QUANTIFICATION VECTORIELLE SOUPLE, BASE DE LA RECONNAISSANCE DE LA PAROLE

G. ZANELLATO

Service d'Electronique et de Techniques Numériques , Faculté Polytechnique de Mons,
Bvd DOLEZ, 31 - B-7000 MONS (BELGIQUE)

RESUME

Dans un système de quantification classique, un vecteur est remplacé par le centroïde qui en est le plus proche. Il est alors impossible de distinguer deux vecteurs appartenant à une même classe. Afin de pallier à cet inconvénient, nous utilisons les deux plus proches voisins ainsi qu'un "degré d'appartenance", calculé en fonction de la distance entre le vecteur et les deux centroïdes. Cette quantification "souple donne de meilleurs résultats dans le cadre de la reconnaissance en système multilocuteur.

Lors de la reconnaissance de grands vocabulaires, on procède à une segmentation en phonèmes et on effectue la décision sur la succession ainsi obtenue. Cependant, une modélisation correcte des phonèmes présente de grandes difficultés. On peut alors utiliser des "variables intermédiaires", qui sont des composants élémentaires des phonèmes, beaucoup plus simples à modéliser. Le choix s'est porté sur un ensemble de 100 "unités acoustiques" définies à partir des centroïdes provenants de la quantification souple. Les résultats observés sont très prometteurs.

SUMMARY

In a classical quantization system, each vector is represented by the nearest centroïd. But it is then impossible to distinguish two vectors belonging to the same class. In order to mitigate this disadvantage, we have taken into account the two nearest neighbour and also a "belonging degree" calculated from the distances between the vector and the two centroïds. In the case of speaker independant speech recognition system, this "fuzzy" quantization gives better results.

For the recognition of large vocabularies, it is usual to perform a phonemic segmentation and then to achieve the matching on the obtained sequence. But it is very hard to set up a good modeling for the phonemes. Nevertheless, we can take into account elementary components of the phonemes, for which simplest models may be used. We investigate a set of 100 "acoustic units" that have been defined from the centroïds of the fuzzy quantization. The results we obtained are very attractive.

1. INTRODUCTION

Il est courant, en reconnaissance de la parole, de faire la distinction entre plusieurs tâches; on peut se proposer la reconnaissance de :

- mots isolés (petits, grands ou très grands vocabulaires) ;
- mots concaténés ;
- courtes phrases (dans un contexte donné) ;
- parole continue.

Il convient aussi de distinguer eux les systèmes dits "monolocuteur", "plurilocuteur" et "indépendant du locuteur".

L'expérience acquise dans notre laboratoire se rapporte principalement aux systèmes de reconnaissance de mots isolés, petit vocabulaire, monolocuteur, plurilocuteur et indépendant du locuteur. Toutefois, nous proposons une méthode qui permet d'aborder la reconnaissance de mots isolés (grand vocabulaire) et ceci indépendamment du locuteur. Son principe peut éventuellement être étendu à la reconnaissance de courtes phrases ou même à celle de la parole continue.

2. CONDITIONS EXPERIMENTALES.

Les conditions d'analyse sont les suivantes :

. fréquence de coupure du filtre de garde = 4625 Hz.;
. fréquence d'échantillonage = 10000 Hz.;
. pré-accentuation du signal $\alpha = 0.95$;
. l'analyse s'effectue sur des fenêtres de 300 échantillons pondérés par une fonction de Hamming;
. le décalage est de 100 échantillons;
. ordre de l'analyse : p = 12;

Le vocabulaire de travail est constitué de 20 mots de la langue française correspondant à la commande d'une minicalculatrice : ZERO , UN , DEUX , TROIS , QUATRE , CINQ , SIX , SEPT , HUIT , NEUF , PLUS , MOINS , FOIS , DIVISE PAR , EGAL , VIRGULE , PARENTHESE , FERMER , CORRECTION , EFFACER.

Cent locuteurs masculins différents ont prononcé chacun une fois cet ensemble de mots.

3. PRINCIPE DE QUANTIFICATION SOUPLE

3.1. INTRODUCTION.

Pour mieux définir la position d'un vecteur spectral (VS) dans l'espace acoustique, nous avons décidé d'organiser la partition de l'espace spectral selon un nouveau concept :

dans la quantification "simple", chaque tranche du signal est représentée par le numéro de la classe dont le centre de gravité (CG) en est le plus proche. Il est cependant possible d'associer à ce numéro un "degré d'appartenance" qui caractérise l'éloignement du VS par rapport au CG ; en outre, pour une meilleure localisation, on va prendre en compte les deux classes les plus voisines.

En définitive, chaque VS X_j est représenté par 4 nombres : les deux classes les plus proches (C_1 et C_2) et les degrés (ou probabilités) d'appartenance respectifs :

$$a_1 = \frac{d(X_j,C_2)}{d(X_j,C_1) + d(X_j,C_2)}$$

$$a_2 = 1 - a_1$$

où

X_j est le VS considéré ;
C_1 est le CG le plus proche de X_j ;
C_2 est le second CG le plus proche de X_j ;
$d(.,.)$ est la distance d'ITAKURA [1] entre 2 VS.

3.2. REALISATION.

En se basant sur ce principe, nous avons effectué une partition de l'espace spectral en 100 classes. La méthode utilisée pour cette partition comporte deux étapes :

1) Initialisation

1°) Division de l'espace spectral en deux classes (correspondant approximativement aux sons voisés et aux sons non voisés), en partant de deux "points germes" définis par les coefficients de réflexion :

K1 = ½ , ½ , ... , ½
K2 = -½ , ½ , ... , ½

2°) Classement de l'ensemble des vecteurs spectraux par rapport à ces 2 centres de gravité, selon le principe de quantification décrit ci-dessus mais en n'utilisant que C_1 et a_1.

3°) Détermination des coordonnées des CG des classes ainsi créées : on effectue une analyse LPC à partir de la fonction d'autocorrélation moyenne $r_k(k)$ des vecteurs de chaque classe :

$$r_k^{(i)}(k) = \frac{1}{N_i} \sum_{Xj \in (i)} \frac{r_k^{(Xj)}(k) * a^{(Xj)}(i)}{\alpha_p^{(Xj)}}$$

avec
i = une des classes formant l'espace spectral ;
$a^{(Xj)}(i)$ = degré d'appartenance du VS X_j à i ;
$\alpha_p{}^{(Xj)}$ = énergie résiduelle de modélisation de Xj
$N_i = \sum_{Xj \in (i)} a^{(Xj)}(i)$ = somme des degrés d'appartenance

4°) Considérant ces nouveaux CG , on retourne en 2°) jusqu'à ce que d'une itération à la suivante, la position des CG ne soit pas trop modifiée (< 1% p. ex.).

2) Incrémentation du nombre de classes

1°) Eclatement de la classe qui présente la distorsion la plus grande : la distorsion moyenne intra-classe D_i est la somme pondérée étendue à tous les éléments d'une classe, des distances entre ces éléments et le CG de la classe :

$$D_i = \frac{1}{N_i} \sum_{Xj \in (i)} d(X_j(i), C_i) * a^{(Xj)}(i)$$

L'éclatement d'une classe se fait en perturbant les coefficients Ki du CG connu (*0.99 et *1.01).

2°) Reclassement de tous les VS par rapport à tous les CG connus, c'est-à-dire ceux qui n'ont pas été modifiés et les deux nouveaux.

3°) Calcul des fonctions $r_x(k)$ correspondantes.

4°) Optimisation du classement (variation faible des coordonnées des CG).

3.3. RESULTATS.

Un des critères utilisés pour définir la qualité d'une quantification est la mesure de la distorsion totale pondérée (D_T) ; elle a pour expression :

$$D_T = \frac{1}{N} \sum_{i=1}^{N} D_i$$ où N est le nombre de classes.

Nous avons effectué la partition d'un espace spectral constitué de 52.580 VS (50 versions du vocabulaire); d'abord en employant la technique classique des éclatements binaires successifs depuis 2 jusqu'à 128 classes (QB7) [2], ensuite en utilisant la méthode décrite ci-dessus, jusqu'à 100 classes (QS100). Dans le premier cas, D_T vaut 0,52 et dans le second, 0,43.

Au niveau de la reconnaissance proprement dite, deux séries de tests ont été réalisées. La première mettait en oeuvre un système monolocuteur basé sur l'algorithme du Dynamic Time Warping (DTW) [3] et utilisait quatre versions du vocabulaire; les résultats sont quasi identiques pour les deux types de quantification retenus (≈99%). La seconde se rapportait à un système multilocuteur basé sur la modélisation par automates probabilistes [4]. Dans ce cas, deux types de modèles furent utilisés : les Modèles Markoviens Classiques [5] (cas 1), et ceux pour lesquels les probabilités de transition ont été supprimées (on définit 2 sous-états par état du modèle) et où on a introduit les "probabilités de présence" [6] (cas 2). Dans chaque cas, 50 versions (T1) ont servi à l'entraînement des modèles et 50 autres (T2) uniquement aux tests (les 50 locuteurs de T2 sont différents des 50 de T1). Les résultats sont repris dans le tableau suivant; on a indiqué le nombre d'erreurs pour 1000 tests :

	QB7,cas1	QS100,cas1	QB7,cas2	QS100,cas2
T1	9	11	6	6
T2	34	20	37	17

Ces résultats montrent aisément l'intérêt de cette nouvelle méthode pour les systèmes multilocuteurs.

4. PRINCIPE DE LA RECONNAISSANCE PAR UNITES ACOUSTIQUES

4.1. INTRODUCTION.

Les systèmes de reconnaissance de mots isolés, petit vocabulaire (20 à 50 mots) offrent actuellement de bonnes performances. L'expérience nous a montré que l'algorithme du DTW en monolocuteur et l'algorithme basé sur la modélisation par automates probabilistes en multilocuteur conduisent à des taux d'erreurs réduit (1% à 2%) et des temps de réponse court (≈ 1 à 2 fois le temps réel si l'on emploie des micro-processeurs spécialisés). Cependant, lorsque la taille du vocabulaire croît, les temps de réponse et le taux d'erreurs augmentent rapidement.

Il convient donc ce cas d'adopter une autre donc se proposer de segmenter les mots à re-connaître en phonèmes et de "comparer" chaque segment ainsi déterminé à des modèles de référence représen-tant chacun un phonème de la langue. On obtient alors une suite de "phonèmes reconnus" qui constitue une re-présentation phonétique du mot; un dictionnaire de correspondance en fournit la (ou les) représenta-tion(s) orthographique(s) (problème des synonymes).

Cette méthode, simple en son principe, ne donne malheureusement pas de bons résultats, ceci étant principalement dû au fait que la forme sonore d'un phonème est essentiellement évolutive et fortement influencée par le contexte phonémique dans lequel il est placé (problème de la co-articulation).

On peut cependant concevoir un phonème comme étant composé d'une séquence de "sons élémentaires", produits, chacun, par une succession de configurations différentes du conduit vocal. Or, il existe une cor-respondance bi-univoque entre une position donnée de ce conduit et la position d'un VS dans l'espace acous-tique. La succession des VS représentant un son élémentaire dessine donc un chemin dans l'espace spectral; ce chemin peut traverser plusieurs classes. Différents chemins peuvent correspondre au même son élémentaire; leur ensemble constitue ce que nous appelons une "unité acoustique" (UA); celle-ci peut donc être considérée comme l'image d'un son élémentaire. D'autre part, des successions de sons élémentaires différents peuvent correspondre au même phonème placé dans des contextes différents.

Il s'ensuit que le choix des composants élémen-taires d'un phonème comme unités de base pour la seg-mentation et la reconnaissance constitue une solution intéressante. Il est également avantageux de pouvoir automatiser la détermination de ces UA .

A cette fin, nous utilisons comme unités acous-tiques, les représentants (ou centres de gravité) des 100 classes obtenues lors du partitionnement de l'espace spectral par la méthode décrite ci-dessus (Cf. "Principe de quantification souple"). Il est évident que l'on pourrait utiliser les classes de quantification obtenues d'une autre façon (p.ex., par la technique des éclatements binaires successifs).

En résumé, ce principe permet donc de segmenter un signal vocal en une succession d'UA. Ces unités partagent l'espace acoustique en "zones" caractérisées par des centroïdes définis par quantification vecto-rielle. En réalité, ces zones peuvent se recouvrir; elles correspondent donc à un concept plus large que celui de "classes" de quantification.

4.2. MODELISATION DES UNITES ACOUSTIQUES.

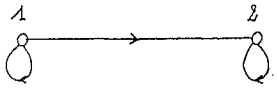
Tout comme dans le cas des mots isolés, l'en-traînement des modèles s'effectuera en deux phases : initialisation puis entraînement proprement dit.

4.2.1. INITIALISATION.

Il s'agit, d'abord, de fixer la "taille" des modèles. Dans un premier temps, nous avons utilisé des automates probabilistes simples constitués de deux états : le premier caractérisant au mieux le son cor-respondant à l'UA considérée, c'est-à-dire que l'on peut le voir comme étant une image de la région de l'espace spectral correspondant à cette UA, et le se-cond caractérisant un son aussi différent que possible de celui représenté par le premier; il s'agit en fait de la région complémentaire, par rapport à l'espace spectral utilisé, de celle associée au premier état.

On pourra éventuellement essayer par la suite 3 ou 4 états ou même un nombre d'états dépendant soit de la taille du "nuage de vecteurs spectraux" formant chacune des UA , soit de sa compacité (distribution des distances), soit des deux.

Le modèle probabiliste d'une UA a, dans notre cas, la forme représentée à la figure suivante.



Le vecteur de probabilités d'émission $b_1(u_k)$ (k=1,2,...,100) associé au premier état, est déduit de l'initialisation et ré-estimé après chaque étape d'en-traînement. Celui associé au second état est calculé en fonction du premier :

$$b_2(u_k) = 1 - b_1(u_k) \qquad k = 1,2,...,100 ,$$

de façon à définir un état "complémentaire" : le "son" correspondant à cet état doit être le plus "éloigné" possible de celui correspondant à l'UA considérée.

Les valeurs à déterminer sont les matrices de production (on ne tient pas compte des matrices de

çon suivante (il faut se référer au "Principe de quantification souple") : chaque vecteur spectral est "quantifié" par un ensemble de quatre nombres : les deux centroïdes les plus proches et les degrés d'appartenance correspondants) :

1°) On utilise l'ensemble des VS quantifiés dont le plus proche voisin correspond à la classe dont on veut modéliser l'UA associée (par exemple l'unité acoustique J).

2°) Chacun de ces vecteurs est considéré faire partie de la classe J pour une quantité égale à son degré d'appartenance à cette classe.

3°) La somme de ces (premiers) degrés d'appartenance représente la "quantité" (nombre réel) de VS affectés à la classe J .

4°) On considère également le second plus proche voisin (classe L) et le degré d'appartenance correspondant de chacun des VS définis en 1°).

5°) Pour chaque classe L , on fait la somme de tous les degrés d'appartenance, ce qui fournit la "quantité" (nombre réel) de VS affectés à cette classe pour le modèle de l'UA J .

6°) La somme de toutes les "quantités" (pour toutes les classes utilisées) est égale au nombre N de VS définis en 1°) .

7°) Le rapport entre la "quantité" Q_k afférante à une classe k et le nombre N fournit la "probabilité de production" du son k dans le premier état du modèle de l'UA considérée.

8°) De façon à s'assurer qu'aucune probabilité n'est nulle, on affecte aux sons jamais rencontrés une probabilité égale à la moitié de la plus faible déterminée en 7°).

9°) Ce calcul terminé, on pondère toutes les valeurs de sorte que la probabilité de production du son correspondant à l'UA que l'on vient de modéliser (c'est-à-dire le son J pour l'UA J) soit égale à une valeur donnée (0.9 par ex.). La somme des probabilités relatives à un état n'est alors plus unitaire.

On obtient de la sorte la matrice de production initiale.

4.2.2. ENTRAINEMENT.

Cette seconde phase nécessite, pour son déroulement, l'application du principe de la reconnaissance de mots (UA) enchaînés. Nous allons développer ici une telle méthode.

4.2.2.1. RECONNAISSANCE DE MOTS ENCHAINES.

Le terme "mots" désigne dans ce cas tout autant des mots (complets) que des phonèmes, diphones ou "unités acoustiques".

Le signal à reconnaître étant composé d'une succession de mots dans un ordre quelconque, il est nécessaire de le comparer à l'ensemble des modèles (de mots) de référence et ceci d'une façon continue, c'est-à-dire du début à la fin du signal.

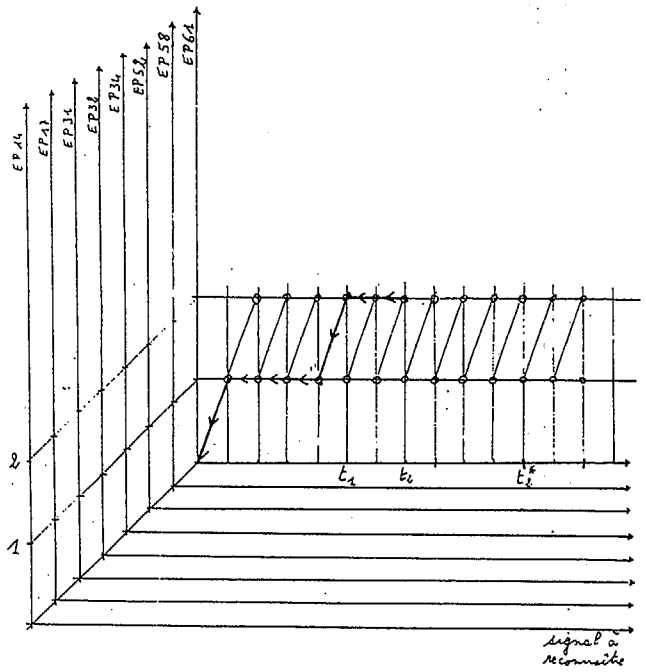
Cependant, le fait d'utiliser comme UA les (100) représentants servant à la quantification présente un avantage important : on n'effectue la comparaison qu'avec les modèles des représentants réellement présents dans le signal! Ainsi, par exemple, s'il se présente sous la forme quantifiée suivante (pour le plus proche voisin uniquement) :

17-34-34-31-34-34-34-34-34-52-17-58-58-61-14-32-14-14,

alors les modèles utilisés pour la reconnaissance seront ceux des UA 14, 17, 31, 32, 34, 52, 58 et 61, c'est-à-dire huit modèles au lieu de 100 !

De plus, lorsque le signal de parole est beaucoup plus long (quelques centaines de tranches), ce choix des modèles retenus n'est pas effectué sur son entièreté (auquel cas, on serait parfois obligé d'utiliser la quasi-totalité des modèles) mais sur de petites parties (une quinzaine de tranches); dans ce cas, des modèles de "silence" doivent cependant être pris en compte. Cette technique est applicable car chaque fois que l'on a "trouvé" une UA , on se replace dans des conditions initiales (Cf. plus loin).

Le premier objectif est la détermination des modèles d'UA à utiliser (on se base sur le signal quantifié). Il faut ensuite organiser plusieurs exécutions de l'algorithme de VITERBI en parallèle (il y en a un par UA conservée (Cf. la figure suivante état (c'est-à-dire l'état qui permet de "quitter" le modèle : il est aussi "éloigné" que possible de l'UA considérée (Cf. supra)). Pour cela, on est obligé d'effectuer un "back-tracking" à chaque tranche du signal à reconnaître (pour chacun des modèles).On considère que l'on "quitte" le modèle lorsque l'on obtient un chemin horizontal dans le second état (3, 4 (ou plus) points successifs) car cela signifie que pour les tranches en question, on est éloigné du modèle de l'UA considérée.



Dans la figure ci-dessus, le back-tracking effectué à partir de la tranche t_1 du signal est tel que l'on peut supposer que l'on est toujours dans le modèle alors qu'à la tranche t_2 , on peut supposer qu'on le quitte (on a choisi 3 points successifs).

On note la valeur de la probabilité maximale obtenue en (t_2-3) , c'est-à-dire quand on quitte le modèle. On effectue le même calcul de parcours optimal pour tous les modèles de UA (les t_2 seront différents) et on décrète comme "reconnu" celui qui, en "son" (t_2-3) présente la probabilité maximale la plus élevée. Soit t_2^* la tranche du signal qui a fourni ce résultat. On "élague" alors le signal à reconnaître des tranches 1 jusqu'à (t_2^*-3) , c'est-à-dire celles qui, dans la détermination du chemin optimal, ont été affectées au premier état du modèle.

On recommence alors une procédure identique sur les quinze tranches suivant (t_2^*-3) (ou jusqu'à la fin du signal), afin de déterminer l'UA suivante. On regarde d'abord quels sont les numéros de classes (ou UA) qui sont présents dans le signal restant afin de diminuer le nombre d'opérations, puis on procède à nouveau comme expliqué ci-dessus.

On obtient finalement, pour l'entièreté du signal considéré, une succession d'UA (supposons 34 - 58 - 14 dans l'exemple utilisé ici) qui représente en fait la segmentation du mot en UA.

Remarque : il est à noter que le nombre d'opérations nécessaires à la décomposition en UA n'est que lentement croissant en fonction du nombre de modèles d'unités acoustiques et donc du nombre de classes de quantification; le problème de la reconnaissance devient plus intimement lié à celui de la quantification.

A ce stade, l'entraînement diverge de la reconnaissance. Dans le cadre de celle-ci, il "suffit" en effet de comparer la suite des UA obtenue à celles contenues dans un dictionnaire pré-établi.

4.2.2.2. RE-ESTIMATION DES PARAMETRES.

Les modèles d'UA seront, dans un premier temps, entraînés avec un ensemble de 1000 mots isolés (50 locuteurs) Chacun de ces mots sera décomposé en une succession d'UA .

On connaît, grâce à la méthode exposée au § 4.2.2.1., le groupe de VS initiaux (c'est-à-dire formant le mot isolé) associé à chaque UA de chaque succession. Il "suffit" donc, *pour chaque modèle d'UA*,

état de chaque modèle se fait alors classiquement mais en tenant compte du fait que la contribution de chaque VS est égal à sa "probabilité d'appartenance" à la classe considérée. Ainsi, par exemple, si un vecteur spectral affecté par l'entraînement au premier état du modèle associé à l'UA n° 30 , est quantifié de la façon suivante (Cf. Principe de quantification souple) :

C1 : 17 ; a1 : 0.8
C2 : 21 ; a2 : 0.2

alors, la contribution de ce vecteur au calcul des probabilités d'émission est telle que la "quantité" afférante à la classe 17 est augmentée de 0.8 et celle afférante à la classe 21 est augmentée de 0.2.

On s'arrange cependant pour qu'aucune valeur ne soit nulle et on pondère ensuite les probabilités comme expliqué plus haut.

Les probabilités relatives au second état sont déduites de celles-ci d'une façon identique à celle décrite au § 4.2.1. .

4.3. PRINCIPE DU PASSAGE DES UA AUX PHONEMES.

La seconde partie de la méthode consiste en l'identification des phonèmes à partir des successions d'unités acoustiques obtenues.

On peut estimer qu'un phonème est constitué d'une succession de VS formant, dans l'espace acoustique, un chemin traversant plusieurs UA. L'effet de la co-articulation se traduit par la traversée potentielle de différentes UA en fonction des différentes élocutions d'un phonème donné. La prise en compte des différentes successions possibles permet de créer un modèle de phonème, celui-ci étant entraîné par un ensemble d'UA , tout comme les modèles d'UA sont entraînés par un ensemble de VS.

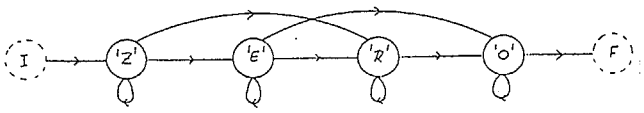
Afin de déterminer la ou les successions représentant un phonème donné, nous réutilisons les mots dont nous disposons. A chaque mot du vocabulaire est associé un automate probabiliste comportant (dans un premier temps) un nombre d'états égal au nombre de phonèmes constituant le mot. Nous allons entraîner ces automates avec les mots segmentés en UA , de façon à déterminer, pour chaque état, un vecteur de probabilités d'émission d'UA.Il est à prévoir qu'après cet entraînement, chaque état d'un modèle sera spécialisé dans "l'émission" d'un phonème (c'est-à-dire une succession d'UA).

En résumé, la méthodologie qui est proposée consiste à prévoir une étape intermédiaire avant la segmentation d'un mot en phonèmes. L'intérêt est une automatisation totale des procédures.

Actuellement, la reconnaissance d'un mot s'effectue de la façon suivante : le signal vocal est d'abord segmenté en unités acoustiques par comparaison aux modèles d'UA prédéfinis; ensuite, la succession de segments obtenus est traitée par les automates représentant chacun un des mots du vocabulaire. A ce stade, le mot est ou non décrété reconnu.

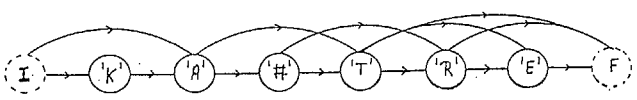
4.4. MODELISATION DES MOTS.

Dans ce cas, le nombre d'états d'un modèle est égal au nombre de phonèmes du mot auquel il est associé (compte tenu des silences éventuellement présents à l'intérieur du mot) et les transitions permises dépendent de la structure du mot. Ainsi, par exemple, le mot 'ZERO' sera-t-il représenté par l'automate suivant :



c'est-à-dire que toutes les versions débutent par 'Z', que le 'E' et le 'R' peuvent être évités, mais pas le 'O',

et le mot 'QUATRE' , par le suivant :



c'est-à-dire que l'on peut éventuellement éviter le 'K' du début du mot, mais pas le 'A' , que le silence ('#'), le 'T' , le 'R' et le 'E' peuvent être évités, et que le mot peut se terminer sur 'T' , sur 'R' ou sur 'E' .

4.5. RESULTATS.

En employant ces deux derniers concepts et sur une base de 1000 mots (50 versions) ayant servi à l'entraînement des modèles d'UA *et* des modèles de mots, on obtient un taux d'erreurs de :

- 1 % (10 erreurs) si on reprend ces mêmes mots comme tests

- 2 % (19 erreurs) si on utilise 1000 autres mots (50 nouveaux locuteurs).

5. BIBLIOGRAPHIE.

[1] F.ITAKURA, "Minimum Prediction Principle Applied to Speech Recognition", IEEE Trans., ASSP-23, February 1975, pp. 67-72.

[2] R.M.GRAY, "Vector Quantization", IEEE ASSP Magazine, April 1984, pp. 4-29.

[3] H. SAKOE, S. Chiba, "Dynamic Programming Algorithm for Spoken Word Recognition", IEEE Trans., ASSP-26, February 1978, pp. 43-49.

[4] F.JELINEK, "Continuous Speech Recognition by Statistical Methods", Proc. IEEE, Vol. 64 (April 1976), pp.532-556.

[5] L.R.RABINER, S.E.Levinson, and M.M.Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independant Isolated Word Recognition", B.S.T.J., V62, N°4, April 1983, pp.10-1105.

[6] R.BOITE, H.Leich, G.Zanellato "Isolated Word Recognition by Hidden Markov Model", EUSIPCO-86, pp. 541-544.

L.E.BAUM, T.Petrie, G.Soules, and N.Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabailistic Functions of Markov Chains", ANN. Math. Stat., 41 (1970), pp.164-171.