الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

13/88

وزارة التعليم و البحث العلمي
MINISTERE DE L'ENSEIGNEMENT ET DE LA RECHERCHE SCIENTIFIQUE

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT ELECTRONIQUE

PROJET DE FIN D'ETUDES

SUJET

RECONNAISSANCE DES VOYELLES EN MODE MONOLOCUTEUR SUR MICRO VAX 2

Proposé par : M^r N.BENIDDIR

Etudié par : M^elle H. KARKAR

Dirigé par : M^r N_BENIDDIR

PROMOTION : JANVIER 1988

E.N.P. 10, Avenue Hacen Badi – EL-HARRACH – ALGER

D E D I C A C E S

A ma délicieuse mère,

A mon merveilleux père,

A mon frère,

A mes soeurs,

A tous ceux qui me sont chers,

et

Ceux à qui je suis chère.

REMERCIEMENTS

Je remercie vivement Monsieur N. BENIDDIR pour sa précieuse aide et son soutient le long de de ce projet, et surtout pour le choix du sujet qui m'a permis de découvrir le monde informatique.

Mes vifs remerciements vont à ma soeur LILIA pour son soutient moral et son aide, je ne pouvais m'en passer.

Que messieurs H. BEN MESSAOUD et M. SEFSAF trouvent ici ma gratitude pour la sympathie qu'ils m'ont témoigné.

Toute ma reconnaissance à Monsieur D. SAIS qui ne m'a pas privé de son savoir.

CHAPITRE I : APPAREIL PHONATOIRE

1. Introduction.
2. Description et fonctionnement de l'appareil phonatoire.
3. Le mouvement vibratoire.
4. Les types d'excitation.

CHAPITRE II : ANALYSE DE LA PAROLE.

1. La prédiction linéaire.
   - 1.1. Principe
   - 1.2. Optimisation des coefficients prédicteurs.
   - 1.3. Principales méthodes de la prédiction linéaire.
   - 1.4. Mise en oeuvre de la L.P.C.
2. Analyse cepstrale.
   - 2.1. Introduction
   - 2.2. Le traitement homomorphique.
   - 2.3. Cepstre.
   - 2.4. Echelle de Mel.
   - 2.5. Mise en oeuvre de l'analyse cepstrale.
   - 2.6. Simulation des signaux des voyelles.
   - 2.7. Mise au point du programme analyse.

CHAPITRE III : RECONNAISSANCE DE LA PAROLE.

1. Introduction.
2. Les grandes approches.
3. Reconnaissance des mots isolés.

4. Les sources de variabilités.
5. La programmation dynamique.
6. Application de la programmation dynamique à la reconnaissance.
7. Contraintes locales.
8. Contraintes Globales.
9. Fonction de coïncidence.
10. Fonction de pondération w(K)
11. Choix de la distance.
12. Programme de la reconnaissance.
13. Algorithme de la D.T.W.
14. Résultats.
15. Domaine d'application.

ORGANIGRAMMES.

CONCLUSION.

ANNEXES

BIBLIOGRAPHIE.

INTRODUCTION.

Si communiquer avec une machine était le rêve, autrefois, d'un écrivain de science fiction, c'est devenu aujourd'hui une réalité.

L'homme grâce à son ambition, sa volonté et ses efforts, a pu franchir les obstacles qui le gênaient. En effet, tant de problèmes rencontrés, tant de difficultés inattendues, ne l'ont pas découragé pour autant. Ainsi depuis 50 ans des chercheurs tentant de concevoir des dispositifs permettant l'interaction vocale entre l'homme et la machine et ces efforts n'ont pas été en vains: aujourd'hui certaines machines parlent (synthèse vocale), d'autres entendent (reconnaissance vocale).

Nous nous interessons dans ce mémoire à la reconnaissance automatique de la parole (R.A.P). La reconnaissance de mots isolés (voyelles) en mode monolocuteur notamment grâce aux méthodes d'analyse, la prédiction linéaire et la cepstrale et aux méthodes issues de la programmation dynamique appliquée aux contraintes de Sakoe et Shiba.

CHAPITRE I

APPAREIL PHONATOIRE

1. Introduction:

L'expérience montre qu'une meilleure reconnaissance de la parole nécessite de bonnes connaissances en phonétique, Ce qui a poussé les chercheurs à mener des études plus ou moins profondes dans ce domaine, suivant l'objectif à atteindre (synthèse ou reconnaissance).

Pour notre part, nous ne rappelons dans ce chapitre, que les éléments essentiels pour la production de la parole ayant une relation directe avec notre travail.

2. Description et fonctionnement de l'appareil phonatoire:

Il est essentiellement constitué des organes suivants:

a. Les poumons: jouent le rôle de générateur d'air. En effet au cours de la phonation, la diminution du volume thoracique tend à classer l'air emmagasiné dans les poumons en quantité variable suivant les besoins.

Le larynx est alors alimenté par de l'air à une pression atmosphérique, la pression subglottique, qui met en vibration les cordes vocales.

b. Les cordes vocales: jouent le rôle d'un excitateur. Elles sont attachées à la base du larynx. Lorsqu'elles sont aux repos, la glotte est normalement ouverte, permettant la respiration.

c. Le conduit vocal: Il agit comme caisse de résonnance pour les sons émis par le larynx, il est capable d'amplifier ou d'amortir certains sons. Il est composé de deux parties:

- Le conduit nasal: il est formé des forces fixes nasales qui sont deux cavités de forme fixe dont la communication avec la cavité orale est commandée par le voile du palais, prolongement mobile du palais dur.

- Le conduit oral: possède un volume et une géométrie extrèmement variables grâce à la grande mobilité de la langue essentiellement, et du maxillaire inférieur.

Lors de la production d'une voyelle, le conduit vocal se déforme très peu et ses fréquences de résonnance sont très stables.

3. Le mouvement vibratoire

Il est caractérisé par son amplitude à laquelle est relié le niveau sonore perçu par le locuteur, et par son spectre fréquentiel.

La fréquence fondamentale conditionne la hauteur de la voix.

La présence d'harmoniques caractérise son timbre; plus le rapport entre les durées d'obturation et d'ouverture de la glotte est grand, et plus le timbre comportera d'harmoniques.

4. Les types d'éxcitation:

Il en éxiste deux:

- Le premier type: Il est produit par la vibration des cordes vocales sous l'action de la pression de l'air en provenant des poumons. L'onde produite est assimilée à un train d'impultions appelée fréquence de mélodie (le pitch) qui varie en moyenne:

- de 100 à 150 Hz chez les hommes.

- de 200 à 250 Hz chez les femmes.

Les sons produits par ce type d'excitation sont dits voisés.

- Le deuxième type: Il est provoqué par la génération d'un écoulement d'air turbulent dans le canal vocal.

Les sons ainsi produits sont dits non-voisés comme certaines consonnes "s"; "ch" et "f".

Néanmoins certains sons necessitent la combinaison des deux phénomènes comme "Z"; "J"; "V".

Par ailleurs avant la mise en équation d'un phénomène physique, il est nécessaire de l'étudier puis de lui trouver un modèle mathématique qui le représente.

Ainsi l'appareil phonatoire a été modélisé comme l'indique la figure (1.b) (1;C) (11) (19)

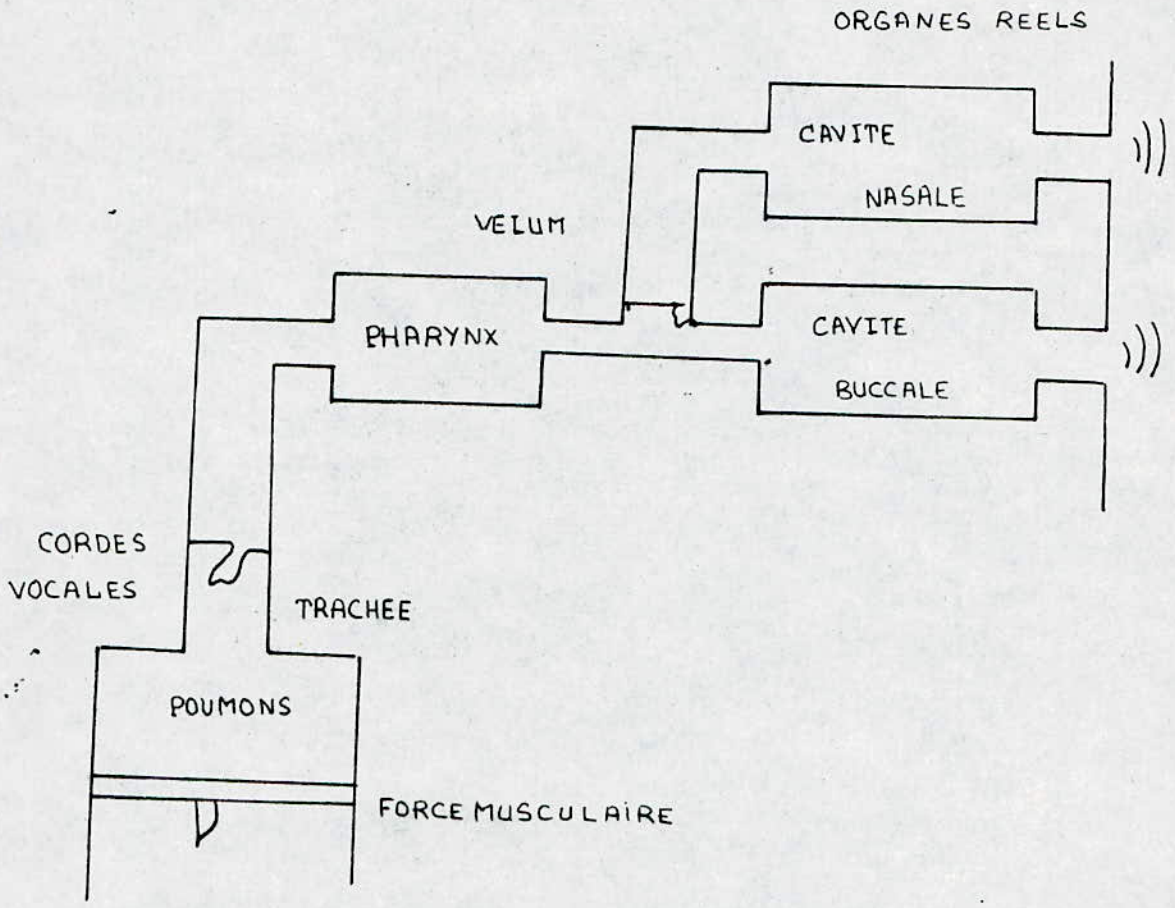
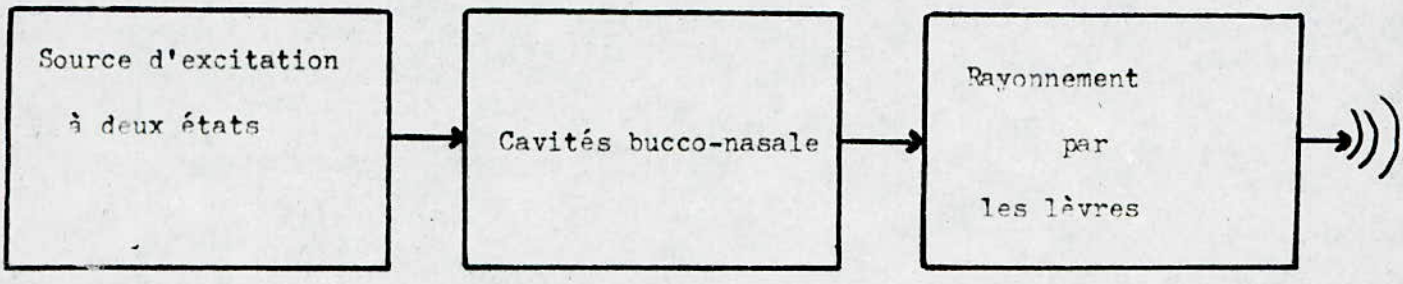


fig (1.a)

MODELE EQUIVALENT

A L'APPAREIL PHONATOIRE



conduit vocale
assimilé à un filtre

MODELE SIMPLIFIE

fig (1,c)

CHAPITRE II

ANALYSE DE LA PAROLE

II. Analyse de la parole:

Introduction:

Le signal de la parole est très redondant(il transporte beaucoup plus d'informations que nécessaire).

L'analyse est l'ensemble de toutes les techniques succeptibles d'optimiser l'information et donc d'extraire de la masse de données disponibles un ensemble de paramètres pertinents que nous exploitons lors de la reconnaissance. Ce qui ne va sans difficultés, lorsqu'on sait la variabilité et la complexité du signal de la parole.

Les téchniques d'analyse du signal permettent d'effectuer une réduction d'information sans trop la dégrader.

Deux grandes classes de méthodes sont en oeuvre parfoissimultanément:

- Des méthodes spécifiques fondées sur une modélisation du signal de production de la parole, codage prédictif linéaire (LPC, également utilisé en synthèse de la parole), analyse ceptrale.

- Des méthodes générales (valables pour tout type de signal permett ant d'extraire des paramètres temporels (énergie, nombre de passages par zéro du signal) ou fréquentiels (bans de filtres, transformée de Fourier)

Notre choix s'est porté sur deux techniques spécifiques au signal de la parole:

- L'analyse par prédiction linéaire: c'est l'une des plus puissantes techniques qui consiste en la recherche d'un modèle se rapprochant le plus possible du signal original.

- L'analyse cepstrale: qui consiste en la déconvolution du signal modulé de son enveloppe, générée au niveau du conduit vocal.

1. La prediction linéaire:

1.1 Principe:

La méthode est basée sur le fait qu'un échantillon de parole S(nT) peut être prédit par une combinaison linéaire d'un certain nombre d'échantillons prélevés à des instants précédents.

La valeur prédite de $\hat{S}(n)$ s'obtient par une somme pondérée linéairement de P échantillons.

$$\hat{S}(n) = - \sum_{k=1}^{P} a(k).\ S(n-k)$$

avec S(n): l'échantillon prédit

P : l'ordre de prédiction.

L'idée essentielle de cette analyse est de considérer les signaux comme réponse d'un filtre linéaire.

La prédiction linéaire permet d'écrire:

$$\text{II.2} \qquad S(n) = -\sum_{k=1}^{P} a(k).S(n-k) + G.\,U(n).$$

Où $a(k)$: représente les coefficients des filtres prédicteurs.

G : représente un facteur de gain.

De II.2, on a: $S(n) + \sum_{k=1}^{P} a(k) \cdot S(n-k) = G.U(n)$.

En utilisant la transformée en Z [10] on a:

$$S(Z) + \sum_{k=1}^{P} a(k).\,S(Z).\,Z^{-k} = G.U(Z)$$

D'où:

$$\text{II.3} \qquad H(Z) = \frac{S(Z)}{U(Z)} = \frac{G}{1 + \sum_{k=1}^{P} a(k)\,Z^{-k}} = \frac{G}{\sum_{k=0}^{P} a(k)\,Z^{-k}}$$

Avec:

$$a(0) = 1$$

D'après les équations II.2 et II.3, le signal est modélisé comme étant la sortie d'un filtre de fonction de transfert H (Z) avec une entrée U(Z).

Pour assurer au modèle une existance physique, il faut que les filtres soient causals et stables.

De plus, ce modèle ne sera unique que si le filtre est stable et à phase minimale.

1.2 Optimisation des coefficients predicteurs:

Optimiser les coefficients du predicteur a(K) revient à trouver les a(K) qui permettent d'avoir en sortie du modèle un signal à analyser ou du moins le plus proche possible.

Pour cela on cherche à minimiser l'erreur entre le signal S(n) (réel) et le signal prédit S(n).

L'erreur de prédiction est définie comme suit:

II.4 $C(n) = S(n) - \hat{S}(n)$

L'erreur quadratique totale est définie comme suit:

II.5 $E(P) = \sum_{n=n0}^{n1} e^2(n)$

Où encore:

$$E = \sum_{n=no}^{n1} \left(S(n) + \sum_{k=1}^{P} a(k).\ S(n-k) \right)^2$$

Les a( ) qui minimisent E(P) vérifient:

II.6 $$\frac{\partial E(P)}{\partial a(k)} = 0$$

C'est à dire:

$$\frac{\partial}{\partial a(k)} \left(S(n) + \sum_{k=1}^{P} a(k).\ S(n-k) \right)^2$$

Qui devient:

$$\sum_{n=o}^{n1} S(n) \cdot S(n-j) + \sum_{n=no}^{n1} \sum_{k=1}^{P} a(k).S(n-k).S(n-j) = 0$$

Mieux encore:

II.7 $$\sum_{k=1}^{P} a(k).C(j,k) = -C(j,0)$$

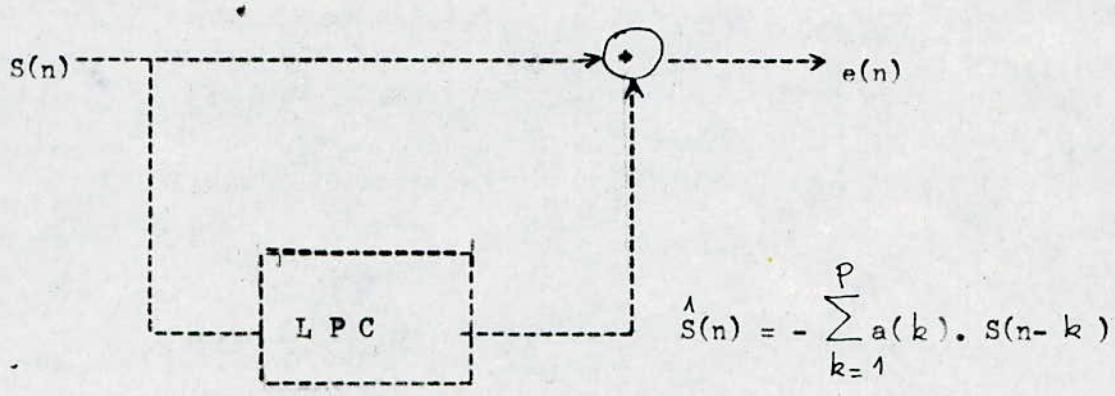
Telle que

$$c(j,k) = \sum_{n=no}^{n1} S(n-k) \cdot S(n-j)$$

$$c(j,0) = \sum_{n=no}^{n1} S(n).\ S(n-j)$$

Cette équation II.7 constitue un ensemble de "P" équations à "P" inconnus que l'on peut résoudre pour obtenir les coefficients prédicteurs.

Compte tenu des équations II.1 et II.4, nous pouvons schématiser le modèle L P C comme suit:



1.3. Principales méthodes de la prédiction linéaire:

Il existe plusieurs possibilités de déterminer les coéfficients a(k), dont chacune se distingue des autres par ses hypothèses. Parmi lesquelles nous citons:

a. La méthode d'autocorrelation:

L'hypothèse de base dans cette méthode, est de considérer le signal comme stationnaire dans un intervalle fini (intervalle d'étude). D'où nécessité d'un fenêtrage.

Quant au but d'utilisation d'une telle méthode est de séparer le signal utile du bruit dont il est entraché en comparant ce signal composité avec son propre double qui est progressivement retardé.

L'autocorrelation de S(n) est définie par:

$$\text{II.8} \quad R(j) = \sum_{n=-\infty}^{+\infty} S(n).\ S(n+j)$$

Comme le signal est limité par une fenêtre et $S(n) = 0$ si $n \notin [0,\ N-1]$ alors :

$$\text{II.9} \quad R(j) = \sum_{n=0}^{N-j-1} S(n)\ .\ S(n+j) \qquad 1 \leq j \leq P$$

$$\text{II.10} \quad R(j) = \sum_{k=1}^{P} a(k)\ .\ R(|j-k|)$$

Le système s'écrit sous la représentation matricielle suivante :

II. Analyse de la parole:

$$\text{II.10}\quad \underbrace{\begin{bmatrix} R(0) & R(1) & \cdots & R(P-1) \\ R(1) & R(2) & \cdots & R(P-2) \\ \vdots & & & \vdots \\ R(P-1) & R(P-2) & \cdots & R(0) \end{bmatrix}}_{R} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(P) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(P) \end{bmatrix}$$

La matrice R est une matrice de Toeplitz

b. La méthode de Covariance:

Le signal est défini pour N échantillons consécutifs et l'erreur quadratique totale est minimisée sur les N - P derniers échatillons.

$$\text{II.11}\quad E(P) = \sum_{P}^{N-1} e^2(n)$$

II.11 d'où:

$$\text{II.12}\quad C(j,k) = \sum_{n=P}^{N-1} S(n-j).S(n-k) \qquad j,k = 0, \ldots P$$

Le système s'écrit sous la représentation matricielle suivante:

$$\begin{bmatrix} C(1,1) & C(1,2) & \cdots & C(1,P) \\ C(2,1) & & & \\ \vdots & & & \\ C(P,1) & C(P,2) & \cdots & C(P,P) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \\ a(P) \end{bmatrix} = \begin{bmatrix} C(1,0) \\ C(2,0) \\ \\ C(P,0) \end{bmatrix}$$

II; Analyse de la parole:

Pour notre part, nous avons opté pour la méthode d'auto-correlation pour les raisons que nous citons:

- Calcul moindre pour les coefficients de prédiction dans la méthode d'auto-corrélation que dans celle de covariance.

- L'erreur totale minimale de prédiction est calculée en même temps que les coefficients de prédiction.

- La méthode d'auto-corrélation assure la stabilité (théorique) du filtre de prédiction.

Cependant, dans le cas de la méthode de covariance, on peut réduire les possibilités d'instabilité du prédicteur en augmentant le nombre d'échantillons contenu dans la séquence de signal à analyser, pour cela, il faut augmenter la valeur de N car la valeur de P est normalement fixée.

A TAL et HANUVER (9) décrivent une méthode permettant de corriger la position des pôles situés à l'extérieur du cercle unité.

Tableau récapitulatif:

Analyse	Prédiction linéaire	
Méthode	Autocorrélation	Covariance
Fenêtrage	Nécessaire	Pas nécessaire
Stabilité	Théoriquement assuré	non

II. Analyse de la parole:

Détermination des coefficients a(k) par la méthode de DURBIN:

Il éxiste plusieurs méthodes qui permettent de résoudre le système. II.10 , parmi lesquelles nous citons

- Méthode de Gauss - Seidel
- Méthode de Gauss - Jordan
- Méthode de Jacobi
- Méthode de Durbin

Nous utilisons la méthode de Durbin pour ses performances.

- Rapidité d'éxécution
- Calcul moindre

L'Algorithme de DURBIN : ce fut LEVINSON qui l'a développé en 1947, et appliqué au calcul des coefficients de prédiction par DURBIN.

Les conditionsinitiales:

$$K(1) = R(1)/R(0) \quad ; \quad E(0) = R(0)$$

$$A(1,1) = K(1)$$

$$E(1) = (1 - D(1) \cdot D(1)) \cdot R(0)$$

Pour $I > 2$, , $I = 2, \ldots P$

$$K(I) = R(I) - \sum_{j=1}^{I-1} A(J, I-1) \cdot R(J - I) / E(I - 1)$$

II. Analyse de la parole:

Où K(I) sont les coefficients de réflexions

Coefficient de prédiction :

$$A(J,I) = A(J,I-1) - K(I) \cdot A(I-J, I-1)$$

$$A(I,I) = K(I)$$

Erreur quadratique

$$E(I) = (1 - K(I)^2) \cdot E(I-1)$$

Ainsi pour calculer les coefficients A(J) d'ordre P. A(J,P), nous devons calculer tous les coefficients d'ordre $< P$; les A(J,I) sont donc liés d'une manière récurssive.

La méthode de Durbin, en plus des avantages cités précédement permet de calculer les coefficients de réflexion $\hat{K}(I)$ qui interpretent la stabilité.

1.4 : Mise en oeuvre de la LPC:

Cette méthode est devenue une technique prédominante pour éstimer les paramètres de base de la parole : les formants, le spectre, la fréquence fondamentale, les coefficients de prédiction..

Cependant, un signal, avant que l'on puisse éxtraire ses paramètres, doit subir un prétraitement que l'on peut diviser en plusieurs parties:

Analyse de la parole:

a. Echantillonnage:

Les méthodes d'analyse sont du type numérique, ce qui fait que le signal avant d'être analysé doit être converti de l'analogiqueen numérique, pour pouvoir éxtraire les paramètres qui le caractérise, d'où la nécéssité de l'échantillonner.

L'éxpérience montre qu'un signal de parole à 6 K Hz conserve ses caractéristiques et demeure intelligible, c'est pourquoi une fréquence d'échantillonnage de 12,8 K Hz est satisfaisante, selon le théorème de Shanon:

$$Fe \geqslant 2\ F_S$$

avec Fe : fréquence d'échantillonnage

Fs : fréquence max du signal de parole.

b. Préaccentuation:

Le conduit et le milieu extérieur sont deux milieux différents, chacun possède sa propre impédance mécanique, ce qui fait que le son sortant du conduit vocal, arrivé au contact des deux milieux (au niveau des lèvres), il subit une baisse d'énergie de 6dB . Pour établir l'adaptation, on est amené à faire une préaccentuation de 6dB .

Il faut noter, que l'utilisation du filtre de préaccentuation est nécessaire dans le cas des sons voisés, pour les sons non voisés, son effet n'est pas génant.

Analyse de la parole:

Pour notre part, nous avons exploité l'expression mathématique permettant d'interpréter ce phénomène physique.

c. Fenêtrage:

Le conduit vocal, dont l'ensemble des coefficients prédicteurs est caractéristique, a une forme et des dimensions continuellement variables dans le temps.

Cependant, ces variations sont lentes. Il est donc possible de les considérer (forme, dimensions et coefficients prédicteurs) comme constants sur des durées de 10 à 25 ms, c'est ce qu'on apelle fenêtrage temporel. Ainsi le signal pris dans chacun de ces intervalles est considéré comme stationnaire.

On le multiplie par la fenêtre de pondération

$$SF(n) = SP(n) \cdot W(n)$$

Avec:

SF(n) : échantillons détenus après fenêtrage

SP(n) : échantillons préaccentués

W (n) : échantillons de la fenêtre utilisée

Analyse de la parole:

- Choix de la fenêtre: il existe plusieurs types de fenêtres; néanmoins le choix est dicté par l'application et l'objectif à atteindre. Nous citons:

  * Fenêtre rectangulaire
  * Fenêtre triangulaire
  * Fenêtre de Hamming
  * Fenêtre parabolique

L'existence d'un aussi grand nombre de fenêtres est dûe au fait que l'on veuille améliorer l'éfficacité de l'une par l'utilisation d'une autre.

Sur ce point, la documentation consultée [8] et [9] précise que la fenêtre de Hamming donne de bons résultats pour le traitement de la parole.

Nous citons certains de ses avantages:

- Niveau du premier lobe secondaire est à 43,9 au dessous du lobe principal.

- 99,96% d'énergie est concentrée dans le lobe principal.

L'équation de la fenêtre de Hamming

$$W(n) = \begin{cases} 0{,}54 - 0{,}46 \cos(2\pi . n/N) & \text{pour } 0 < n < N-1 \\ 0 & \text{ailleurs} \end{cases}$$

Avec N le nombre total d'échantillons contenu dans la fenêtre.

Analyse de la parole:

II.2: Analyse cepstrale:

2.1: Introduction:

Le signal de la parole est considéré comme étant le signal de sortie d'un système linéaire (le conduit vocal) dont les propriétés varient lentement avec le temps.

Le conduit vocal module la source d'exitation, il y a cependant combinaison par convolution de l'exitation et de la réponse impulsionnelle du conduit vocal.

L'étude d'une telle analyse se résume en la séparation bruit - signal utile c'est à dire aux composantes convoluées: la déconvolution.

fig II. Ce qui peut être réalisé grâce aux calculs cepstraux.

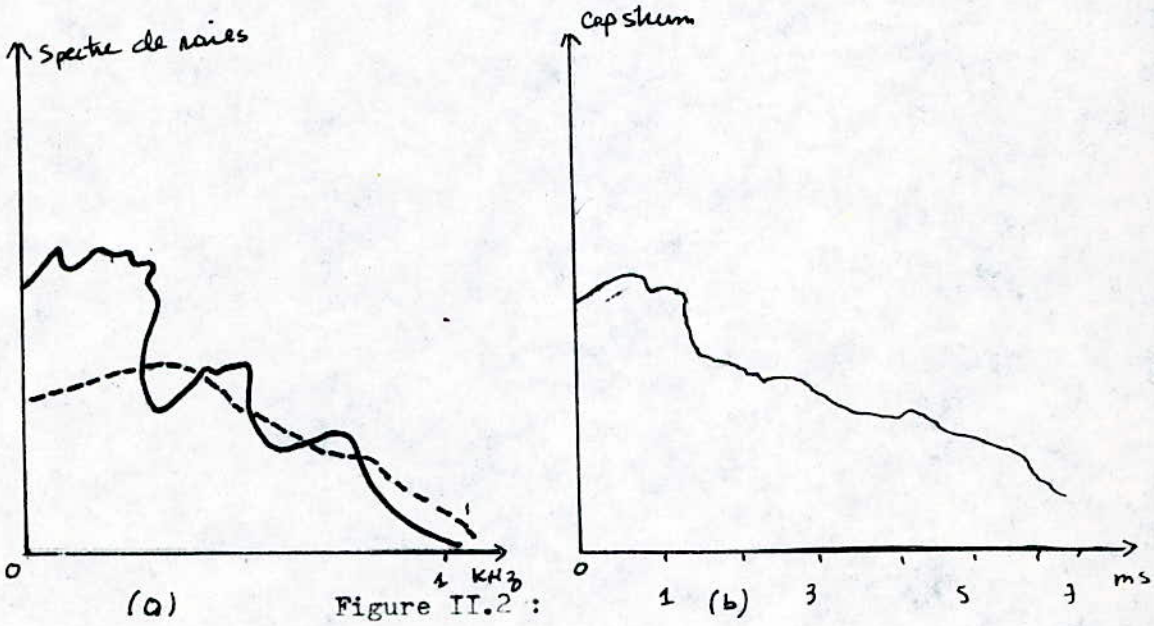


Figure II.2 :

a. convolution de deux signaux

b. déconvolution et obtention de l'enveloppe.

Analyse de la parole:

2.2 Le traitement homomorphique:

Le traitement homomorphique permet, dans le cas de l'analyse de la parole, de séparer les informations prosodiques, des informations phonétiques, combinées par convolution.

En effet, les basses fréquences sont porteuses d'indices concernant le locuteur et peuvent servir dans un système d'identification du locuteur (déterminer qui est-ce qui parle). Alors que les hautes fréquences permettent de connâître la forme du conduit vocal donc porteuse d'informations sur ce qui est dit.

Ainsi, la déconvolution permet-elle de ne retenir que les hautes fréquences nécéssaires à un système de reconnaissance de la parole.

En effet, soient h(n) le signal issu de la source d'excitation et x(n), la fonction de transfert du conduit vocal indépendante de la première. Ces deux fonctions convoluent et donnent le signal sonore y(n) qui déclanche de l'appareil phonatoire et dont l'équation est comme suit :

$$\text{II 13} \quad y(n) = \sum_{k=-\infty}^{+\infty} h(n-k) \cdot x(k) = h(n) * x(n)$$

Le symbole * représente le produit de convolution.

L'opération D de déconvolution vérifie l'équation suivante:

$$\text{II 14} \quad D(y(n)) = D(h(n) * x(n))$$
$$= D(h(n)) + D(x(n))$$

Et par passage à la transformée de Fourrier le produit de convolution. se transforme en un produit simple.

Analyse de la parole:

$$y(n) = h(n) * x(n) \quad \Longrightarrow \quad y(f) = H(f) \cdot X(f)$$

OPPENHERM (19) a démontré qu'il est possible de représenter un système homomorphique par trois systèmes en série dont celui du milieu est un système linéaire conventionnel: c'est la forme dite canonique des systèmes homomorphiques.

On peut schématiser le processus de transformation du signal dans le but de séparation, comme suit

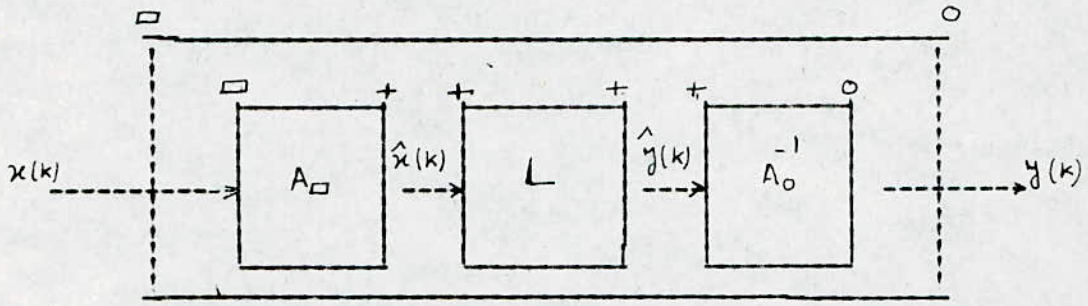


Figure II.3: Déconvolution homomorphique

La fonction logarithme satisfait bien une telle relation, pour autant que les signaux H(f) et X(f) soient strictement positifs.

On a alors:

II 15. $$\text{Log } y(f) = \text{Log } (H(f) \cdot X(f))$$

$$= \text{Log } H(f) + \text{Log } X(f)$$

Dans le cas où ces signaux sont complexes, le logarithme précedent possède une partie imaginaire indéfinie.

$$\text{Log } y(f) = \text{Log } (|y(f)| + j \arg(y(f)))$$

Analyse de la parole:

Dans le cas où il est possible de calculer le logarithme complexe, sa transformée de Fourier inverse donne le cepstre complexe.

2.3 Le cepstre:

C'est la transformée de Fourier inverse du logarithme du module du spectre du signal.

Il est défini par :

$$\text{II 16.} \qquad y(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} /y(e^{jw}) /. e^{jwn} \, dw$$

$$\text{II 17.} \qquad \mathbb{C}(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \text{Log} \, / \, y(e^{jw}) /. e^{jwn} \, dw$$

où

$$\text{II 18.} \qquad c(n) = \frac{y(n) + y(-n)}{2}$$

d'où:

$$\text{II 19.} \qquad c(n) \frac{1}{2\pi} \int_{-\pi}^{+\pi} \text{Log} /y(e^{jw}) / \, . \cos(wn) \, dw$$

On calcule ainsi la transformée de Fourier inverse du logarithme du module de la transformée de Fourier du signal.

Analyse de la parole:

Par définition du cepstre:

$$II\,20. \quad C_p(n) = \frac{1}{N} \sum_{K=0}^{N-1} \text{Log} \, / Y_p(f) / \, e^{j2\pi fn/N}$$

Avec: N nombre de points dans une fenêtre

n rang du coefficient cepstral

Remarque:

- Les coefficients cepstraux décroissent en $1/N$, d'où l'utilisation d'un nombre réduit, suffit pour caractériser le signal.

- Le coefficient d'ordre zéro est une mesure de l'énergie du spectre.

- Les coefficients cepstraux sont indépendants de l'énergie.

L'étude physiologique et perceptive de l'oreille indique qu'elle est sensible à une échelle semi-logarithmique de fréquence linéaire sur le premier K Hz et logarithmique au délà, d'où utilisation de "l'échelle de Mel".

2.4 Echelle de Mel:

C'est une echelle semi-logarithmique, linéaire sur le premier K Hz et logarithmique au delà.

Pour obtenir la linéarité sur une partie et le logarithme sur l'autre partie du signal, on peut modéliser le conduit vocal comme étant:

Une succession de filtres triangulaires décalés les uns des autres de la moitié de la largeur des filtres (d'où le 1/2 vocodeur). Les 10 premiers filtres ont une largeur uniforme, les autres ont une largeur qui suit une progression géométrique.

Le cepstre s'écrit alors:

$$C(n) = \frac{1}{NF} \sum_{K=1}^{NF} \log(E(K)) \cdot \cos(n(K - 1) \quad /NF) \qquad \text{II 21.}$$

où:

n: le rang du cpefficient cepstral

K: le numéro du filtre d'énergie E(K)

NF: le nombre total des filtres triangulaires

NF est généralement compris entre 16 et 25.

2.5: Mise en oeuvre de l'analyse cepstrale:

Les étapes du traitement sont résumées par la figure :

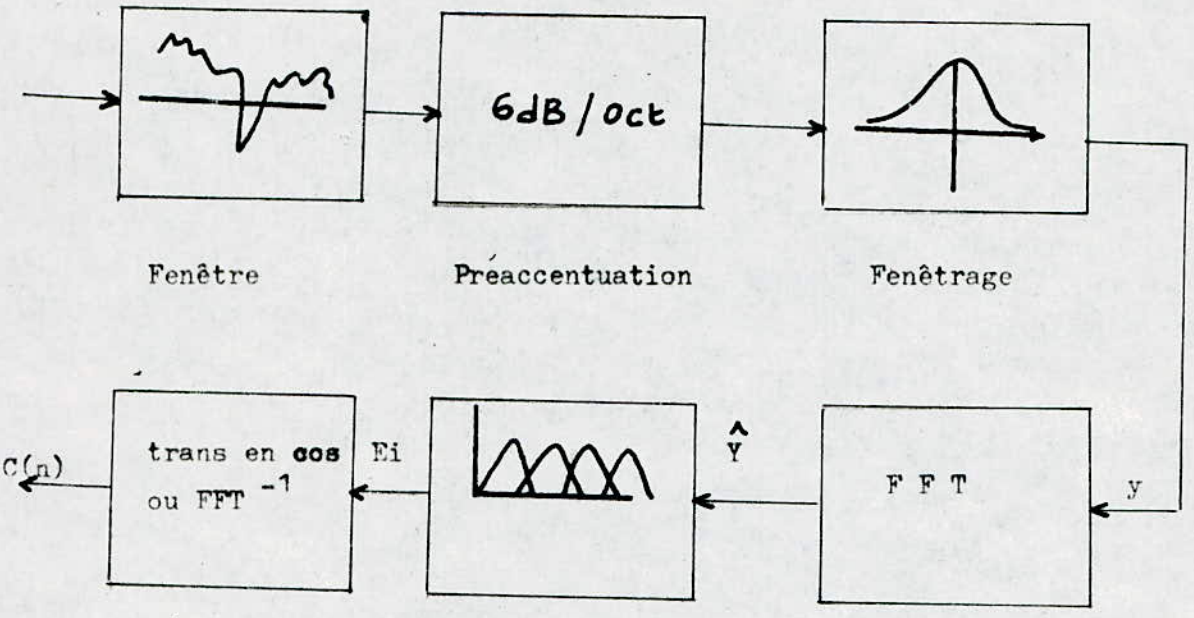


Figure II.4 : processus de transformation du signal de parole.

Cette série de transformation se résume comme suit:

a. Une fenêtre d'analyse dans laquelle le nombre d'échantillons à traiter est bien fixe et doit être fonction de la fréquence d'échantillonnage, et de la durée du signal.

$\hat{n}$

Dans ce travail, cette fenêtre correspond à la limitation du signal de la parole à une durée où on peut le considérer stationnaire.

b. Préaccentuation: comme dans le cas de la PLC, un filtre de préaccentuation de la forme $P(Z) = 1 - bZ^{-1}$ $(b = 0,96)$ est necessaire pour compenser les effets dûs à la source d'excitation du conduit vocal et au rayonnement des lèvres.

c. Fenêtrage de Hamming pour éliminer l'effet de trancature.

d. une transformée de Fourrier rapide (FFT)

e. Calcul de l'énergie en sortie de chaque filtre du banc.

f. Transformation en cosinus pour l'obtention des différents coefficients cepstraux.

Le logiciel que nous avons développé comprend toutes ces étapes successivement. Cependant, l'obtention des coefficients cepstraux telle qu'elle est exprimée par la formule (II 21.) est bien possible, en utilisant un vocodeur à canaux dont la réparation des canaux suivrait celle de l'echelle "Mel" des fréquences. Et la seule transformation à réaliser serait la transformation en cosinus sur les énergies délivrées par le vocodeur. Ceci moyennant une réalisation concrête de la butterie de filtres et des détecteurs d'enveloppe.

2.6 Simulation des signaux des voyelles:

L'handicapede ne pouvoir introduire la parole dans la machine et pour tester les logiciels que nous avons développé sur le MICROVAX, en langage FORTRAN 77, nous avons eu recours à la génération de 6 voyelles suivant le principe ci-dessous.

Le son de chaque voyelle correspond à une position particulière de la bouche, de la langue, des lèvres et du voile du palais.

Chaque voyelle a ses formants caractéristiques. De basse fréquence pour (U), (O), de moyenne fréquence pour (a) de basse et haute fréquence pour (e), (i).

Ces formants sont à l'heure actuelle le seul élément presque invariable que l'on ait trouvé pour caractériser ces voyelles qui donnaient des oscillogrammes si divers. (2)

En outre, on peut assimiler la forme du conduit vocal à une succession de cavités resonnantes que amplifient l'énergie à certaines fréquences et l'atténuent à d'autres d'où l'obtention du signal de parole (voyelle) par somme de sinusoides dont, chacune correspond à une harmonique.

Mais il serait fastidieux si nous devions retrouver le signal harmonique par harmonique, cependant, nous ne prenons en considération que quelques harmoniques (celles dont l'énergie est supérieur à 3 db)

Chaque voyelle ayant ses propres formants, le tableau suivant présente les résultats expérimentaux des formants fondamentaux de ces voyelles obtenus par EMRTT (2)

Formants en Hz / Voyelle	F 1	F 2
U	250	600
O	375	750
a	750	1200
e	375	2200
i	250	1500
j	250	1800

S'inspirant du triangle de DELATRE, deux formants suffisent pour caractériser une voyelle. Il n'en n'est pas de même pour les consonnes où à quatre (4) formants sont nécéssaires pour transporter toute l'information.

Mise au point du programme analyse

Le signal de la parole est caractérisé en premier lieu par ses échantillons. Ceux-ci ne peuvent être exploités directement, vu leur taille et donc la place mémoire qu'ils peuvent occuper.

Partant de ce principe plusieurs techniques d'analyse ont été appliquées à la parole (dans le but de réduire la place mémoire, entre autre).

Les échantillons sont traités par parties "fenêtres" dont la durée peut varier entre 10 ms et 40 ms, ce qui correspond au nombre d'échantillons dans une fenêtre de 2^7 à 2^9.

Dans ce travail il a été retenu N = 256 échatillons par trame.

A la fin de l'analyse le nombre N sera représenté par P coefficients (une trame) où P = 12 pour la prediction linéaire et 8 pour la cepstrale.

Constitution du dictionnaire

L'importance de celui-ci est primordiale, en effet la parole constituant le dictionnaire doit être prononcée dans de très bonnes conditions de telle façon à éviter d'introduire dans le vocabulaire de mauvaises références.

Pour la réalisation d'un système de reconnaissance de la parole, cette phase dite aussi apprentissage exige un soin particulier.

Ainsi la reconstitution du signal d'une voyelle repose essentiellement sur une étude plus ou moins profonde en phonétique. En effet, en assimilant le conduit vocale à une succession de filtre passe bas et connaissant les formants de la voyelle dont nous voulons extraire le signal, nous pouvons passer du domaine fréquentiel au domaine temporel ou chaque harmonique représente une sinusoide et la somme de toutes les sinusoides donne le signal désiré.

Notons cependant que la longueur d'un mot est supposé normale, si le nombre de trames contenu dans le signal caractéristique est:

$$N \geq 16$$

Pour tester nos programmes nous avons opté pour N = 16 trames, dont chacune comporte 256 échantillons.

$$t = N/f = \frac{256}{12,8.10^3} = 20 \text{ ms}$$

D'où nous tirons ladurée totale du signal.

$$t_t = t \times N = 20 \times 16 = 0,325.$$

CHAPITRE III

RECONNAISSANCE DE LA PAROLE

La reconnaissance de la parole:

1. Introduction:

La reconnaissance de parole, c'est pour une machine la faculté de reconnaissance puis de réagir à la voix d'un ou de plusieurs locuteurs.

Cependant, après plus de trente ans de recherche fondamentale et appliquée, le traitement automatique de la parole a donné lieu à des applications pratiques importantes. Mais ces applications souvent spectaculaires ne doivent pas masquer l'ampleur des problèmes rencontrés et ceux qui restent encore à résoudre, d'ailleurs il a fallu attendre l'arrivée des ordinateurs et l'aboutissement des études de physiologie en phonétique pour traiter la parole d'une manière plus fine. Et plus récement, des progrès importants ont été réalisés.

La reconnaissance des mots isolés ne constitue plus un problème et ce depuis près de quinze ans. Mais les difficultés sont encore nombreuses en ce qui concerne la reconnaissance des mots enchainnés.

2. Les grandes approches:

Le signal est non seulement variable d'un locuteur à un autre, mais aussi pour un locuteur donné (état émotif, fatigue, ...), ce qui rend très ardu le problème de la reconnaissance d'un méssage, indépendamment de son locuteur.

Du fait de ces difficultés, deux voies ont été suivies simultanément, la première ayant été historiquement plus explorée que la seconde:

Reconnaissance de la parole:

- La première approche, paragramatique ou "globale" : consiste à appliquer des hypothèses simplificatrices au problème de façon à le rendre plus abordable, cette approche a conduit à la reconnaissance de mots isolés par des méthodes de reconnaissance des formes globales.

- La seconde approche, analytique: consiste à segmenter le message et à identifier ses constituants: phonèmes.

3. Reconnaissance de mots isolés:

L'opération de segmentation d'un message en ses constituants n'est pas aussi facile que son nom l'indique.

En effet, elle repose sur des règles empiriques qui décrivent les transitions possibles entre mots, elles sont fondées sur des critères d'énergie , de voisement,..., de telles lois sont très difficilement transposables et généralisables par passage d'un vocabulaire à un autre , ce qui diminue l'éfficacité, lorsque ces derniers sont importants et complexes.

D'où une reconnaissance appliquée à une mauvaise segmentation aura sans nul doute un résultat érroné.

Pour éviter cette opération délicate qu'est la segmentation, de nombreux projets de la R.A.P se sont intéréssés à la reconnaissance de mots prononcés isolement. Le processus devient alors plus aisé.

Reconnaissance de la parole:

Les mots constituants le vocabulaire à reconnaitre sont au cours d'une phase préalable et stockée en mémoire sous forme de paramètres cepstaux ou predicteurs (selon l'analyse faite précedement). Un mot à reconnaitre est considéré comme une forme globale, dont on mesure la "distance à l'aide d'une mesure adequate, aux différentes formes de référence stockées en mémoire .

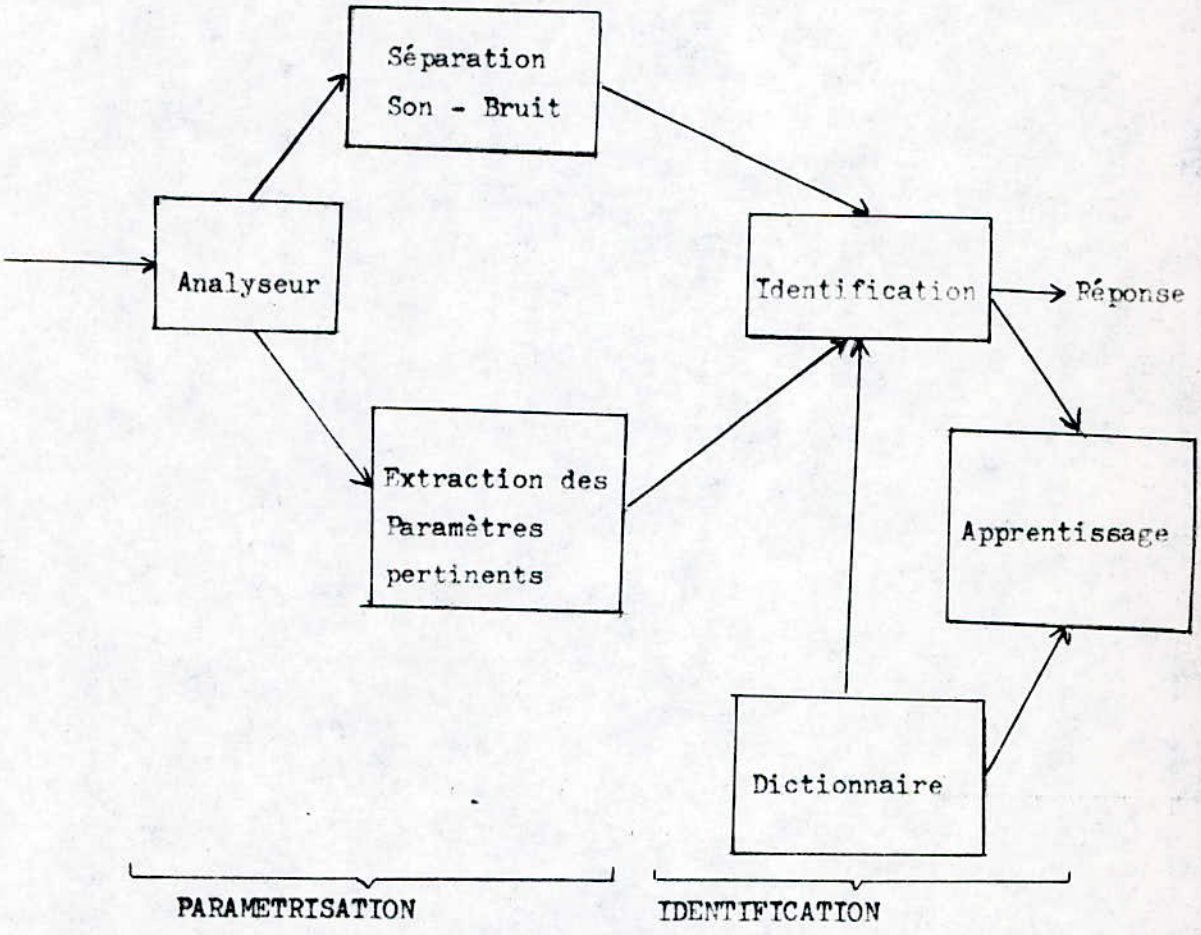


Figure III.1: Schéma d'un système de reconnaissance de mots

Reconnaissance de la parole:

4. Les sources de variabilités:

La difficulté de la variabilité du signal vocal qui affecte les mots prononcés, ces variabilités ont plusieurs sources parmi lesquelles nous citons:

- Les variabilités energétiques: le niveau sonore de la parole affecte le spectre des sons ce qui, fait que les spectres de mots identiques mais prononcés dans des états différents sont différents dans les détails.

- Les variabilités temporelles: la vitesse d'élocution d'un même locuteur est variable. Le contexte, l'intonnation et l'état de la personne sont autant de paramètres qui influent sur la longueur temporelle des mots à comparer.

En d'autres termes, ces variabilités se manifestent par des variations non linéaires de rythme et de durée des mots à comparer. Ce que lors de la reconnaissance, le taux de reconnaissance n'est jamais de 100%.

Plusieurs techniques concernant la reconnaissance ont été élaborées puis développées et enfin appliquées à la parole. Parmi lesquelles nous citons:

- La programmation dynamique qui s'avère très éfficace pour les mots isolés et très prometteuse pour les mots liés.

- Les processus de MARKOV, qui sont de plus en plus utilisés dans le domaine de la parole.

Reconnaissance de la parole:

Bien que ces techniques sont toutes fondées sur un même principe: la comparaison et l'optimisation, chacune a ses propriétés spécifiques.

La technique de la comparaison dynamique a fournit dès le début des années 1970, unesolution optimale à ce problème (13).

5. La programmation dynamique:

La recherche d'une solution optimale pour un problème quelconque fut l'objet de l'étude de Jean Bernoulli, **euleur**, Lagrange, Hamilton, et celui d'un grand nombre de mathématiciensmodernes.

Cependant la programmation dynamique est l'une des techniques ayant répondu aux besoins de ces **chercheurs**, c'est une méthode d'optimisation à aspect séquentiel, qui fait intervenir le temps d'une manière décisive. **E**lle est b**asée** sur le principe d'optimalité introduit par R. Bellman, en 1957. (5)(6)

Enoncé du principe:

Une politique est optimale si, à une période donnée quelque soit les décisions précendantes, les décisions qui restent à prendre constituent une politique optimale en regard du résultat des décisions précédentes.

Reconnaissance de la parole:

Les problèmes traités par programmation dynamique doivent satisfaire les conditions suivantes:

- La mise du problème sous forme sequentielle

- La fonction d'optimalité doit être décomposable

6. Application de la programmation dynamique à la reconnnaissance:

Le principe de la reconnaissance automatique de la parole basé sur la programmation dynamique consiste à chercher la meilleure supperposition de deux mots.

La comparaison se fait globalement sur les trames des mots. S'ils sont différents, la distance attendue est plus grande que celle qui découle de deux mots identiques.

Le champ de comparaison est un rectangle dont chaque côté est gradué par les trames des mots à comparer. Figure III.2

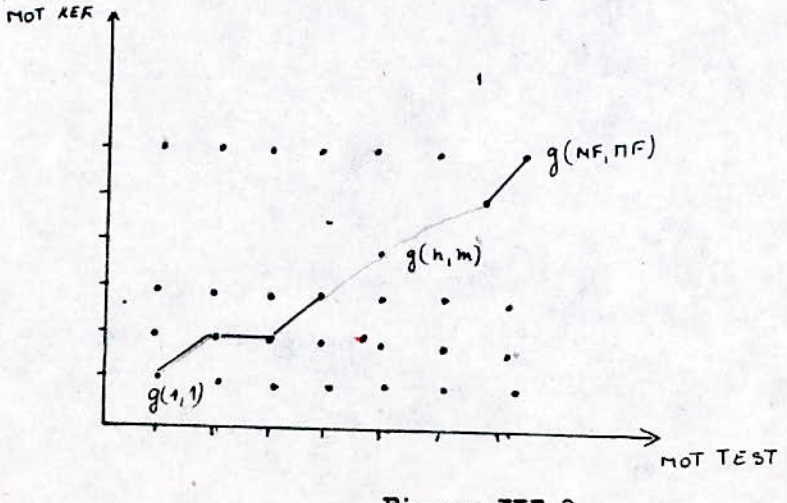


Figure III.2

CHAMP DE COMPARAISON, SANS AUCUNE RESTRICTION

Reconnaissance de la parole:

Où: NF : la longueur du mot référence (le nombre de trames)

MF : la longueur du mot test (le nombre de trames)

Localement la comparaison se fait entre coefficients des deux trames d'où la notation;

$$R = R(n) = (r1(n), (r2(n), \ldots\ldots rP(n)),$$
$$n = 1, \ldots\ldots NF$$
$$T = T(m) = (t1(m), t2(m), \ldots\ldots tP(m)),$$
$$m = 1, \ldots\ldots MF$$

Avec R : mot référence

T : mot test

P : le nombre de coefficients pour chaque trame, 12 pour la prédiction linéaire et 8 pour la cepstrale.

7. Contraintes locales:

Pour passer d'une trame à une autre, les chemins autorisés sont représentés sur la Figure III.3

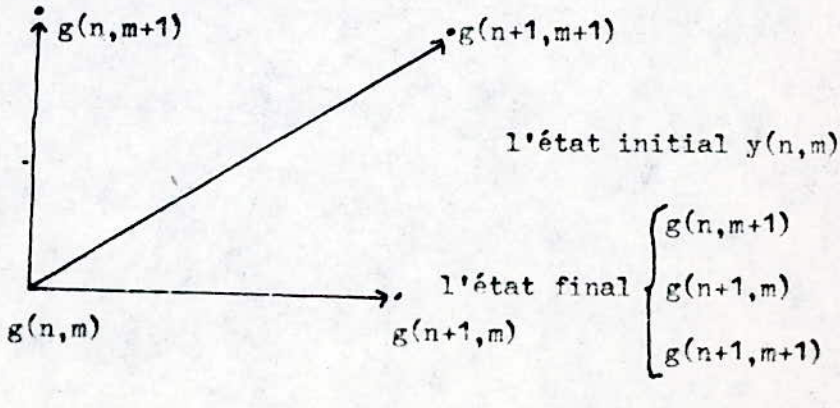
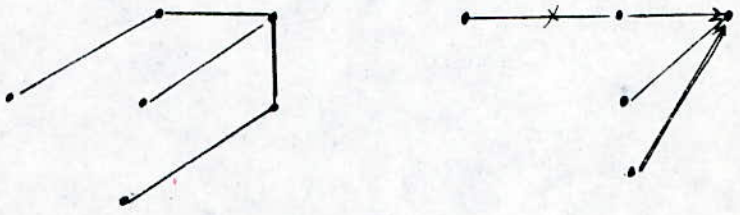


fig: III.3 - REPRESENTANT LES CHEMINS ADMISSIBLES POUR PASSER D'UNE TRAME A UNE AUTRE.

Reconnaissance de la parole:

Mais il a été remarqué q'une telle procédure conduit à un traitement de certains cas inutils (calcul de certaines distances inutilement). Pour éviter ce problème qui agit principalement sur le temps de réponse du système, Sakoe et Shiba et Itakura proposent deux types de contraintes représentées en figure III.4, appelés respectivement type I ou contraintes de Sakoe et Shiba, et type II ou contraintes d'Itakura.



Type I. Type II.

Figure III.4: LES DEUX TYPES DE CONTRAINTES

En effet ces contraintes sont non seulement simples à mettre au point, mais aussi performantes. Elles diminuent le nombre d'opération à traite et par conséquent améliore le temps de réponse du système.

Chaque contrainte peut être utilisé avec l'une des quatre (04) pentes suivantes: P=0, P=½, P=1 et P=2.

Le tableau suivant donne les equations recurrentes symétriques et asymétriques avec les contraintes de pente citées.

	Symétrique / Asymétrique		Equations recurrentes
P=0	Symétrique	min	$g(n,m-1) + d(n,m)$ $g(n-1,m-1) + 2d(n,m)$ $g(n-1,m) + d(n,m)$
	Asymétrique	min	$g(n,m-1)$ $g(n-1,m-1) + d(n,m)$ $g(n-1,m) + d(n,m)$
P=½	Symétrique	min	$g(n-1,m-3) + 2d(n,m-2) + d(n,m-1) + d(n,m)$ $g(n-1,m-2) + 2d(n,m-1) + d(n,m)$ $g(n-1,m-1) + 2d(n,m)$
			$g(n-2,m-1) + 2d(n-1,m) + d(n,m)$ $g(n-3,m-1) + 2d(n-2,m) + d(n-1,m) + d(n,m)$
	Asymétrie	min	$g(n-1,m-3) + (d(n,m-2) + d(n,m-1) d(n,m)/3$ $g(n-1,m-2) + (d(n,m-1) + d(n,m))/2$ $g(n-1, m-1) + d(n,m)$ $g(n-2,m-1) + d(n-1,m) + d(n,m)$ $g(n-3,m-1) + d(n-2,m) + d(n-1,m) + d(n,m)$
P=1	Symétrie	min	$g(n-1,m-2) + 2d(n,m-1) + d(n,m)$ $g(n-1, m-1) + 2d(n,m)$ $g(n-2, m-1) + 2d(n-1,m) + d(n,m)$
	Asymétrie	min	$g(n-1,m-2) + d(n,m-1) + d(n,m)/2$ $g(n-1,m-1) + d(n,m)$ $g(n-2,m-1) + d(n-1,m) + d(n,m)$
P=2	Symétrie	min	$g(n-2,m-3) + 2d(n-1,m-2) + 2d(n,m-1) + d(n,m)$ $g(n-1,m-1) + 2d(n,m)$ $g(n-3,m-2) + 2d(n-2,m-1) + 2d(n-1,m) + d(n,m)$
	Asymétrie	min	$g(n-2,m-3) + 2d(n-1,m-2) + d(n,m-1) + d(n,m)/$ $g(n-1,m-1) + d(n,m)$ $g(n-3,m-2) + d(n-2,m-1) + d(n-1,m) + d(n,m)$

Tableau : ILLUSTRANT LES CONTRAINTES DE SAKOE ET SHIBA

Reconnaissance de la parole;

La contrainte de pente montre les chemins admissibles arrivant en un point considéré, elle est évaluée par le rapport P = n/m.

Comme le montre le tableau I, le nombre de possibilités offertes dans la recherche optimale différe d'une pente à une autre.

- Pente 0 : il n'y a pas de contraintes de pente, il y a trois (03) chemins possibles .

- Pente ½ : il y a cinq (05) chemins possibles.

- Pente 1 et 2 : il y a trois (03) chemins possibles.

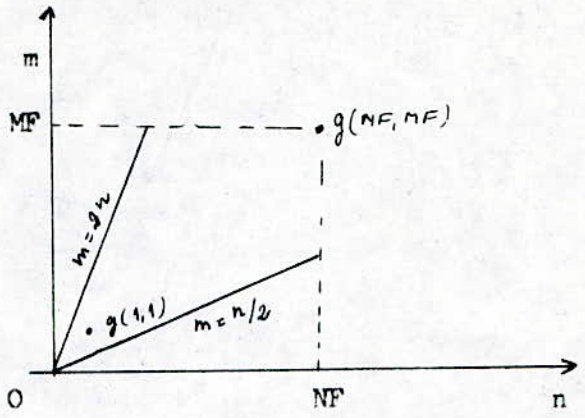
1 2 La contrainte globale:

C'est un moyen de rejection des références qui sont: soit trop longues, soit trop courtes par rapport au mot à reconnaitre. Elle est généralement prise dans un rapport extrême ck 2 où ½.

Néanmoins, il est possible de remédier au cas des comparaisons de mots très courts et ceci en utilisant la contrainte dans un rapport extrême de 3 ou 1/3, soit une contrainte globale, moins rigoureuse. Dans ce cas la reconnaissance se paie par un nombre de calcul supplémentaire (20).

Reconnaissance de la parole:

Le champ de comparaison devient alors un rectangle dont les côtés sont limités 2 et 1/2 de la **contrainte** globale, comme l'indique la figure III.5.



avec MF : nombre total de trames pour le mot test

NF : nombre total de trames pour le mot referme

Fig III.5.

Fonction de coïncidence:

Lors de la phase de reconnaissance le problème consiste à éliminer les différences temporelles entre les deux mots et en déduire la distance résiduelle après normalisation. Cette distance représente l'écart effectif entre les deux mots référance et test.

Les différences temporelles peuvent être traduites par la séquence de point:

$$F = F(1), F(2), \ldots, F(K)$$

où $F(1) = g(1,1)$, $F(K) = g(N,M)$

Reconnaissance de la parole:

Cette fonction de déformation F dite aussi fonction coïncidence, s'écarte plus ou moins de la diagonale et tend à s'en éloigner si les différences temporelles entre les deux mots augmentent.

Une façon d'évaluer la fonction de coïncidence F est de calculer la somme des distances pondérées le long du chemin associe à F:

$$S(F) = \sum_{1}^{K} g(n(K),m(k) \cdot w(k)$$

La distance globale entre l'état initial et l'état final sera normalisée par rapport au temp.

$$D(R,T) = \underset{k,(n,m)}{\text{Min}} \left[\frac{\sum_{}^{k} g(n(k),m(k)) \cdot W(k)}{\sum_{1}^{k} w(k)} \right]$$

Notons cependant que la fonction de coïncidence doit satisfaire les conditions suivantes:

Monotonie $\quad n(k-1) \leqslant n(k)$

$\quad m(k-1) \leqslant m(k)$

Continuité $\quad n(k) - n(k-1) \leqslant 1$

$\quad m(k) - m(k-1) \leqslant 1$

Reconnaissance de la parole:

10 - Fonction de pondération W(k)

Elle dépend de la forme à laquelle elle est associée:

- La forme symétrique:

$$W(k) = n(k) - n(k-1) + m(k) - m(k-1)$$

Elle correspond donc à une intégration le long d'un axe commun, à tout instant, aux deux mots.

- La forme asymétrique

$$W(k) = n(k) - n(k-1) \quad \text{ou} \quad W(k) = m(k) - m(k-1)$$

Elle correspond à une intégration le long de l'axe associé à R (Référence) ou à T (Test).

$W = W(k) = NF + MF$ Pour la forme symétrique

NF Pour la forme asymétrique par rapport à R.

Le problème d'optimisation devient

$$D(R,T) = \frac{1}{W} \underset{k(n,m)}{\text{Min}} \frac{K}{2} g(n(k),m(k) \cdot W(k)$$

Reconnaissance de la parole:

3- Choix de la distance:

Notion mathématiques sur les distances:

On appelle distance entre deux éléments X et Y d'un ensemble E, l'application définie de EXE dans R^+:

$$(X,Y) \text{ ------ } d(x,y) \quad R^+$$

Et vérifiant les propriétés suivantes:

$$d(x,y) = 0 \quad \Longleftrightarrow \quad x = y$$

$$d(x,y) = d(y,x) \qquad \forall x \in E, \quad \forall y \in E$$

$$d(x,y) \leq d(x,z) + d(z,y) \qquad \forall x \in E, \forall y \in E, \forall z \in E$$

Exemples de distances:

si $X = (x, \ldots\ldots, x_n)$

$Y = (y, \ldots\ldots, y_n)$

$d(x,y) = \sum (x_i - y_i)$ distance de Minkowsky

$d(x,y) = (\sum (x_i - y_i)^2)^{1/2}$ distance enclidienne

$d(x,y) = \text{Max} |x_i - y_i|$ distance de Chebytchef

Pour notre part, on se contentera de la distance euclidienne pour évaluer la similitude ou la dissimilitude de deux mots. Toutefois, il sera intéressant d'étudier l'influence du choix de la distance sur le taux de reconnaissance.

Reconnaissance de la parole:

8- Programme de la reconnaissance:

La reconnaissance proprement dite commence par le choix de la voyelle à traiter. Celle-ci est représentée par ses P coefficients sur les NF trames, qui sont préalablement stockés dans un fichier dit fichier test de la même manière que pour la constitution du dictionnaire.

La reconnaissance de la parole se base sur la comparaison et l'optimisation, c'est à dire que la mot (voyelle) test est comparé à tout le vocabulaire du dictionnaire.

La plus petite distance doit necessairement correspondre à la comparaison du mot (voyelle) test avec son correspondant dans le dictionnaire, néanmoins si celui-ci n'existe pas, le mot test est confondu avec un autre du dictionnaire. Pour éliminer ce risque, une distance seuil est imposé de telle manière que le mot test soit rejeté.

En fait, la comparaison entre mots est du point de vu global. Localement, elle se fait entre coefficient de deux trames: l'une appartenant au test l'autre au dictionnaire, autorisée par le fenêtre et la pente proposées par Sakoe et Schiba.

Notre programme utilise comme pente P = 1, celle-ci est la plus performante car, elle donne le plus grand taux de reconnaissance (20), et comme fenêtre M = 2N (limite supérieure) et M = N/2 (limite inférieure) car une plus petit de rejeter un mot du dictionnaire et une plus large entraine un nombre de calcul plus grand, agissant ainsi sur le temps de réponse du système.

Reconnaissance de la parole:

Ainsi tous les chemins lient deux trames contenues dans le champ de comparaison sont calculés, mais seuls les chemins optimaux sont pris en considéretion.

Enfin, lorsque la décision concernant le mot à reconnaitre ne peut pas être prise, celui-ci est à rejeter. Il existe deux sortes de rejet:

- Ceux qui sont rejetés avant la comparaison avec le dictionnaire de référence. C'est le cas d'un mot très court ou très long.

- Ceux qui sont rejetés après la comparaison avec le dictionnaire de référence, ce sont ceux dont la distance globale optimale (minimale) entre le mot à reconnaitre et la totalité du dictionnaire est tellement grande que nous nous permettant de penser que le mot à reconnaitre et soit une détection d'un bruit soit un mot qui fait parti du vocabulaire mais dans un état méconnaissable (20).

Le rejet se fait à la suite de la détermination d'une distance seuil, distance globale optimale (minimale) entre mots. Celle-ci s'obtient par renouvelement de l'opération (reconnaissance) pour une masse très importante de mots de telle manière à ne pas rejeter un mot qui peut être reconnu et à ne pas identifier un autre qui doit être rejeter d'où la reconnaissance dépend énormement de cette étape.

Néanmoins, il est plus tolérable de rejeter un mot dont l'identification est peu sûre plutôt que de comettre des erreurs flagrantes (20).

Reconnaissance de la parole:

1 3 Algorithme de la normalisation par la programmation dynamique:

- Fixation des paramètres initiaux (état initial)

$(n,m) = (1,1)$

$g(1,1) = 2 \times d(1,1)$

dans le cas des contraintes symétriques.

- Application de la contrainte globale (fenêtre) ayant pour limites:

$m = 2n$ limite supérieure

$m = n/2$ limite inférieure

- Si le point considéré est à l'intérieur de la fenêtre alors:

* Si les chemins admissibles (admit par la contrainte locale) (la pente) arrivant en ce point, partantd'autres points situés à l'intérieur du champ de comparaison alors:

Calculer la distance.

$$g(n,m) = \begin{cases} g(n-1,m-2) + 2d(n,m-1) + d(n,m) \\ g(n-1,m-1) + 2d(n,m) \\ g(n-2,m-1) + 2d(n-1,m) + d(n,m) \end{cases}$$

si non, Ou si le point considéré est à l'extérieur de la fenêtre incrementer m, incrementer n.

Refaire toutes les étapes avec comme point de départ $g(n-1, m-1)$ ou $g(n, m-1)$ ou alors $g(n-1,m)$ suivant le traitement precedent et ce pour la pente P=1.

Reconnaissance de la parole:

14 Résultats:

Rappelons que l'objectif de notre étude est la mise au point d'un logiciel de reconnaissance des mots isolés (voyelles) et ceci en utilisant deux méthodes d'analyses: la prediction linéaire (LPC) et et la cepstrale, et comme méthode de normalisation temporelle la programmation dynamique.

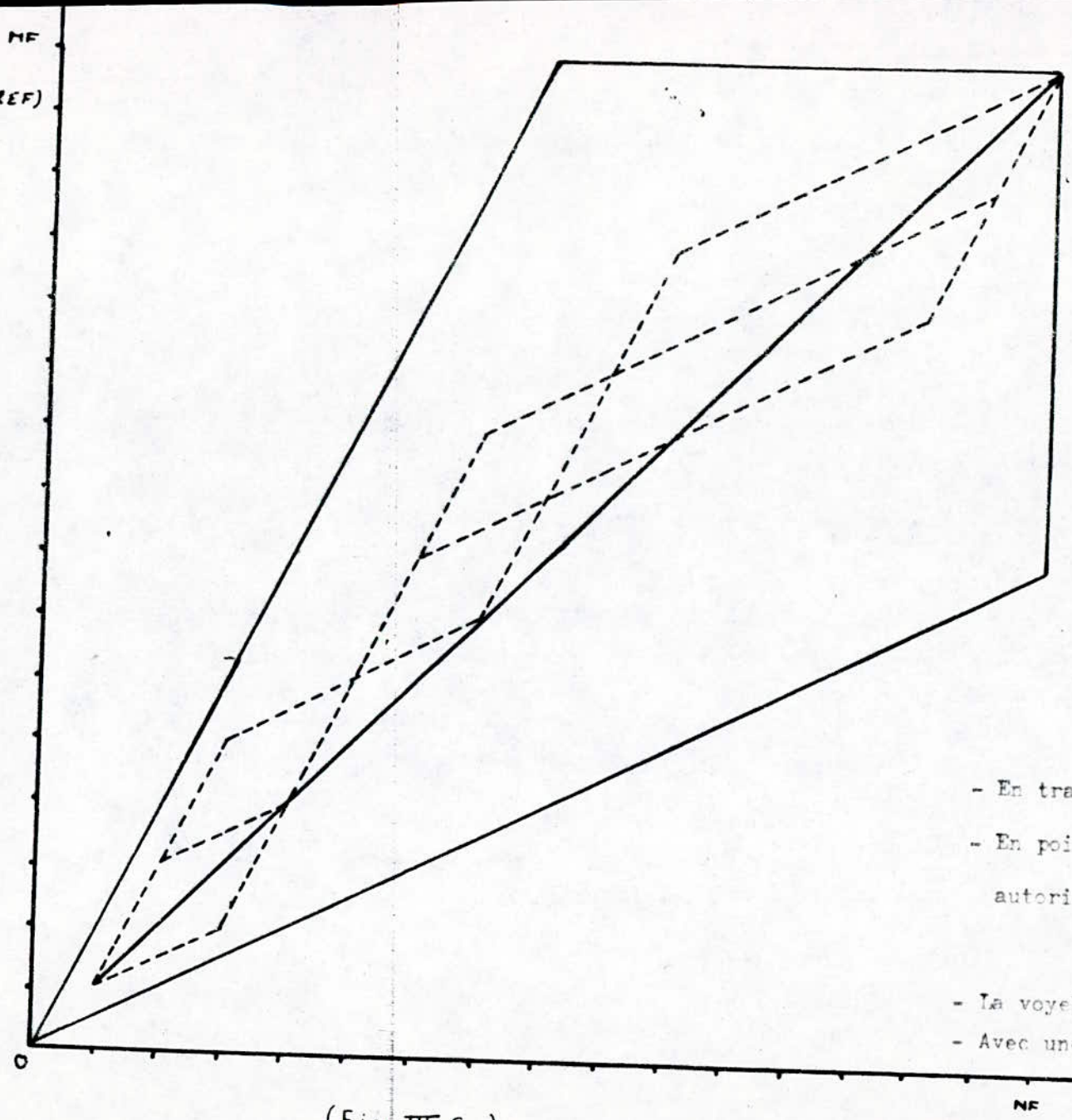
Les figures III,6 a, III 6 b, III 6 c, III 6 d, III 6 e, III 6 f et III 7 a, III 7 b, III 7 c, III 7 d, III 7 e, III 7 f, représentent chacune la comparaison d'une voyelle test avec tout le vocabulaire du dictionnaire qui est dans ce cas au nombre de six (6).

Ainsi nous pouvons constaté (comme le vérifie la (théorie) que le chemin global optimal (en trait plein), coincide avec la première bissectrice.

Nous pouvons expliquer la présence des distorsions à la fin de chaque chemin par l'acumulation des erreurs provenant de l'arondissement des chiffres appelé: bruit de calcul. Mais étant donné que cet arrondissement se fait sur le dernier chiffre uniquement, nous pouvons remédier relativement à ce problème, en prenant le format sur lequel s'écrit le chiffre, le plus grand possible, toutefois sans dépasser la limite. Tendant ainsi l'erreur vers zéro et donc évitant l'influence de celle-ci sur l'intelligibilité de la voyelle et sa reconnaissance.

Les résultats obtenus par la LLC sont plus satisfaisant que ceux qui découlent de la cepstrale.

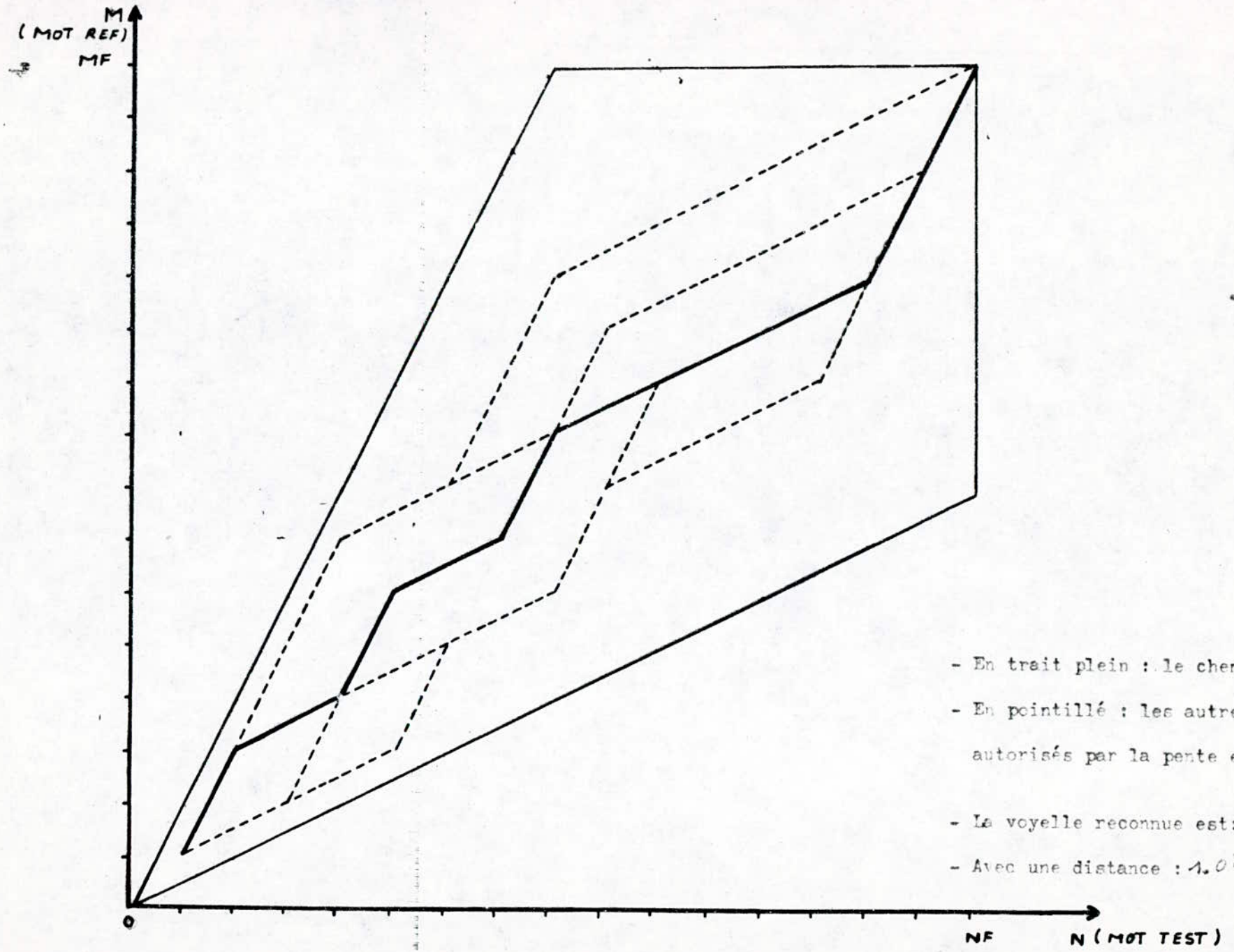
RESULTATS
DE LA
PREDICTION LINEAIRE



- En trait plein : le chemin optimal
- En pointillé : les autres chemins autorisés par la pente et la fenêtre

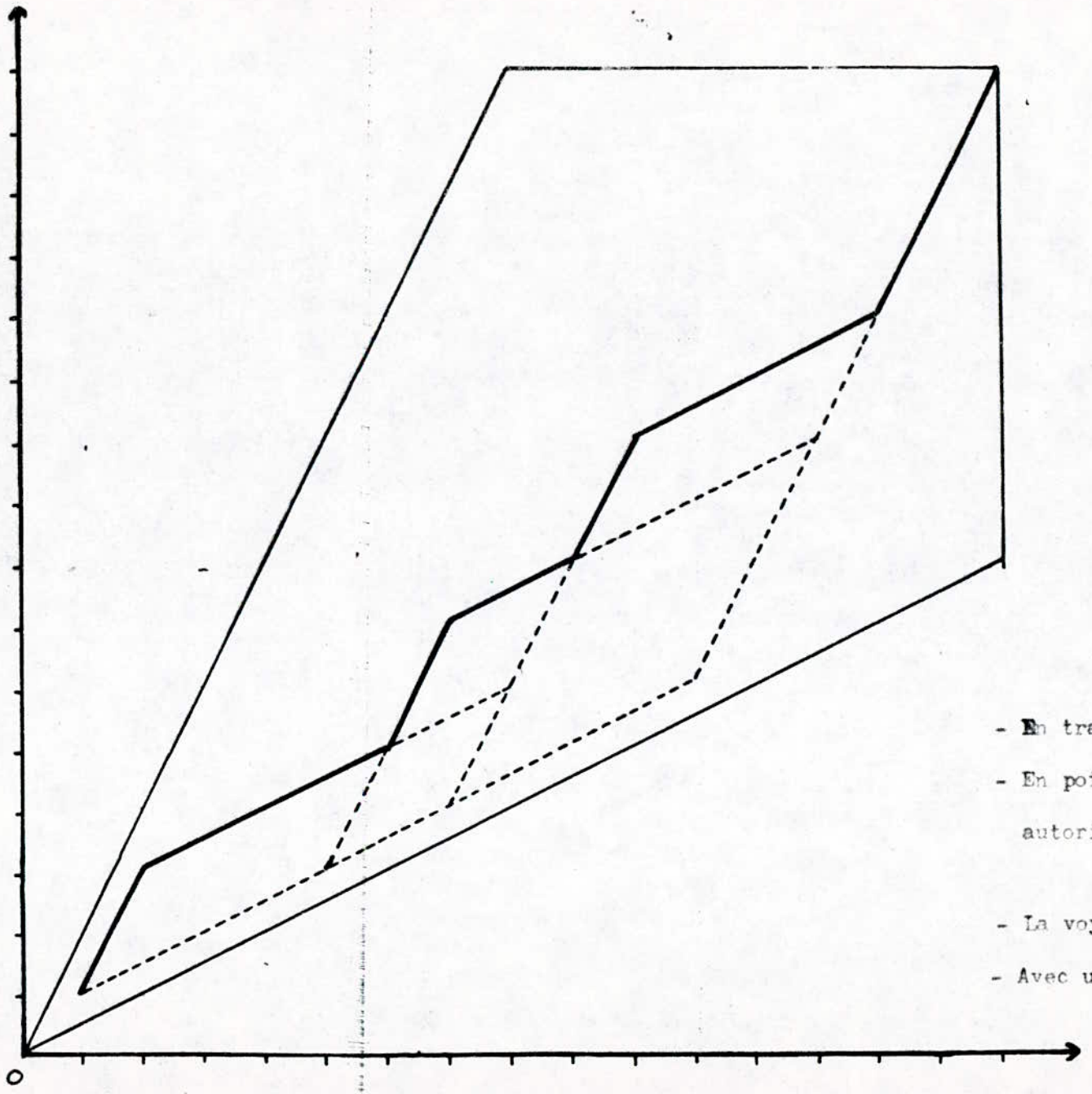
- La voyelle reconnue est : U
- Avec une distance : 2.3111121E-07

(Fig III 6a) : CHEMIN DE DEFORMATION DE LA VOYELLE : U



- En trait plein : le chemin optimal
- En pointillé : les autres chemins autorisés par la pente et la fenêtre
- La voyelle reconnue est: O
- Avec une distance : 1.08055428E-07

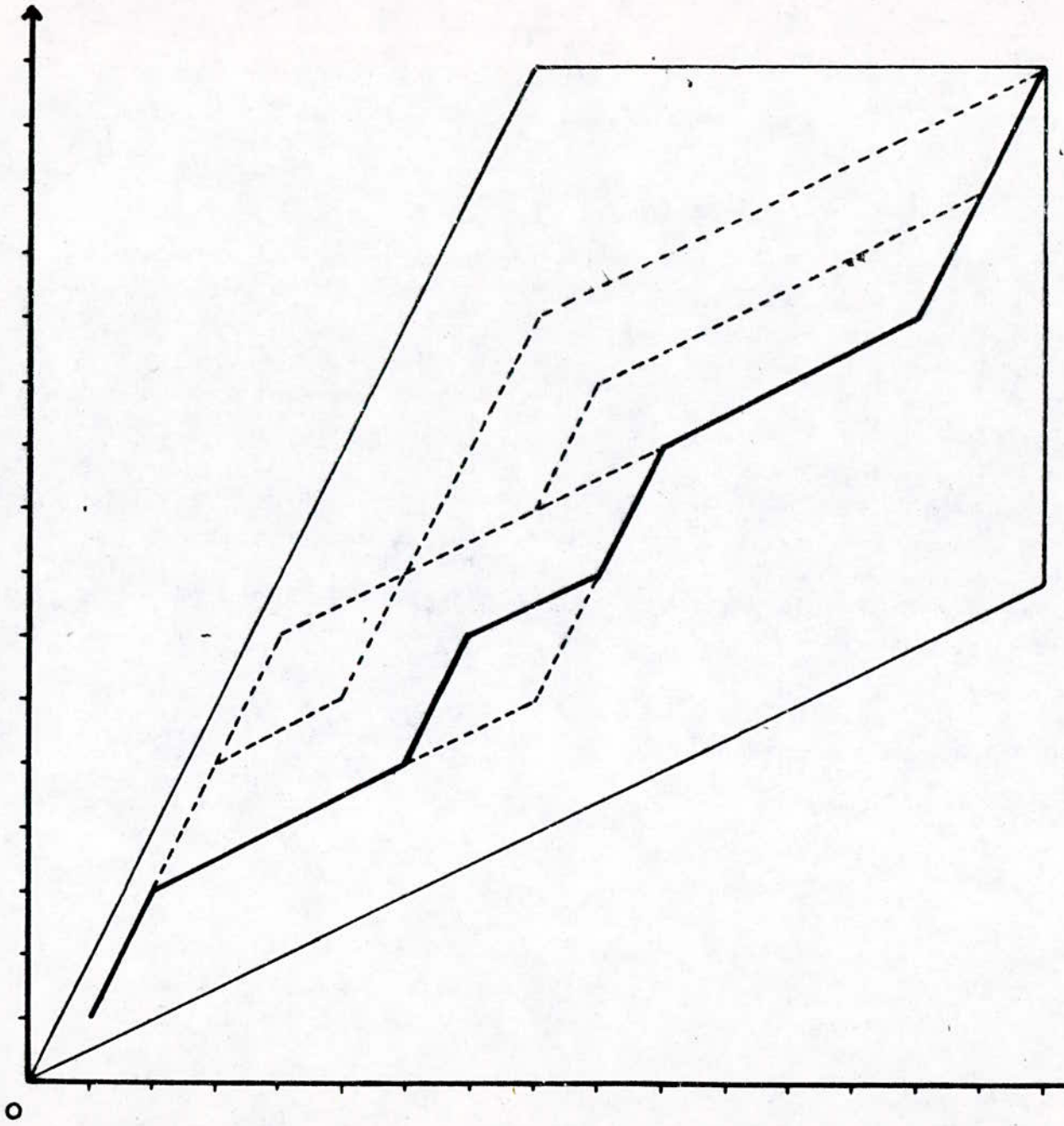
(fig III 6.b) : CHEMIN DE DEFORMATION DE LA VOYELLE : O



- En trait plein : le chamin optimal
- En pointillé : les autres chemins autorisés apr la pente et la fenêtre

- La voyelle reconnue est : E
- Avec une distance : 5.960464455E-08

(fig III 6 c) CHEMIN DE DEFORMATION DE LA VOYELLE : E

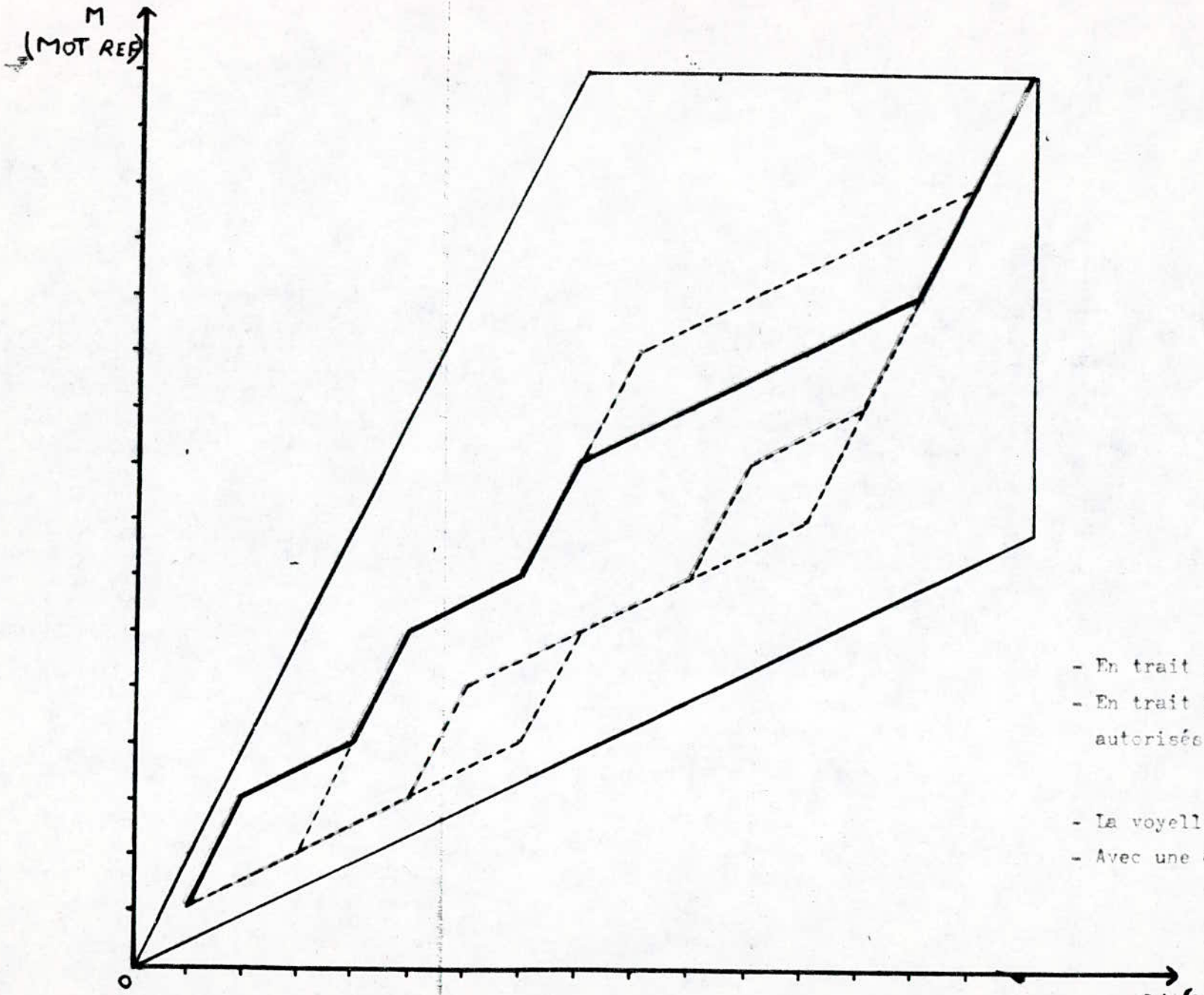


- En trait plein : le chemin optimal
- En pointillé: les autres chemins autorisés par la pente et la fenêtre.

- La voyelle reconnue est A

  Avec une distance: 0

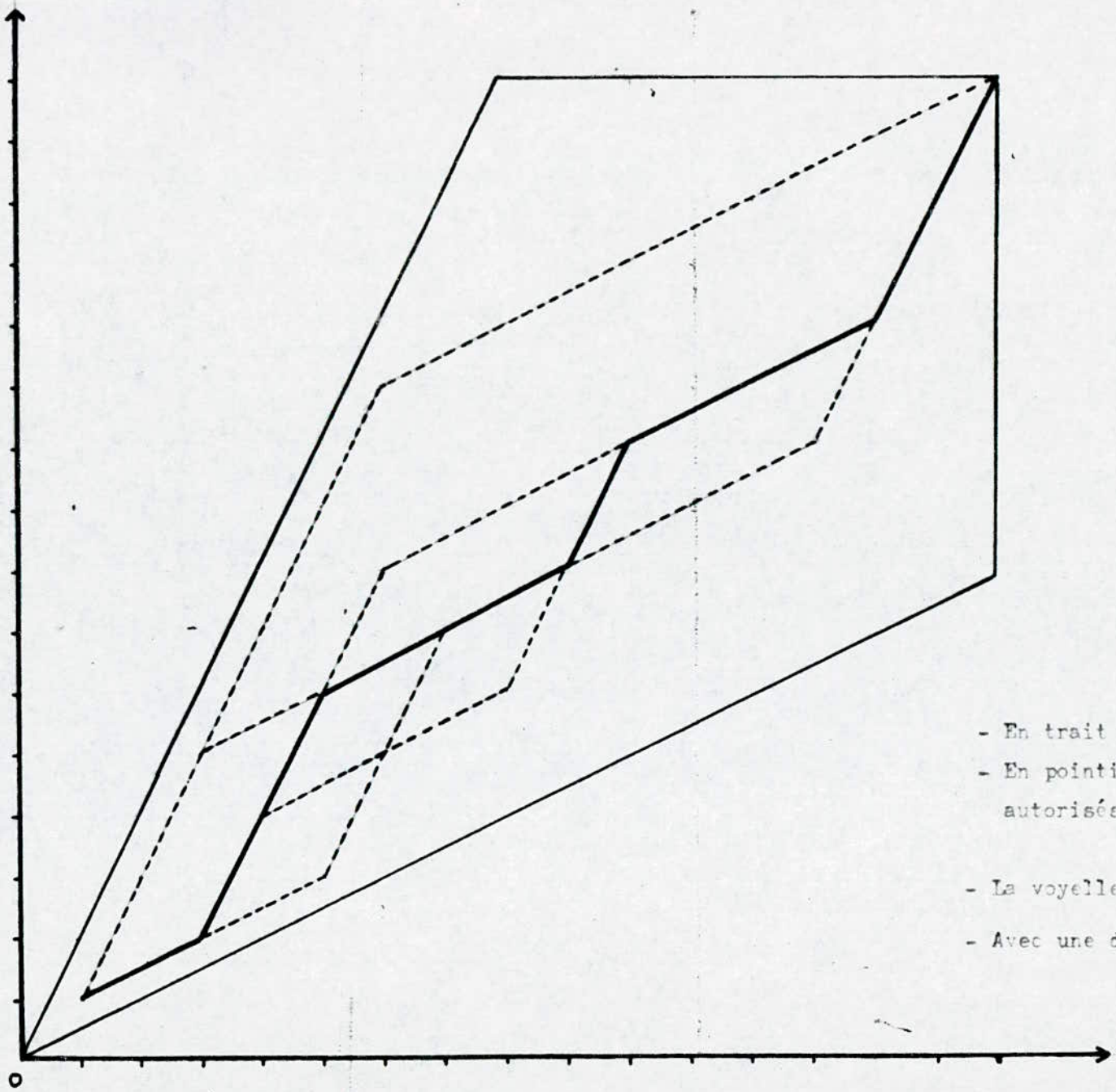
(fig III 6 d) Chemin de Déformation de la Voyelle A



- En trait plein: le chemin optimal
- En trait pointillé: autres chemins autorisés par la pente et fenêtre

- La voyelle reconnue est I
- Avec une distance: 0

(fig III 6 e) : CHEMIN DE DEFORATION DE LA VOYELLE I

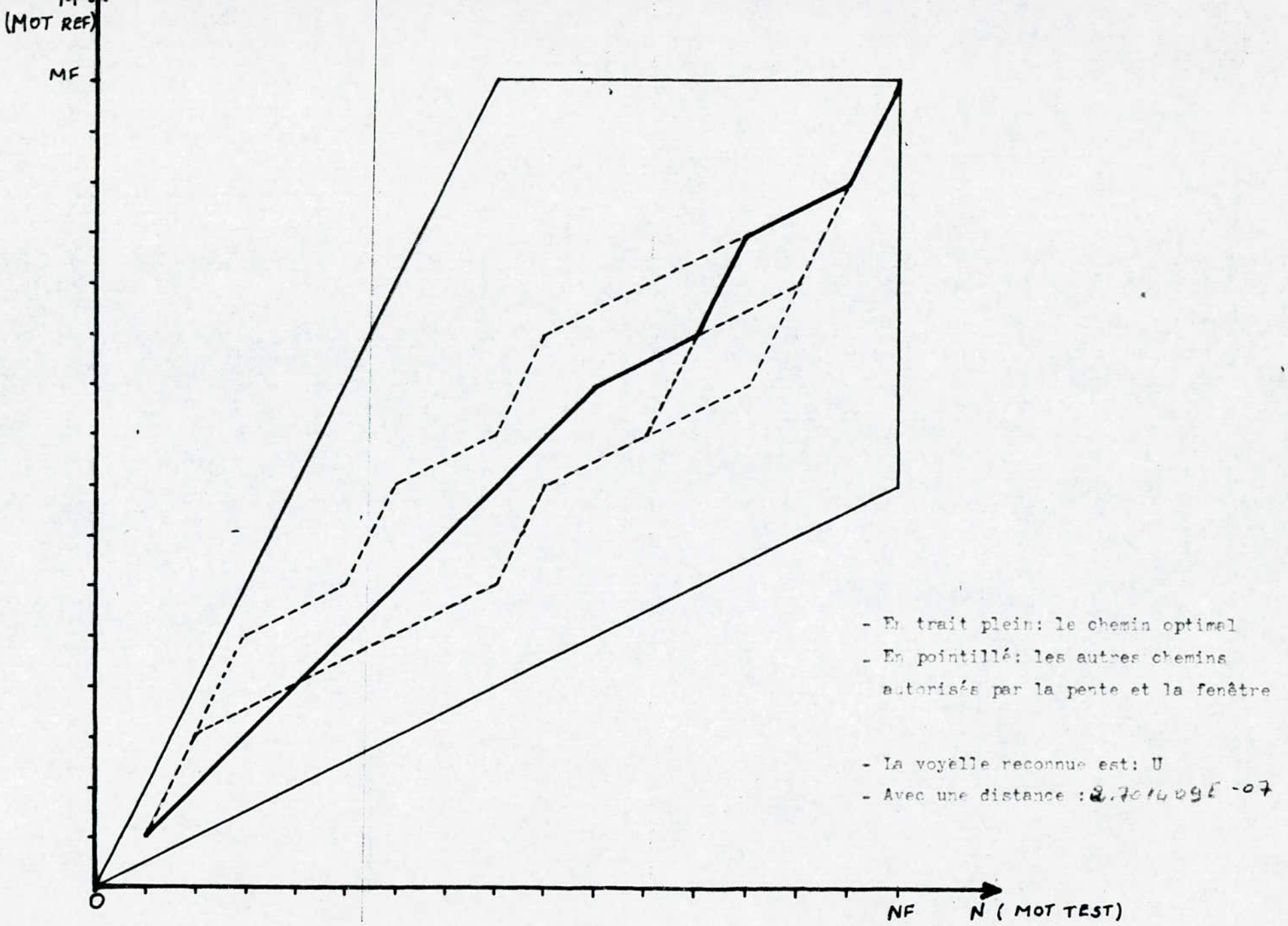


- En trait plein : le chemin optimal
- En pointillé : les autres chemins autorisés par la pente et la fenêtre

- La voyelle reconnue est : Y
- Avec une distance: 0

(fig III 6 f) CHEMIN DE DEFORMATION DE LA VOYELLE Y

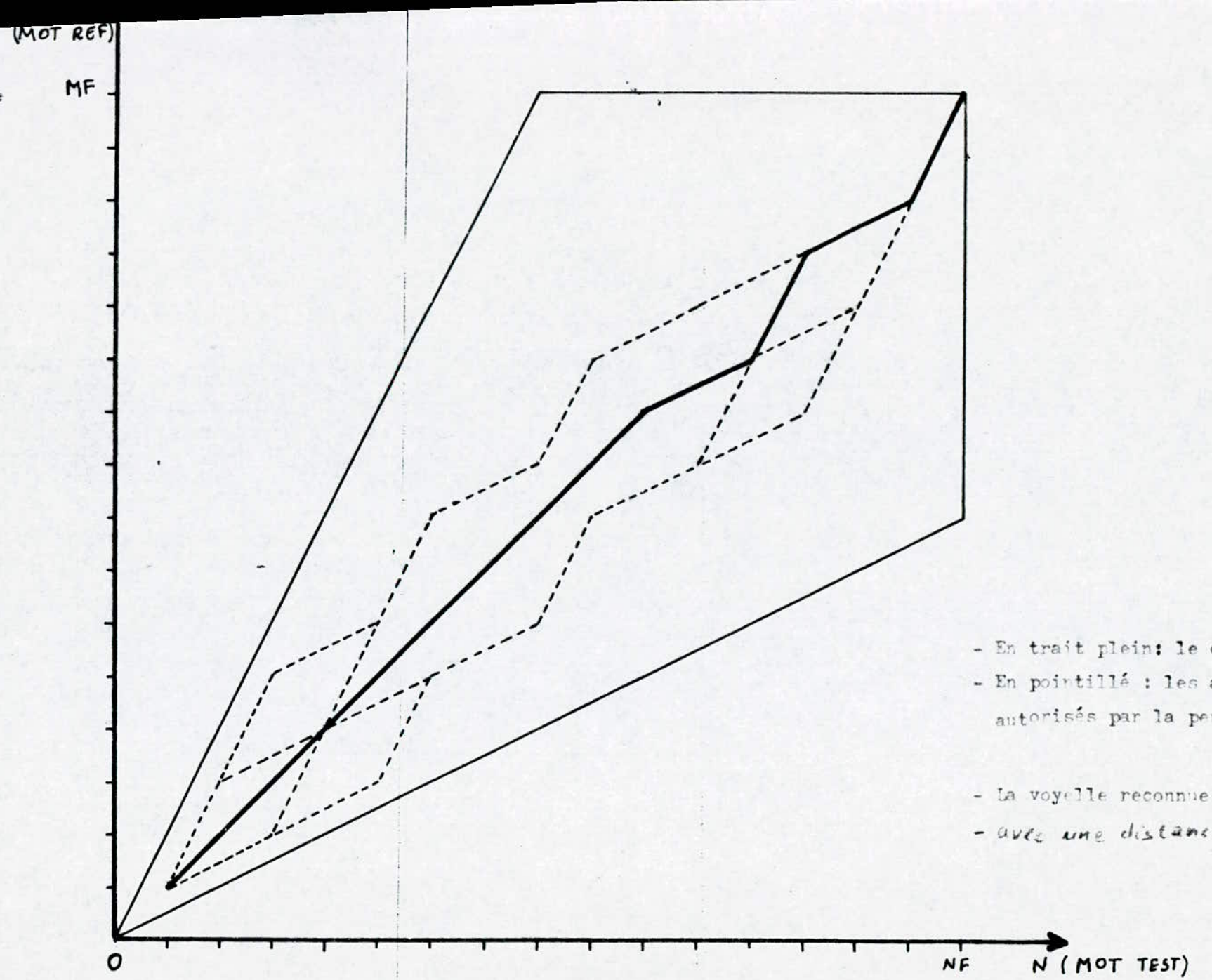
RESULTATS DE LA CEPSTRALE



- En trait plein: le chemin optimal
- En pointillé: les autres chemins autorisés par la pente et la fenêtre

- La voyelle reconnue est: U
- Avec une distance : 2.7016096E-07

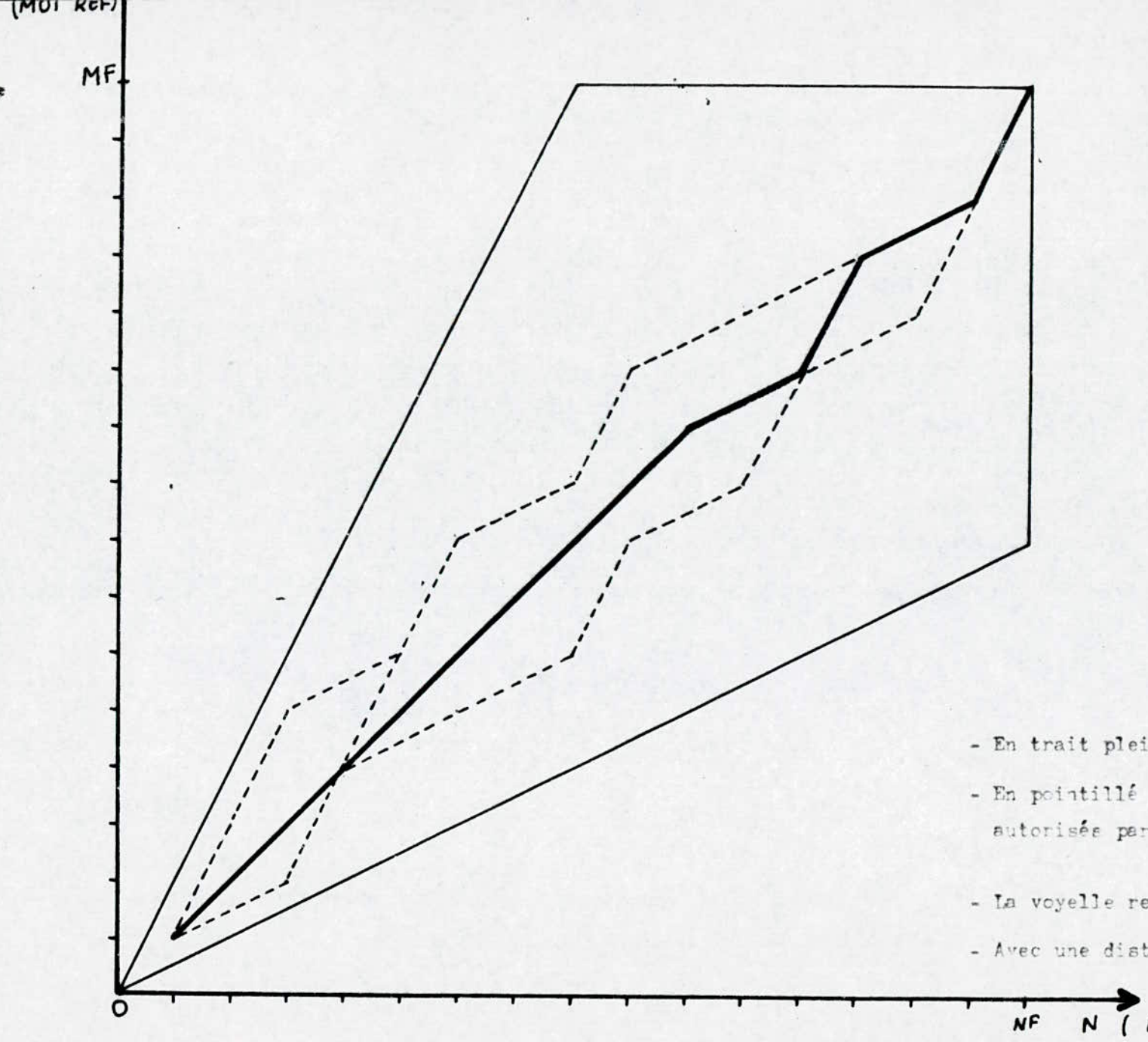
(fig III 7 a) chemin de DEFORMATION DE LA VOYELLE / U



- En trait plein: le chemin optimal
- En pointillé : les autres chemins autorisés par la pente et la fenêtre

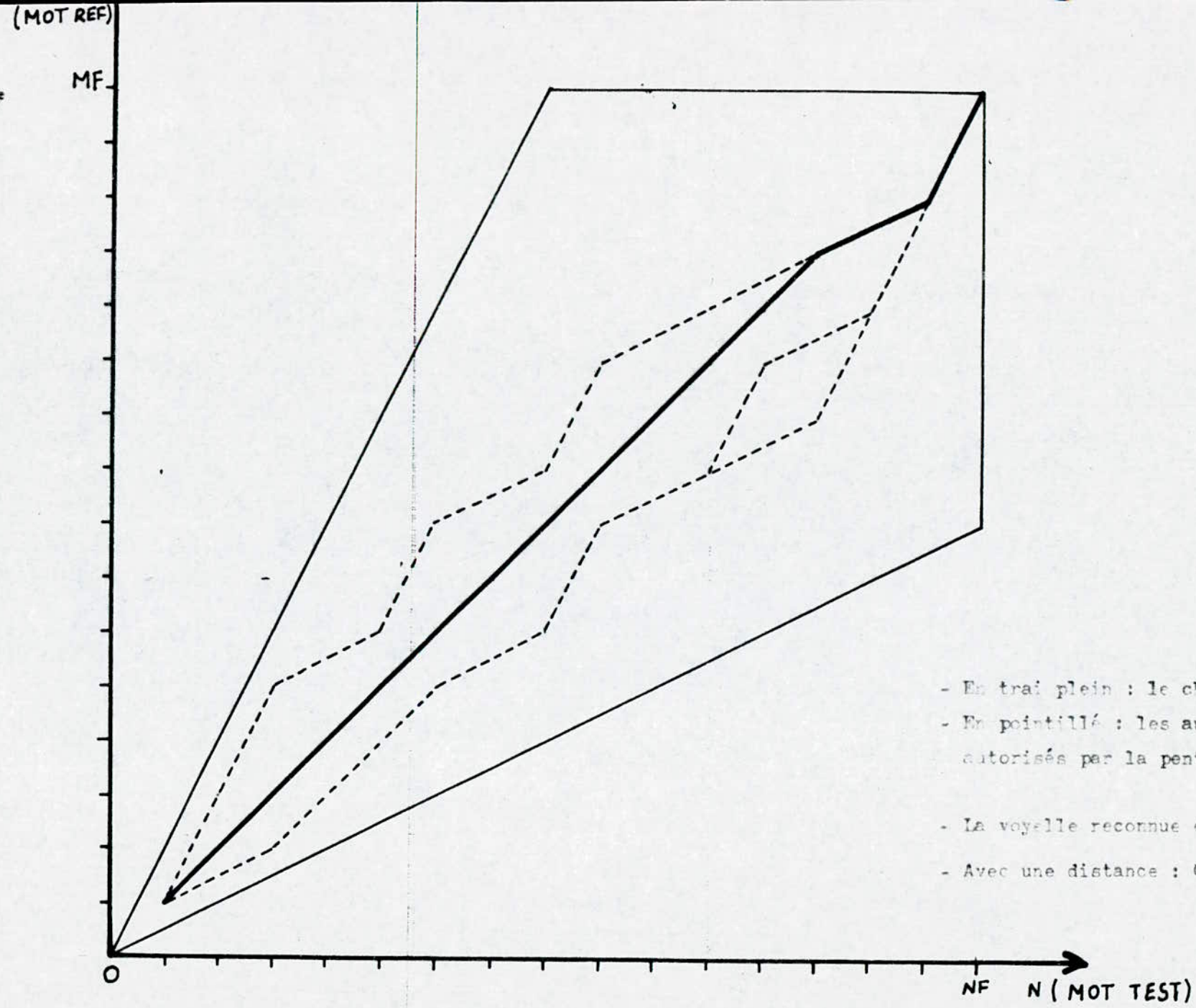
- La voyelle reconnue est : O
- avec une distance : 0,22 40023

(fig III. 7b) CHEMIN DE DEFORMATION DE LA VOYELLE: O



- En trait plein : le chemin optimal
- En pointillé : les autres chemins autorisés par la pente et la fenêtre
- La voyelle reconnue est : E
- Avec une distance : 0,2000054

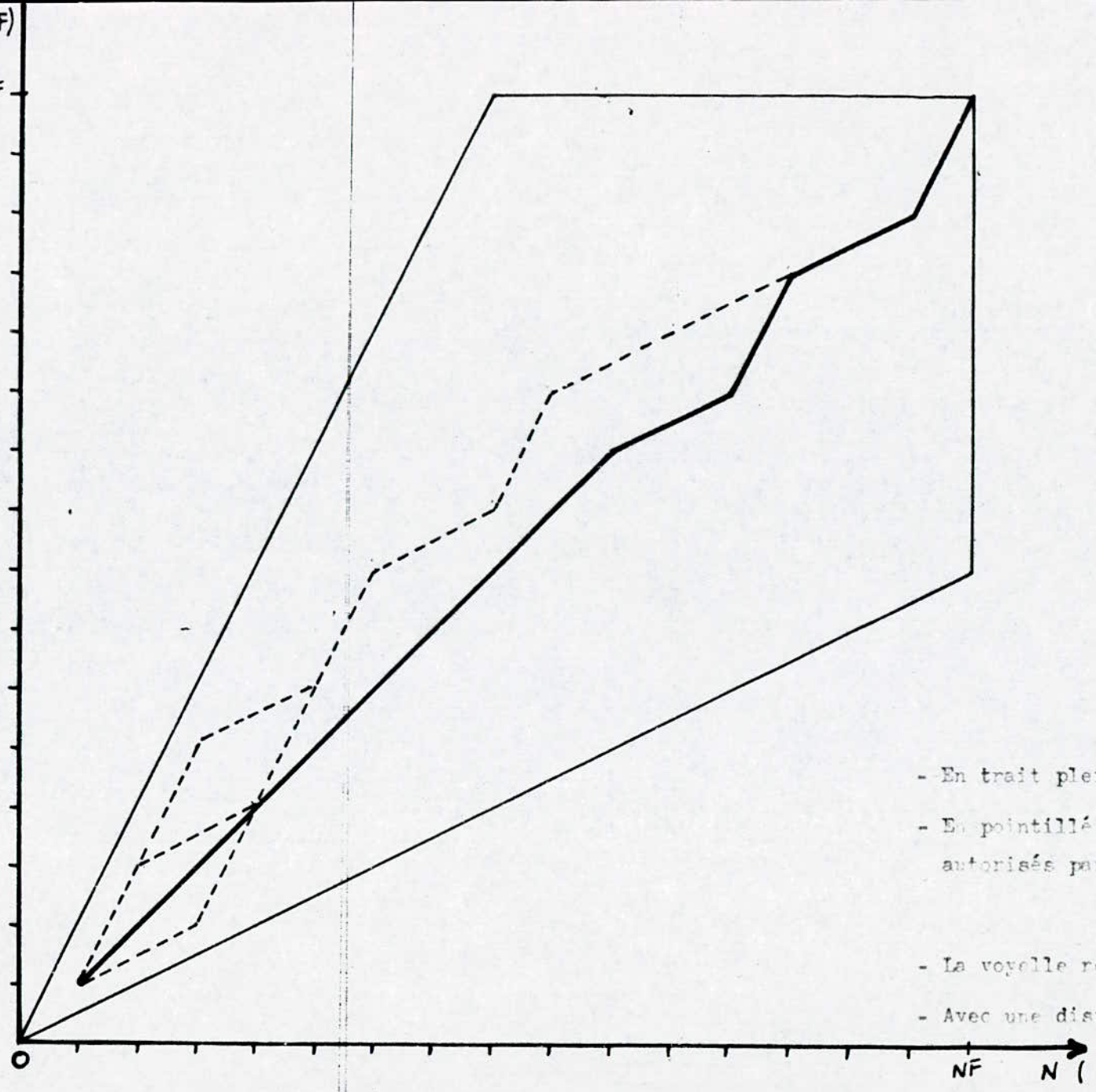
(fig III 7c) CHEMIN DE DEFORMATION DE LA VOYELLE : E



- En trai plein : le chemin optimal
- En pointillé : les autres chemins autorisés par la pente et la fenêtre

- La voyelle reconnue est : A
- Avec une distance : 0.190287

(fig III 7d) CHEMIN DE DEFORMATION DE LA VOYELLE : A

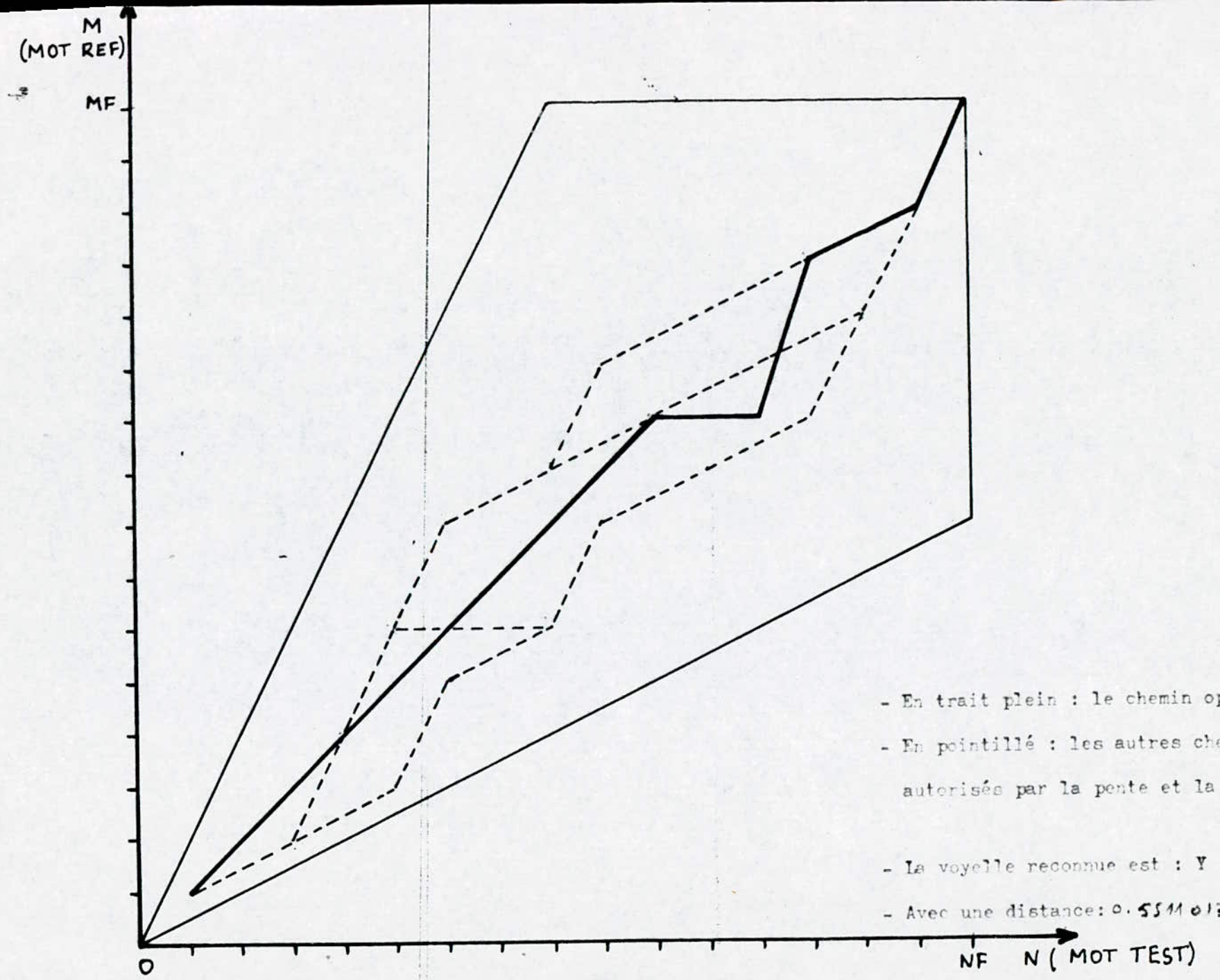


- En trait plein : le chemin optimal
- En pointillé : les autres chemins autorisés par la pente et la fenêtre.

- La voyelle reconnue est : I
- Avec une distance : 0 2663034

(fig III 7 e)

CHEMIN DE DEFORMATION DE LA VOYELLE : I



- En trait plein : le chemin optimal
- En pointillé : les autres chemins autorisés par la pente et la fenêtre

- La voyelle reconnue est : Y
- Avec une distance: 0.5511013

(fig III 7f) CHEMIN DE DEFORMATION DE LA VOYELLE: Y

Une erreur sur le format du dictionnaire, nous a mené à des résultats erronés en utilisant l'analyse cepstrale.

Le manque de temps ne nous a pas permis de faire apparaitre la correction du logiciel.

Les figures citées precedemment montrent que les distances sont très petites, négligeables même, (Dans la majorité des cas nulles pour la PLC.)

Comparaison de mots identiques et représentation formantique différentes:

Bien que la représentation formantique de deux mots identiques soit différente, le mot est reconnu. Chose attendue est que la distance globale optimale entre un mot et son correpondant dans le dictionnaire passe de zéro (évident pour une simulation) à une valeur supérieure (environ 3 pour U et 0, 5 pour E et 8 pour I) lorsque les formants se déplacent (glissement vers les hautes ou les basses fréquences).

Mais un cas exceptionnel se présente pour la voyelle y.

Notons cependant que si les formants sont très différents le mot n'est pas reconnu.

Distance seuil

Sans aucune condition cancernant la distance optimale, certains mots test n'ayant pas de correspondant dans le dictionnaire, au lieu d'être rejetés ils sont confondus à d'autres. Pour éviter ce risque la condition de seuil est établie. Néanmoins celle-ci est fixée arbitrairement.

Conclusion:

Une meilleure reconnaissance se base non seulement sur une bonne analyse mais aussi sur un choix préalable de la distance seuil.

La valeur du seuil doit être fixée expérimentalement suivant le lexique et l'application choisie.

Comparaison de mot ayant la même représentation formantique et de durées différentes.

Etant donné que la durée d'un mot varie d'une prononciation à une autre, nous avons choisi de faire ce test pour déterminer l'effet de la DTW sur notre travail.

En prenant deux voyelles identiques, mais de durées différentes, peut-on prévoir le résultat? Le mot test sera t-il identifié ou non?

La programmation dynamique a pour propriétés de fixer les extrémités du chemin optimal. L'état final du mot peut donc être à l'intérieur ou à l'extérieur du champ de comparaison. De ce fait si celui-ci est à l'extérieur, ce mot test est rejeté. Par contre s'il est à l'intérieur il reste déterminer s'il n'y a pas de confusion entre mots et ceci peut-être résolu par un choix adéquat du seuil.

Dans ce cas, nous obtenons un taux de reconnaissance de 33 %. Il serait interressant de ~~tester le~~ taux de reconnaissance en utilisant:

- D'une part d'autres contraintes pour l'algorithme de Sakoe-Chiba.

- D'autre part, d'autres algorithmes de la programmation dynamique.

Celle-ci est confondu avec la voyelle U quand ses formants glissent vers les hautes fréquences.

Les tableaux suivants résument les résultats obtenus en utilisant la LPC.

Tableau I : donne les distances pour chaque voyelle dans le cas où il n y a pas de chagement des formants:

VOYELLE	U1	O1	E1	A1	I1	Y1
U1	2.311121E-07	6.471391	36.96753	80.71133	22.62682	76.20394
U1	6.471392	1.0803342E-07	41.24354	81.97436	22.71299	74.92952
E1	36.96753	41.24354	5.9604645E-08	70.71635	34.92131	75.64623
A1	80.71133	81.97436	70.71625	0	69.11840	66.46940
I1	22.62682	22.71299	34.92131	69.15801	0	63.44027
Y1	72.20394	75.64623	66.46940	63.44027	63.44027	0
DECISION	RECONNUE	RECONNUE	RECONNUE	RECONNUE	RECONNUE	RECONNUE

Tableau II: donne les distances entre voyelle test et le contenu du vocabulaire lorsque les formants sont légèrement différents.

VOYELLE	U2	O2	E2	A2	I2	Y2
U1	2.778804	8.894033	76.43878	35.23437	14.26310	12.08919
O1	4.229023	2.422642	77.70181	42.51035	14.51272	15.27293
E1	37.478530	42.48460	66.07392	2.622848	34.17380	37.54332
A1	80.34079	82.51811	5.200924	71.29201	73.28974	81.10836
I1	21.82866	22.68513	64.84586	35.65385	8.363717	30.15358
Y1	55.09417	74.27103	61.766664	75.32768	67.95938	88.29312
DECISION	RECONNUE	RECONNUE	RECONNUE	RECONNUE	RECONNUE	NON RECONNUE

Reconnaissance de la parole:

1.- Domaine d'application:

L'utilisation de la parole comme mode de communication avec une machine présente des avantages certains notamment dans trois cas alliant souvent reconnaissance et synthèse de la parole.

- Utilisateur occasionnel, non spécialiste d'un système.

- Utilisateur ayant déjà les mains ou la vue occupées.

- Accès à distance (téléphone radio comunication)

Le champ des applications potentielles et donc très vaste. Néanmoins, certains domaines nécessite encore des travaux de recherche. Les applications actuelles relèvent maintenant de la reconnaissance de mots (isolés ou enchainés).

Les applications avancées de dialogue sont pour l'an 2000, et concernent essentiellement les domaines militaire (notamment avionique) et industriel (notamment automobile (13).

La parole permet aussi à un handicapé moteur un côntrole efficace de son environnement. Ce créneau d'application très limité devrait également être développer à l'avenir.

Reconnaissance de la parole:

La voix a également servi au contrôle d'un bras articulé lors d'une mission de la navette spatiale Américaine.

L'utilisation de la parole continue permettant dans de telles applications d'avoir un dialogue plus riche.

Conclusion:

De façon générale, le choix d'une application doit faire l'objet d'une étude attentive, fondée sur un ensemble de critères objectifs. En particulier, il est important d'examiner si la voix apporte véritablement un accroissement des performances ou un meilleur confort d'utilisation.

Par ailleurs, il ne faut pas trop attendre de la commande vocale et ne la considérer, en tout état de cause, que comme un moyen complémentaire, d'autres moyens de communication plus traditionnels.

Enfin, l'insertion d'un système de reconnaissance automatique de la parole R.A.P dans son environnement réel d'utilisation pose des problèmes très importants dont dépend pour une large part le succès final de l'opération ergonomie du système, niveau de dialogue, ...)

ORGANIGRAMME
DE LA
PREDICTION LINEAIRE

Definition des parametres utilisés dans les organigrammes.

organigramme : Extraction des parametres pertinents par la prediction lineaire (PLC)

NT : Le nombre total d'echantillon contenu dans le signal de parole.

N : Le nombre d'echantillon par fenêtre.

NF : Le nombre de fenêtre, Le nombre de trames.

K : permet Le deplacement de La fenêtre.

X(M) : Tableau de frequence caracteristique pour chaque voyelle.

R(M) : Tableau des amplitudes correspondantes.

SI(I) : Tableau du signal de parole.

W(I) : Tableau du signal de la fenêtre de Hamming.

R(L) : Tableau des coefficients d'autocorrelation.

D(I) : Tableau des coefficients de reflexion

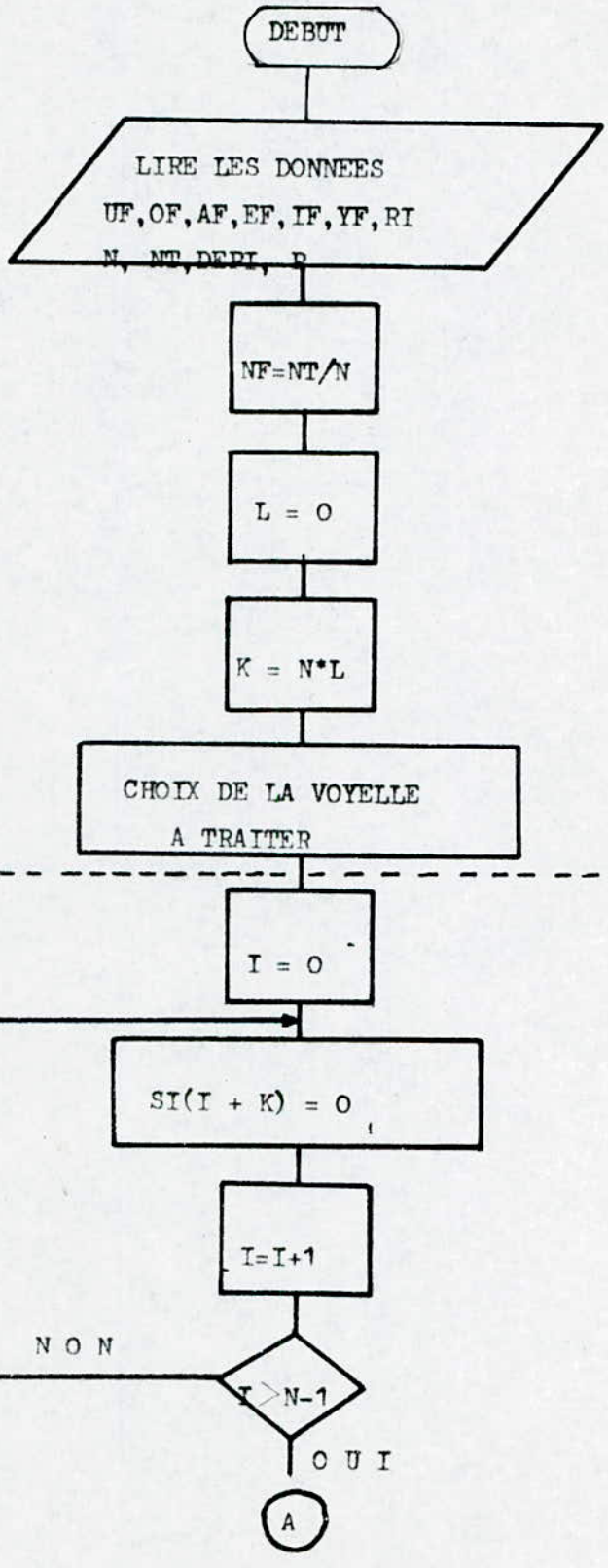
E(I) : Tableau des erreurs quadratique

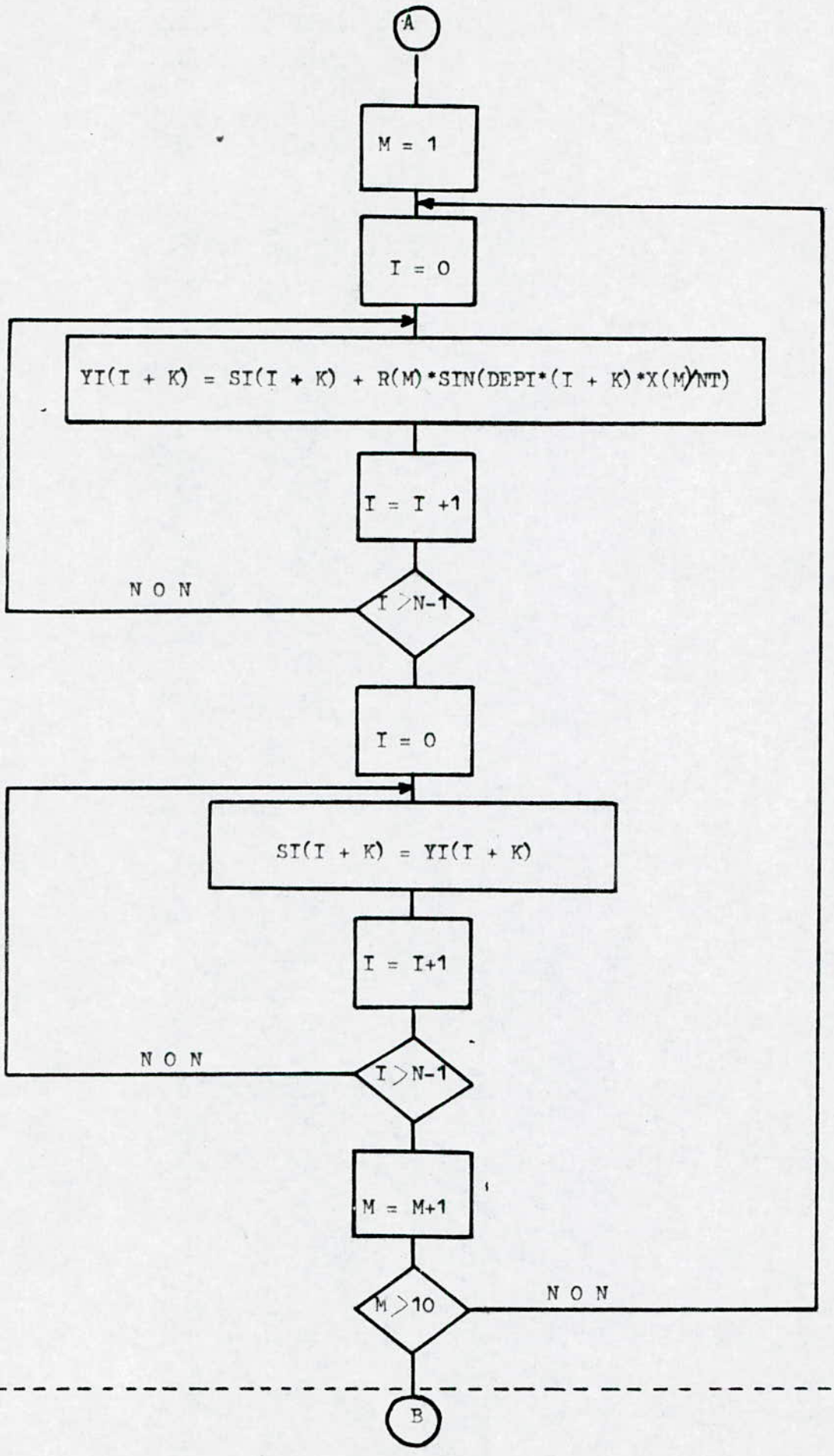
A(J,I) : Matrice des coefficients predicteurs.

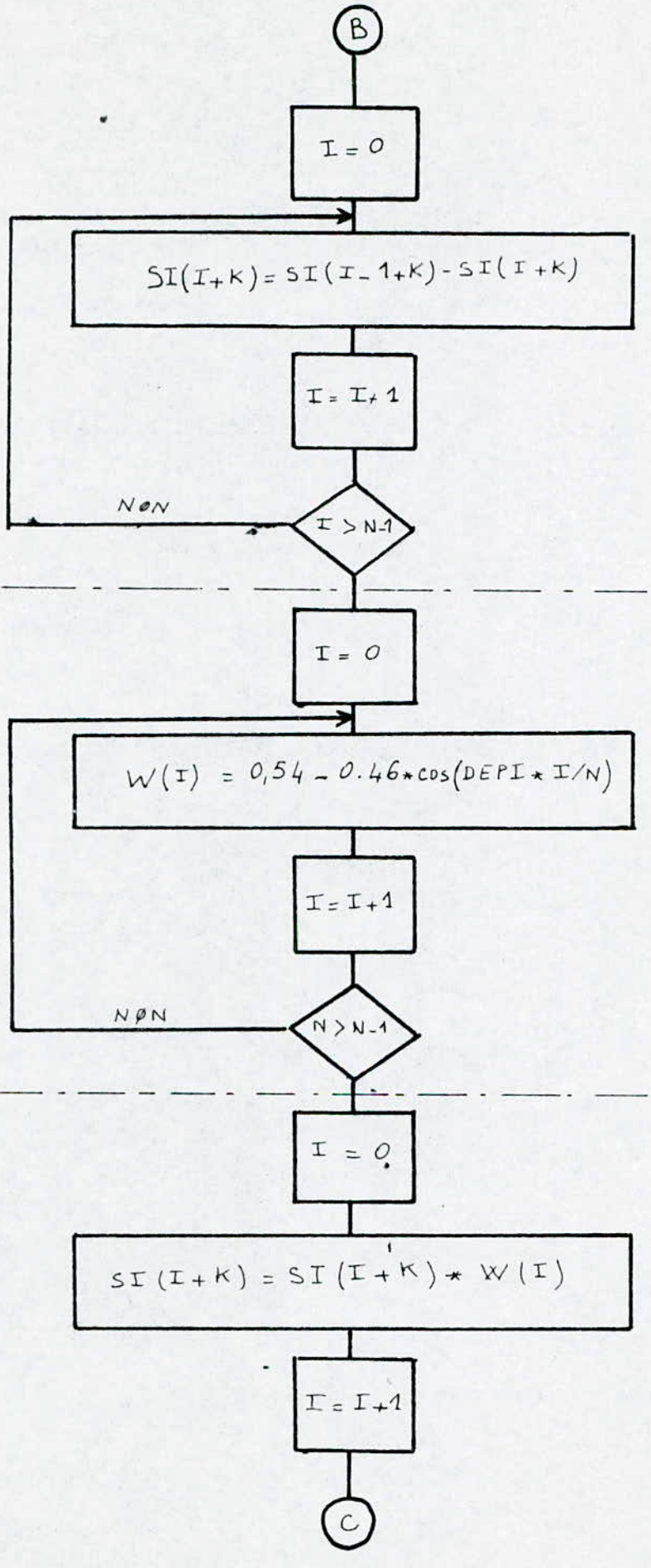
EXTRACTION DES PARAMETRES

PERTINANTS PAR LA PREDICTION

LINEAIRE



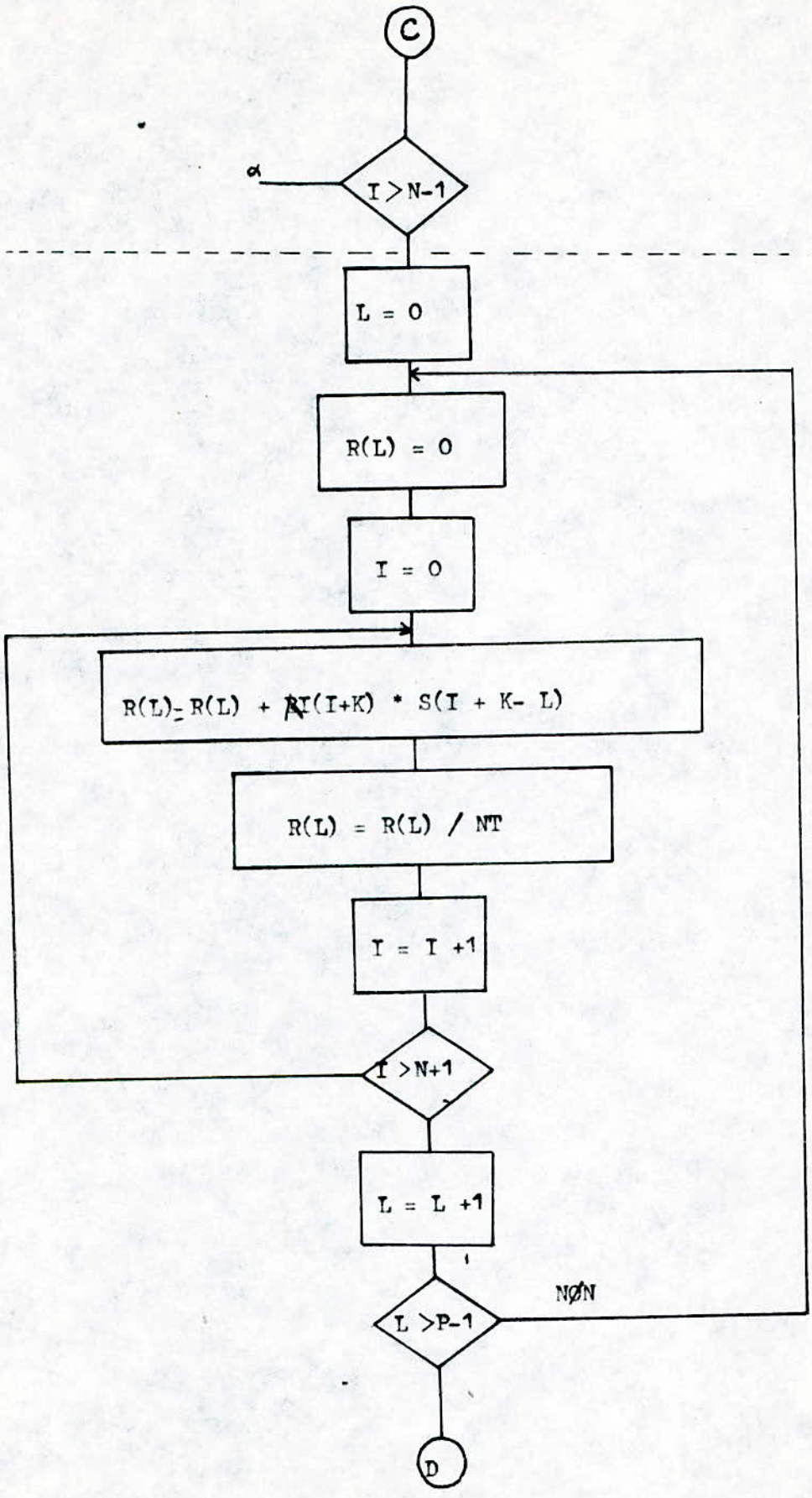

A
M = 1
I = 0
YI(I + K) = SI(I + K) + R(M)*SIN(DEPI*(I + K)*X(M)/NT)
I = I +1
N O N
I >N-1
I = 0
SI(I + K) = YI(I + K)
I = I+1
N O N
I >N-1
M = M+1
M >10
N O N
B



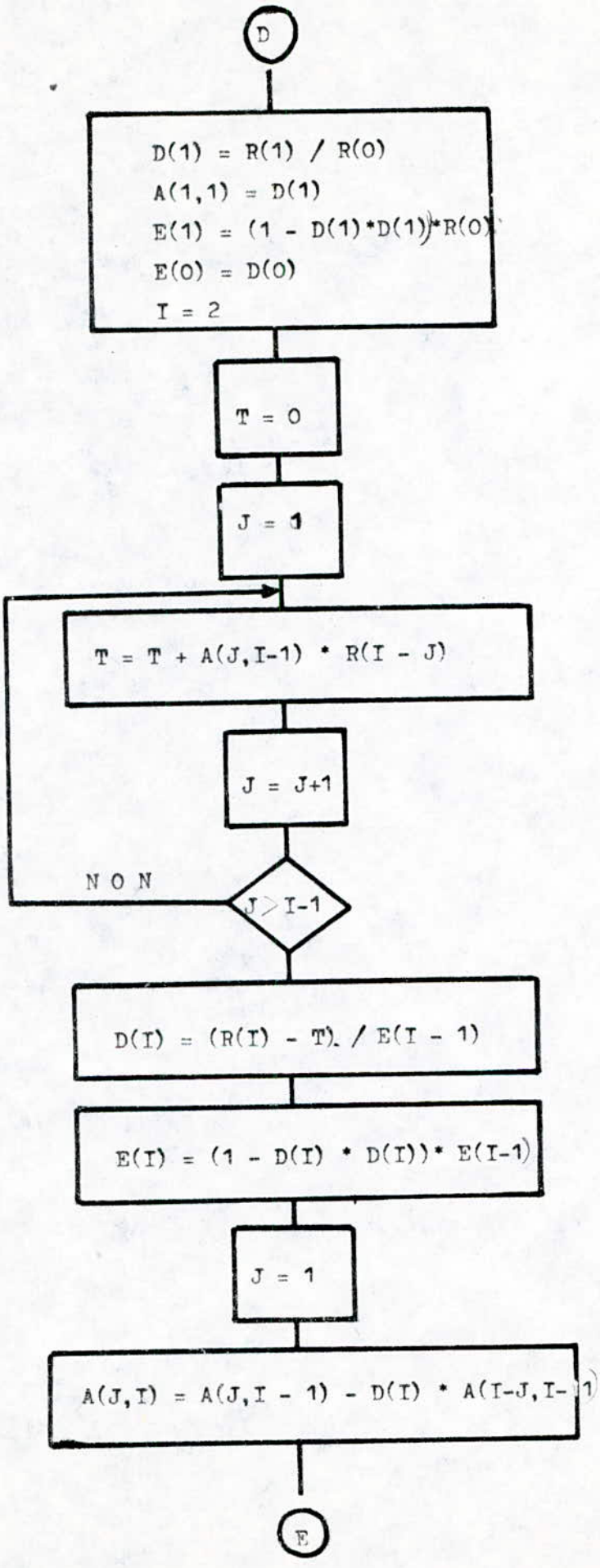
PREACCENTUATION

FENETRE DE HAMMING

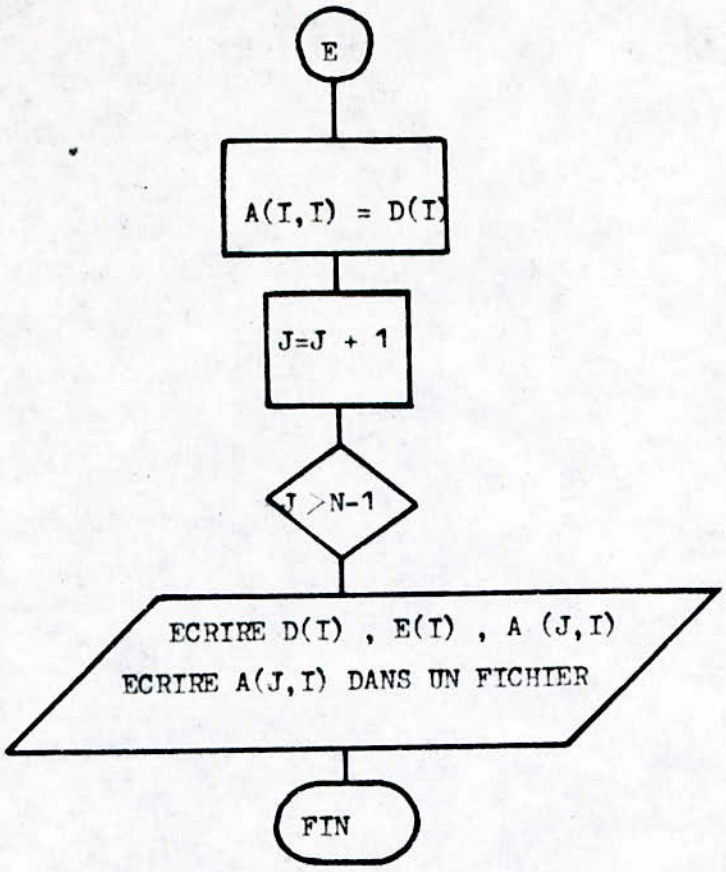
FENETRAGE



DETERMINATION DES COEFFICIETS D'AUTOCORRELATION.



METHODE DE - DURBIN-


E
A(I,I) = D(I)
J=J + 1
J >N-1
ECRIRE D(I) , E(I) , A (J,I)
ECRIRE A(J,I) DANS UN FICHIER
FIN

ORGANIGRAMME DE LA CEPSTRALE

Parametres utilisés dans les organigrammes de la cepstrale

FECH : frequence d'echantillonnage

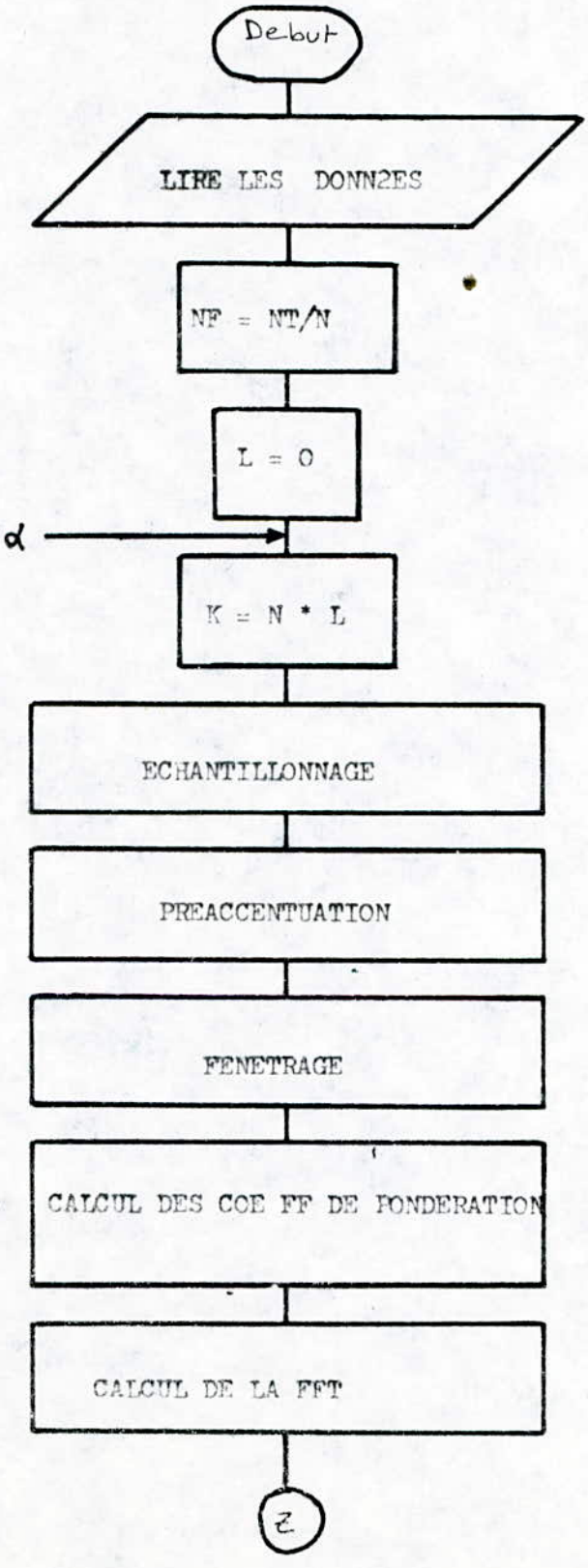
FCOUP : frequence de coupure

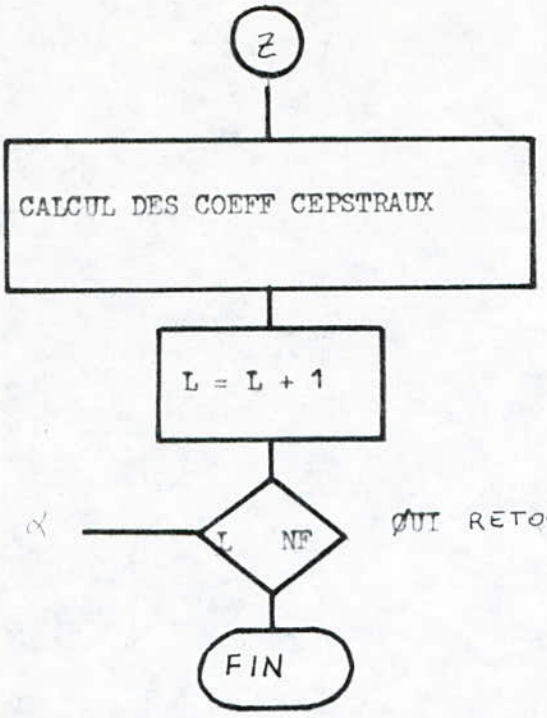
X(I) : coefficients de ponderation

CI(I) : coefficients cepstraux

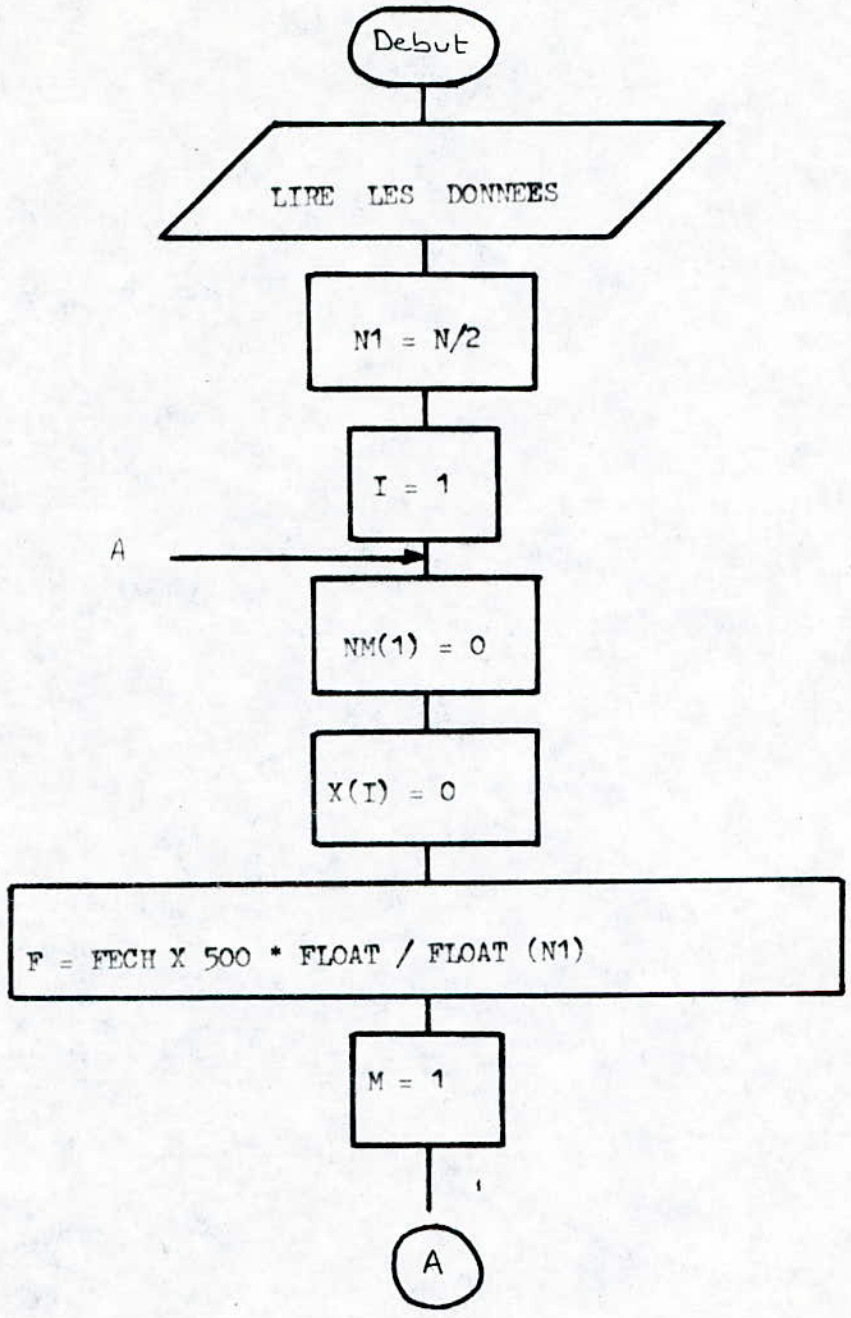
EXTRACTION DES PARAMETRES PERTINENTS

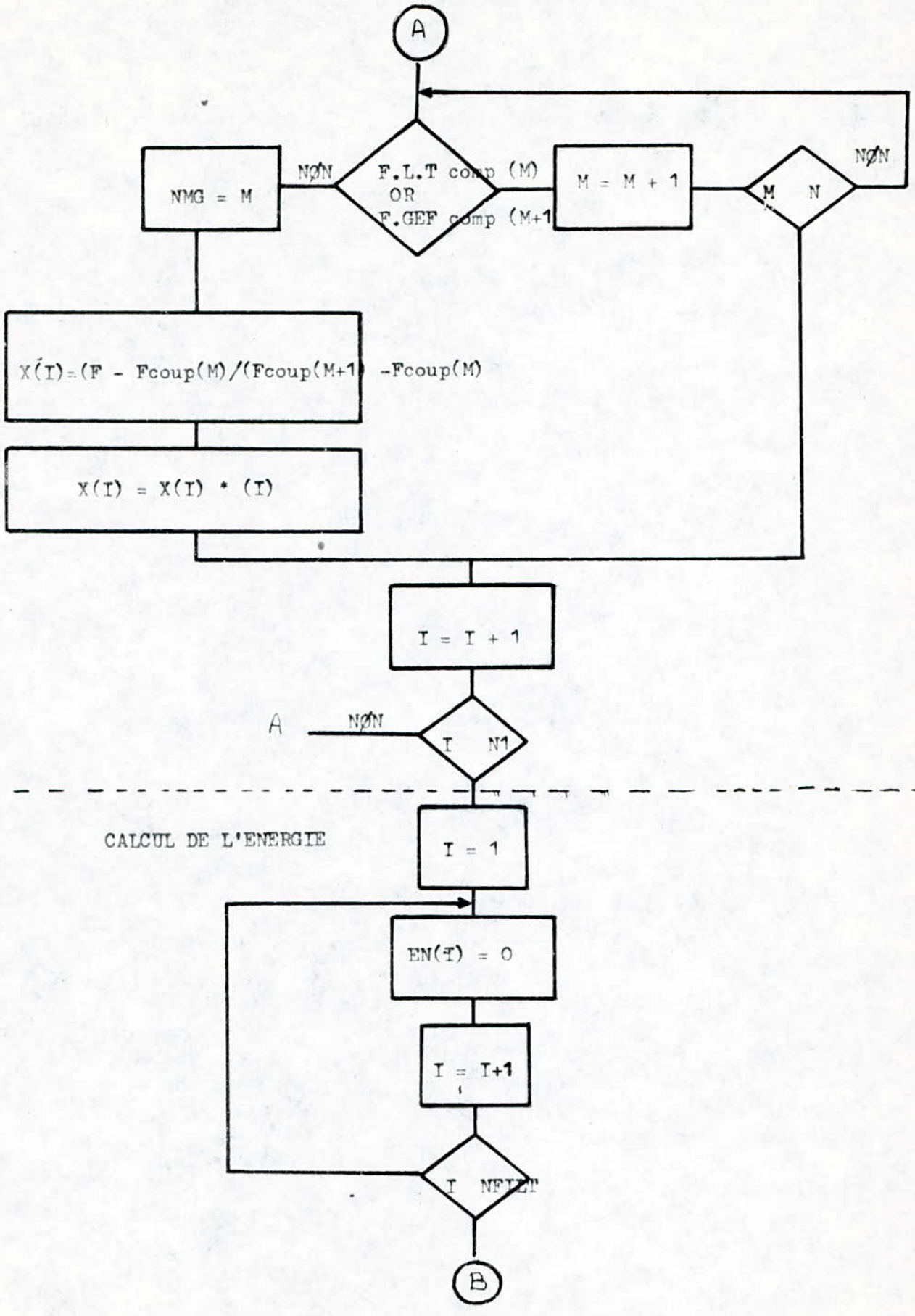
PAR L'ANALYSE CEPSTRALE

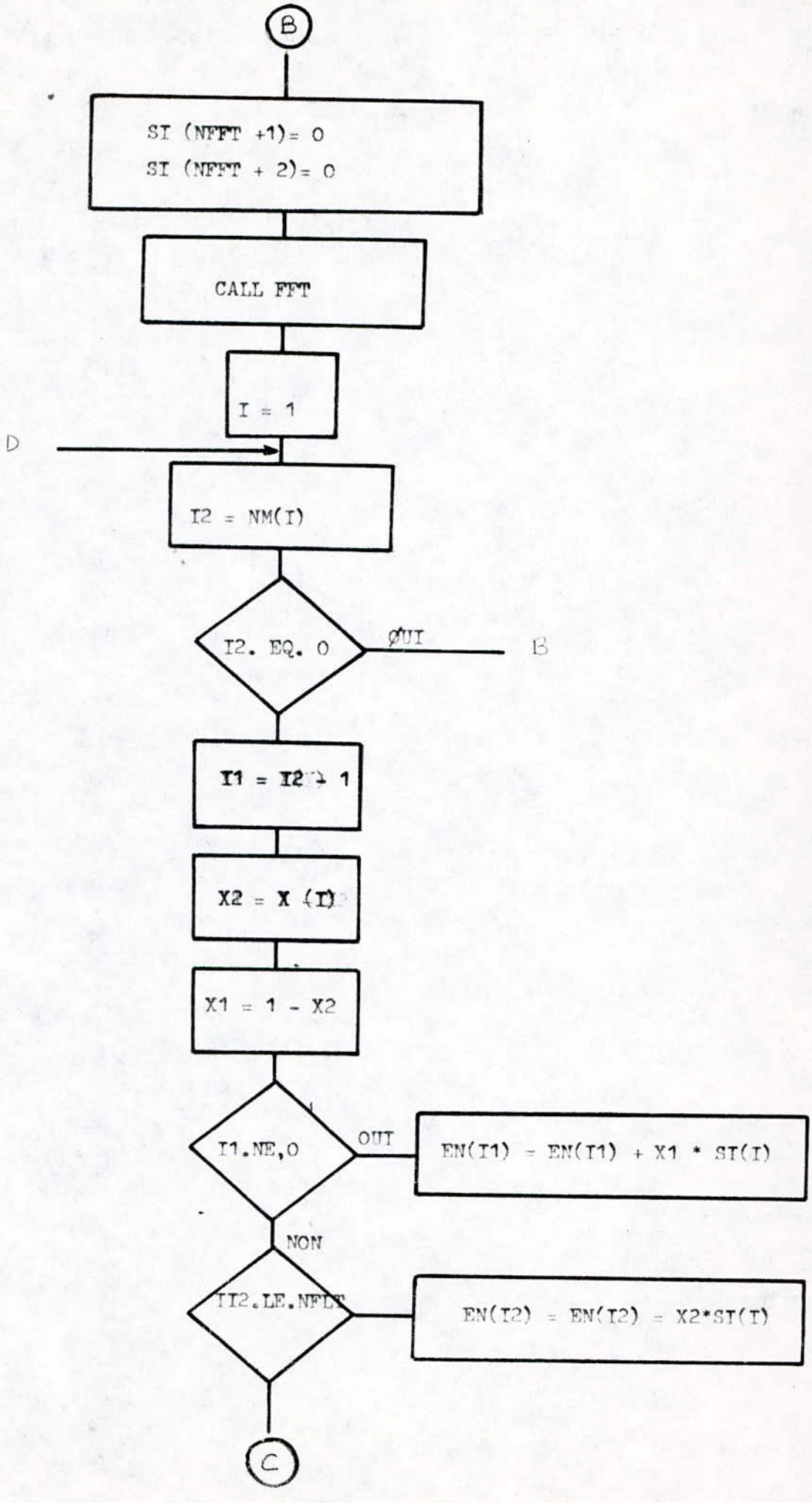


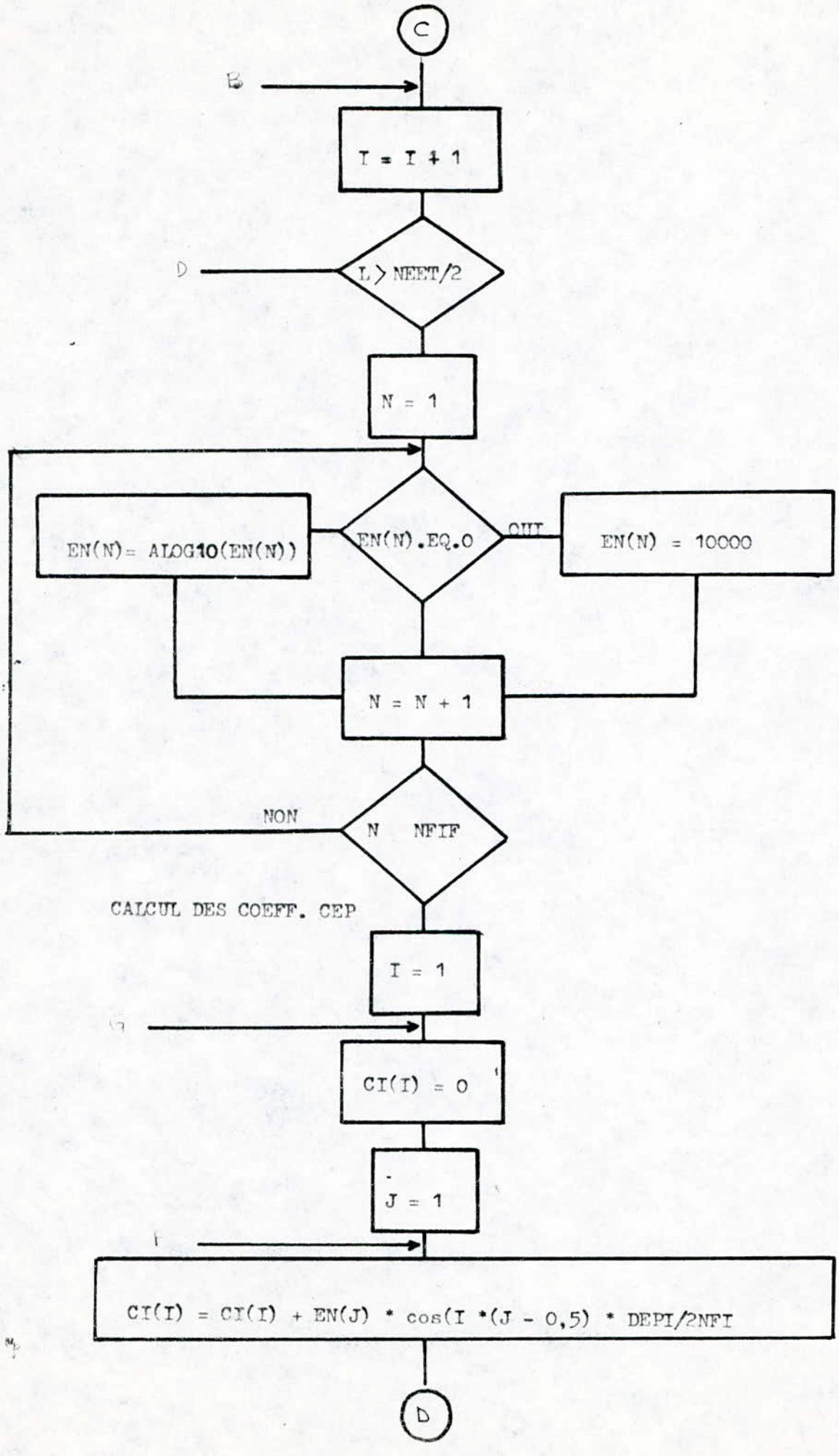

2
CALCUL DES COEFF CEPSTRAUX
L = L + 1
L NF
OUI RETOUR AU PROGRAMME PRINCIPAL.
FIN

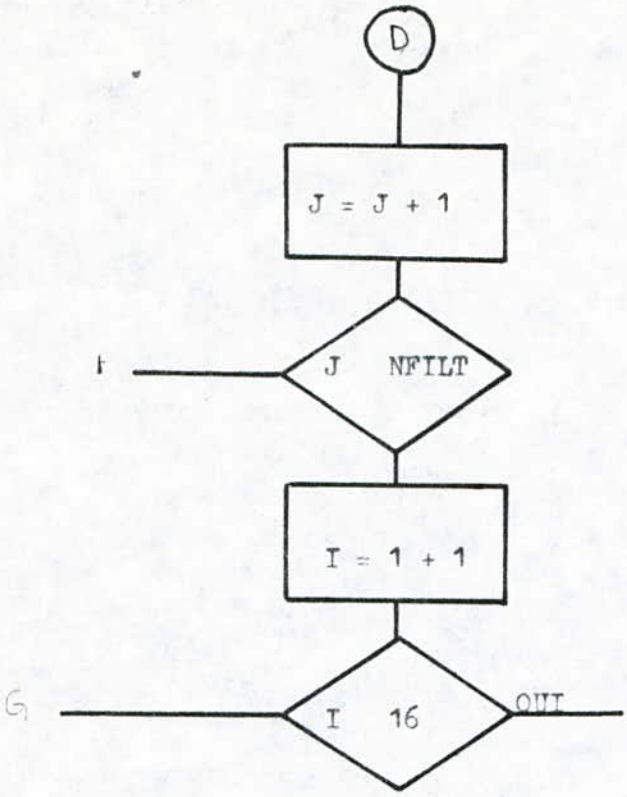
calcul des coefficients de ponderation




A
NMG = M
NØN
F.L.T comp (M)
OR
F.GEF comp (M+1
M = M + 1
M N
NØN
X(I)=(F - Fcoup(M)/(Fcoup(M+1) -Fcoup(M)
X(I) = X(I) * (I)
I = I + 1
A
NØN
I N1
CALCUL DE L'ENERGIE
I = 1
EN(I) = 0
I = I+1
I NFIET
B


B
SI (NFFT +1)= 0
SI (NFFT + 2)= 0
CALL FFT
I = 1
D
I2 = NM(I)
I2. EQ. 0
OUI
B
I1 = I2 + 1
X2 = X (I)
X1 = 1 - X2
I1.NE,0
OUI
EN(I1) = EN(I1) + X1 * SI(I)
NON
I2.LE.NFFT
EN(I2) = EN(I2) = X2*SI(I)
C


C
B
I = I + 1
D
L > NEET/2
N = 1
EN(N) = ALOG10(EN(N))
EN(N).EQ.0
OUI
EN(N) = 10000
N = N + 1
NON
N NFIF
CALCUL DES COEFF. CEP
I = 1
CI(I) = 0
J = 1
CI(I) = CI(I) + EN(J) * cos(I *(J - 0,5) * DEPI/2NFI
D


D
J = J + 1
J NFILT
I = 1 + 1
I 16
OUI
Retour au programme principale.

ORGANIGRAMME
DE LA
RECONNAISSANCE DE LA PAROLE

Parametres utilisés dans les organigrammes de la reconnaissance

NF : nombre de trames pour le mot test

nF : nombre de trames pour le mot reference

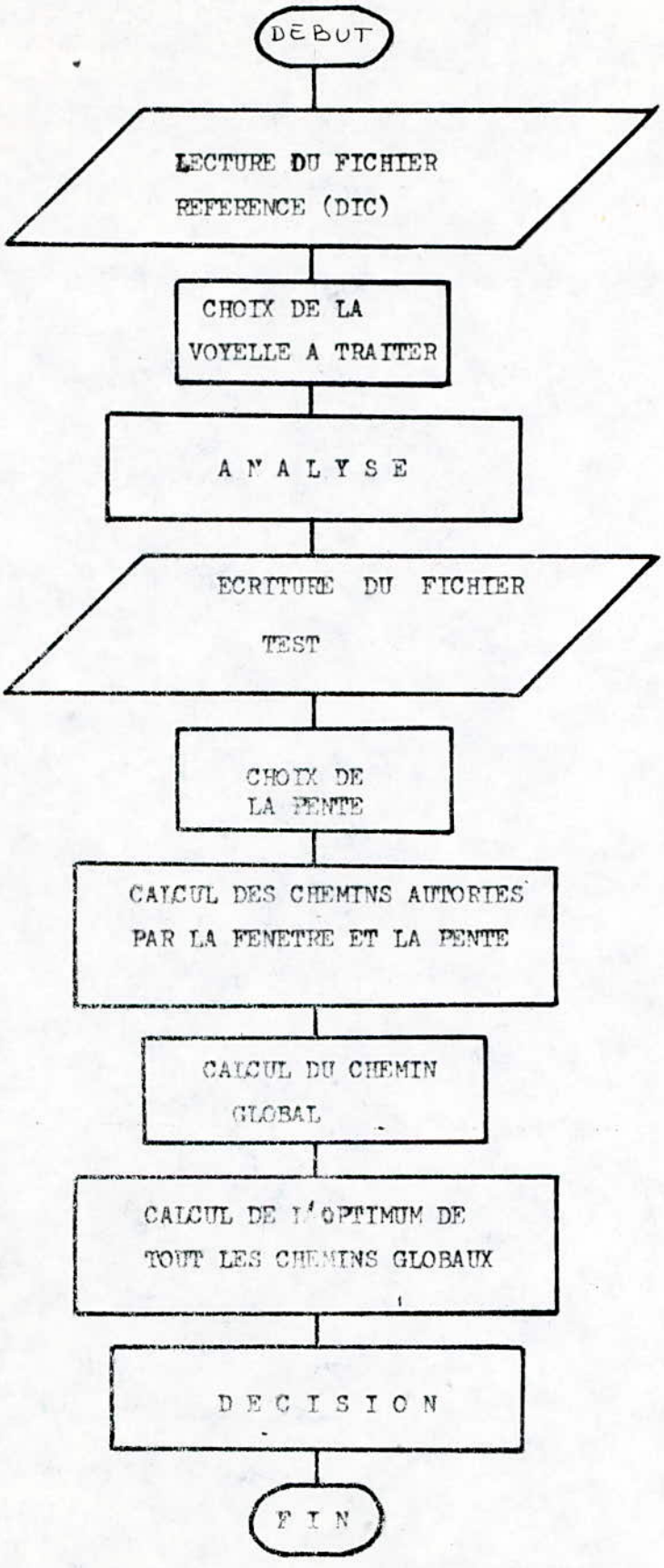
DI : Distance entre coefficients de deux trames l'une du test l'autre du dictionnaire.

G(N,n) : Distance entre trames

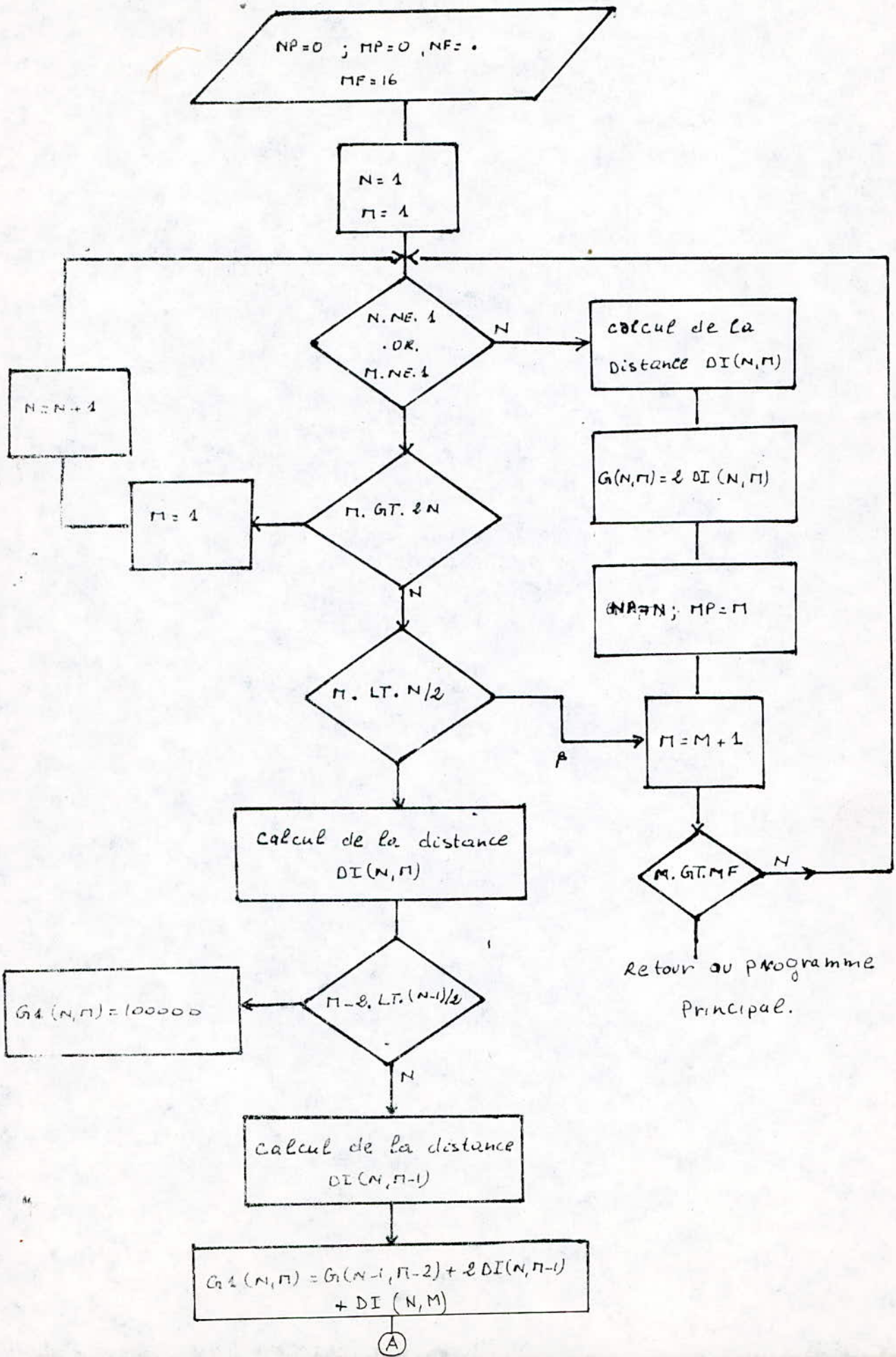
CoGL(N,n) : chemin global pour une voyelle traitée

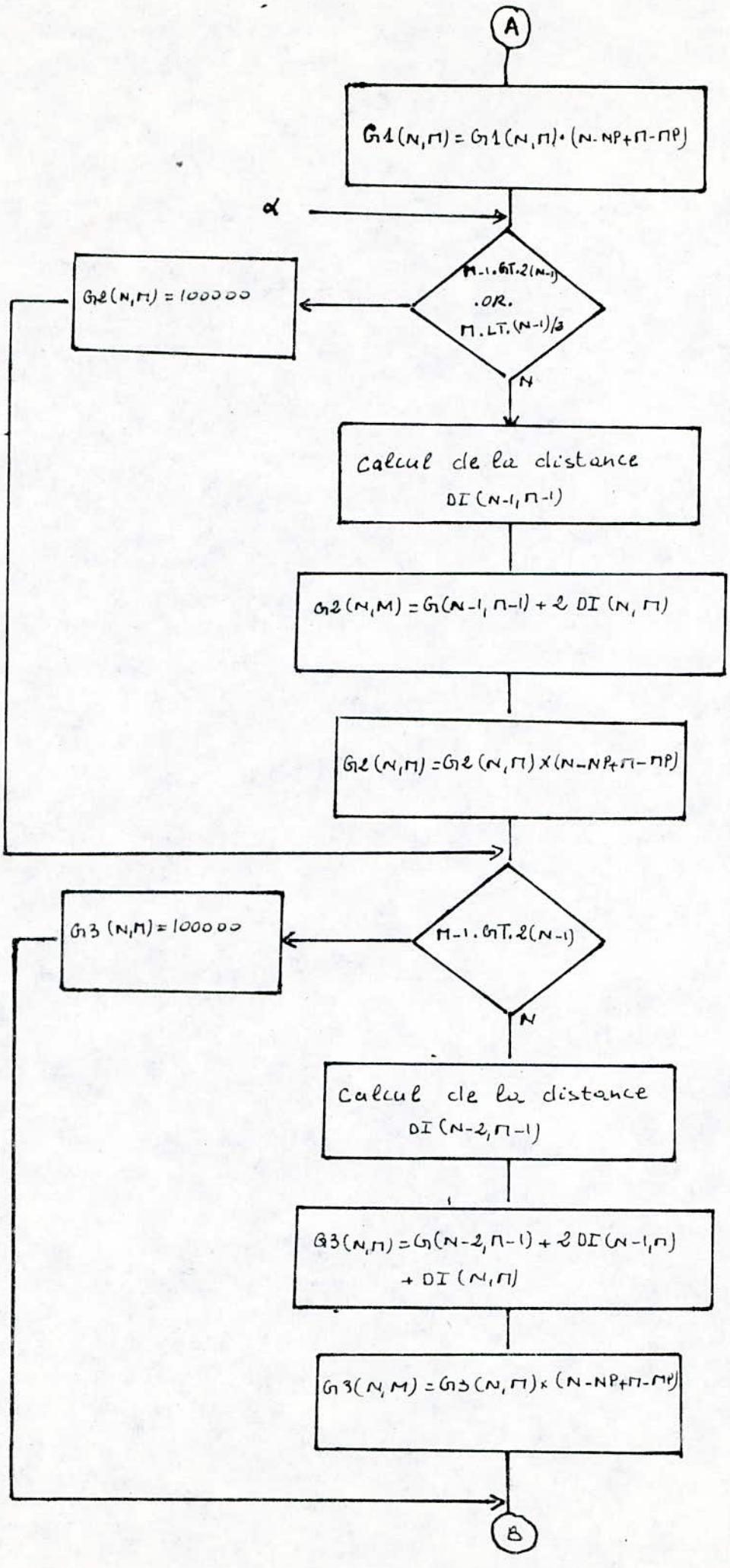
CMIN : le minimum des chemins globaux

CoG : le chemin de deformation en un point considéré

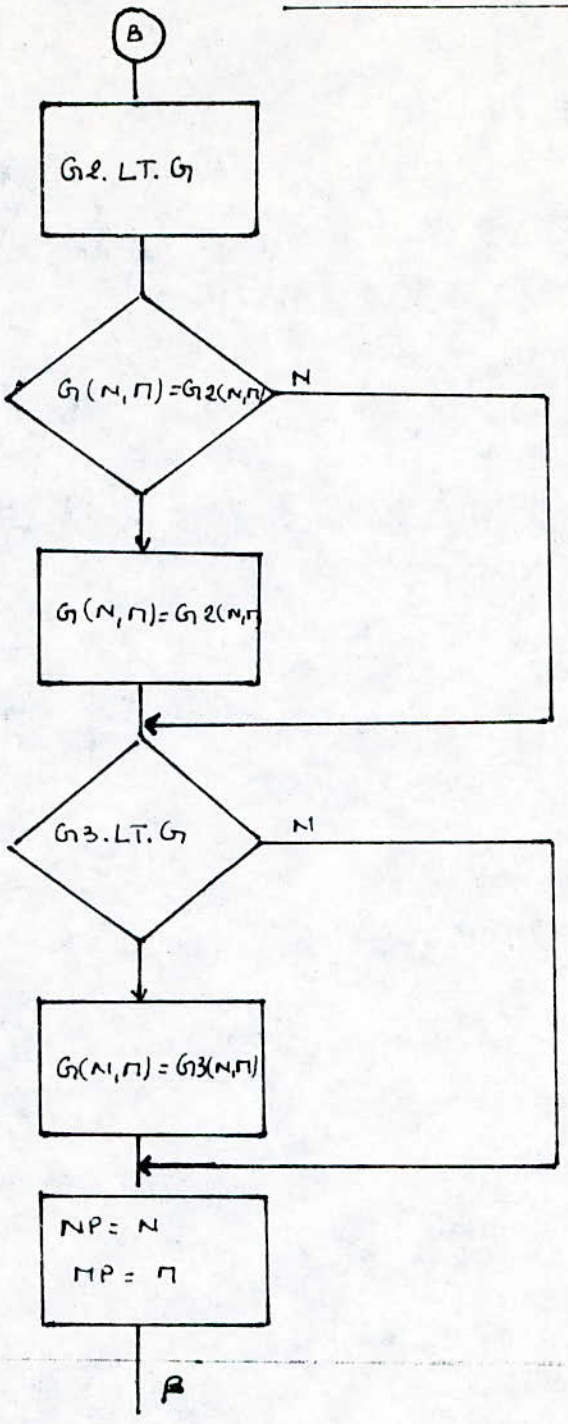

DEBUT
LECTURE DU FICHIER REFERENCE (DIC)
CHOIX DE LA VOYELLE A TRAITER
A N A L Y S E
ECRITURE DU FICHIER TEST
CHOIX DE LA PENTE
CALCUL DES CHEMINS AUTORIES PAR LA FENETRE ET LA PENTE
CALCUL DU CHEMIN GLOBAL
CALCUL DE L'OPTIMUM DE TOUT LES CHEMINS GLOBAUX
D E C I S I O N
F I N

CALCUL DES CHEMINS AUTORISES PAR LA FENETRE ET LA PENTE.

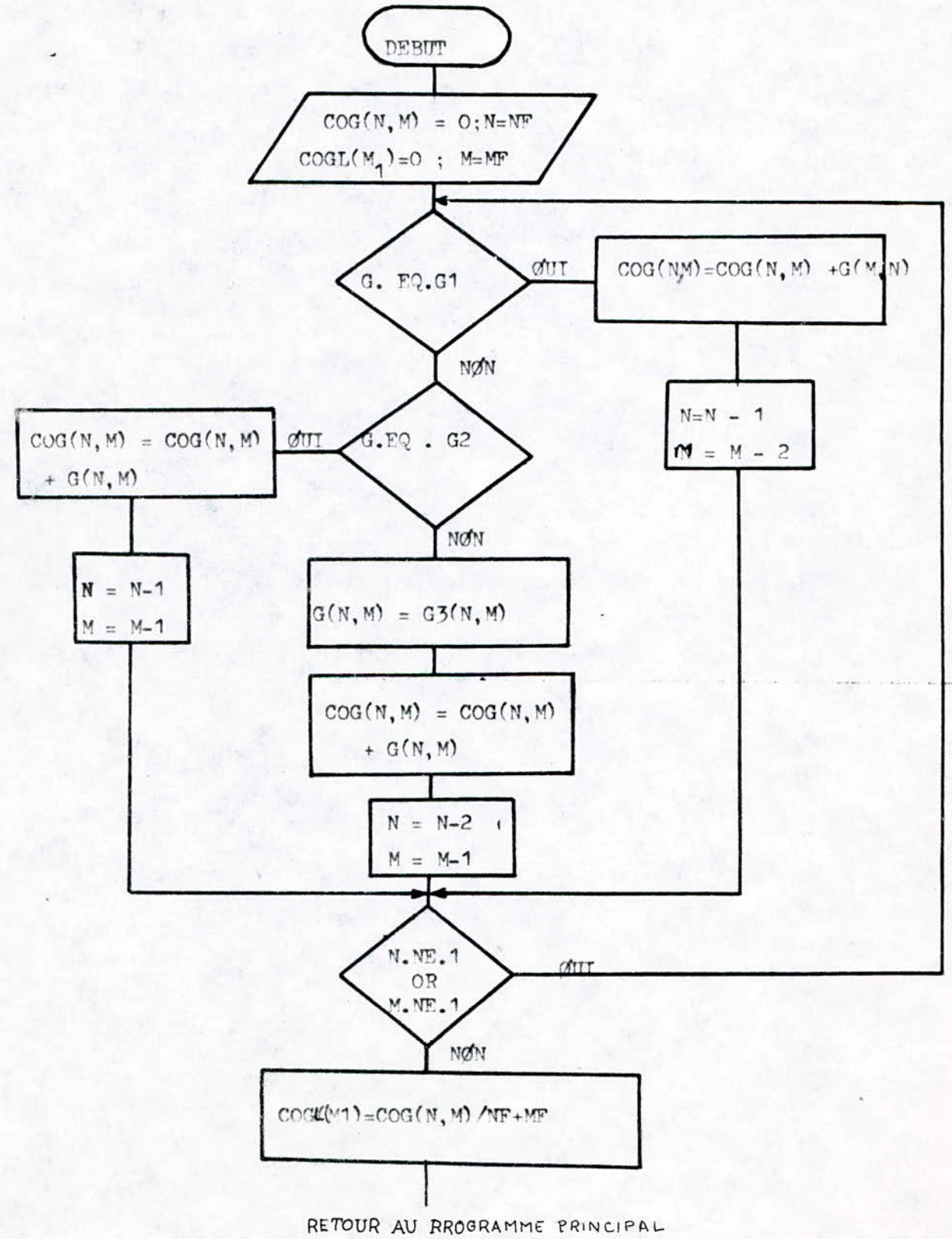



A
G1(N,M) = G1(N,M) * (N-NP+M-MP)
α
M-1.GT.2(N-1)
.OR.
M.LT.(N-1)/2
G2(N,M) = 100000
N
Calcul de la distance
DI(N-1,M-1)
G2(N,M) = G1(N-1,M-1) + 2 DI(N,M)
G2(N,M) = G2(N,M) X (N-NP+M-MP)
M-1.GT.2(N-1)
G3(N,M) = 100000
N
Calcul de la distance
DI(N-2,M-1)
G3(N,M) = G1(N-2,M-1) + 2 DI(N-1,M)
+ DI(N,M)
G3(N,M) = G3(N,M) x (N-NP+M-MP)
B

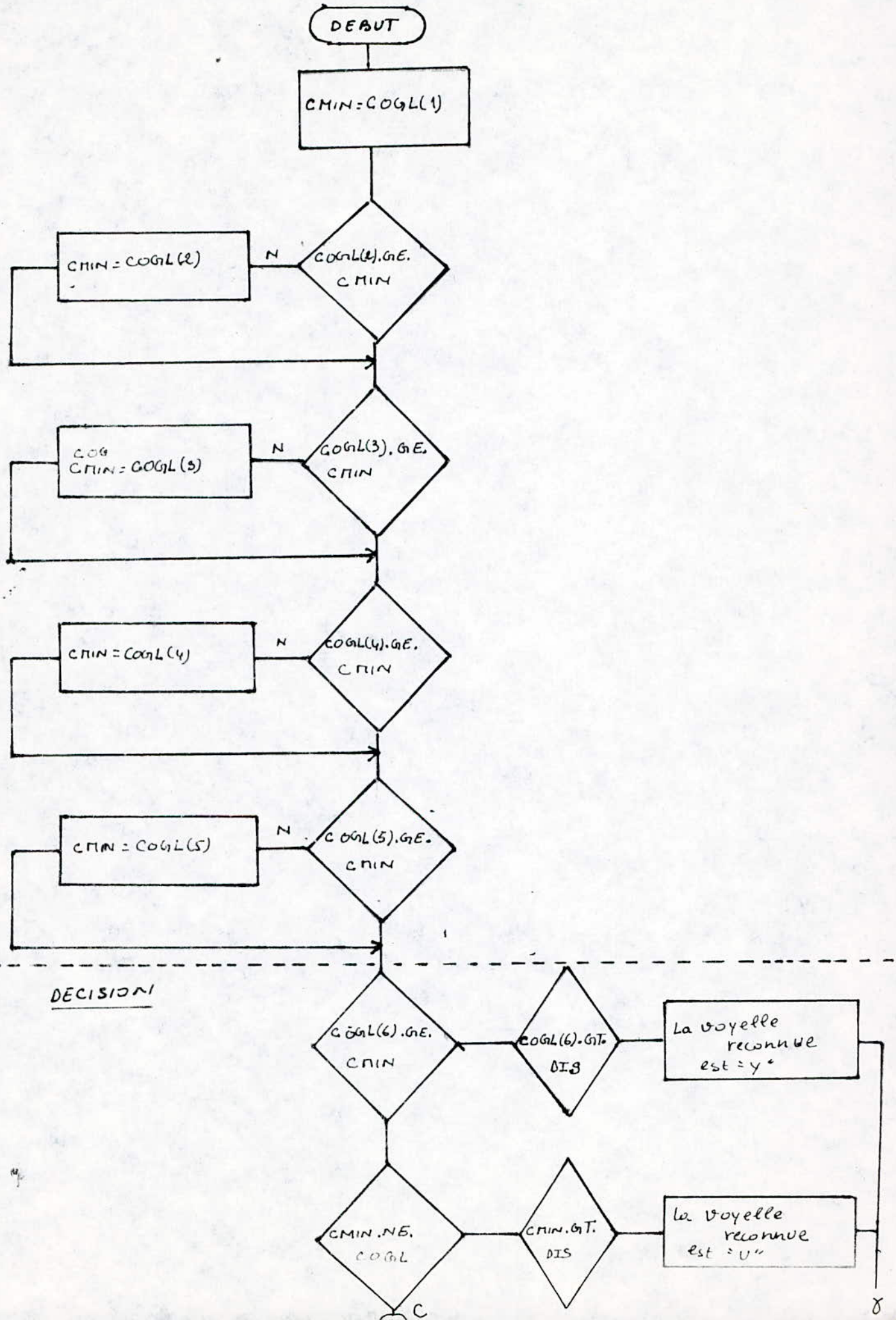
CALCUL DU CHEMIN OPTIMAL

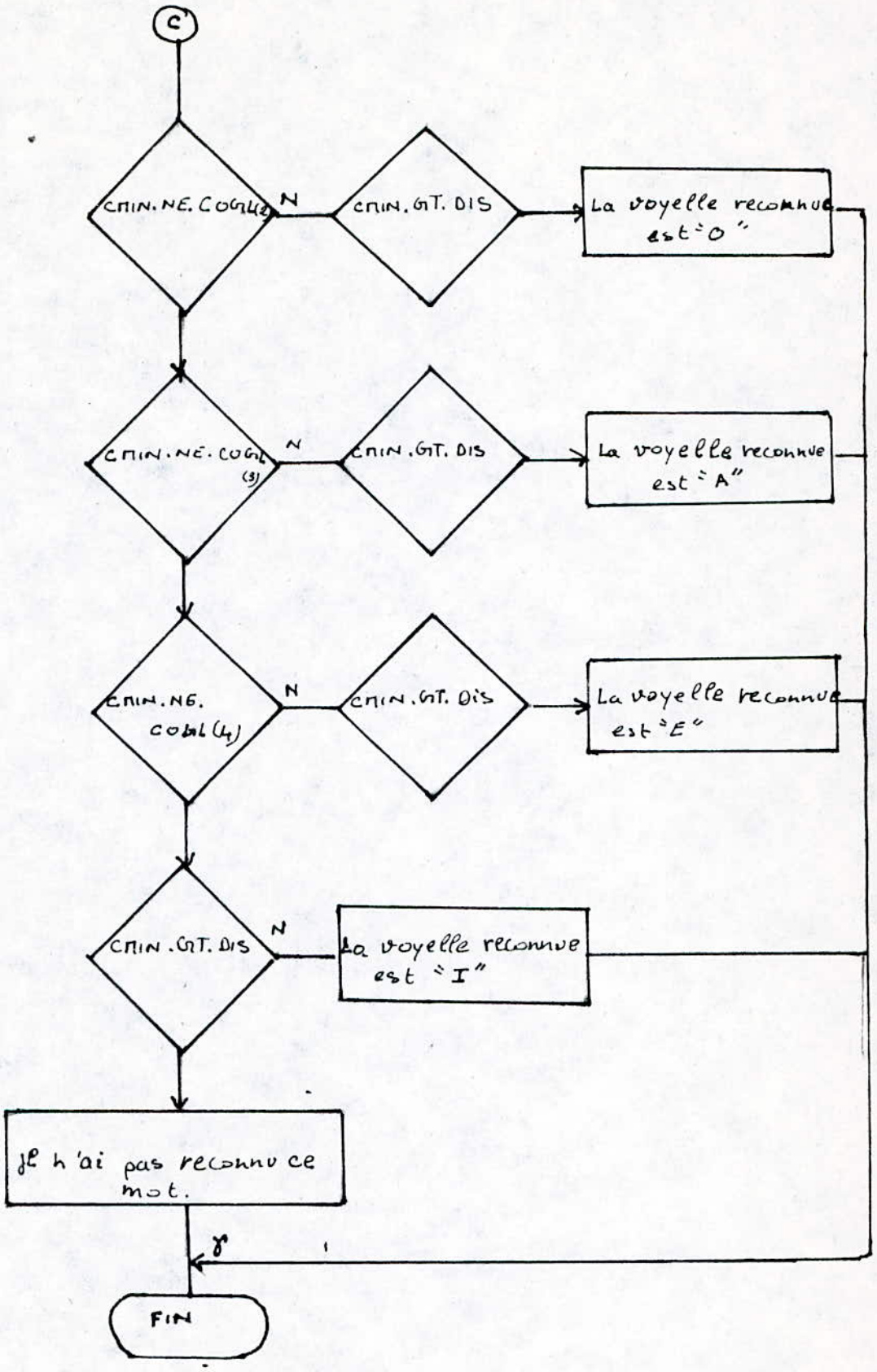


CALCUL DU CHEMIN GLOBAL OPTIMAL



CALCUL DE L'OPTIMUM DE TOUT LES CHEMINS GLOBAUX.




C'
CMIN.NE.COGL(2)
N
CMIN.GT.DIS
La voyelle reconnue est "O"
CMIN.NE.COGL(3)
N
CMIN.GT.DIS
La voyelle reconnue est "A"
CMIN.NE. COGL(4)
N
CMIN.GT.DIS
La voyelle reconnue est "E"
CMIN.GT.DIS
N
La voyelle reconnue est "I"
Je n'ai pas reconnu ce mot.
FIN

CONCLUSION GENERALE

Cette étude nous a permis de travailler un sujet d'actualité et d'exploiter les connaissances acquises pendant notre formation.

La mise au point d'algorithme de reconnaissance necessite une bonne connaissance en traitement de signal et en phonetique.

Un taux de reconnaissance satisfaisant passe par la maîtrise d'un certain nombre de paramètres (nombre de coefficients d'analyse, seuil, contraintes), pour la DTW et le test de plusieurs algorithmes (techniques d'analyse ...) et de DTW) avant d'opter pour le meilleur.

L'aspect temps réel repose sur un choix adéquat du microprocesseur (TMS 320).

Nous pouvons dire que notre objectif est atteint puisque nous sommes parvenus à mettre au point un logiciel de reconnaissance de la parole qui pourrait être transferer sur une carte autonome à base de microprocesseur.

Mais ceci ne doit pas masquer l'ampleur des problèmes rencontrés particulièrement la non disponibilité d'une acquisition de parole. En effet, pour tester nos programmes nous avons eu recours à une simulation du signal vocal, qui ne nous permet pas d'arriver à des résultats necessairement valables pour la parole réelle.

Recommendation.

Notre travail a été réalisé en langage évolué "FORTRAN 77" sur le mini-vax, cependant, il peut être concrétisé en implantant les logiciels mis au point à un micro - processeur, or il s'avère necessaire que ces derniers soient convertis en assembleur.

Le micro - processeur le plus performant, répondant aux conditions exigés par la reconnaissance de la parole est le"TMS 320 M 10" dont les caractéristiques sont les suivantes :

- Aspect temps réel
- Taux de reconnaissance raisonnable($\simeq$ 95%)
- Un jeu d'instruction simple
- Un multiplexeur cablé
- Possibilité d'une extension de mémoire (aussi bien RAM que ROM).

ANNEXE

Transformée de Fourier

La transformée de Fourrier d'un signal x(K) est définie par:

$$X(f) = \sum_{k=-\infty}^{+\infty} x(k)\, e^{-j2\pi fk}$$

La fonction X(f) est périodique de période 1 et est généralement fonction complexe de la variable f (f une variable discrète).

La transformée de Fournier inverse est donnée par

$$x(k) = {}^{1}/_{N} \sum_{n=-N/2}^{N/2-1} X(n) \exp(+j\, 2\pi\, nk/N)$$

Pour faciliter la notation on pose:

$$W_N = \exp(j\, 2\pi /N)$$

Ainsi, le signal complexe : $\exp(j\, 2\pi \cdot n \frac{k}{N})$ est dénoté par:

$$W_N^{nk} = \exp(j\, 2\pi\, nk/N)$$

Propriétés de la transformée de Fournier: séparabilité

$$- W_N^{K+1} = W_N^K \cdot W_N^1$$

Périodicité:

$$- W_N^{k + lN} = W_N^{K \bmod N}$$

Cas particulier:

$$- W_N^{ln} = 1$$

$$- W_N^{N/2} = -1$$

$$- W_N^{K+N/2} = W_N^K$$

$$- W_N = W_{N/2}$$

La FFT est un moyen très rapide permettant le passage du domaine temporel au domaine fréquentiel et vise versa.

Ce fut cooley le premier ayant conçu un algorithme ayant pour objetif la rapidité d'exécution entre autres Sande améliora cet algorithme par quelque changement et récement Radix propose un deuxième changement.

Fenêtrage

Pour étudier le signal temporel sur une durée limitée, il faut le multiplier par une fenêtre temporelle. Cette limitation en temps implique une distorsion au niveau spectral : apparition des lobes secondaires et élargissement du lobe principal.

Il existe plusieurs types de fenêtres:

Hamming, Hamming et rectangulaire (voir J.Max)

Le choix d'une fenêtre plutot qu'une autre est compromis entre l'élargissement du lobe principal et l'énergie des lobes secondaires.

BIBLIOGRAPHIE

BIBLOGRAPHIE

Livres:

1. Application of digital signal processing
   Alan V. Oppenheim, editor - 1978.

2. Cours de phonétique acoustique
   E. EMERIT - 1977.

3. Digital signal processing
   Alan V. Oppenheim/Ronald W. SCHAFER - 1975.

4. Digital Processing of spech signals
   L.R. Rabinier/R.W. Schafer - 1978.

5. La programmation dynamique.
   A. CHEVALIER Dunod - 1977.

6. La programmation dynamique et ses applications
   R.E. BELLMAN S S.E DREYFUS - 1965.

7. Les méthodes rapides de transformation du signal:
   Fornier Wlsb, Hadomand, Haar.
   J. LIFERMANN MASSON - 1980.

8. Méthodes et techniques de traitement du signal
   et application aux mesures physiques.
   J.MAX, MASSON Tome I - 1981.

9. Traitement numérique des signaux
   M. KUNT Dunod - 1981.

Revues

10. Applications industrielles de la reconnaissance vocale.
Electronique automatique et informatique industrielle,
Revue bimensuel - 1 er Décembre 1980.

11. Dynamic programming algorithm optimization for spoben word recognitroh.
by Hiroabi SAKOE and Seibi CHIBA.
IEEE transaction on acoustics, speed, and signal processing
Vol ASSP - 26.N°1, FEBRUARY - 1978.

12. Minimum prediction residuol principle applied to speech recognition.
by Fumitada ITAKURA.
IEEE transaction on acoustic, speech, and signal processing
FEBRUARY - 1976.

13. Principes et techniques de la reconnaissance de parole.
J.P HATON
Journée d'étude "interaction homme machine et IA"
CRIN/INRIA, NANCY.
Toulouse, octobre - 1986.

14. Recueil de publication et communication en analyse perception, synthèse et reconnaissance de parole
Edition de décembre 1983.
CNET LANNION.

15. SERAPHINE : Reconnaissance de parole continue par méthodes globales.
Note technique NT/LAA/TSS/79.
CNET (LANNION)
AOUT 1981.

16. Signal processing with the TMS 320 family.
Lee V. Kaplan.
Applications engineer - Texas instruments.
signal processing products and technology
February 24, - 1982.

Thèses:

17. Application de l'algorithme D.T.W à la reconnaissance de la parole
L. FERHAT. HAMIDA et B.FERGINI
Projet de fin d'étude
B.Z JUIN 1987.

18. Etude S realisation d'un synthétiseur
M. OTMINI et N. HASSAINE
Projet de fin d'étude
ENP - JANVIER - 1986

19. Reconnaissance automatique de la parole par la méthode globale. Application à des particularités linguistiques de l'arabe standars.
B. BOUSSEKSOU
Titre de magister université d'Alger - 1983.

20. Reconnaissance de la parole en mode multilocuteur par des méthodes globales (mots isolés)
A. MENACER.
Thèse Docteur - ingénieur, Université de Rene - 1986.

21. Simulation de l'analyse de la parole à l'aide du microprocesseur TMS 320 M10
D.E TALBI H.OUAHAB
Projet de fin d'étude
ENP - JUIN 1987.