formes à reconnaître en termes de flou et de définir un mot.

Un observateur expérimenté cherchant à reconnaître un mot d'après sa forme binaire, extrait l'information utile par l'estimation au jugé de son allure. Plus précisément, il effectue sur l'ensemble du mot des évaluations floues que, pour prendre une décision, il compare à celles des formes de référence mémorisées au cours d'une construction préalable pour la reconnaissance automatique.

Le principe de la méthode proposée est d'effectuer une telle comparaison par le calcul d'un indice de similarité entre un mot prononcé et chaque mot d'un lexique préalablement enregistré, suivie de la sélection du mot reconnu qui correspond à l'indice le plus élevé.

Cet indice de similarité s'exprime par la distance calculée entre les deux sous-ensembles flous des mots comparés dont il faut fixer les positions relatives.

Le rang de chaque échantillon est mémorisé en valeur relative par rapport à une référence RF qui est la durée limite autorisée par le sous-programme d'acquisition d'un mot. On arrondit aux valeurs entières les positions relatives puisque les zones temporelles comportent un nombre entier d'échantillons.

Durant la construction du lexique, on ne retient, pour un mot, que la position relative NR_j de fin de chaque zone temporelle :

$$NR_j = \left[\frac{N_j \times RF}{ECH} \right] \qquad (68)$$

où RF est la longueur standard d'un mot prise à 120 échantillons,

N_j la position absolue de la fin de la zone temporelle j du mot,

ECH le nombre d'échantillons pour le mot,

où les crochets indiquent la valeur arrondie à l'entier le plus proche.

Si n_{ijk} est le nombre de "1" présents dans le pavé d'indice i, j et k avec

$i \in I$ $\quad I = \{1, \ldots, CRIT\}$ où CRIT est le nombre de zones fréquentielles

$j \in J$ $\quad J = \{1, \ldots, l, \ldots, NZ\}$

k est le numéro du mot

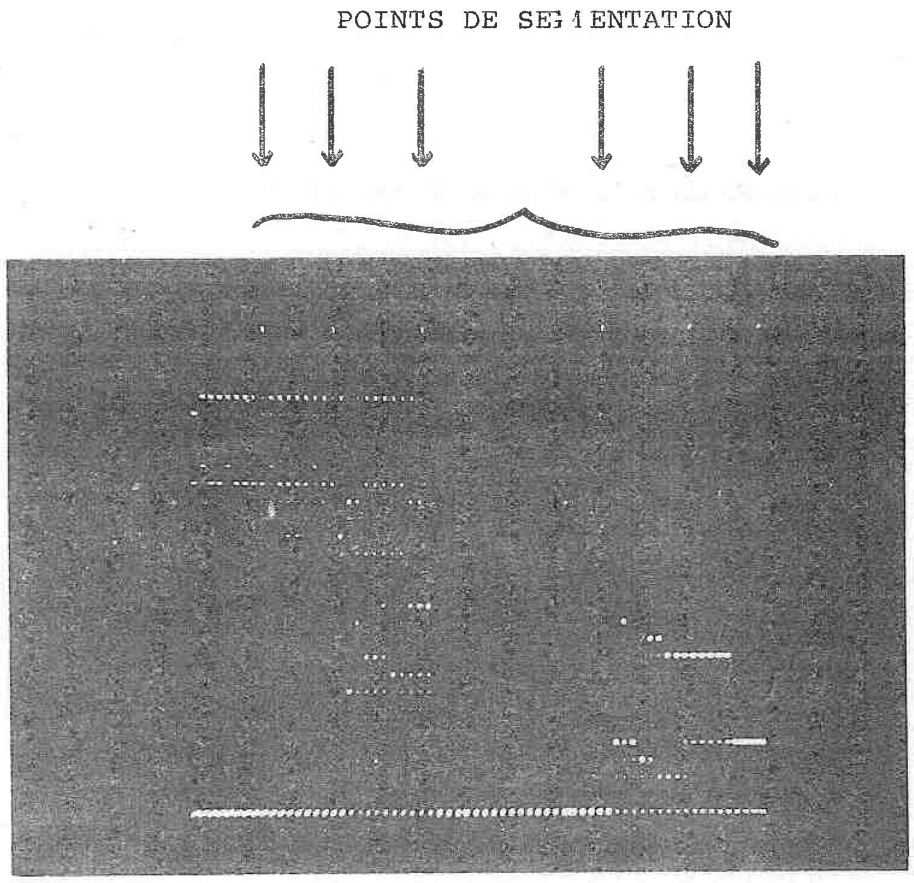
et si VM est le nombre de filtres inclus dans une zone fréquentielle, les degrés d'appartenance non nuls relatifs au mot k, se calculent alors par

$$\mu_{\underset{\sim}{A}_k}(x_{ilk}) = \frac{n_{ilk}}{(N_l - N_{l-1}) \times VM} \qquad (69)$$

B - Construction du lexique

On mémorise dans le calculateur une occurrence d'un mot k du lexique de M mots sous la forme d'un sous-ensemble flou $\underset{\sim}{A}_k$ avec $k \in K$, $K = \{1, \ldots, m, \ldots, M\}$.

Chaque mot, segmenté automatiquement, est visualisé sous la forme binaire avec l'indication des points de segmentation (figure 6). Si la segmentation est jugée représentative, on calcule la position relative des points de segmentation et le degré d'appartenance de chaque pavé.



Visualisation du mot "Jambon"

Figure 6.

Ces résultats, ainsi que le libellé du mot, sont stockés en mémoire sous la forme suivante :

Une zone temporelle décrite par 4 critères demande 6 mots - mémoire, soit :

1 mot pour la position relative,
1 mot pour le code du mot prononcé,
4 mots pour les 4 degrés d'appartenance.

Le libellé de chaque mot est rangé dans un tableau à part.

Sous cette forme le lexique de référence demande nettement moins de place en mémoire que sous la forme binaire. Ainsi, un mot de 500ms et avec 4 zones temporelles exige 24 mots mémoire alors qu'en représentation binaire il en faut 100 dans les conditions actuelles de l'analyse et de l'acquisition.

Nous verrons plus loin qu'il est possible d'augmenter la taille du lexique pour la même capacité de mémoire.

Le mode opératoire de l'acquisition et de la construction des mots du lexique est résumé par l'organigramme de la figure 7 a et b.

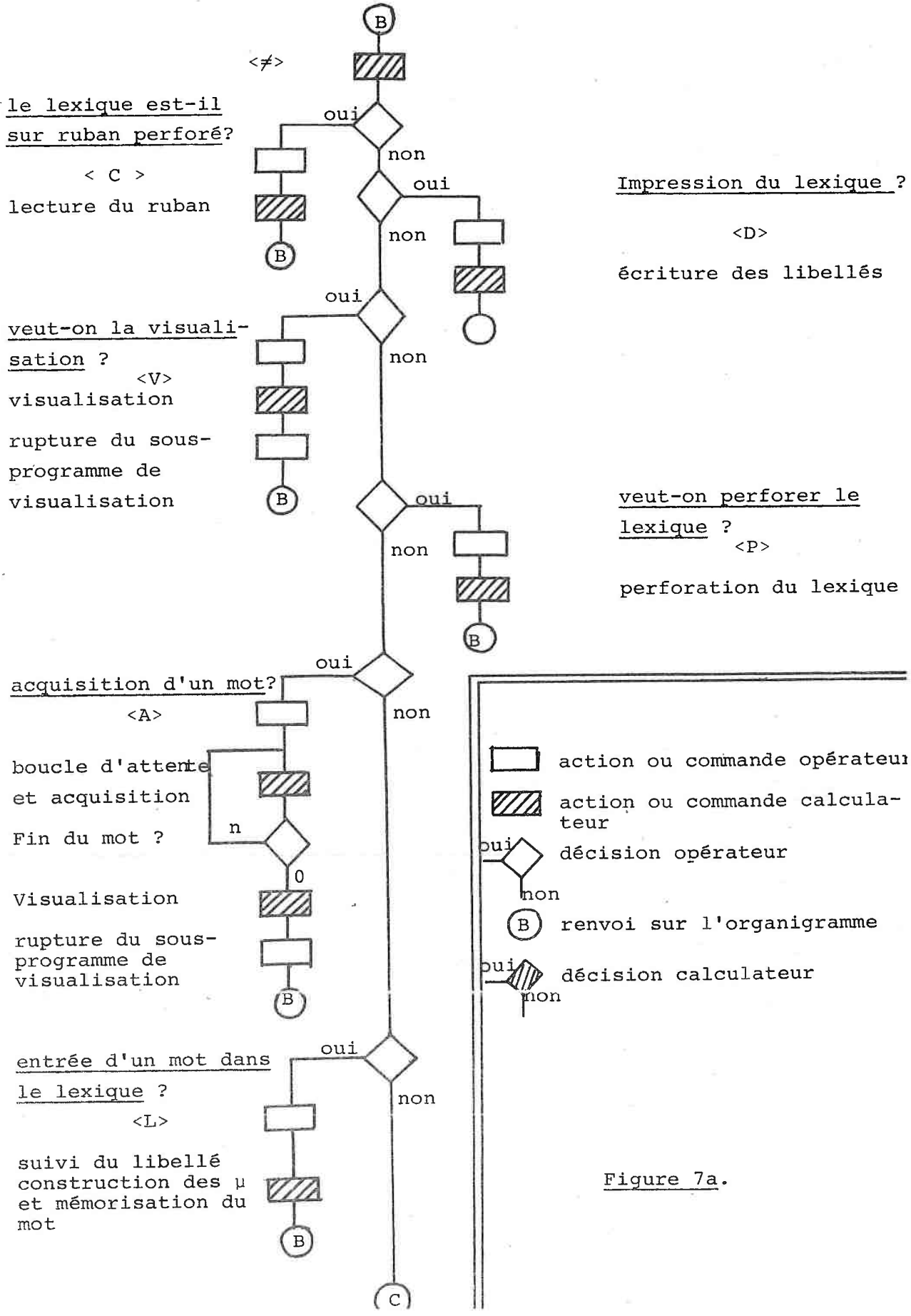

B
<≠>
le lexique est-il sur ruban perforé?
oui
non
< C >
lecture du ruban
B
oui
Impression du lexique ?
<D>
écriture des libellés
non
veut-on la visualisation ?
oui
<V>
visualisation
rupture du sous-programme de visualisation
B
non
veut-on perforer le lexique ?
oui
<P>
perforation du lexique
B
non
acquisition d'un mot?
oui
<A>
boucle d'attente et acquisition
Fin du mot ?
n
0
Visualisation
rupture du sous-programme de visualisation
B
non
entrée d'un mot dans le lexique ?
oui
<L>
suivi du libellé construction des μ et mémorisation du mot
B
non
C
action ou commande opérateur
action ou commande calculateur
oui
décision opérateur
non
B
renvoi sur l'organigramme
oui
décision calculateur
non

Figure 7a.

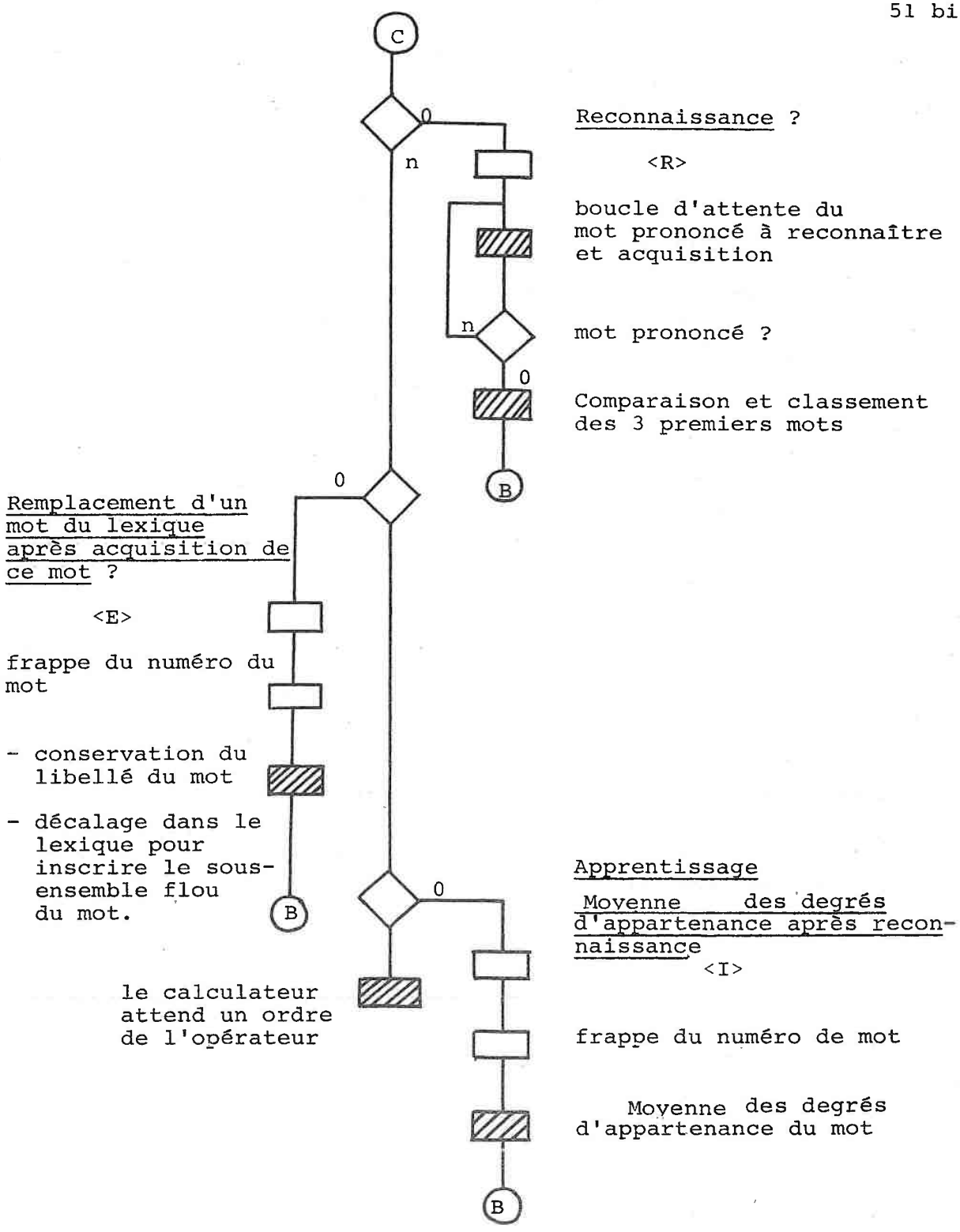

C
0
n
Reconnaissance ?
<R>
boucle d'attente du mot prononcé à reconnaître et acquisition
n
mot prononcé ?
0
Comparaison et classement des 3 premiers mots
B
0
Remplacement d'un mot du lexique après acquisition de ce mot ?
<E>
frappe du numéro du mot
- conservation du libellé du mot
- décalage dans le lexique pour inscrire le sous-ensemble flou du mot.
B
0
Apprentissage
Moyenne des degrés d'appartenance après reconnaissance
<I>
le calculateur attend un ordre de l'opérateur
frappe du numéro de mot
Moyenne des degrés d'appartenance du mot
B

Figure 7b.

La construction du lexique se déroule en mode conversationnel avec possibilité à tout moment d'acquérir un nouveau mot (commande A), de le visualiser (V), d'en inscrire le libellé et de construire les degrés d'appartenance (L), d'imprimer le lexique (D) et enfin, de perforer le lexique avec libellé (P).

La commande A déroute le programme vers une boucle qui permet d'attendre l'échantillon en provenance de l'analyseur-codeur dès qu'il se présente. Le stockage de la forme binaire commence dès l'apparition du premier échantillon non nul et se poursuit jusqu'à la détection d'un silence d'au moins 300ms qui marque la fin du mot.

Pendant l'acquisition, les deux algorithmes de segmentation agissent simultanément en temps réel, un dispositif annexe de comptage rapide permettant de totaliser les "1".

C - Phase de reconnaissance

La reconnaissance d'un mot s'effectue par la comparaison avec chacun des mots enregistrés du lexique. Les termes de la comparaison sont les deux sous-ensembles flous relatifs à ces deux mots.

Ceci implique :

- d'une part que les éléments préalablement acquis de la comparaison soient en correspondance biunivoque : au pavé d'indice i,j,k d'un mot doit correspondre le pavé de mêmes indices de l'autre mot ; on doit, autrement dit, définir pour le mot prononcé les mêmes positions relatives que celles apparaissant dans le mot de référence ;

- d'autre part, que soit défini, pour la comparaison, un indice de similarité entre les degrés d'appartenance des pavés correspondants de chacun de ces mots considérés comme des sous-ensembles flous.

a) Acquisition :

La saisie des échantillons du mot à reconnaître pris dans la liste de référence se déroule comme dans la phase de construction, à l'exception du calcul des points de segmentation. En effet, nous venons de voir que les pavés du mot à reconnaître et ceux du mot de référence doivent correspondre strictement. Or, les résultats de la segmentation n'étant pas parfaitement reproductibles, il y a obligation, dans la phase de reconnaissance, de mémoriser les données partielles de chaque échantillon et de différer leur exploitation. Ainsi, à chaque instant et pour chaque critère, on cumule et on mémorise les "1" des échantillons précédents ; ces résultats servent par la suite à calculer les degrés d'appartenance après détermination des positions relatives des zones temporelles.

Quand l'acquisition du mot est terminée, on calcule également les positions relatives de chaque échantillon par rapport à la référence RF.

b) Propositions relatives :

Soit B l'ensemble ordonné des positions relatives des zones temporelles d'un mot du lexique, comparé au mot à reconnaître et b_j un élément de cet ensemble

$$b_j \in B$$

$$j \in J \quad si \quad J = \{1, \ldots, l, \ldots, NZ\}$$

Soit A l'ensemble ordonné des positions relatives r_k de chaque échantillon du mot prononcé :

$$r_k \in A$$

$$k \in K \quad si \quad K = \{1, \ldots, ECH\}$$

en particulier $r_{ECH} = RF$.

Le problème qui se pose est de construire un sous-ensemble ordonné $B' \in A$ dont les éléments sont :

$$b'_j = \sup \{r_{k-1}, r_k\} \quad si \quad b_j \in]r_{R-1}, r_R]$$

avec $j \in J$, $J = \{1, \ldots, l, \ldots, NZ\}$.

En d'autres termes, la comparaison de A relatif au mot à reconnaître à B relatif au mot du lexique se fait aux seules positions relatives de A les plus proches de celles de B. La figure 8 illustre cette explication.

Ce procédé revient à "transcrire" sur le mot prononcé la segmentation acquise sur le mot du lexique, pour s'affranchir ainsi du "flou" des limites temporelles.

GENERATION DES ELEMENTS DE B'

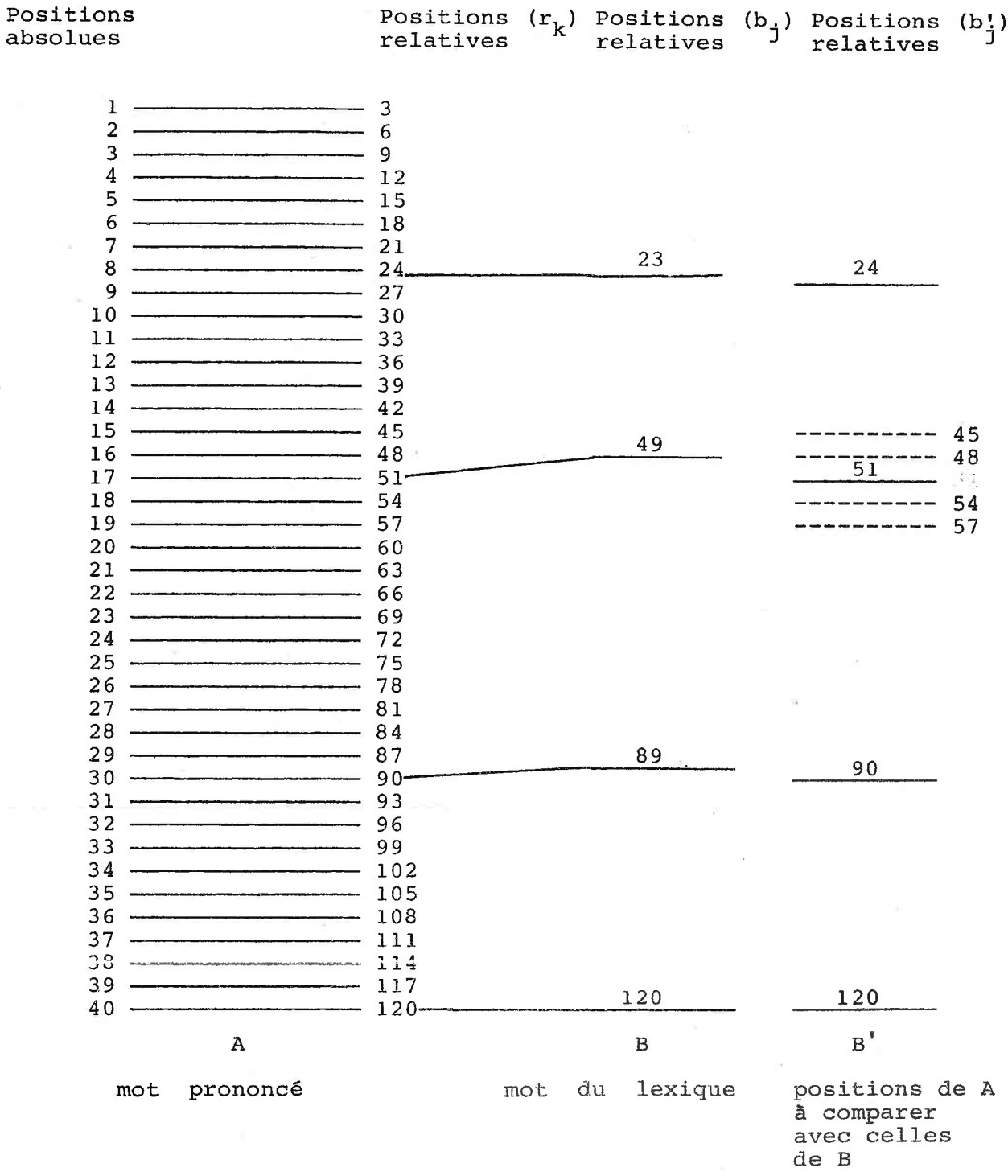


Figure 8.

On remarque que pour tenir compte des fluctuations des positions relatives lors des diverses occurrences , on aurait pu comparer un élément de B avec ceux de A situés "au voisinage". Par exemple, sur la figure 8 , la position relative 49 pourrait être comparée également aux positions 45, 48 54 ou 57 de A. On retiendrait alors de ces 5 cas celui de la plus forte similarité, méthode qui s'appliquerait au détriment de la vitesse de calcul.

c) Indice de similarité :

Les pavés placés en b'_j pour le mot prononcé sont ensuite directement comparés aux pavés en b_j à l'aide de l'indice de similarité que nous allons calculer.

Soient un ensemble $X = \{x_1,\ldots,x_i,\ldots,x_n\}$, un ensemble de propriétés $P = \{p_1,\ldots,p_k,\ldots,p_m\}$ et une matrice booléenne M(n x m) d'éléments m_{ik} tels que $m_{ik} = 1$ ou 0 selon que l'élément x_i jouit ou non de la propriété p_k.

Pour évaluer la similitude entre deux éléments x_i et x_j, on considère un indice de similarité $S(x_i,x_j)$ qu'avec ROY [R 1] nous définirons par :

$$S(x_i,x_j) = 1 - \frac{|A(x_i) \Delta A(x_j)|}{m} \qquad (70)$$

où $A(x_i)$ est l'ensemble des propriétés que possède x_i et où Δ désigne la différence symétrique des ensembles $A(x_i)$ et $A(x_j)$.

Pour cet indice, on a :

a) $0 \leqslant S(x_i,x_j) \leqslant 1$

b) $S(x_i,x_j) = 1$ ssi $A(x_i) \Delta A(x_j) = \emptyset$

c) $S(x_i,x_j) = 0$ ssi $A(x_i) \Delta A(x_j) = P$

ce qui signifie que x_i possède une propriété ssi x_j ne la possède pas.

$$\begin{cases} \text{d)} & S(x_i,x_j) = S(x_j,x_i) \\ \text{e)} & S(x_i,x_j) \text{ décroît lorsque } |A(x_i) \Delta A(x_j)| \text{ croît.} \end{cases}$$

Mais comme dans notre cas il n'est pas possible d'affirmer qu'un élément x_i a ou n'a pas une propriété p_k, nous affectons au couple (x_i,p_k) un coefficient pour apprécier à quel point l'élément x_i possède la propriété p_k. Le coefficient $\mu_k(x_i)$ est le degré d'appartenance de la propriété p_k à l'élément x_i, tel que :

$$0 \leqslant \mu_k(x_i) \leqslant 1 \quad , \quad \forall k \text{ et } \forall i$$

La nouvelle matrice M, obtenue avec $\mu_k(x_i)$, définit une relation floue dans (X x P).

Nous définissons ainsi l'indice de similarité floue suivant qui revient à utiliser la distance de Hamming :

$$S(x_i,x_j) = 1 - \frac{\sum_{k=1}^{k=m} |\mu_k(x_i) - \mu_k(x_j)|}{m} \qquad (73)$$

Si les propriétés p_k n'ont pas la même importance, on peut les pondérer et prendre

$$S(x_i,x_j) = 1 - \frac{\sum_{k=1}^{m} C_k|\mu_k(x_i) - \mu_k(x_j)|}{C} \qquad (74)$$

avec $C = \sum_{k=1}^{m} C_k$.

Cet indice traduit l'appartenance floue par l'intermédiaire des degrés d'appartenance μ et évalue la similarité entre deux éléments, en soulignant l'importance relative des différentes propriétés.

Soient alors $\underset{\sim}{B}$ le sous-ensemble flou du mot du lexique et $\underset{\sim}{B}'$ le sous-ensemble flou du mot prononcé pour lequel on a défini les NZ positions relatives.

Pour déterminer la similarité d'un mot m du lexique et du mot prononcé, nous ferons la moyenne des indices de similarité de tousles pavés d'indices i,j,k avec k = m seulement,sans avoir à retenir les pavés des sous-ensembles flous $\underset{\sim}{B}$ et $\underset{\sim}{B}'$ dont les degrés d'appartenance sont nuls et qui n'apportent donc aucune contribution de ressemblance.

Comme la seule propriété de chaque pavé x_{ijm} est l'importance des "1", c'est elle qui détermine le degré d'appartenance $\mu(x_{ijm})$ qui donne, pour la similarité entre 2 pavés correspondants, des sous-ensembles flous $\underset{\sim}{B}$ et $\underset{\sim}{B}'$:

$$S(x_{ijm},\underset{\sim}{B} \ , \ x_{ijm},\underset{\sim}{B}') = 1 - | \ \mu_{\underset{\sim}{B}}(x_{ijm}) - \mu_{\underset{\sim}{B}'}(x_{ijm}) | \qquad (75)$$

On note que cet indice de similarité ne répond pas à la définition (70) puisqu'il concerne la similarité d'appartenance au référentiel d'un même élément de deux sous-ensembles flous distincts, et pas la ressemblance de deux éléments d'un même ensemble.

Pour les deux mots à comparer, nous définissons la similarité comme la moyenne des indices de similarité de tous les pavés ou score :

$$S(\underset{\sim}{B},\underset{\sim}{B}') = 1 - \frac{\sum_{i,j} | \mu_{\underset{\sim}{B}}(x_{ijm}) - \mu_{\underset{\sim}{B}'}(x_{ijm}) |}{NZ \times CRIT} \qquad (76)$$

Les degrés d'appartenance dans cette dernière expression se calculent avec les données de chaque échantillon enregistrées dans un tableau au cours de l'acquisition. Ce tableau comporte ECH lignes et 2 + CRIT colonnes qui mémorisent :

a_k : numéro absolu de l'échantillon

r_k : numéro relatif de l'échantillon

C_{ik} : cumul des "1" depuis le début du mot sur les CRIT bandes de fréquences.

Le degré d'appartenance $\mu_{B'}(x_{ilm})$ pour le critère i, la zone temporelle l et le mot m est donné par :

$$\mu_{B'}(x_{ilm}) = \frac{C_{i,b'_l} - C_{i,b'_{l-1}}}{(a_{b'_l} - a_{b'_{l-1}}) \text{ X VM}} \qquad (77)$$

où : b'_l est la position relative dans le mot prononcé en en rapport avec la $l^{ème}$ position relative du mot de référence m ;

$a_{b'_l}$ est le numéro absolu de l'échantillon de position relative b'_l ;

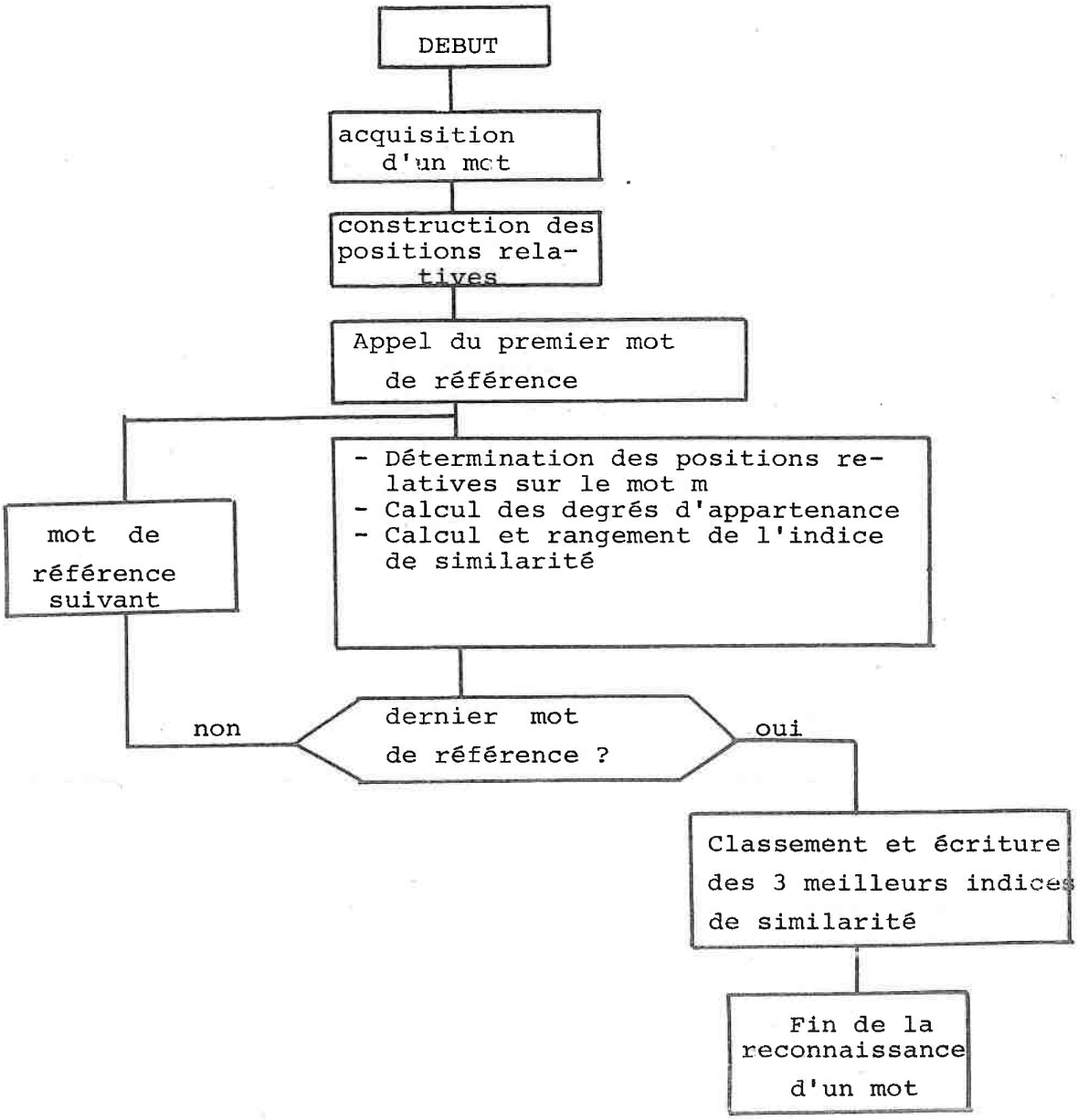
VM est le nombre de filtres dans la $i^{ème}$ bande de fréquences.

Le degré d'appartenance $\mu_B(x_{ilm})$ est pris directement dans le tableau du lexique de référence.

d) Décision :

Après la comparaison du mot prononcé avec tous les mots du lexique, on classe par ordre décroissant les trois meilleurs scores (76) dont le premier désigne le mot reconnu.

L'organigramme de la figure 9 décrit le processus de reconnaissance pour un mot.



ORGANIGRAMME DU PROCESSUS DE RECONNAISSANCE POUR UN MOT

Figure 9.

Le mode opératoire en phase de décision est décrit sur la figure 7 b où la commande (R) appelle le sous-programme de reconnaissance.

6) *MISE EN OEUVRE ; RESULTATS* [B3 , B4].

Le travail en temps réel a lieu dans le local même de l'ordinateur et de ses périphériques où le niveau de bruit est d'environ 65 décibels (Figure 10).

Jusqu'au chapitre V les résultats ne concernent qu'un seul locuteur.

Dans l'ordinateur T 2000 utilisé, de 8k mots (de 19 bits) de mémoire centrale, le programme écrit pour un mode conversationnel occupe environ 2000 mots, et la forme binaire acquise du mot à reconnaître, 300. Le reste de la mémoire est donc disponible pour le lexique, au maximum de 200 mots, trois quarts pour leur forme et un quart pour leur libellé.

Pour étudier d'abord l'influence du nombre de critères, nous avons construit une première séquence d'un lexique L19 de 19 mots. Puis, pour un cycle complet de reconnaissance de ces 19 mots, nous avons enregistré une seconde séquence, puis une troisième.

Pour manipuler, sur des séquences immuables dans ces expériences préliminaires, les enregistrements ont été transcrits sur bandes perforées.

L'examen de ces enregistrements (tableau 2) montre que la durée d'élocution par rapport à la référence peut varier de ±20%.

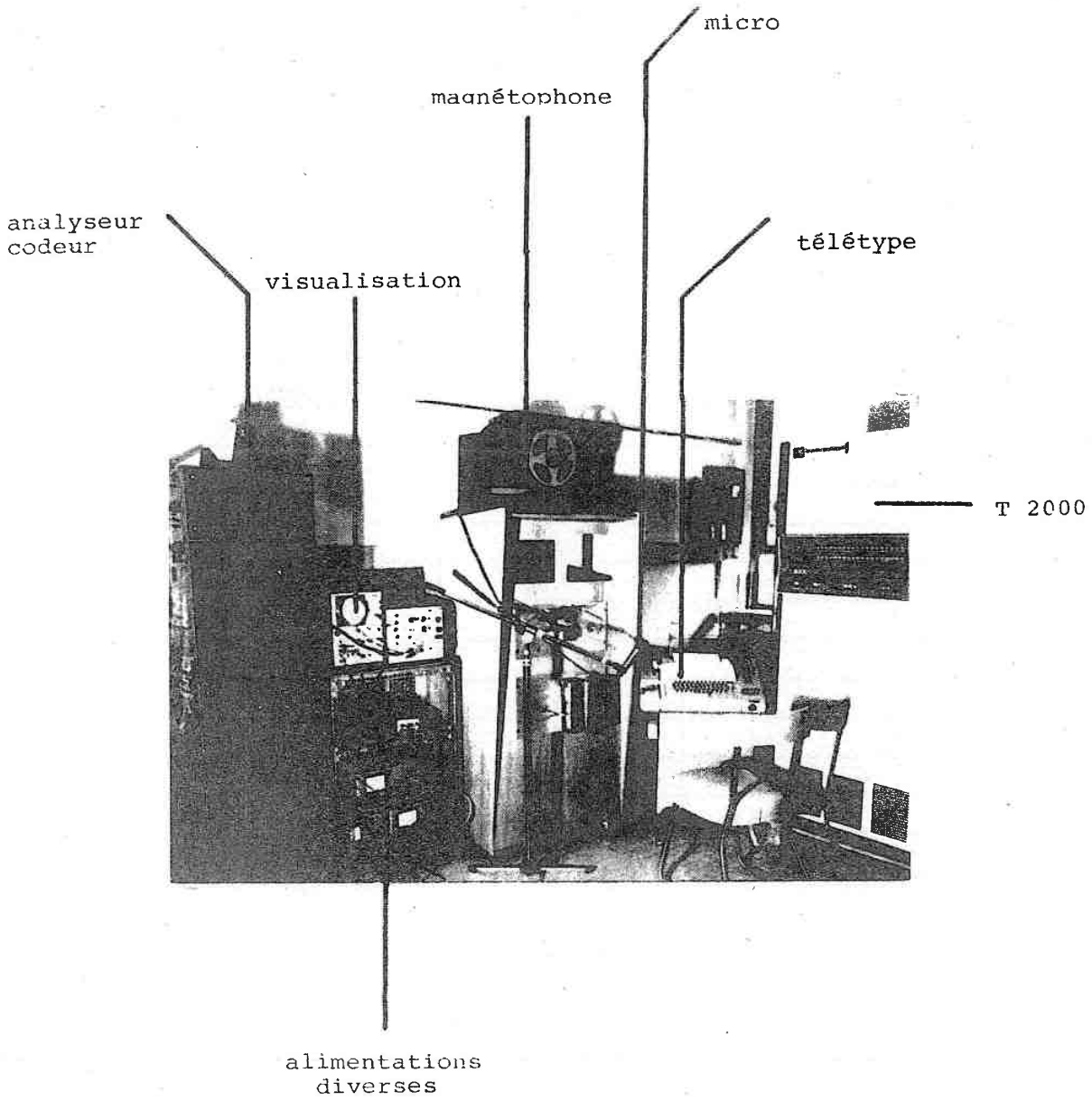

micro
magnétophone
analyseur
codeur
visualisation
télétype
T 2000
alimentations
diverses

Figure 10.

lexique L19	Durée (ms) 1ère séquence	2e séquence	3e séquence
1 ANATOLE	610	570	600
2 BERTHE	400	410	400
3 CELESTIN	660	660	690
4 DESIRE	480	500	470
5 EUGENE	450	370	370
6 GASTON	450	440	430
7 HENRI	320	310	310
8 IRMA	380	370	370
9 JOSEPH	380	420	410
10 KLEBER	440	460	440
11 LOUIS	240	270	310
12 MARCEL	540	500	560
13 NICOLAS	420	430	450
14 OSCAR	550	550	540
15 PIERRE	340	360	350
16 QUINTAL	510	490	520
17 RAOUL	500	440	520
18 SUZANNE	450	440	560
19 THERESE	480	480	480

Tableau 2.

Le taux de reconnaissance moyen pour les deuxième et troisième séquences a ensuite été étudié en retenant 1, 2, 3, 4, 8 puis 12 critères, d'autant de zones fréquentielles groupant successivement chacune 24, 12, 8, 6, 3 et 2 filtres.

Comme le montre la figure 11, il est inutile de retenir plus de 4 critères dans le point de vue considéré.

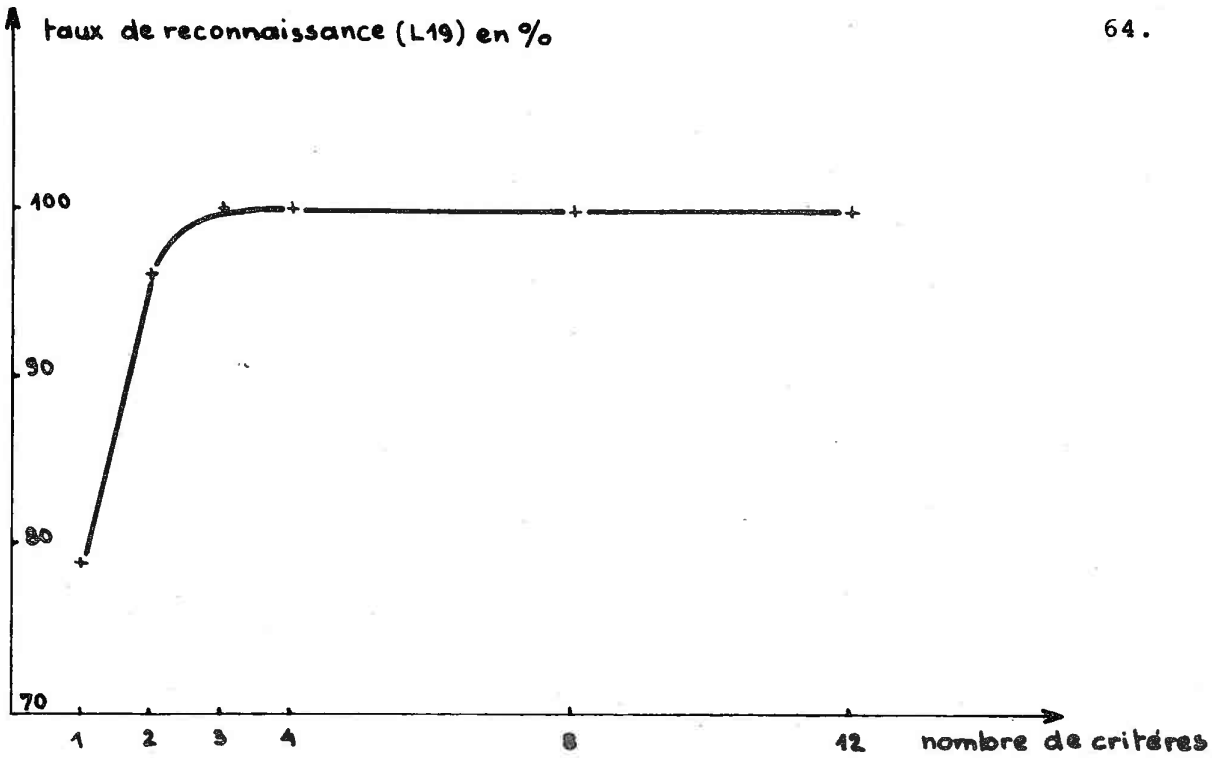


Figure 11.

Puis, pour chaque mot, nous avons relevé les indices de similarité avec les 19 mots du lexique pour un nombre croissant de critères. Concernant par exemple le mot "Célestin" et 1, 2, 4 et 12 critères, les résultats obtenus sont ceux de la figure 12 qu'il convient d'interprêter, pour la discrimination, en différence de score : par exemple pour "Célestin" par rapport à "Marcel", la valeur significative est :

0,18 pour 1 et 2 critères
0,22 pour 4
0,27 pour 12.

La modicité de cette valeur pour un petit nombre de critères résulte évidemment de la faible prise en compte de la distribution fréquentielle, bien qu'il faille souligner qu'une augmentation du nombre de critères ne donne pas une amélioration spectaculaire dans le cas d'un lexique réduit.

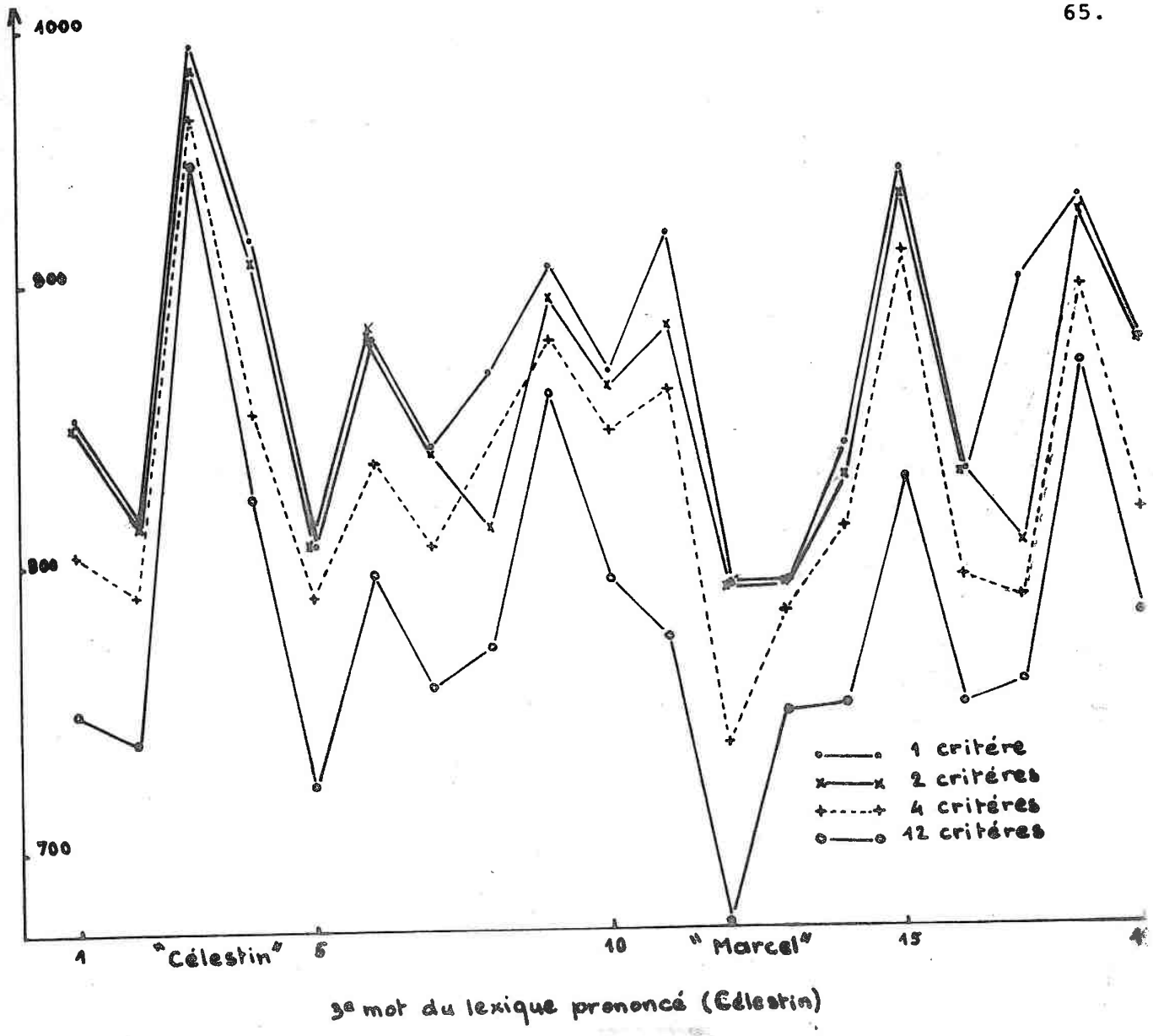


Figure 12.

Quoiqu'il en soit, le gain au-delà de 4 critères, nombre déjà retenu, n'est pas assez significatif pour justifier l'encombrement croissant du calcul.

Le nombre de critères fixé, nous avons établi une liste constituée au hasard, en prenant dans le dictionnaire "Petit Robert", toutes les 10 pages, le dernier mot de cette page, à l'exception cependant de quelques mots trop longs.

Cette liste (tableau 3) et le programme lui-même utilisé alors remplissent complètement la mémoire du calculateur.

ACCES
ADAPTATION
APIOL
APRES
ARQUE
ASSASSINAT
ATTACHER
AUTODAFE
AVOCAT
BALISTIQUE
BASTRINGUE
BESOIN
BLANCHISSEUR
BOTTE
BRAILLEMENT
BRUN
CADRAN
CANDIDEMENT
CARLINE
CAVALCADE
CHAHUTER
CHERIR
CHRONIQUEUR
CLASSIQUE
COLLANTE
CONCILIAIRE
CONTINUEL
COTER
COURTINE
CREVETTE
CUIRE
DECHARGE
DEFAIRE
DELAVER
DERACINEMENT
DESIR
DEUX-DEUX
DISCUSSION
DIVISION
DOUBLE
DUPLICATA
ECHEC
EDITION
ELEMENT
EMONDES
ENTREMISE
EPICIER
ERAFLURE
ESSAIM
ETENDU
EUNUQUE
EXCITE
FALLACIEUX
FAYOTTER
FEUILLETER
FINITION
FOI
FORME
FOUTRE
GARE
GOBELIN
GRANDIR
HABITUDE
HEBDOMADAIRE
HYPNOTISER
INERTIE
INNOVATION
JONQUILLE
KHAN
LANCEE
LEGERETE
LICTEUR
LIVRE
LOUVRE
MALE
MANUSCRIT
MARTRE
MECHANT
MENTHE
MINIATURE

LEXIQUE L 80

TABLEAU 3.

Pour des sous-ensembles successifs de 20, 40, 60 et 80 mots de la liste, une séquence enregistrée a été prise comme lexique et pour 4 séquences suivantes on a testé la reconnaissance.

Le tableau 4 donne pour le mot à reconnaître le taux de reconnaissance obtenu, dans le cas où ce mot est classé en première position, dans les deux premières positions ou dans les trois premières.

La variation du taux en fonction de la taille du lexique, dans les trois éventualités précédentes, est représentée par les courbes de la figure 13.

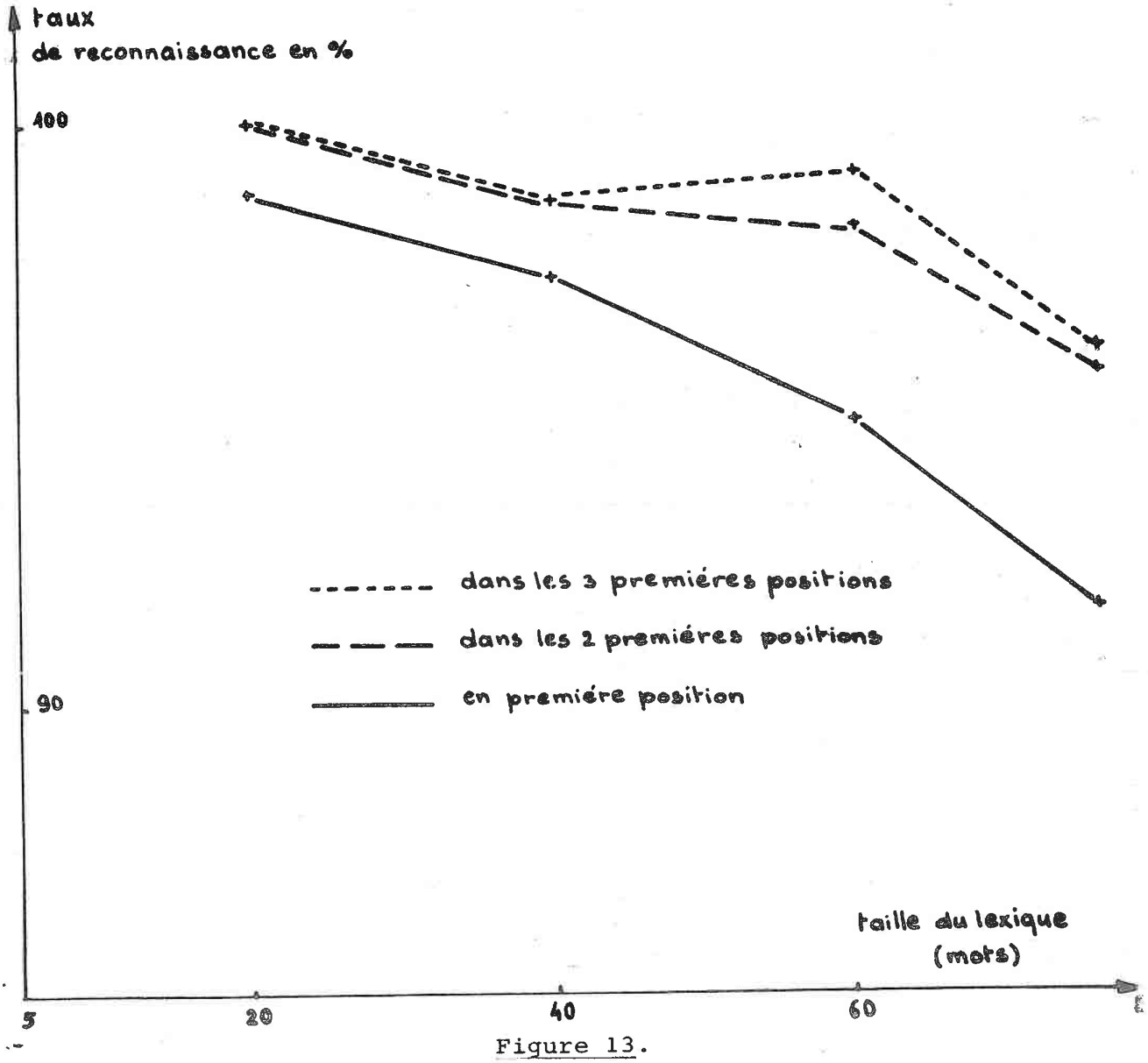


Figure 13.

taille du lexique	n° de la séquence	En 1ère position	Dans les 2 premières	Dans les 3 premières	% moyen En 1ère position	% moyen Dans les 2 premières	% moyen Dans les 3 premières
20	2	100	100	100	98,8	100	100
	3	100	100	100			
	4	100	100	100			
	5	95	100	100			
40	2	100	100	100	97,4	97,4	98,7
	3	97,4	97,4	100			
	4	94,7	94,7	97,4			
	5	97,4	97,4	97,4			
60	2	96,6	98,3	100	94,9	98,3	99,2
	3	93,1	98,3	98,3			
80	2	91,9	95,6	96,2	91,7	95,8	96,2
	3	90	92,5	97,5			
	4	91,2	96,2	97,5			
	5	93,7	98,7	98,7			

TABLEAU 4.

Taux de reconnaissance en %

Des décisions incorrectes de reconnaissance, donc des confusions, peuvent se produire évidemment pour des mots semblables comme "Fayotter" et "Feuilleter".

Cependant, certains mots, bien que de longueurs et d'aspects fréquentiels très différents, peuvent être confondus, par exemple "Braillement" comparé à la référence "Brun" pour laquelle la segmentation a donné une seule zone temporelle.

Pour calculer la similarité, les deux mots doivent avoir le même nombre de zones temporelles avec approximativement les mêmes valeurs de positions relatives. Dans le cas cité, "Braillement" est donc traité en un seul segment, et comme son aspect fréquentiel est tel que, globalement, la densité des "1" dans chaque bande est semblable à celui de "Brun", il y a confusion. Cette confusion pourraît être levée en tenant compte grossièrement de la longueur des mots.

Par contre, il n'y a aucune difficulté si "Brun" est à reconnaître et "Braillement" la référence.

La durée moyenne des mots enregistrés a tendance à diminuer pour un même locuteur, lorsque la longueur totale de l'enregistrement augmente, sans doute à cause de la fatigue et de la routine dans la répétition d'un même lexique (figure 14). Cet effet est cependant sans conséquences sur la reconnaissance.

==========

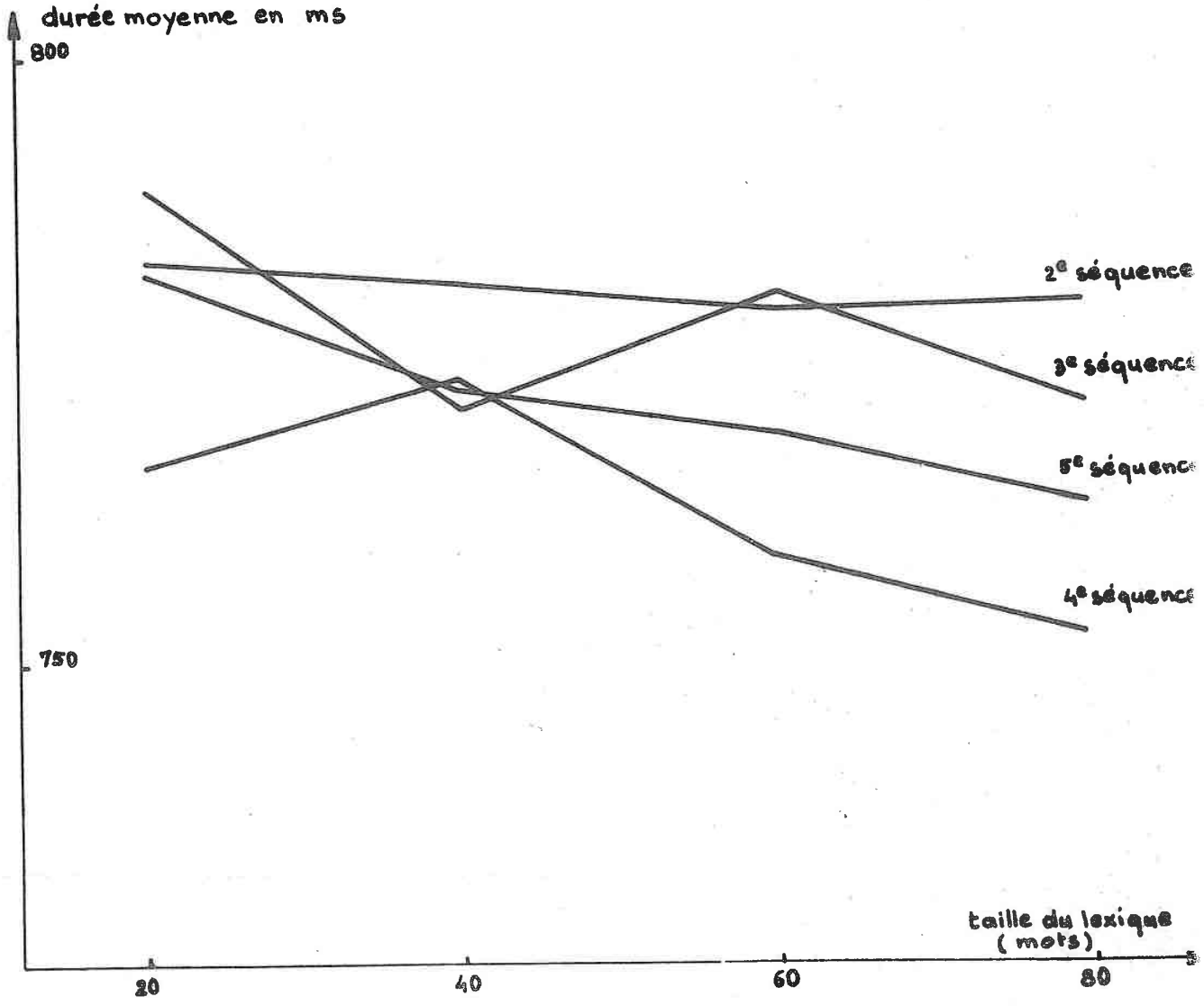

durée moyenne en ms
800
750
2e séquence
3e séquence
5e séquence
4e séquence
taille du lexique
(mots)
20
40
60
80

Fig. 14

CHAPITRE IV

VALIDITE ET EXTENSION DE LA METHODE

Les résultats précédents peuvent être considérablement améliorés en ce qui concerne la reconnaissance proprement dite et l'utilisation optimale des mémoires pour le stockage du lexique.

1) *SEGMENTATION AUTOMATIQUE, SEGMENTATION A VUE ET TAUX DE RECONNAISSANCE.*

Dans tout ce qui précède, nous avons admis le principe de la segmentation automatique qui libérait l'opérateur d'une tâche délicate et fastidieuse, pour la détermination des positions relatives en phase de construction. Il est cependant important de comparer la segmentation automatique à la segmentation à vue en ce qui concerne les résultats définitifs de la reconnaissance.

Pour effectuer cette comparaison, le locuteur a enregistré deux séquences d'un même lexique L50 de 50 mots choisis dans les codes téléphoniques d'identification anglais et français, pour leurs aspects jugés caractéristiques.

La première séquence a été segmentée à vue, donnant 226 segments pour les 50 mots, puis automatiquement avec 204 segments. La superposition des segments est satisfaisante puisque 152 segmentations coïncident à 1 échantillon près.

50 mots

ALPHA	WHISKY	PIERRE
BRAVO	X RAY	QUINTAL
CHARLIE	YANKEE	RAOUL
DELTA	ZOULOU	SUZANNE
ECHO	ANATOLE	THERESE
FOX	BERTHE	URSULE
HOTEL	CELESTIN	VICTOR
INDIA	DESIRE	WILLIAM
KILO	EUGENE	XAVIER
LIMA	GASTON	YVONNE
MIKE	HENRI	ZOE
NOVEMBER	IRMA	PARIS
PAPA	KLEBER	NANCY
QUEBEC	LOUIS	NICE
ROMEO	MARCEL	TOULOUSE
SIERRA	NICOLAS	AMIENS
TANGO	OSCAR	

Lexique L 50.

TABLEAU 5

Cette première séquence, segmentée à vue et automatiquement, servant de lexique, la reconnaissance a été faite pour la deuxième séquence enregistrée.

Globalement, les résultats sont comparables puisqu'on relève 90, 96 et 98 % de mots reconnus en première, dans les deux premières et dans les trois premières positions.

La phase de reconnaissance effectuée a été mauvaise pour les mots inscrits dans le tableau 6 alors que les autres mots sont reconnus correctement.

	mot reconnu dans les "I" premières positions / mot prononcé	I = 1	I = 2	I = 3
Segmentation auto. Appréciation à vue	ALPHA	Papa 932	Alpha 917	écho 884
	DELTA	écho 898	Delta 898	William 888
	FOX	Nancy 913	Paris 902	Berthe 895
	QUEBEC	Thérèse 921	Yvonne 911	Québec 905
	DESIRE	Eugène 927	Désiré 919	William 892
	FOX	Paris 909	Nancy 907	Fox 893
	KILO	Suzanne 930	Kilo 913	Victor 896
	NOVEMBER	Alpha 918	november 904	Quintal 895
	DESIRE	Eugène 914	Désiré 905	William 899
	YVONNE	kilo 913	Suzanne 901	Quintal 893

N.B. : les indices sont en %.

TABLEAU 6.

Les mots "Fox" et "Désiré" sont mal reconnus les deux fois.

Pour "Fox" on note que, dans les deux cas, la confusion vient surtout de la présence en construction d'une trace de la fricative qui n'est pas relevée par l'analyseur en reconnaissance. C'est ce que montre la figure 15 qui représente la juxtaposition de la forme "Fox" et des formes "Nancy" et "Paris" avec lesquelles il y a confusion dans la sixième ligne du tableau 6.

Les formes binaires de ces mots qui ont servi à la construction du lexique sont surmontées pour "Fox" de deux tableaux A) et B). La première colonne de ces tableaux indique respectivement en A) les positions absolues et en B) les positions relatives à 120 en segmentation automatique. Les quatre colonnes suivantes sont en A), pour les pavés constitués, le cumul des "1", et en B) le degré d'appartenance en %.

La forme de droite est celle de "Fox" en reconnaissance, pour laquelle l'imprimante écrit le tableau C) qui place "Fox" en troisième position avec un score de 893 malgré la perte de l'information essentielle de la fricative.

Pour "Désiré", bien que la segmentation à vue donne un segment supplémentaire, la reconnaissance fournit le même classement.

Pour "Alpha" et "Québec" où les nombres de segments sont nettement différents dans les deux méthodes, les résultats sont bons avec la segmentation automatique dont le choix est justifié par sa constante efficacité et sa nécessité dans l'automatisation complète.

2) *DISCRIMINATION PAR UN INDICE MODULE*.

A - Généralités

Pour réduire les confusions en reconnaissance, nous avons introduit un indice de similarité modulé sur le mot de référence.

L'indice de similarité précédemment utilisé, ou indice brut, fait la comparaison entre des sous-ensembles flous se correspondant "pavé à pavé" ; pour le lexique L 50, il donne un taux de reconnaissance de 86 %.

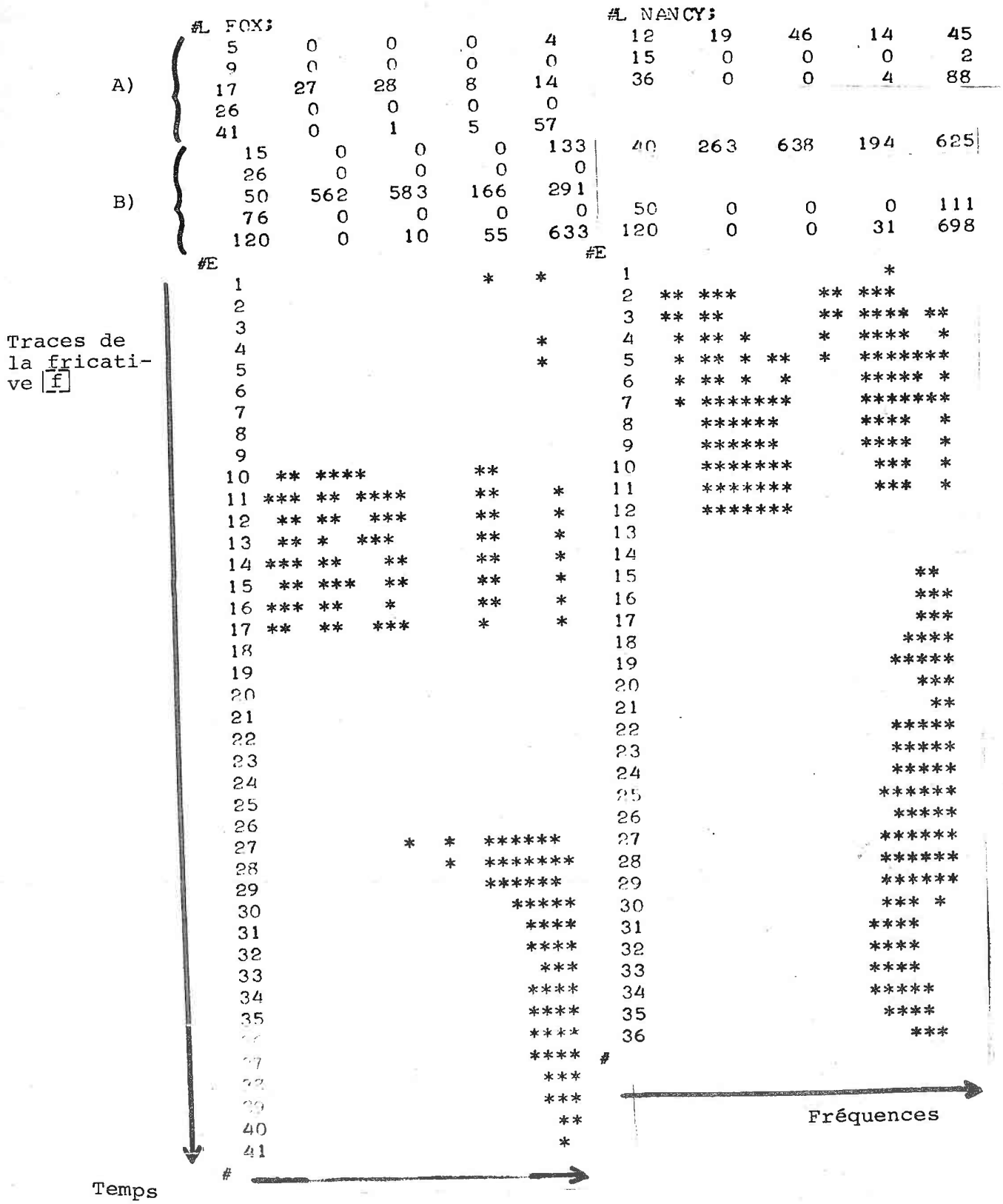


Figure 15 (début)

CONSTRUCTION

```
#L PARIS;
   9     16     36      9     17
  15      0      0      0      0
  24      0      0     16     25
    45    296    666    166    314
    75      0      0      0      0
   120      0      0    296    462
#E
   1   **  ******        **
   2  ***  **  ***       **     **
   3  ***  ****          **      *
   4       ****   *      **      *
   5       ****          **      *
   6       ******        **      *
   7       *****         **      *
   8       *****         **      *
   9        ****         **
  10
  11
  12
  13
  14
  15
  16                   ****
  17                  ******
  18                  ** ***
  19                   ******   *
  20                   *****    *
  21                    ****
  22                    *****
  23                      ***
  24                       *
#
```

Fréquences

RECONNAISSANCE

```
#E
 35
  1       *******        **
  2       * *****         *
  3  ***  * *****        **      *
  4  ***  **    **        **      *
  5   **  **   ***        **      *
  6   **  **    **        **     **
  7   **  **   *          **     **
  8   **  ** **           **    * *
  9   **  ** *            **  ****
 10   **                        ***
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21                  *      *
 22                *** ****
 23                   * ****
 24                   ******
 25                *  ******
 26                   ******
 27                     ***
 28                    ****
 29                    ****
 30                    ****
 31                    ****
 32                     ***
 33                     ***
 34                     ***
 35                     **
```

Fréquences

C)

```
#R
 35 PARIS;    909  NANCY;    907  FOX;    893
```

NOTA : On remarque l'absence de la fricative [f] d'où résulte la ressemblance avec "Nancy" et "Paris".

FIGURE 15. (fin)

En fait, comme le confirme l'expérience, cet indice est un médiocre discriminateur pour les pavés des zones temporelles étroites qui sont affectées davantage que les zones larges par l'imprécision des limites. On doit donc accorder plus d'importance à l'indice de similarité pour les zones larges.

Nous avons ainsi pondéré les degrés d'appartenance des pavés par un coefficient proportionnel à la longueur de la zone considérée. Ce coefficient, qui marque la confiance accordée à la participation de cette zone pour le calcul de la similarité, intervient selon l'expression (74) du nouvel indice.

Cette procédure revient à moduler l'indice de similarité brut proportionnellement à la durée relative des zones temporelles du mot pris sous la forme stockée en référence.

Pour un mot, l'indice brut (76) de comparaison entre les sous-ensembles flous $(\underset{\sim}{B})$ et $(\underset{\sim}{B}')$ est, avec le nombre l de zones temporelles j de n_j échantillons

$$S(\underset{\sim}{B},\underset{\sim}{B}') = 1 - \frac{\sum_{i,j} |\mu_{\underset{\sim}{B}}(x_{ijm}) - \mu_{\underset{\sim}{B}'}(x_{ijm})|}{l \times CRIT}$$

$$= \frac{1}{l} = \left[1 - \frac{\sum_{i,j} |\mu_{\underset{\sim}{B}}(x_{ijm}) - \mu_{\underset{\sim}{B}'}(x_{ijm})|}{CRIT} \right]$$

$$= \frac{1}{l} \sum_{j=1,l} S_j(\underset{\sim}{B},\underset{\sim}{B}') = \frac{1}{l} \sum_{j=1,l} q_j \qquad (78)$$

où $S_j(\underset{\sim}{B},\underset{\sim}{B}')$, abrégé selon q_j, est l'indice de similarité pour la zone j.

Dans la modulation proposée, nous remplaçons S

par l'indice modulé

$$S'(\underset{\sim}{B},\underset{\sim}{B}') = \sum_{j=1,l} \frac{n_j\, q_j}{n} \qquad (79)$$

n étant le nombre total des échantillons du mot.

Pour apprécier l'amélioration apportée par cette modulation, considérons le cas où, pour simplifier, elle est appliquée à une seule zone temporelle t d'un mot, dont par ailleurs les autres zones sont supposées égales entre elles.

L'indice modulé S' améliore S si S< S', ce qui donne dans ce cas :

$$(q_t + \sum_{j=1,l} q_j) < \frac{n_t\, q_t}{n} + \sum_{\substack{j=1,l \\ j\neq t}} \frac{n_j\, q_t}{n}$$

En posant $K_1 = \sum_{\substack{j=1,l \\ j\neq t}} q_j$

et $K_2 = \sum_{\substack{j=1,l \\ j\neq t}} \frac{n_t\, q_t}{n}$

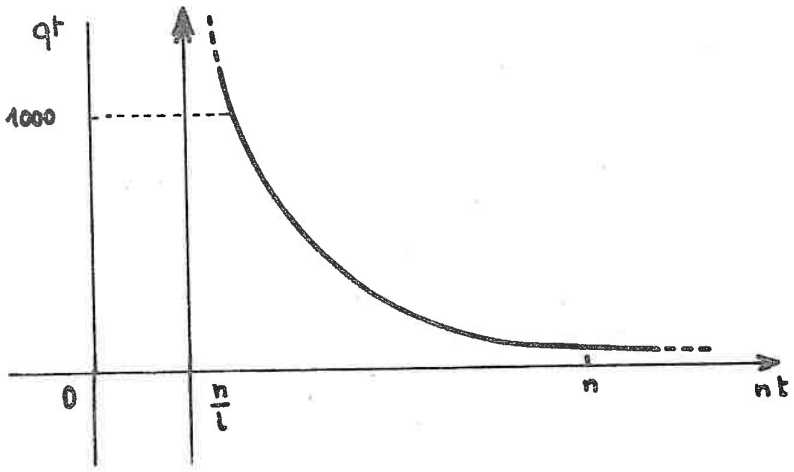
on en déduit :

$$\frac{K_1 + q_t}{l} < K_2 + \frac{n_t\, q_t}{n}$$

soit :

$$q_t > \frac{(K_1 - lK_2)\, n}{l n_t - n} \qquad (80)$$

La partie utile de l'hyperbole de discrimination correspondante est représentée sur la figure 16 avec q_t en %.



Cette partie utile est évidemment limitée par la valeur maximale de l'indice et $n_t = n$.

Pour le mot de longueur 17 et dont les zones temporelles ont pour longueur et indice

$$n_t = \begin{cases} 7 & \text{et} & 950 \\ 1 & \text{et} & 900 \\ 6 & \text{et} & x \\ 3 & \text{et} & 800 \end{cases}$$

x étant la valeur minimale recherchée pour qu'il y ait amélioration.

On a :

$K_1 = 2650$
$l = 4$
$K_2 = 585,3$

d'où $\underline{q_t = 775}$

Au cas où la comparaison porte sur le même mot, une telle valeur a toutes les chances d'être dépassée, de telle sorte que la reconnaissance correcte est ainsi renforcée par la modulation. En-dessous de cette valeur, l'absence d'amélioration est sans conséquence puisqu'alors la discrimination est trop faible pour être significative.

Si, au contraire, la comparaison s'effectue sur des mots différents, un indice de similarité inférieur est fréquent et la modulation, en accentuant la différence, augmente la discrimination.

La retouche de l'indice de similarité par la modulation reste numériquement toujours faible puisqu'avec

$q_t = 600$ S = 815 et S' = 797
et avec $q_t = 900$ S = 887 et S' = 903.

Elle est cependant suffisante en pratique pour améliorer la discrimination comme le prouvent les résultats suivants.

B - Amélioration apportée par la modulation

L'amélioration apportée par la modulation a été appréciée avec le lexique L50 enregistré sur deux séquences.

Dans deux cas, le mot à reconnaître, qui n'était pas même en troisième position avec l'indice brut, apparaît en première position avec l'indice modulé.

Pour le détail qui suit, nous nous limitons aux cas des mots classés dans les trois premières positions.

A titre d'exemple, le mot "November" est mal reconnu avec l'indice brut, classé deuxième seulement après "Alpha" et à peu près à égalité avec "Quintal". Avec l'indice modulé, la reconnaissance est bonne puisque "November" repasse en première position.

Le tableau 7 explicite pour ce cas le détail de l'amélioration apportée, par la mise en évidence de la longueur relative des segments, des indices bruts et modulés correspondants et des indices globaux. En encadré figure pour "November" et "Alpha" le segment le plus principalement concerné par la modulation, une fois dans un sens et une fois dans l'autre.

La figure 17,qui représente le cas précédent de la même manière que la figure 15, montre que, si la confusion de "November" avec "Alpha" est concevable, il n'en est pas de même pour "Quintal" ; il s'agit là d'une faiblesse locale de l'utilisation des indices bruts lorsque les écarts entre degrés d'appartenance des pavés se compensent en moyenne, mais ce cas est rare.

MOT	Segmentation relative	Durée du mot en ms	Durée relative	Indice brut par zone temporelle	Indice brut moyen	Indice modulé par zone temporelle	Indice modulé
"November"	16	530	16	886	907	117	898
	29		13	916		104	
	48		19	890		134	
	79		31	970		256	
	102		23	916		173	
	118		16	869		114	
"Alpha"	28	390	28	906	926	209	896
	49		21	899		161	
	62		13	983		101	
	89		27	950		219	
	117		28	895		206	
"Quintal"	38	350	38	882	908	277	884
	75		37	913		288	
	117		42	929		319	

TABLEAU 7.

Pour le mot "Fox" déjà rencontré mal reconnu (indice brut 882) derrière "Nancy" (908) et "Paris" (896), l'indice modulé lui donne la première position (901) devant "Nancy" (881) et "Paris" (848), la perte de l'information sur la fricative étant moins élevée.

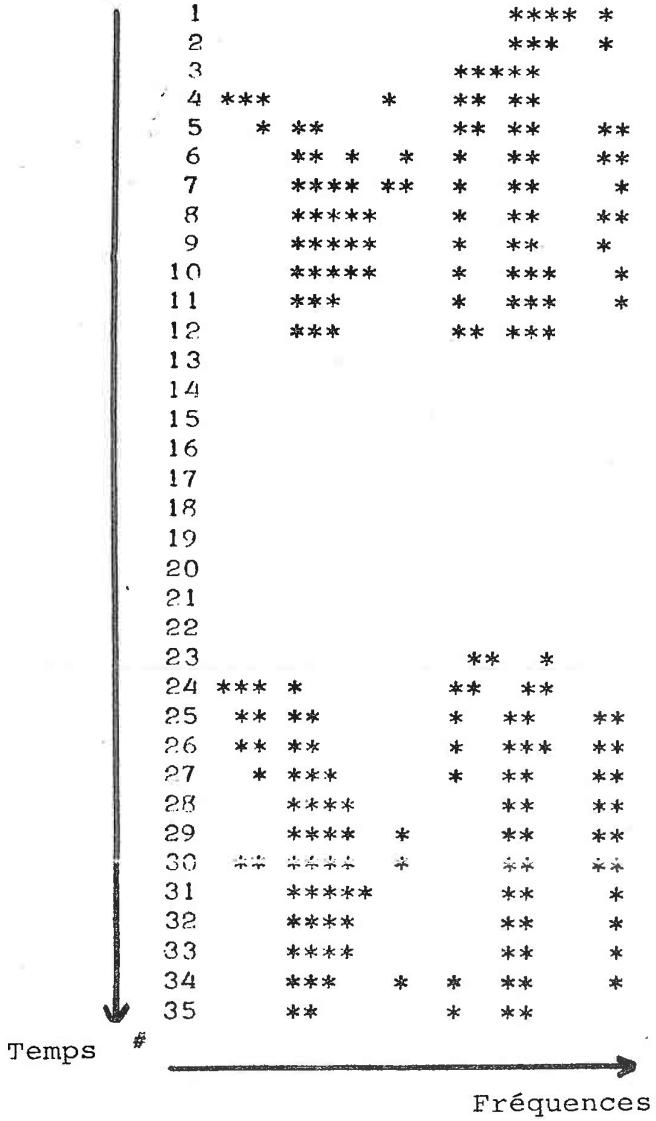
En résumé, les cas de dégradation de la reconnaissance étant rares, en particulier aucun déclassement depuis la première position n'ayant été relevé, la reconnaissance globale avec l'indice modulé est de 96 % alors qu'avec l'indice brut, il n'était que de 86 % pour le même lexique.

CONSTRUCTION

#L QUINTAL;

A	2	0	0	2	7
	12	12	26	25	23
	22	0	0	0	0
	35	22	29	20	31
B	7	0	0	166	583
	41	199	433	416	383
	75	0	0	0	0
	120	281	371	256	397

#E



Temps #

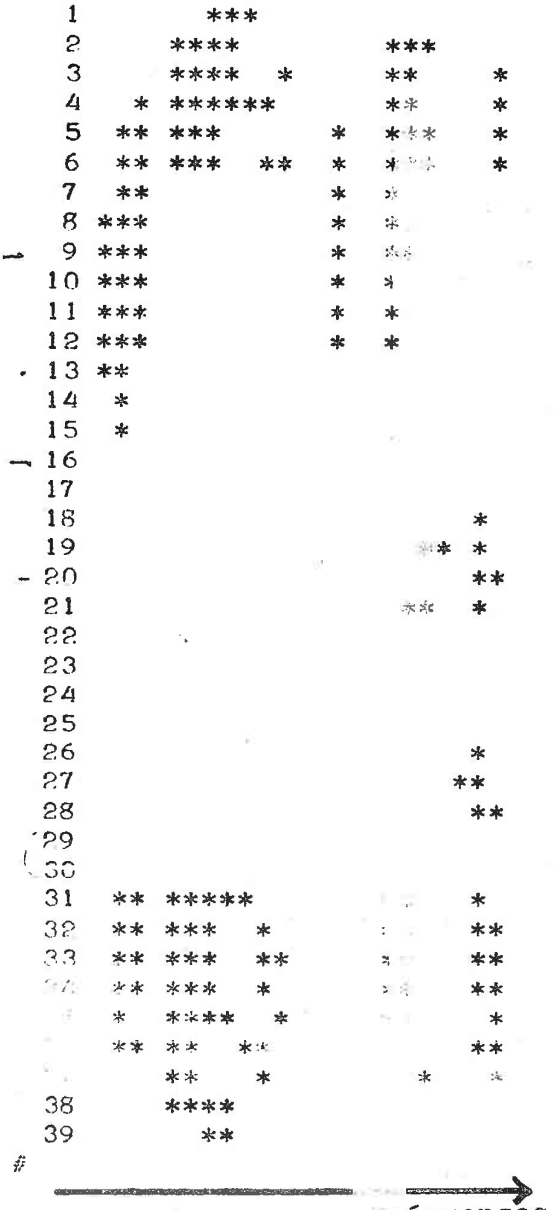
Fréquences

CONSTRUCTION

#L ALPHA;

6	10	21	7	12
13	19	0	12	5
18	2	0	0	1
21	0	0	0	8
26	0	0	0	1
29	0	0	0	4
30	0	0	0	0
39	19	28	8	22
18	277	583	194	333
40	452	0	285	118
55	66	0	0	33
65	0	0	0	444
80	0	0	0	33
89	0	0	0	222
92	0	0	0	0
120	351	518	147	407

#E



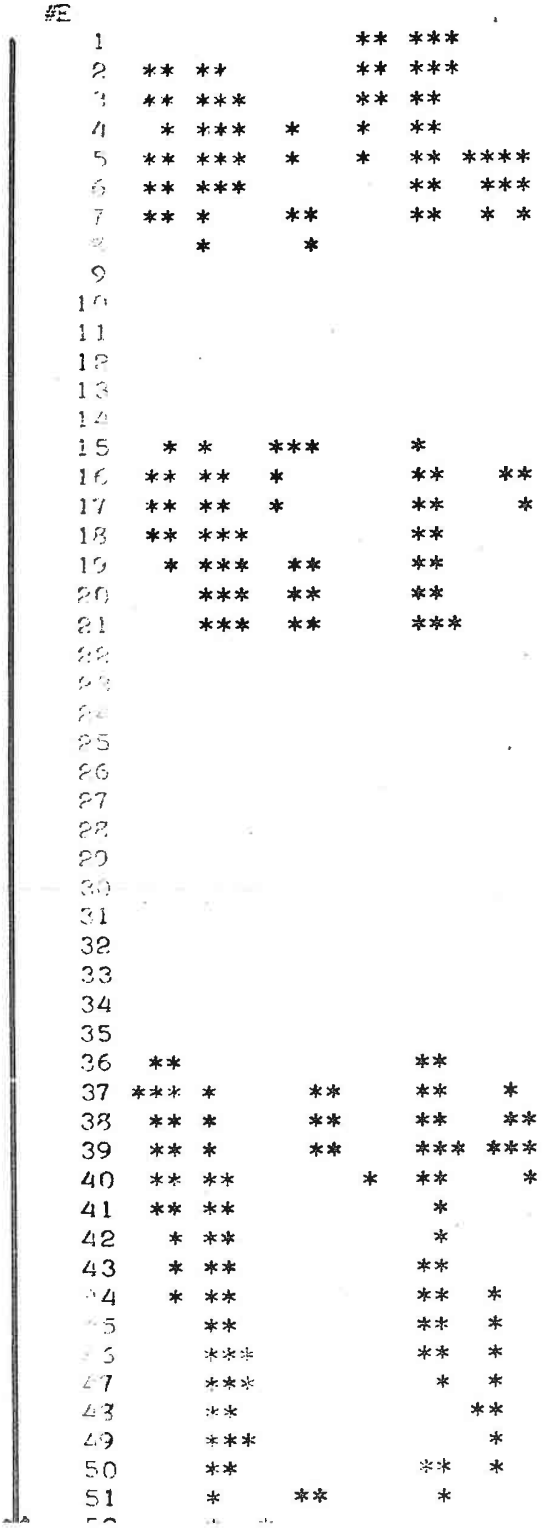
#

Fréquences

Figure 17 (début)

L NOVEMBER : CONSTRUCTION

8	18	14	15	18
14	0	0	0	0
21	15	21	7	10
35	0	0	0	0
42	24	8	11	18
56	8	12	3	12
77	375	291	312	375
98	0	0	0	0
119	356	500	166	238
140	0	0	0	0
161	444	147	203	333
182	166	250	62	250



RECONNAISSANCE

54 ALPHA 926
QUINTAL908
NOVEMBER 907

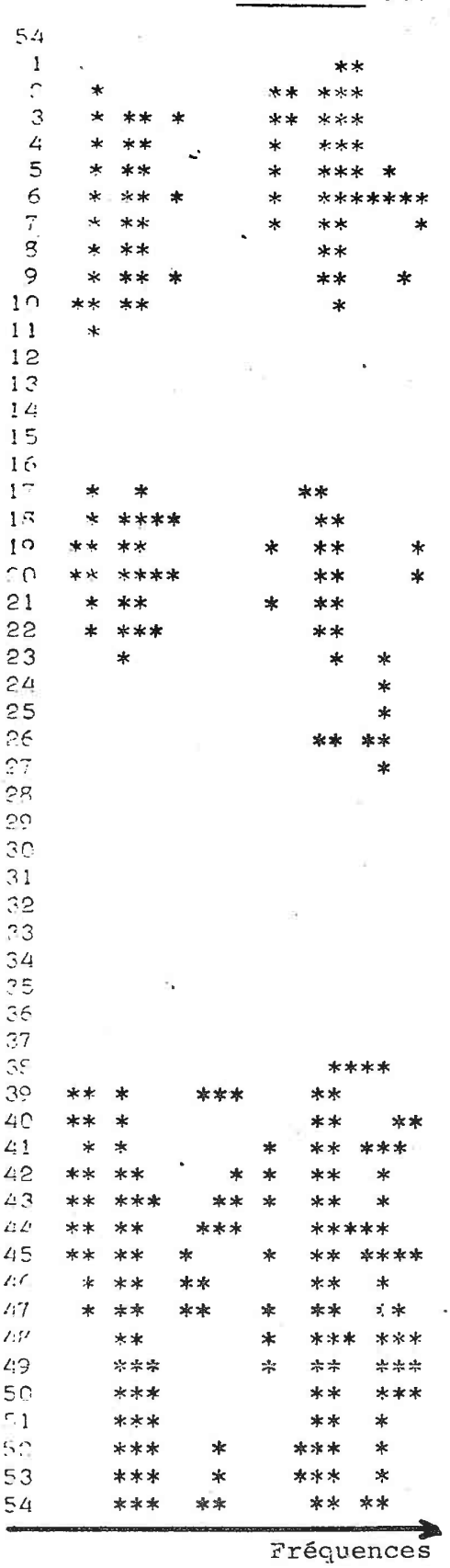


FIGURE 17 (fin)

Par ailleurs, les 20 mots les plus mal reconnus avec l'indice brut, repris en lexique exclusif avec l'indice modulé, ont été reconnus à 100 % suivant les résultats du tableau 8.

ALPHA;	927	DELTA;	917	LAPIN;	907
BRAVO;	882	TANGO;	864	LAPIN;	836
DELTA;	920	LAPIN;	891	ALPHA;	872
DESIRE;	936	KILO;	801	HOTEL;	796
EUGENE;	885	JULIETTE;	886	YVONNE;	839
FOX;	867	PARIS;	853	NANCY;	833
GOLF;	898	TANGO;	837	BRAVO;	827
HOTEL;	906	QUINTAL;	860	YVONNE;	823
JAMBON;	918	SAPIN;	819	LAPIN;	818
JULIETTE;	936	YVONNE;	855	EUGENE;	848
KILO;	871	SAPIN;	833	DESIRE;	797
LAPIN;	935	DELTA;	906	NOVEMBER;	849
NANCY;	926	PARIS;	835	GOLF;	808
NOVEMBER;	898	LAPIN;	857	BRAVO;	845
PAPA;	913	LAPIN;	880	DELTA;	876
PARIS;	898	NANCY;	842	FOX;	838
QUINTAL;	892	PAPA;	877	ALPHA;	877
SAPIN;	919	JAMBON;	875	LAPIN;	838
TANGO;	904	BRAVO;	870	GOLF;	813
YVONNE;	913	JULIETTE;	832	JAMBON;	820

TABLEAU 8.

3) DISTANCE DE HAMMING ET DISTANCE EUCLIDIENNE.

L'indice $S(\underset{\sim}{B},\underset{\sim}{B}')$ de similarité entre B et B' écrit en (76) avec la distance de Hamming et utilisé tel quel depuis, peut aussi s'exprimer avec la distance euclidienne (58),

ce qui donne :

$$S(\underset{\sim}{B},\underset{\sim}{B}') = 1 - \frac{1}{CRIT \times l} \sum_{ij} |\mu_B(x_{ijm}) - \mu_{B'}(x_{ijm})|^2 \quad (81)$$

Si, dans les calculs sur les vecteurs à composantes binaires, il est indifférent d'utiliser la distance euclidienne ou la distance de Hamming, il n'en est plus de même pour les vecteurs flous comme c'est le cas ici.

Nous avons donc repris la reconnaissance avec la distance euclidienne ; avec le lexique L80 du tableau 3, nous avons constitué cinq séquences, l'une de construction, les autres servant de reconnaissance en utilisant 4 critères, puis 2 et 1.

Le tableau 9 résume les résultats ainsi obtenus et montre qu'ils sont absolument comparables aux précédents.

Pour la suite, nous retenons donc l'indice modulé déterminé par la distance de Hamming dont le calcul est plus rapide.

	taux de reconnaissance en % / mot reconnu dans les I premiers classés	Indice modulé distance de Hamming	Indice modulé - distance Euclidienne
4 critères	I = 1	95	95
	I = 2	96,9	97,5
	I = 3	97,2	97,5
2 critères	I = 1	93,8	90,2
	I = 2	95	96,3
	I = 3	96,3	96,3
1 critère	I = 1	80	56,1
	I = 2	91,3	73
	I = 3	91,3	80

TABLEAU 9

4) NOMBRE DE CRITERES ET TAUX DE RECONNAISSANCE.

Les résultats du tableau 9 montrent aussi que la reconnaissance correcte est à peine dégradée pour le lexique L80, quand on passe de 4 à 2 critères.

Il est même nécessaire de remarquer qu'avec un seul critère, les résultats restent très satisfaisants en distance de Hamming surtout si la décision s'effectue sur le classement dans la première ou dans les deux premières positions.

Cette dernière conclusion souligne l'importance de la seule prise en considération des contours globaux de la forme à reconnaître, sans qu'interviennent les détails dans le "flou" général.

Remarque : le résultat médiocre donné par l'usage de la distance euclidienne avec 1 critère est sans doute à attribuer aux troncatures numériques dans les élévations au carré.

5) TRONCATEUR DES DEGRES D'APPARTENANCE.

Pour gagner de la place en mémoire et au prix d'une troncature, nous avons placé les degrés d'appartenance μ pour les critères retenus au nombre de 4, sur un seul mot-mémoire (au lieu de 4 de 12 bits chacun). La troncature correspondante est représentée sur la figure 18 en fonction du nombre de bits utilisés pour chaque μ .

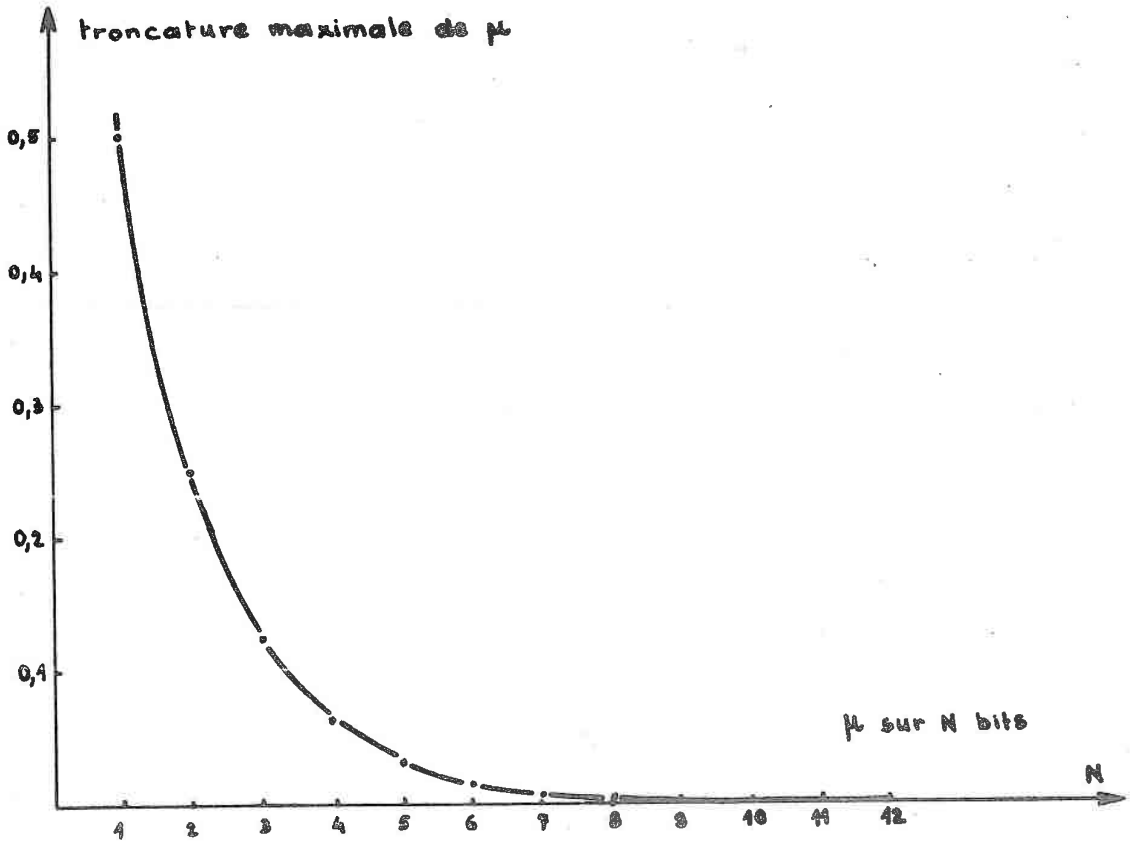


Figure 18.

On peut penser qu'une telle troncature, même assez sévère, n'apporte pas de perturbations notables dans la reconnaissance, puisque pous les sous-ensembles flous à reconnaître, les degrés d'appartenance sont justement assez grossièrement appréciés comme résultants d'une description sommaire des pavés ordonnés par définition.

Pour apprécier les conséquences de la troncature, nous avons effectué la discrimination avec l'indice brut, l'effet sur les résultats qui restent toujours excellents pour l'indice modulé n'étant pas suffisamment significatifs pour en tirer une conclusion nette.

Un extrait des résultats pour le lexique L80 est représenté sur le tableau 10.

nombre de critères	nombre de bits pour μ	taux de bonne reconnaissance en %
4	12	91,3
4	4	91,3
4	3	90
4	2	87,5
4	1	33,8
2	12	86,3
2	4	86,3
2	3	86,3
2	2	57,5
2	1	17,5

TABLEAU 10

Nous opérons dans la suite avec la distance de Hamming pour 4 critères et des degrés d'appartenance de 4 bits.

Dans ces considérations, un seul mot mémoire stocke l'information pour les quatre pavés d'une zone temporelle. Le gain de place qui en résulte permet d'opérer sur des lexiques relativement importants. C'est ainsi que, pour le lexique L175 de 175 mots (tableau 11), nous avons obtenu 90,3 % de mots reconnus en première position. Une seconde séquence effectuée aussitôt après n'a cependant donnée que 85,7 %, diminution à attribuer à la seule fatigue de l'opérateur.

Les résultats concernant L175 permettent d'étendre les courbes de variation du taux de reconnaissance moyen en fonction de la taille du lexique comme sur la figure 19.

6) TEMPS MOYENS DE REPONSE.

Le temps de réponse en fonction de la longueur du mot reconnu est donné par la figure 20 pour les lexiques L175 et L80, avec la mise en évidence de l'incidence très faible du calcul par l'indice modulé.

La figure 21 compare ces résultats pris en moyenne avec ceux de J.-P. HATON en programmation dynamique dans les domaines communs de l'importance des lexiques respectifs, et dans les mêmes conditions d'expérimentation et de calcul avec, dans notre cas, le gain d'un facteur 5 sur l'importance du lexique stocké.

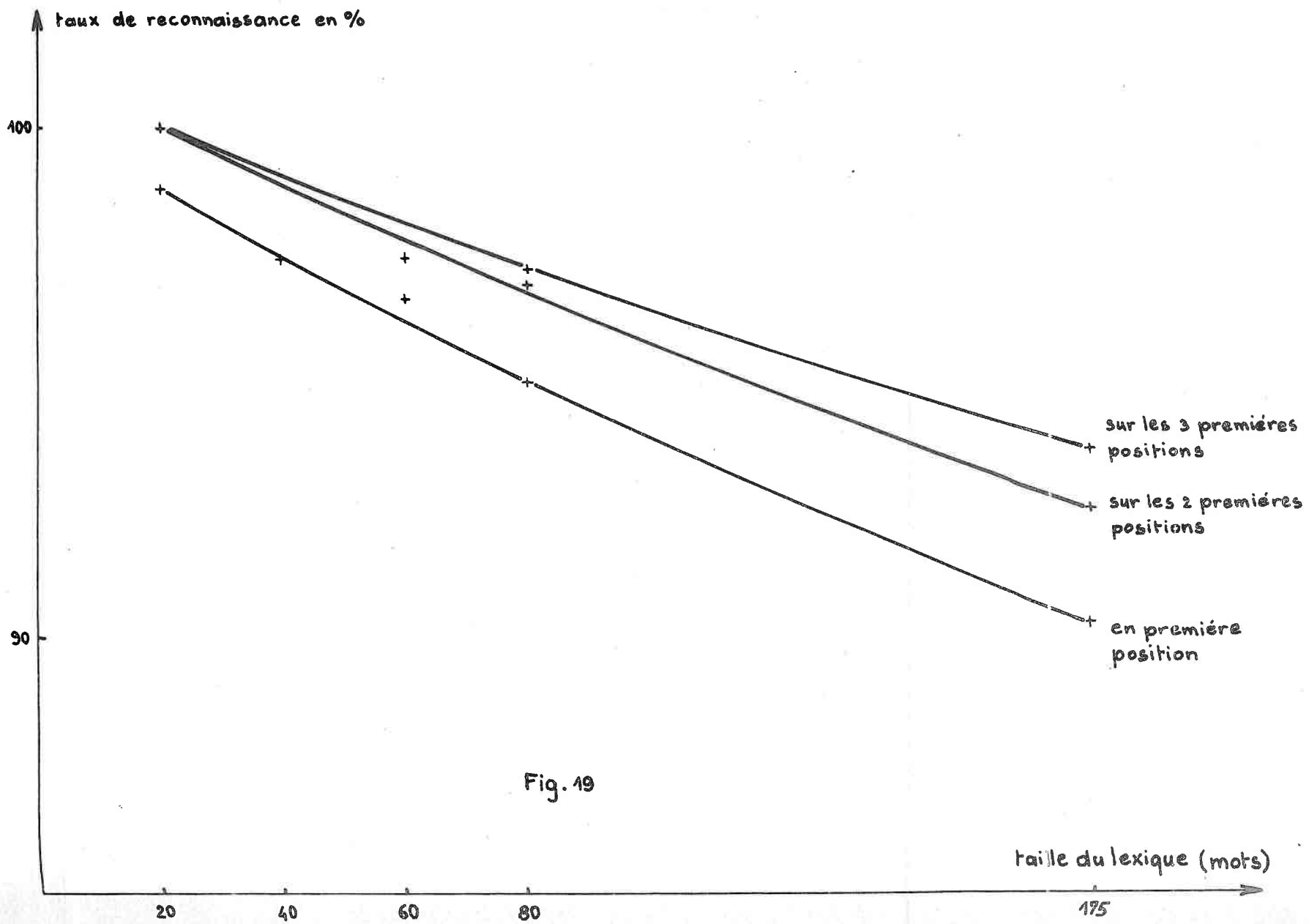

taux de reconnaissance en %
100
90
sur les 3 premières positions
sur les 2 premières positions
en première position
taille du lexique (mots)
20
40
60
80
175

Fig. 19

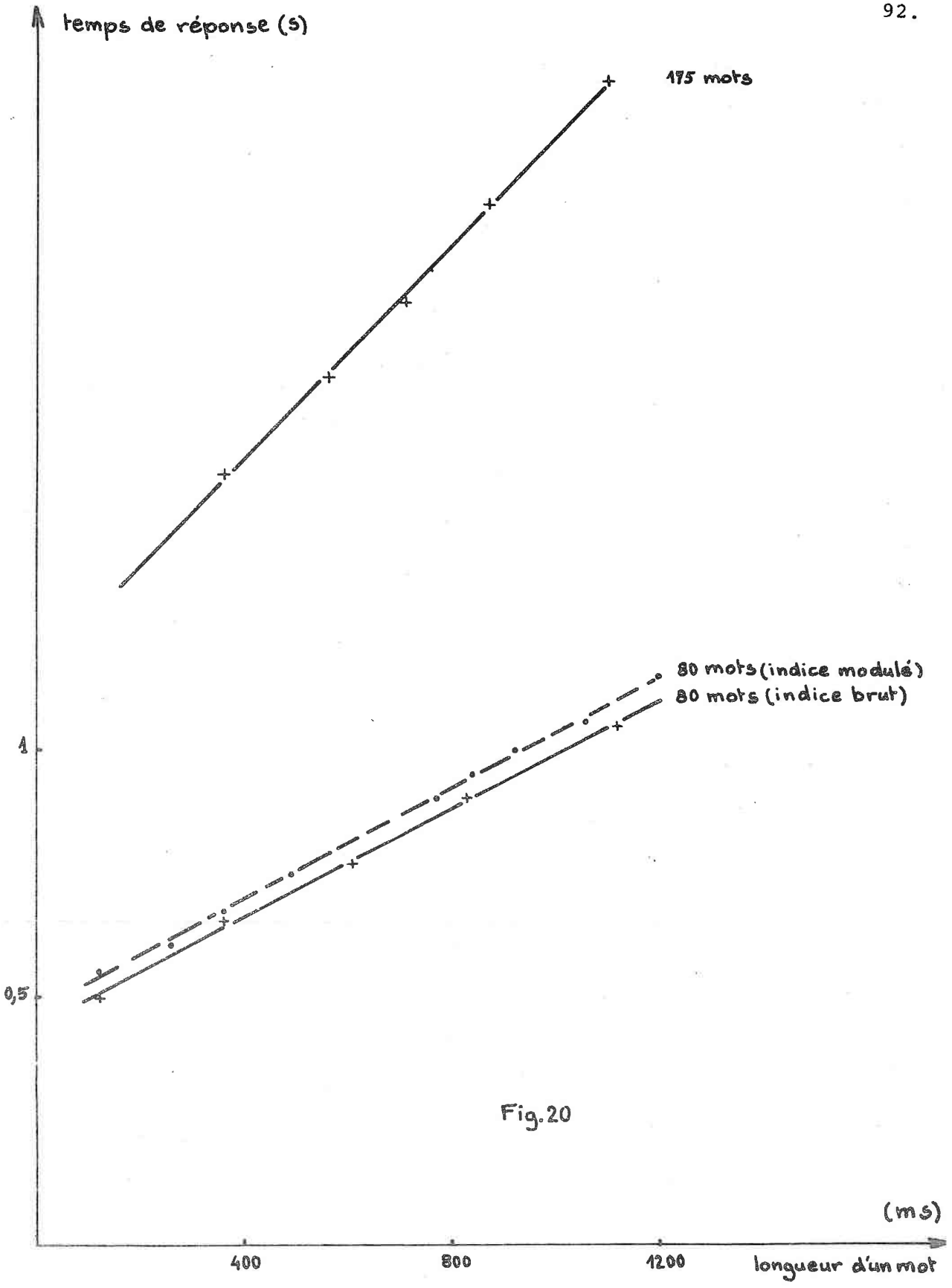

temps de réponse (s)
175 mots
80 mots (indice modulé)
80 mots (indice brut)
1
0,5
(ms)
400
800
1200
longueur d'un mot

Fig. 20

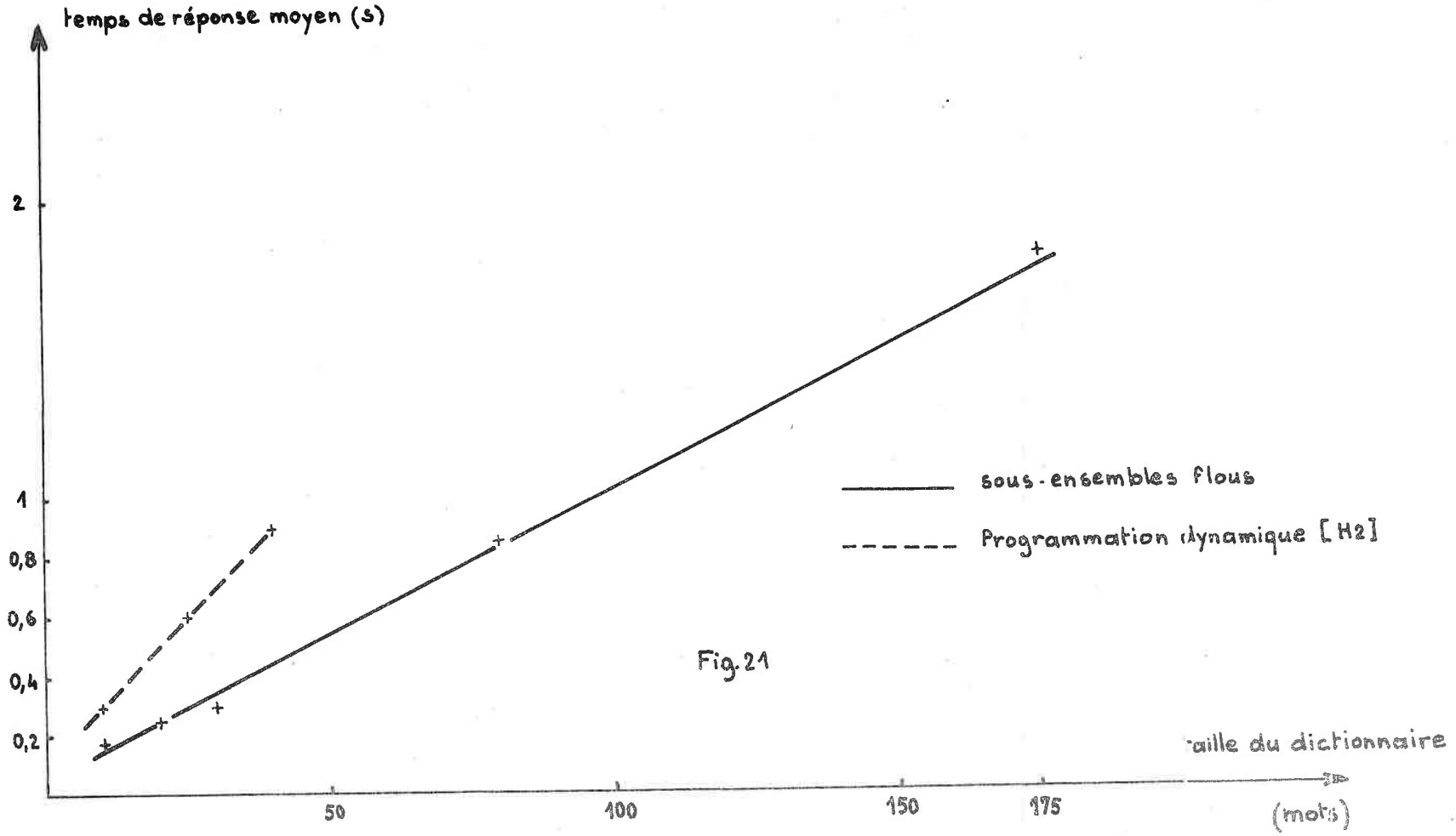

temps de réponse moyen (s)
2
1
0,8
0,6
0,4
0,2
50
100
150
175
(mots)
sous-ensembles flous
Programmation dynamique [H2]

Fig. 21

D
AIGUILLON; ALBERVILLE; ALENCON; AMBOISE; AMIENS;
ANTIBES; ARCIS; ARGENTAT; ARPAJON; AUBUSSON;
AUTUN; AUXERRE; AZAY; BAGNOLS; BAR LE DUC;
BARENTON; BEAUGENCY; BEAULIEU; BEAUVAIS; BEAUVAL;
BERGERAC; BESANCON; BONNEVILLE; BOULOU; BOURDEAU;
BREHAT; LABRESSE; CALVI; CANCALE; CANNES;
CARCASSONNE; CARNAC; CAVAILLON; CERNAY; CHALAIS;
CHALONS; CHAMBON; CHAMONIX; CHAMPAGNOLE; CHAPELLE;
CHARITE; CHATEAUBRIAND; CHAUNAY; CHINON; CLUNY;
COMMERCY; CONCARNEAU; CONDE; CORCIEUX; CORNIMONT;
COTEAU; COURTISOLS; CRAVANT; DEAUVILLE; DIJON;
DINAN; DINARD; DORMANS; EPERNAY; ETRETAT;
FALAISE; FERTE; FONTAINE; FUMEL; GERARDMER;
GRANDPRE; GRANVILLE; GRIMAUD; GUERET; HENDAYE;
HIRSON; ILE D YEU; ISSOIRE; ISSOUDUN; JOIGNY;
JOYEUSE; LACANAU; LANGEAIS; LANGRES; LAPALISSE;
LASALLE; LAVANDOU; LESPERON; LIVERDUN; LUNEVILLE;
LUXEUIL; MARCON; MARSEILLE; MASSIAC; MATIGNON;
MAUBEUGE; MAUREPAS; MAYENNE; METZ; MEURSAULT;
MONSEGUR; MONTARGIS; MONTAUBAN; MONTESQUIEU; MONTJOUX;
MONMEDY; MONTRESOR; MORLAIS; MORTAIN; NANCAIS;
NANTUA; NEVERS; NICE; NIORT; NUBECOURT;
ORCHAMPS; PAIMPOL; PALAVAS; PARIS; PERIGUEUX;
POISSON; POITIERS; PONCHON; PONCIN; PONTARLIER;
PONCHATEAU; PORNIC; PORNICHET; PRIVAS; QUIBERON;
QUIMPER; QUINTIN; REVIN; ROANNE; ROCHE;
ROHAN; ROUSSES; SABLES; SAILLANT; SAINTES;
SALBRIS; SALLANCHES; SALON; SAMOENS; SAMOREAU;
SAVENAY; SAXEL; SEDAN; SELESTAT; SERANON;
SEYSSEL; SIGEAN; SIOUVILLE; SISCO; SOUSTONS;
STRASBOURG; SULLY; TANINGES; TARASCON; TERRASSON;
THEOULE; THIONVILLE; THOUARS; TILLOU; TONNERRE;
TOULOUSE; TREVES; TRILPOT; USSEL; USSY;
VALADY; VALENCE; VALLAURIS; VAUCOULEURS; VENCE;
VERNOUILLET; VERSAILLES; VIERZON; VILLEFRANCHES; VILLERS;

#

Lexique L175

TABLEAU 11

CHAPITRE V

GENERALISATION AU CAS DE PLUSIEURS LOCUTEURS

Jusqu'à présent, c'est le même locuteur qui opérait pendant les phases de construction et de reconnaissance. Nous nous proposons maintenant d'étudier les conséquences de l'intervention de plusieurs locuteurs,d'abord avec une seule construction préalable, puis avec un apprentissage en cours de reconnaissance.

1) CAS D'UNE CONSTRUCTION.

Le sous-ensemble flou d'une seule occurrence de chaque mot du lexique L80 prononcé par le locuteur 1 est rangé en mémoire. Puis 7 autres locuteurs masculins et 2 féminins répètent la liste.

Les résultats de la reconnaissance sont consignés dans le tableau 12 reproduit sous la forme de la courbe de la figure 22.

Pour le locuteur 1 qui a construit le lexique de référence, la reconnaissance est franche.

Pour l'apprentissage fait par le locuteur 1, variation des taux de reconnaissance pour les 40 locuteurs / 80 mots

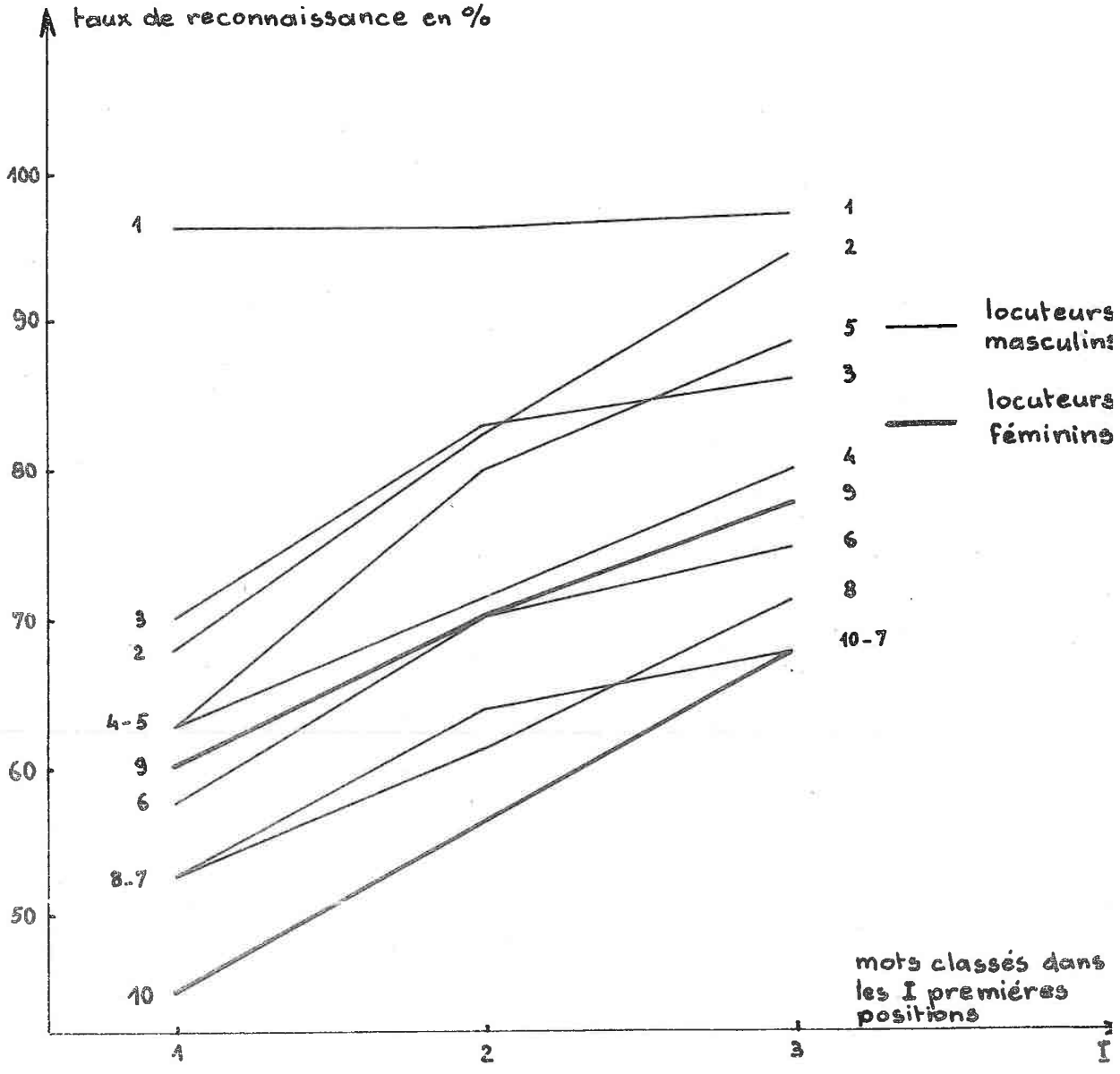


Fig. 22

locuteurs non entraînés

Mots classés dans les I premières positions / locuteur	+ 1	+ + 2	+ + + 3	4	5	6	7	8	+ + + + 9	10	Sous-ensemble de L 80
I = 1	100	75	90	80	80	85	75	65	60	70	20 mots
I = 2	100	90	95	85	95	85	80	70	80	75	
I = 3	100	90	95	85	100	85	80	70	85	75	
I = 1	97,4	72,5	80	70	70	65	65	55	72,5	45	40 mots
I = 2	97,4	87,5	90	82,5	85	74,5	70	67,5	85	62,5	
I = 3	97,4	90	90	85,5	95	75	70	80	85	65	
I = 1	96,6	70	73,3	63,3	66,7	61,7	48,3	61,7	61,7	36,7	60 mots
I = 2	97,4	88,3	80	73,3	88,3	71,7	65	75	75	55	
I = 3	97,4	91,7	81,7	80	93?3	76,7	65	80	78,3	65	
I = 1	96,2	67,8	70	62,5	62,5	57,5	52,5	52,5	60	45	80 mots
I = 2	96,2	82,5	82,5	71,3	80	70	63,8	61,3	70	56,3	
I = 3	97,5	92,5	86,3	80	88,8	75	67,5	71,3	77,7	67,5	

TABLEAU 12

++++ Locuteurs féminins -

+ La construction du lexique a été effectuée par ce locuteur.-

+ + Locuteurs particulièrement entraînés -

+++ Il s'agit de 2 séquences successives de reconnaissance effectuée par un même locuteur -

Les locuteurs entraînés 2 et 3 ont ensuite les meilleurs résultats.

Le locuteur 4-5 et le locuteur féminin 10, bien que non entraînés, les suivent de près.

Par ailleurs, la dégradation observée pour la reconnaissance lorsque la taille du lexique augmente, paraît plus rapide au début qu'ensuite.

Pour des lexiques supérieurs à 80 mots, il semble que la moyenne du taux de reconnaissance, pour l'ensemble des locuteurs, tende vers une limite d'environ 60 % (figure 23).

Pour apprécier la netteté de la reconnaissance, nous avons calculé pour l'ensemble des mots bien reconnus par chacun des locuteurs, l'indice moyen et sa dispersion donc en première position, et de même pour les mots donnés alors dans les deux et trois premières positions (tableau 13).

locuteurs	première position		2 premières positions		3 premières positions	
	moyen	dispersion	moyen	dispersion	moyen	dispersion
1	918	19	857	21	844	23
2	883	23	856	18	844	16
3	883	24	855	19	845	24
4	878	23	852	24	845	23
5	886	23	855	24	845	13
6	878	24	853	20	840	18

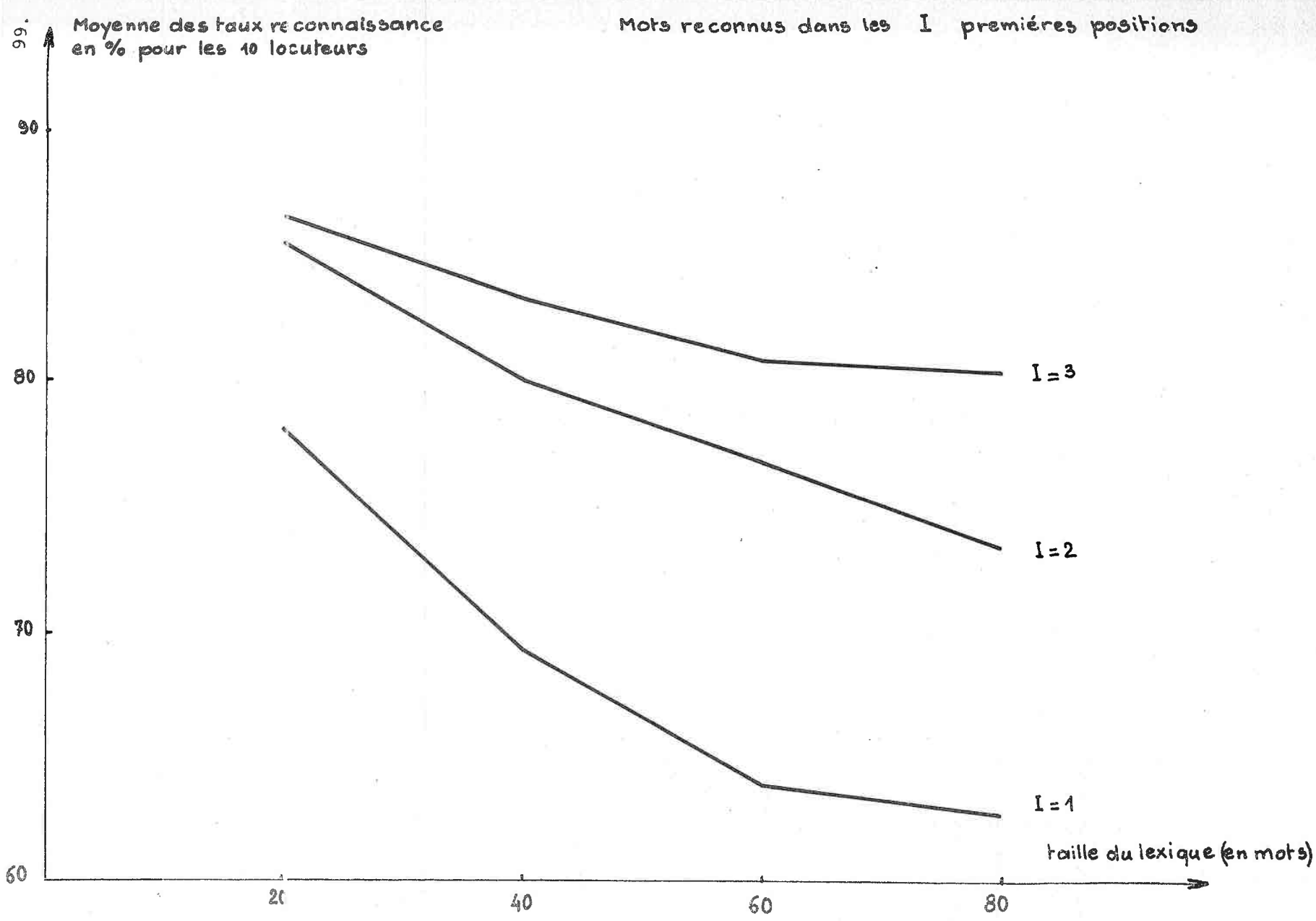

Moyenne des taux reconnaissance
en % pour les 10 locuteurs
Mots reconnus dans les I premiéres positions
99
90
80
70
60
I = 3
I = 2
I = 1
taille du lexique (en mots)
20
40
60
80

Fig. 23

7	879	24	859	30	842	21
8	876	22	852	17	842	17
9	880	21	859	27	844	17
10	886	24	851	22	841	22

TABLEAU 13.

Ces valeurs reportées sur la figure 24 montrent que la discrimination est : (1) ; (2,3,5,6,8) ; (4,7,9,10). A l'exception du locuteur 1, la discrimination n'est pas liée au taux de reconnaissance pour les autres locuteurs puisqu'elle n'en suit pas la variation et que rien, autre que ce taux, ne marque la distinction entre des locuteurs entraînés au pas ; ce fait souligne la complexité du problème de reconnaissance avec plusieurs locuteurs.

J.-P. HATON, en programmation dynamique, sur 20 mots reconnus à 100 % par un locuteur, a obtenu pour les pourcentages de reconnaissance cumulés jusqu'à 4 autres locuteurs reconnus de moins en moins bien, la courbe (1) de la figure 25.

Reprenant nos valeurs relevées dans les mêmes conditions, nous avons tracé pour le locuteur 1, suivi des autres locuteurs, la courbe (2).

Dans les conditions les plus défavorables, avec le locuteur 1 et nos 4 locuteurs les moins bien reconnus, nous avons encore obtenu la courbe 3.

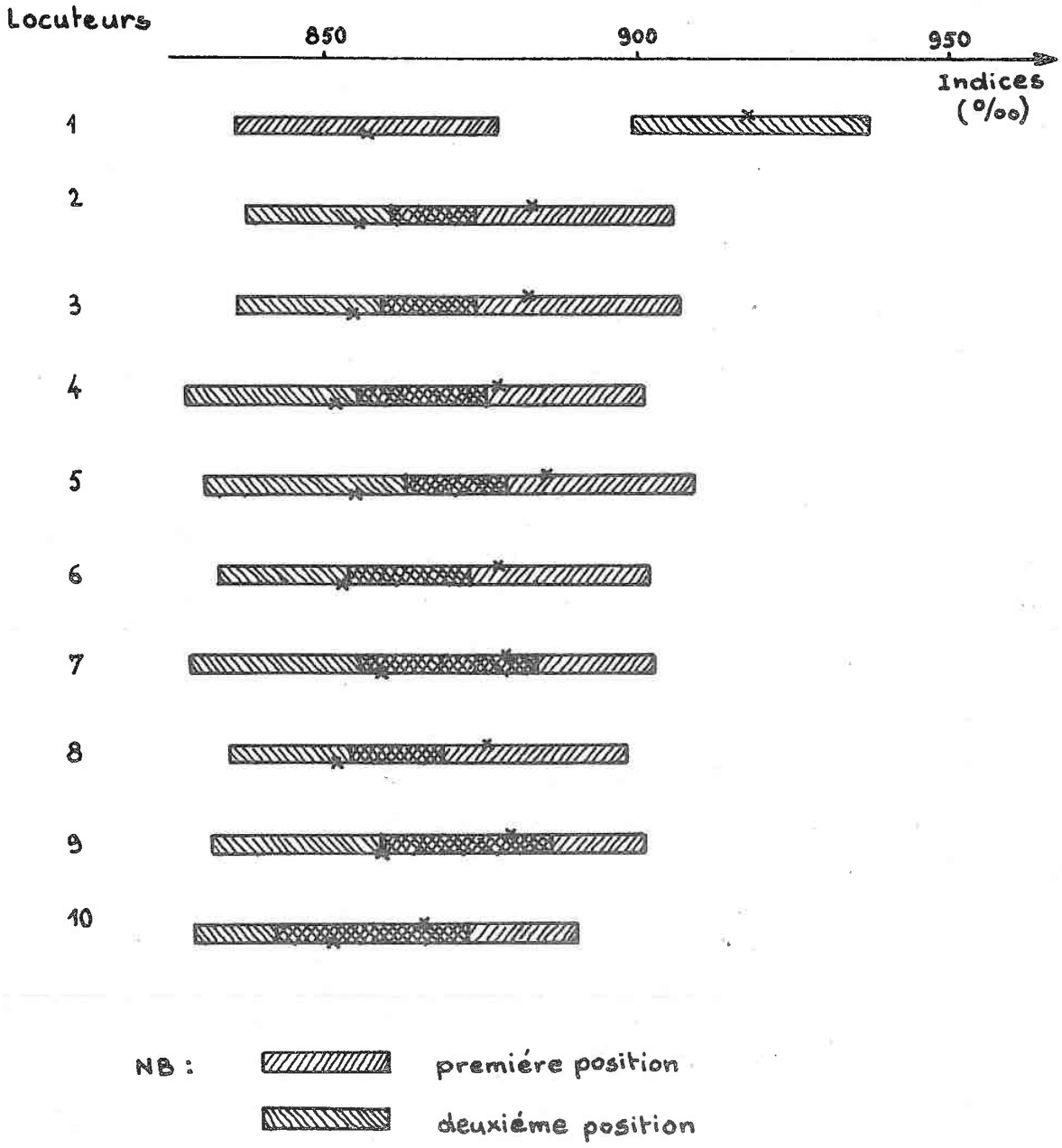

Locuteurs
850
900
950
Indices (‰)
1
2
3
4
5
6
7
8
9
10
NB : première position
deuxiéme position

Fig. 24

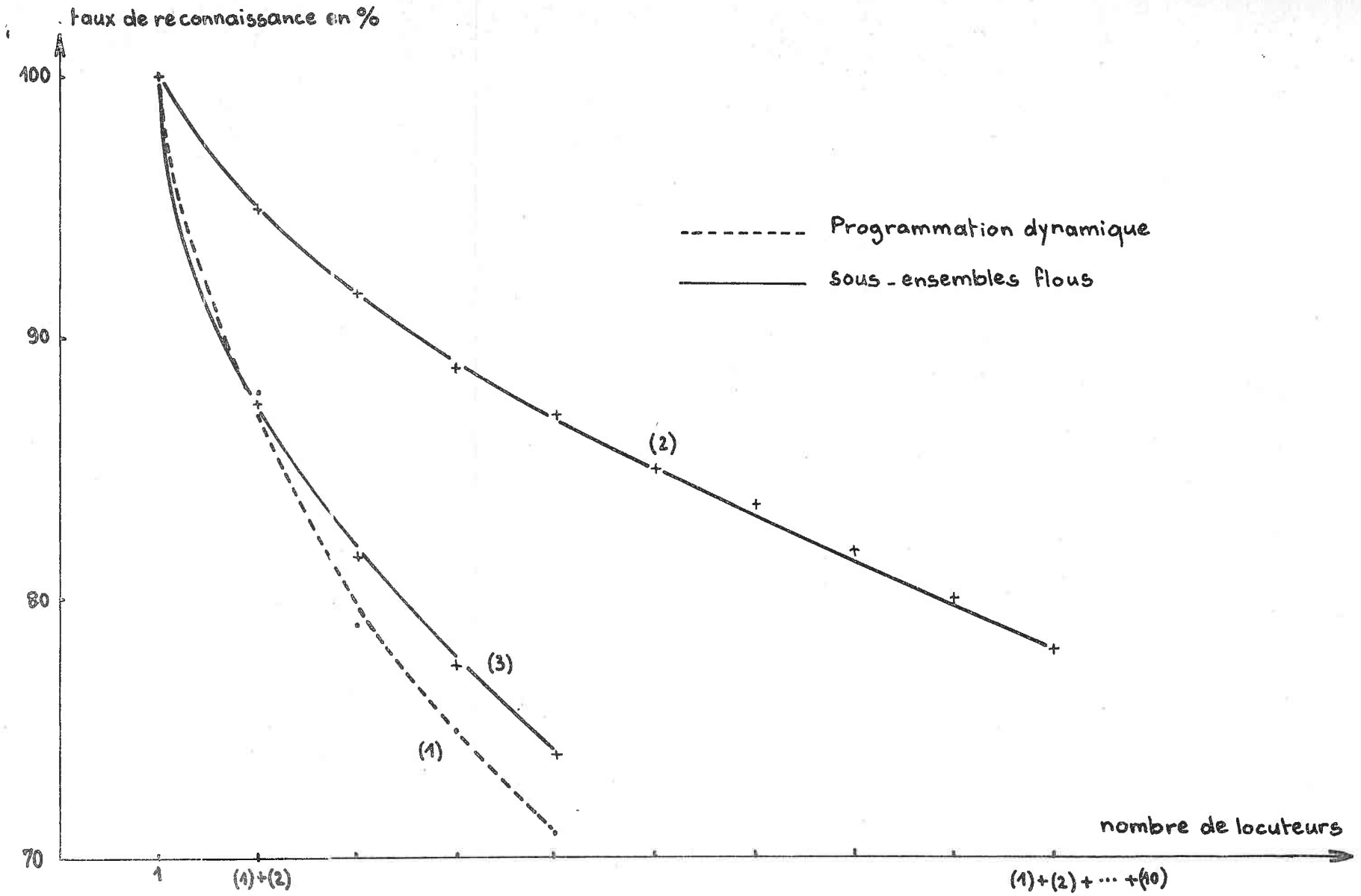

taux de reconnaissance en %
100
90
80
70
Programmation dynamique
sous-ensembles flous
(1)
(2)
(3)
nombre de locuteurs
1
(1)+(2)
(1)+(2)+ ··· +(10)

Fig. 25

La reconnaissance par les sous-ensembles flous donne donc des résultats nettement supérieurs.

2) ADAPTATION.

Pour améliorer la reconnaissance, en particulier pour plusieurs locuteurs, la méthode qui vient à l'esprit consiste à construire un lexique par locuteur, ou pour un seul locuteur à construire un lexique comportant plusieurs occurrences du même mot.

Mais, même dans le cas où on limite cette opération aux seuls mots mal reconnus, cette méthode encombre la mémoire au détriment évident de l'importance effective du lexique, sauf si la première occurrence du lexique. est purement et simplement remplacée par une autre.

Dans une autre méthode, nous proposons qu'en cas de reconnaissance incorrecte, à la constatation de l'erreur, l'opérateur agisse sur le système afin d'augmenter les chances d'une reconnaissance ultérieure correcte.

Pour celà, nous gardons la constitution de l'ensemble des pavés de référence, auxquels nous affectons de nouvelles valeurs des degrés d'appartenance, prises comme moyennes entre les degrés initiaux et les degrés pour le sous-ensemble flou du mot incorrectement reconnu. Ainsi, si ce mot est une nouvelle fois prononcé de la même manière, le sous-ensemble flou correspondant sera plus proche de la nouvelle référence.

Ce processus qui peut se répéter revient à modifier l'ensemble de référence dans le sens d'une adaptation, en particulier à la voix d'un nouveau locuteur. Dans ce dernier cas, le taux de reconnaissance pour le constructeur du lexique initial se dégrade évidemment, de telle sorte qu'il conviendra de limiter l'adaptation.

Ce qui précède suppose que la segmentation est invariable. Mais cette hypothèse est incertaine comme le montrent nos dernières expériences effectuées à degrés d'appartenance invariables.

Dans les deux cas, nous avons opéré dans des conditions volontairement médiocres au départ, avec un grand lexique de 175 mots et une élocution peu soignée.

Dans la méthode de remplacement des mots mal reconnus, on obtient avec un locuteur pour 3 séquences de reconnaissance, 66, 78 et 87 % de reconnaissance correcte. Si l'amélioration paraît convenable, elle ne fait que rétablir, après un départ mauvais, des taux convenables avec, il est vrai, un opérateur qui se fatigue.

Avec deux locuteurs, le résultat objectif n'est pas meilleur et fait conclure à l'intérêt pour le deuxième locuteur à construire immédiatement son propre lexique.

Pour l'adaptation par moyennes, le lexique étant construit, l'opérateur a procédé à une séquence de reconnaissance volontairement perturbée, d'abord en niveaux (mais non en durées) par la suppression de nombreux 1 dans les formes, par diminution de l'amplification avant l'analyse (cf. p. 29).

La reconnaissance s'étant ainsi altérée par la modification des degrés d'appartenance, l'opérateur a construit par moyennes une nouvelle forme comme il a déjà été dit. Ce cycle a été répété 5 fois, avec les taux de reconnaissance du tableau 14, où la séquence zéro fournit les résultats avant la perturbation.

séquence	0	1	2	3	4	5
I = 1	89	60	48	61	83	87
I = 2	95	74	65	72	90	95
I = 3	96	80	74	83	93	96

TABLEAU 14.

Pour deux locuteurs A et B, l'appareil en fonctionnement normal, on peut admettre que les densités des 1 sont voisines, comme le confirme la visualisation. La reconnaissance incorrecte par B du lexique de A traduit la segmentation forcée des mots de B par le lexique A. Les conséquences sur la reconnaissance sont corrigées ici encore, par la prise des moyennes des degrés d'appartenance.

Les résultats de ces expériences sont reportées sur le tableau 15 qui montre que l'amélioration limite apportée par cette opération est atteinte. Après un nombre réduit de séquences, après que, comme dans le cas précédent, on ait noté au passage de la première à la deuxième séquence une désorganisation du système.

Taux de reconnaissance en %

		Séquences locuteur B/locuteur A					
	Locuteur A/Locuteur A	1	2	3	4	5	6
I = 1	86	67	42	62	80	77,2	75
I = 2	91	82,5	62,5	73,7	86,2	86,2	
I = 3	95	88,7	70	78,7	88,7	93,7	

TABLEAU 15.

CONCLUSION.

D'un point de vue nouveau dans ce domaine, nous avons étudié la reconnaissance de la parole en considérant les formes à identifier, de contours essentiellement incertains suivant les occurrences ou les locuteurs, sous l'aspect de sous-ensembles flous.

Une formulation convenable permet alors de déterminer la similarité entre ces formes et des formes d'un lexique de référence, également prises comme des sous-ensembles flous.

Cette formulation qui est simple ne demande que des moyens modestes de calcul et permet d'opérer en temps réel.

Les formes traitées proviennent d'un analyseur spectral dont quatre zones sont utilisées, le temps intervenant au niveau de la segmentation. Cette segmentation repérée sur les formes de référence est imposée à la forme à reconnaître après normalisation temporelle. Les pavés fréquence-temps ainsi obtenus, flous déjà en ce qui concerne les limites temporelles, sont affectés de degrés d'appartenance fixés globalement par l'amplitude.

La similarité est alors appréciée par un indice global, dont la valeur maximale est l'élément de décision.

Pour un locuteur et un lexique de 175 mots pris au hasard, le taux de reconnaissance est de 90 %, avec un temps de calcul moyen de 1,7s.

Pour 10 locuteurs, le système inchangé donne encore pour 80 mots un taux cumulé de presque 80 %/

Le principe même de la méthode, par l'utilisation du flou des formes, permet de mieux adapter le système au cas de plusieurs locuteurs.

L'optique nouvelle dans laquelle nous avons entrepris ces recherches paraît appropriée à la reconnaissance de la parole et les résultats concrets obtenus portent à la fois sur le taux atteint, sur la vitesse de calcul et les besoins en mémoire.

==========

BIBLIOGRAPHIE

[B 1] J. BREMONT - "Simulation numérique d'un perceptron à défilement d'images" - Thèse de Doctorat en Spécialité Physique - Juin 1968 - Université de Nancy I.

[B 2] J.-F. BARS, J.-Y. GRESSER, M. QUERRE - "Application de la programmation dynamique à la reconnaissance des mots" - Journées d'Etude sur la parole - Bruxelles, Mai 1973.

[B 3] J. BREMONT, M. LAMOTTE - "Reconnaissance globale de la parole en temps réel par calcul d'un indice de similarité floue" - 5ème Journées d'Etude du groupe "Communication parlée" - Orsay, 15-17 Mai 1974.

[B 4] J. BREMONT, M. LAMOTTE - "Reconnaissance de formes. Contribution à la reconnaissance automatique de la parole en temps réel par la considération de sous-ensembles flous" - C.R. Acad. Sciences - 15 Juillet 1974.

[D 1] M.-J. DUGRAVOT - "Prétraitement numérique du signal vocal dans le domaine spectral" - Thèse de Doctorat en Spécialité Automatique - Octobre 1969 - Université de Nancy I.

[F A] K.-S. FU - "Pattern Recognition and Machine Learning" - Plenum Press - New-York - London - 1971.

[G 1] J.-A. GOGHEN - "L-Fuzzy Sets" - Journal of Math. Anal. and Appl. - V18, pp. 145-174, 1967.

[H 1] J.-P. HATON - "Reconnaissance de la parole : bilan de vingt années de recherches et tendances actuelles" - Ann. Télécom. 27, n°3-4, pp. 77-88, 1972.

[H 2] R. HORNUNG - "Dispositif de reconnaissance de la voix parlée" - Thèse de Docteur-Ingénieur - Février 1973 - Université de Nancy I.

[H 3] J.-P. HATON - "Contribution à l'analyse, la paramétrisation et la reconnaissance automatique de la parole" - Thèse d'Etat - 8 Janvier 1974 - Université de Nancy I.

[H 4] R. HUSSON - "Physiologie de la phonation" - Masson et C^{ie} - 1962.

[K 1] A. KAUFMANN - "Introduction à la théorie des sous-ensembles flous" - Masson et C^{ie} - 1973.

[L 1] E. LEIPP - "Images et traitement optique d'images" - Bulletin du groupe d'Acoustique musicale - Juin 1973 - n° 68.

[L 2] M. LAMOTTE, J. BREMONT, J.-P. HATON - "Simulation de la reconnaissance des formes vocales par apprentissage" - C.R. Acad. Sciences - Paris 269 - pp. 286-288, 1969.

[L 3] E. LEIPP - "Mécanique et acoustique de l'appareil phonatoire" - Bull. n° 32 - 5 - 12 - 1967 - GAM - Université de Paris VI.

[M 1] M. MLOUKA, J.-S. LIENARD - "Reconnaissance par mots basés sur les transistors phonétiques" - 5ème Journées d'Etude du groupe Communication parlée" - Orsay - Mai 1974.

[R 1] B. ROY - "Algèbre moderne et théorie des graphes" - 2 tomes - Dunod - 1969.

[V 1] M.-J. VIGNERON, M. LAMOTTE, J.-P. HATON, J. BREMONT - "Recherches actuelles sur l'extraction de caractéristiques et la reconnaissance de la voix parlée" - Automatisme n° 12 - Décembre 1970.

[V 2] C. VOMSCHEID - "Analyseur de courbes - Application à l'étude du signal vocal" - Thèse de Spécialité Automatique - 25 Juin 1969.

[Z 1] L.-A. ZADEH - "Fuzzy Sets" - Information and Control V8 - pp. 338-353 - 1965.

[Z 2] L.-A. ZADEH - "Probability Measures of Fuzzy Events" - Jour. Math. Anal. and Appl. - Vol. 10 - pp. 421-427 - Aug. 1968.

[Z 3] L.-A. ZADEH - "Quantitative Fuzzi Semantics" - Information Sciences - 3 - pp. 0-17 - 1970.

[Z 4] L.-A. ZADEH - "On Fuzzy Algorithms E.R.L." - Memo M325- Univ. of California - Berkeley - Fevr. 1972.

==========

NOM DE L'ETUDIANT : BREMONT Jacques

NATURE DE LA THESE : DOCTORAT D'ETAT MENTION SCIENCES

VU, APPROUVE

& PERMIS D'IMPRIMER

NANCY, le 5 Février 1975

LE PRESIDENT DE L'UNIVERSITE DE NANCY I

UNIVERSITE DE NANCY I

J.B. HELLUY