

LE SYSTEME DE SELECTION ET VALIDATION AUTO-ADAPTATIF VARAP
POUR LA RECONNAISSANCE DE LA PAROLE CONTINUE

ANDREEWSKY A. DESI M. RINGOT P.

CNRS - LIMSI - ORSAY BP 30 91406 ORSAY CEDEX FRANCE

The VARAP system has been designed with the goal of being able to select references by using the representativity in a recognition task on a prelabelled data base. This base is analysed in selfconsistency with its own references. Several situations may occur according to whether the analysed references are recognised or not. The selection obtained improves recognition.

VARAP (Validation Autocoherente pour la Reconnaissance Automatique de la Parole Continue) est un système qui sélectionne dynamiquement les références d'un corpus préétiqueté automatiquement, sans utilisation de seuils, en faisant appel uniquement à la capacité de reconnaissance des dites références.

I. LE CORPUS.

Il comprend 700 phrases qui correspondent à 1500 mots différents. L'étiquetage effectué automatiquement est complété par des corrections manuelles qui portent sur environ 10% du corpus.

II. LA DISTANCE.

La comparaison des références (spectres étiquetés) se fait au sens d'une certaine distance que nous allons définir.

Soient deux spectres, l'un noté O_i et l'autre X_i , i variant de 1 à 16 (O_i , X_i sont les valeurs de chacun des spectres sur 16 canaux).

Formons $d_i = O_i - X_i$ ($i = 1...16$) soit $D_i = d_{i+1} - d_i$, alors la distance entre deux spectres est donnée par la formule: $D = |D_1| + \dots + |D_{16}|$
La distance entre deux références tient compte du fait que chaque référence est constituée de trois spectres consécutifs prélevés au voisinage de l'extremum étiqueté et choisis en minimisant leurs distances respectives, l'extremum faisant toujours parti de la référence ternaire. Cependant la disposition relative des spectres par rapport à l'extremum peut être paramétrée.

III. LE MODE DE SELECTION DU SYSTEME VARAP.

La sélection se fait en deux étapes en comparant le fichier CORPUS à un fichier dit VALID construit dynamiquement à partir de CORPUS. Toutefois, étant donné que les temps de calculs sont longs, le fichier CORPUS est partitionné en fractions de 100 phrases. De plus chaque lot de 100 phrases est à son tour partitionné en deux fichiers de références relatives respectivement aux maximums et aux minimums sélectionnés au cours de l'étiquetage automatique. La comparaison dont la description suit peut-être effectuée sur chacun des fichiers partitionnés et il est possible ensuite de concaténer les résultats des comparaisons et de recommencer la sélection sur le fichier concaténé.

III.1. La comparaison.

Dans ce qui suit, le fichier d'entrée sera appelé CORPUS, indépendamment du fait qu'il peut être relatif aux minimas ou aux maximas. On procède comme suit:

A. La première référence de CORPUS est mise dans VALID.

B. Les références suivantes de CORPUS sont alors comparées une à une à toutes les références de VALID. Dans l'expérience effectuée, on affiche les 15 premières références qui constituent ainsi un treillis. Cela n'est possible qu'à partir du moment où il y a déjà au moins 15 références dans VALID.

C. Sur chaque treillis, un scrutin majoritaire est effectué, qui tient compte du rang des références et de leur nombre dans le treillis. Les quatre premiers phonèmes majoritaires sont ainsi retenus.

D. Si le premier des quatre phonèmes retenus est de même valeur phonétique que la référence corpus analysée, alors une opération dite d'incrémentation est effectuée sur les références qui ont contribué à dégager le premier phonème du scrutin majoritaire. Par exemple, on doit reconnaître le phonème /u/ n°1325, le treillis se présente comme suit:

Phonèmes : /u/ /y/ /u/ /u/ /o/ /y/
N° de ref: 270 323 719 227 520 250

Le résultat du scrutin donne: /u/ n°1325 reconnu 3 fois par /u/, 2 fois par /y/, etc.

Remarque 1.

Le /u/ a été reconnu 3 fois par des /u/ à l'aide de références validées du corpus et qui lui sont antérieures. Cependant des références /y/ par exemple peuvent apparaître par la suite et se révéler plus proches du /u/ n°1325 que les références ayant permis de dégager un scrutin majoritaire favorable. On a donc été contraint de faire un deuxième passage de VARAP afin de pallier à cet inconvénient.

Remarque 2.

Dans VARAP, chaque treillis comprend, outre la suite des références phonétiques, le contexte gauche et droit de chaque référence du treillis, la distance entre le candidat à reconnaître et les références du treillis, les compteurs d'incrémentation des références du treillis indiquant le nombre de fois où les dites références ont participé à une bonne reconnaissance en scrutin majoritaire.

E. Si le premier des quatre phonèmes retenus est de valeur phonétique différente de la référence CORPUS analysée alors:

- ou bien la référence CORPUS à reconnaître ne se trouve pas dans le fichier des références validées, et alors elle est introduite dans le fichier VALID. Cela pourra se produire aussi bien pendant le premier ou pendant le second passage de VARAP.

- ou bien la référence CORPUS qu'il faut reconnaître se trouve déjà dans le fichier VALID, alors il y a confusion. Ce cas ne se produira que dans le second passage de VARAP. On effectue alors un nouveau scrutin majoritaire par souci de cohérence avec les incréments. Si après ce dernier scrutin, la référence est reconnue par un phonème de même valeur phonétique, la validation est dite incrémentale, sinon il y a validation simple.

III.2. Les algorithmes de dépouillement du système VARAP.

Ils comprennent:

A. Un tableau donnant la compression du corpus des références phonétiques, phonème par phonème, après les deux passages de VARAP, ainsi que le nombre total d'incrémentations, de validations, de confusions. Exemple:

Phonème	/i/	/s/
avant	357	400
après	139	163
incrémentations	274	310
validation	32	17
confusion	51	73

B. Un tableau donnant la relation entre la confusion et l'incrémentations par ordre décroissant de cette dernière. Il montre en particulier qu'à partir d'une incrémentations égale à dix, les confusions deviennent très faibles: 12 confusions en tout pour l'ensemble du corpus, et à partir de 20 incréments, il y a rarement plus de une confusion.

C. Une matrice de confusion a été construite automatiquement et nous la commentons brièvement dans la conclusion.

D. Un fichier inverse a été construit automatiquement, permettant de savoir combien de fois chaque référence a participé à une bonne et à une mauvaise reconnaissance.

IV. BILAN GENERAL DU SYSTEME VARAP ET CONCLUSION.

Les expériences de validation par autocohérence ont été faites sur 100, 300, 500 phrases. L'augmentation de la taille du corpus apporte une amélioration régulière. Nous donnons les résultats sur 500 phrases. Après le premier passage de VARAP, sur 9952 références du corpus, il y a eut 4810 incréments et 4002 validations. Après le second passage de VARAP, le nombre de références conservées est de 4375, il y a 5973 incréments. 644 nouvelles validations et 2195 confusions.

La matrice de confusions montre qu'en ce qui concerne les voyelles, si l'on écarte les confusions du type /e/, /ɛ/ ou encore /ə/, /ø/ ou encore /ɔ/, /o/ ou encore /a/, /ã/. alors les confusions sont peu importantes. Pour les consonnes, la matrice fait apparaître des confusions importantes entre /p/, /t/, /k/, d'une part et /b/, /v/, /d/ d'autre part. Cela tient en partie au mode d'obtention des références ternaires qui sont obtenues en regroupant les spectres les plus proches les uns des autres, c'est à dire en conservant la partie du minimum la plus stable au sens de la distance utilisée. Mais dans ce cas, les spectres de la référence sont regroupés autour du silence de l'occlusive. Pour pallier à cet inconvénient, un traitement complémentaire est prévu, permettant dans le cas des occlusives sourdes et sonores, de conserver les références ternaires en remontant vers les transitions, après ou avant le minimum.

Pour terminer, disons que le but du système VARAP est multiple:

- Vérifier la qualité de l'étiquetage automatique utilisé.

- Construire un système de sélection qui ne fasse pas appel à des seuils et qui teste l'aptitude à reconnaître des références étiquetées du corpus.

- Avoir un outil modulaire complet qui, entre la procédure et la constitution du dictionnaire de références, permet de paramétrer toutes les décisions intermédiaires. Entre autre l'adaptation au locuteur, l'utilisation de plusieurs types de distances différentes, enfin la conservation ou la mise à l'écart des références de l'apprentissage en fonction de leur aptitude à reconnaître.

BIBLIOGRAPHIE.

ANDREWSKY A. DESI M. POIRIER F. "Le Système SHERPA - de l'étiquetage automatique à la reconnaissance par analyse ternaire", 5ème Congrès RFIA, 1985.

DESI M. POIRIER F. "Le système SHERPA: étiquetage et classification automatique par apprentissage pour le décodage automatique de la parole continue", Thèse en Sciences, Paris-Sud, Orsay, 1985.

DABOUZ M. "Transmission de la parole à faible débit par vocodeur à classification", Thèse de l'ENST, janvier 1984.

DILL A.M. "Extensions de la méthode des plus proches voisins". Thèse de l'Université Claude Bernard, Lyon, 1978.

MAC QUEEN J. "Some methods for classification and analysis of multivariate observations", The 5th Berkeley Symposium of Mathematical Statistic and Probability, 1967.

MICLET L. VICARD D. "Reconnaissance des parties stables de parole continue pour le décodage acoustico-phonétique", 15ème JEP, Aix en Provence, 27-30 mai 1986.