COLLOQUE NATIONAL SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 26 au 30 AVRIL 1977

# RECONNAISSANCE AVEC APPRENTISSAGE DE SIGNAUX SEQUENTIELS DE PERIODE ALEATOIRE.

Jean-Claude LEVY

Bureau de Psychologie Appliquée de la Direction du Personnel Militaire de la Marine
3 Avenue Octave Gréard 75007 PARIS (France)


## RESUME

L'auteur décrit un système procédant à la reconnaissance de suites de symboles à partir de séquences de caractères.

Au caractère de rang n est associée une distribution de probabilités: $(q(n))$ sur l'alphabet des N symboles. Nous appelerons $(P(n))$ une distribution obtenue par composition des $\bar{q}(n)$ et des $\bar{p}(n,k)$ tels que:

$$\bar{\bar{q}}(n,k)=\bar{P}(n-k).\bar{\bar{M}}(k)$$

où $\bar{\bar{M}}(k)$ est une matrice de Markov de rang: N et où k est limité à une valeur maximale: K.

Dans le cas particulier: $K=N-1$ et moyennant une condition relative aux matrices de transition $\bar{\bar{M}}(k)$, on démontre qu'il est possible de définir $(N-1)!$ ensembles de chacun $(N-1)!$ mots de N-1 symboles structurellement stables.

De plus, il est possible de pondérer les $\bar{P}(n)$ par une fonction du temps et de composer $\bar{P}(n)$ et $\bar{P}(n+1)$ de manière à lever l'ambiguité pouvant résulter de l'oubli d'un caractère et à rattraper le décalage entre les caractères et les symboles aux quels ils doivent correspondre.

Enfin le système peut élaborer les matrices de transition en recevant des séquences de symboles à mettre en mémoire.

## SUMMARY

SELF LEARNING RECOGNITION OF SEQUENTIAL SIGNALS WITH RANDOM PERIOD

A system is described vhich recognises successions of symbols from sequences of characters.

To the nth character is associated a probability distribution $(q(n))$ in the set of N disjointed symbols. Let $(P(n))$ be a distribution obtained by the composition of $\bar{q}(n)$ and $\bar{p}(n,k)$ such as:

$$\bar{p}(n,k) = \bar{P}(n-k).\bar{\bar{M}}(k)$$

where $\bar{\bar{M}}(k)$ is a matrix of Markov and k limited to an upper value K.

In the particular case, when $K=N-1$, and according to a condition concerning the transition matrix: $\bar{\bar{M}}(k)$, we can demonstrate that it is possible to define $(N-1)!$ sets of each $(N-1)!$ sequences of N-1 symbols and that, with structural stability.

More over, we can allow to each $\bar{P}(n)$ a weighting function of time in such a way that we remove the ambiguity which could result from the omission of a character and cancel the shifting between characters and corresponding symbols.

Finaly, the system can create transition matrix by receiving sequences to be memorised.

# RECONNAISSANCE AVEC APPRENTISSAGE DE SIGNAUX SEQUENTIELS DE PERIODE ALEATOIRE.

Jean-Claude LEVY

## Introduction

Le système qui va être décrit ne se suffit pas à lui même, il présuppose l'existence d'un autre système, le quel peut tout aussi bien, suivant les cas, être un système de reconnaissance de caractères alphanumériques ou de phonèmes, ou encore de tout élément de signaux séquentiels d'origine naturelle.

Le système décrit a pour rôle d'effectuer la synthèse du signal, ceci par identification d'une suite de symboles, c'est à dire d'un mot.

Soit: N le cardinal de l'alphabet des symboles, soit: L le nombre de symboles constituant un mot. Il y a $N^L$ possibilités de former un mot, mais, sur ce nombre, seule une petite proportion pourra constituer un ensemble d'éléments discernables, une certaine redondance étant indispensable à la correction d'erreurs.

Le calcul ne sera fait complètement que dans le cas particulier où L=N. La capacité du système est alors: $(L-1)!$, ceci compte tenu du fait qu'un symbole devra être réservé pour reconnaitre le début d'un mot.

## 1°) Reconnaissance préliminaire.

Le système de reconnaissance individuelle de chaque caractère n'est pas sensé effectuer de prise de décision: il se contente d'associer à chaque caractère observé une distribution de probabilités: $(q_i)$ où $q_i$ est la probabilité pour que le caractère traité, $A_n$ corresponde au symbole $s_i$. Seul le cas particulier où:

$$\forall j \neq i \Rightarrow q_j=0 \Rightarrow q_i=1$$

correspond à une prise de décision.

La distribution $(q_i)$ sera représentée par un vecteur $\bar{q}$.

## 2°) Correspondance entre caractères et symboles.

Chaque caractère doit être comparé au symbole occupant la même place dans le mot, cette place sera définie par le rang: n qui doit être connu pour chaque caractère: celà peut se réaliser de trois façons, à savoir:

2,1: Fixation d'une origine et d'une échelle des temps, le caractère de rang n devant être observé à l'époque:

$$t = t_o + n\tau$$

où $\tau$ est la période de récurrence. C'est le décodage synchrone qui nécéssite l'observation rigoureuse de l'échelle des temps.

2,2: Comptage des caractères, ce qui correspond au "décodage arythmique", si un caractère est oublié, tous les suivants sont décalés.

2,3: Indication du rang: n dans le signal représentant le caractère, il s'agit des codes à rattrappage de phase peu aptes au décodage des signaux naturels.

2,4: le système que nous allons décrire tient à la fois du premier et du second procédé. Il s'applique beaucoup mieux aux problèmes de liaison homme-machine, l'homme étant capable d'émettre des signaux à une fréquence à peu près stable, mais l'oubli d'un caractère étant toujours possible. Ce point sera discuté plus loin après la description du système.

## 3°) Description du système. (Figure 1)

Pour chaque caractère de rang n, nous définirons une distribution de probabilités: $(P_i)$ et un vecteur $\bar{P}$. Cette distribution est élaborée à partir de K+1 distributions, soit:

- $\bar{q}_n$ définie au §1
- $\bar{p}(n,k) = \bar{P}(n-k).\bar{\bar{M}}(k)$ (3,1)

où $\bar{\bar{M}}(k)$ est une matrice de MARKOV de rang N. C'est à dire que chaque $p_i(n,k)$ établit une probabilité à priori de reconnaitre le symbole $s_i$ au rang n si le symbole $s_j$ a été reconnu au rang: (n-k): Il s'agit en quelque sorte de la formalisation d'une association d'idées entre $s_j$ et $s_i$.

K est fini et définit l'ordre du système.

La probabilité de trouver K+1 fois le même symbole $s_i$ est donnée par l'équation:

$$P_i^\circ(n) = q_i(n).\prod_{k=1}^{k=K} p_i(n,k) \qquad (3,2)$$

La probabilité conditionnelle de trouver K+1 fois le symbole $s_i$ sachant qu'on doit avoir K+1 fois le même symbole est:

$$P_i(n) = P_i^\circ(n) / \sum_{j=1}^{j=N} P_j^\circ(n) \qquad (3,3)$$

Nous éliminons à priori toute solution correspondant à un produit croisé tel que $p_i p_j$

$P_i(n)$ est donc bien la probabilité cherchée.

## 4°) Condition de discernabilité.

Nous ferons sur la matrice $\bar{\bar{M}}(k)$ l'hypothèse suivante: il y a au moins un coefficient $m_{ij}$ nul pour chaque ligne et pour chaque colonne.

Pour la clarté du langage, nous dirons: "strictement nul" les autres coefficients pouvant être simplement négligeables.

Celà définit une matrice $\bar{\bar{C}}(k)$ telle que:

$$m_{ij} = 0 \Rightarrow C_{ij} = 1 \qquad (4,1)$$
$$m_{ij} > 0 \Rightarrow C_{ij} = 0$$

La matrice $\bar{\bar{C}}(k)$ définit une bijection sur lui même de l'alphabet des symboles.

Si $\bar{\bar{C}}(1)$ est définie, les matrices suivantes sont liées par la condition:

$$\bar{\bar{C}}(k) = \left[\bar{\bar{C}}(1)\right]^k \qquad (4,2)$$

Celà définit des couples de digrammes interdits $s_i s_j$ tel que chaque indice figure une fois en première position et une fois en seconde position. Cette règle s'applique de façon récursive pour des couples de symboles séparés par 1,2,...K places. Il y a N! façons de choisir un ensemble de digrammes interdits.

5°) Théorème de discernabilité.

Nous nous limiterons au cas particulier: $K = N-1$, ce qui définit N-1 transitions entre les N symboles d'un mot. Le théorème s'énonce ainsi:

Si une séquence de N-1 caractères est reconnue sans équivoque et satisfait à la condition des digrammes interdits, l'évolution du processus est déterminée sans ambiguité.

Nous partirons d'une permutation particulière des symboles $s_i$ supposée prise à l'instant $n_o$, en appliquant la bijection définie par la matrice $\bar{\bar{C}}(k)$, la transformée de $s_i$ est $s_j$ que nous appellerons $S_i(n_o+k)$, nous appellerons de même:

$$R_i(n_o+k) = P_j(n_o+k)$$
$$r_i(n_o,k) = p_j(n_o,k) \qquad (5,1)$$

Le but de cette transformation est simplement de représenter d'une façon claire la suite des transformées de l'indice initial: i et ceci en conservant cet indice le long de la chaine.

Considérons maintenant une séquence de N symboles $S_i$ telle que chaque valeur de i et de n soit prise une et une seule fois. Il y a N! façons de choisir cette séquence.

D'après l'hypothèse:

$$R_i(n) = 1 \quad , \quad j \neq i \Rightarrow R_j(n) = 0 \quad (5,2)$$

$R_i(n)$ est la seule composante non nulle du vecteur $\bar{R}(n)$, donc:

$$p_j(n+1,k) = r_i(n+1,k) = 0 \quad \text{ceci } \forall k$$

donc :

$$k < K \Rightarrow R_i(n_o+k) = 0$$

comme ayant au moins un facteur nul.

Le symbole $S_i$ ne pourra pas se retrouver avant le rang: n+N. La figure 2 montre que dans chaque chaine un symbole et un seul peut se trouver à chaque place. Le symbole $S_i$ se retrouve donc sans ambiguité au rang n+N, mais il correspond au même symbole $s_i$, il faudra refaire la transformation inverse et retrouver j en effectuant sur les indices la bijection définie par l'équation:

$$\bar{P}(n) = \bar{R}(n).\left[\bar{\bar{C}}(k)\right]^{-1}$$

où $k = n-n_o$

et les vecteurs $\bar{R}$ et $\bar{P}$ n'ayant qu'une composante différente de zéro.

Remarque:

Le processus se déroulant sans arret, il faudra réserver un symbole pour marquer le début et la fin d'un mot, un blanc par exemple, Il lui correspondra un coefficient $m_{ii}$ nul sur la première diagonale, il restera donc (N-1)! façons de choisir (N-1)! mots discernables entre eux.

6°) Stabilité structurelle.

Supposons l'équilibre structurel atteint, le système donnera des séquences de N symboles définis sans ambiguité.

Introduisons une ambiguité dans la reconnaissance des caractères:

$$\varepsilon \ll 1 \;, q_i = 1-\varepsilon \;, \quad j \neq i, q_j = \varepsilon$$

L'équation 3,2 montre que le coefficient de $q_j(n)$ dans le calcul de $P_j(n)$ est nul: le système n'est pas seulement stable, mais: bloqué, les signaux venus de l'extérieur n'ont plus aucune influence sur son évolution.

Supposons maintenant que le régime stable ne soit pas encore atteint, et que le coefficient $\alpha_{jn}$ de $q_j$ dans le calcul de $P_j$ ne soit pas tout à fait nul, nous allons voir qu'il est au plus égal à l'unité et généralement très inférieur.

L'équation 3,2 montre que $P^o_j(n)$ est une fonction croissante de $q_j(n)$, l'équation 3,3 montre que $P_j(n)$ est une fonction croissante de $P^o_j(n)$, en effet, $P^o_j(n)$ est en facteur au numérateur et ne figure au dénominateur que comme un terme d'une somme dont tous les termes sont positifs: $P_j(n)$ est donc une fonction croissante de $q_j(n)$.

Supposons maintenant que $q_j(n)=1$, c'est à dire que l'interprétation de A(n) est aberrante, comme le système effectue d'autres controles, il serait extrêmement peu probable que

$P_j(n) = 1$, le coefficient de $q_j$ dans $P_j$ est donc toujours inférieur à l'unité.
Il en sera de même de l'influence de $P_j(n)$ dans $P_1(n+1)$, l'écart initial: s'amortira selon une loi sensiblement exponentielle, la remarque faite au début de ce paragraphe laisse prévoir que la phase finale de la convergence sera très rapide.

Remarque:

Cette démonstration n'est pas absolument rigoureuse, et elle ne peut pas l'être par ce qu'il y a N! régimes structurellement stables et que le système ne peut pas converger vers tous à la fois. Toute fois comme la redondance est élevée, la convergence sera facilement obtenue pour un taux normal d'erreurs ou d'ambiguités.

La théorie complète n'est pas encore faite.

7°) Rattrapage de phase.

Nous allons maintenant résoudre le problème posé au §2.

Nous appellerons $P_i'$ ce que nous appelions $P_i$, ou $\bar{P}'$ ce qui était $\bar{P}$ .

Nous considérons le vecteur:

$\bar{p}''(n,k) = \bar{P}(n-k).\bar{M}(k+1)$

et calculons $\bar{\bar{P}}''$ par les formules 3,2 et 3,3 dans les quelles nous remplaçons K par K-1. Remplacer $\bar{P}'$ par $\bar{P}''$ revient à courtcircuiter le symbole de rang n et à rechercher celui de rang n+1.

On se donne alors une période de base: $\tau$ et une fonction du temps: $\varphi(t)$ telle que:

$\varphi(t)+\varphi(\tau-t) = 1$

$\varphi(0) = 1-\alpha$

$\varphi(\tau) = \varphi(-\tau) = \alpha$

$\alpha$ n'est pas négligeable devant l'unité.

Celà fait, il suffit d'écrire:

$\bar{P}(n) = \bar{P}'(n).\varphi(t)+\bar{P}''(n).\varphi(\tau-t)$

où :t est l'époque de réception du vecteur $\bar{q}(n)$.

1er cas: un caractère a été oublié et le caractère suivant A(n+1) est reçu à la place du caractère A(n) manquant:
la composante de $\bar{P}''$ correspondant au symbole $s_i$ de rang n+1 sera présente quoi que multipliée par $\alpha < 1$, il reste une ambiguité qui pourra être rattrapée par la suite.

2ème cas: Le vecteur $\bar{q}(n)$ n'a pas été reçu, le vecteur $\bar{P}(n+1)$ donne pour le symbole correct une faible ambiguité.

3°) Cas intermédiaires

Le signal issu de $A_n$ est reçu avec un retard variant entre 0 et $\tau$ . Au fur et à mesure que ce retard s'accroit, on considère comme de plus en plus probable que le caractère $A_n$ a été perdu et que c'est le suivant qu'il faut prendre en considération.

4°) Cas normal

Si le signal est reçu au voisinage de l'instant zéro, le terme en $\bar{P}''$ introduit une certaine ambiguité, mais comme le digramme constitué par les symboles de rangs n et n+1 ne se reproduira pas aux rangs n+1 et n+2, cette ambiguité sera vite "amortie", comme il l'a été expliqué au §6.

7,1 Synchronisation

A chaque réception d'un signal définissant un vecteur $\bar{q}(n)$, on reinitialise la fonction $\varphi(t)$ à la valeur :$-\tau$ il y a donc synchronisation du processus sur le rythme imposé par l'environnement, ce dernier pouvant varier à peu près de $\tau/2$ à $2\tau$.

Les vecteurs $\bar{P}(n-k)$ sont mis dans des régistres décalés à chaque période:

Le vecteur $\bar{P}(n-k)$ qui était multiplié par la matrice $\bar{M}(k)$ deviendra: $\bar{P}((n+1)-(k+1))$ et sera multiplié par la matrice $\bar{M}(k+1)$.
Si k=K le vecteur $\bar{P}(n-k)$ sera effacé de la mémoire.

8°) Auto-apprentissage.

Le principe de l'auto-apprentissage consiste à déterminer les coefficients $m_{ij}(k)$ qui sont les probabilités conditionnelles d'observer le symbole $s_j$ après avoir observé $s_i$. On injectera donc dans le système les mots à reconnaitre en répétant les mots les plus usuels. Pour chaque digramme $s_is_j$ correspondant à deux symboles séparés par k rangs, on incrémentera un compteur :$f_{ij}(k)$.

On s'efforcera alors de rechercher les digrammes les plus rares de façons à définir ceux qu'il faudra éliminer. S'il le faut on décrémentera les compteurs correspondant aux mots supprimés.

Celà fait, il suffira de réduire chaque ligne des matrices $\bar{\bar{F}}(k)$ par la formule:

$$m_{ij}(k) = f_{ij}(k)/_{j}\ f_{ij}(k)$$

Remarque 1

Cette opération ne sert qu'à faciliter la reconnaissance des mots et des digrammes les plus usuels. En l'absence d'idées à priori,

il suffira de faire:
$\forall k,i,j$ , $m_{ij}(k) = 1/(N-1)$
ceci sauf si $m_{ij}(k) = 0$
Autrement dit, le système est entièrement déterminé par la matrice $\bar{C}(1)$.

Dans ce cas, la synthèse d'un mot ne se fait pas par associations d'idées mais par éliminations successives selon le principe des digrammes interdits exposé au §5.

Remarque 2

Dans le cas du système à auto-apprentissage, il serait possible de s'affranchir de la sujetion des digrammes interdits en "ébasant" les distributions de probabilités, opération, peu catholique, qui consiste à annuller toute probabilité inférieure à un seuil donné puis à majorer celles qui restent pour retrouver une partition d'un ensemble de mesure unité.

On s'écarte du modèle mathématique, mais les résultats ne devraient pas différer beaucoup des performances théoriques.

C'est alors que les différences entre les probabités conditionnelles : $m_{ij}(k)$ jouent un rôle prépondérant et que le processus d'auto-apprentissage trouve toute sa signification.

Remarque 3

Le système décrit ici dérive d'un modèle de simulation des réseaux nerveux qui prend encore un peu plus de libertés avec la rigueur mathématique:

-- La composition des signaux se fait non par multiplication mais par addition et par ébasage. Il en résulte que le système ne peut pas se bloquer complètement mais qu'il peut aussi prendre des décisions aberrantes si une composante se trouve être d'une amplitude anormale.

-- On ne peut plus parler de distributions de probabilités parce que la somme des composantes n'est qu'approximativement égale à l'unité, il s'agit de "variables d'état" plus que de grandeurs mathématiques. (Signalons que cette somme est régulée par un asservissement rétroactif dont les oscillations ressemblent aux enrégistrements électro-encéphalographiques réels.)

Les performances de ce système plus proche de la réalité devraient être comparables à celles du modèle théorique.

9°) Conclusion

Il ne semble pas que, dans le domaine des applications informatiques, le système faisant l'objet du présent papier puisse être compétitif avec les systèmes existants.

C'est dans le domaine de la reconnaissance des langages naturels: écriture, parole, et même de phrases musicales, qu'il pourrait trouver son intéret. Les problèmes de liaison homme-machine sont en effet de plus en plus à l'ordre du jour, et c'est probablement dans l'optique de l'imitation de processus naturels, c'est à dire de la "bionique" que pourraient être trouvées des solutions à de tels problèmes.

Notre but n'était pas ici de traiter complètement tous les problèmes théoriques de convergence et de conditions aux limites relatifs à ce processus, nous avons simplement voulu montrer que son pouvoir de résolution était suffisant pour en justifier l'intéret.

BIBLIOGRAPHIE

J.K.HAWKINGS:"Self Organisings systems. A rewiew and Commentary". Proceedings IRE Janvier 1961 Vol.49 n°1

D.O.HEBB:"The Organisation ofBehaviour, a Neurophysiological Theory" N.Y Juin 1961

KLEENE:"Representation of Events in Nerve Nets and Finite automata" Automata studies, Shannon, Mac Carty 1956

J.VON NEUMANN:"Probabilistic Logics"Automata Studies, Shannon, Mac Carty 1956

J.VON NEUMANN:"The Computer and the Brain" Yale University Press, 1958

J.C.LEVY, D.BERTAUX, J.GROLLEAU:La structuration par apprentissage d'une mémoire à association sémantique", Symposium sur l'informatique médicale et les intelligences artificielles, Toulouse, Mars 1967

J.C.LEVY J.GROLLEAU:"Reconnaissance des configurations dynamiques par les réseaux nerveux"Symposium I.F.A.C. Yerevan (URSS) Octobre 1968

J.C.LEVY:"Reconnaissance des configurations spatio-temporelles"Colloque international de reconnaissance des formes. Grenoble 1968

J.C.LEVY:"Reconnaissance des configurations dynamiques. Système mixte, structuré et auto-adaptatif" Automatisme 11 novembre 1970 TOME XV

Figure 1

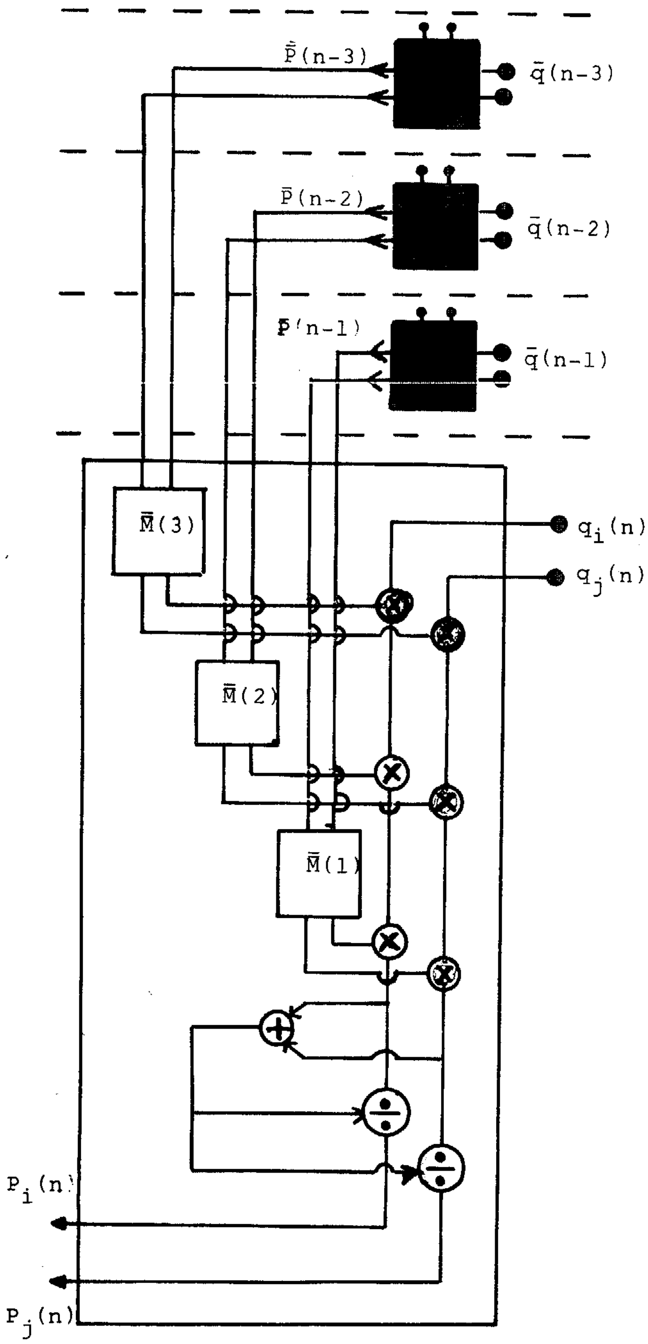
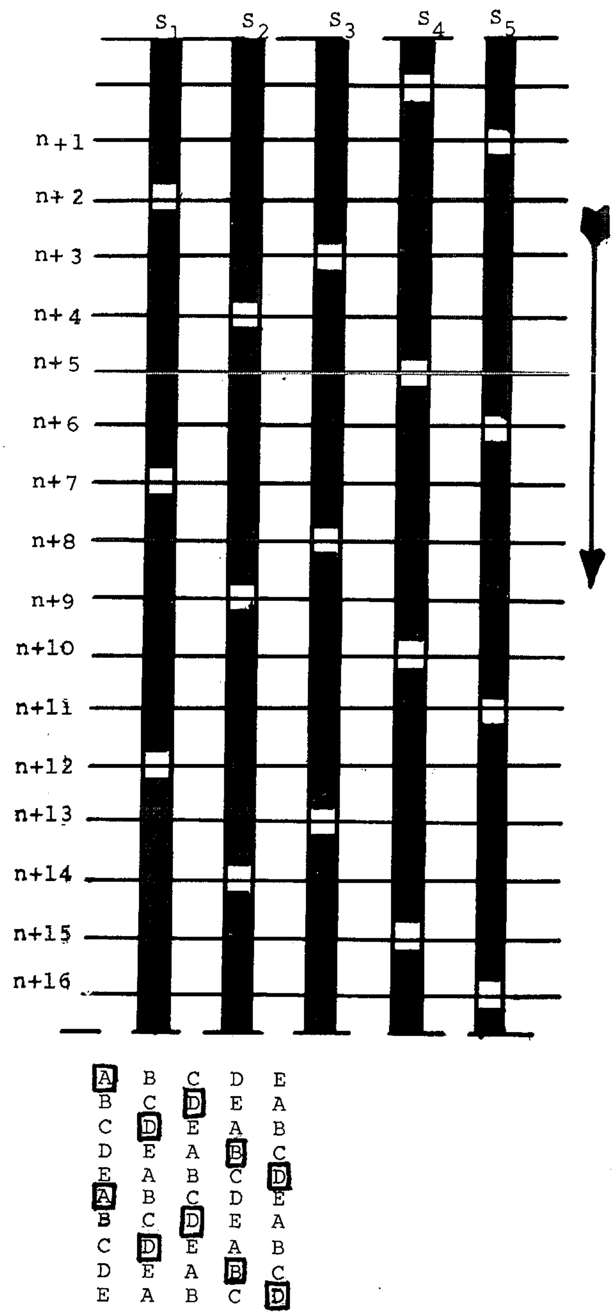



Figure 2




Exemple dans le cas particulier de cinq lettres, les digrammes interdits étant:
AB-BC-CD-DE-EA
Le mot reconnu étant :A-D-D-B-D