HIERARCHICAL RECOGNITION OF FRENCH VOWELS BY EXPERT SYSTEM IROISE-SERAC

Anne Bonneau, Mario Rossi

Institut de Phonétique, Aix-en-Provence (France)

Guy Mercier

CNET, Lannion (France)

ABSTRACT

We are presenting here an implementation of a French vowel recognition program under IROISE, an expert system for acoustic-phonetic decoding, used in CNET. The rules for recognition are based on polycontextual non-formantic cues ; the data are output from a 14-channel vocoder. The algorithm is represented by a binary tree with 37 hierarchized cues.

A rule under IROISE represents a branch of the tree. The first one follows the branch defined only by positive cues ; the second one puts the list of the first rule in its contextual part by eliminating the last cue. If the rule is applied we know that this cue is negative, because the preceeding rule was not set off, and we modify the cue's polarity. With this method, only the cues tested in the recognition phase will have the value "false".

Under IROISE, all cues are systematically tested even if they are not all used in any particular execution of the program. Then we call the algorithm in which every rule represents a branch.

We furnish the recognition results using this program on an initial corpus of 330 words pronounced by five male speakers and the results using rules under IROISE on digits pronounced by other speakers.

1. PRESENTATION DE SERAC-IROISE

SERAC est le module de reconnaissance acoustico-phonétique utilisant le langage du système-expert IROISE.

SERAC réécrit, en utilisant au mieux les possibilités de formalisation des connaissances offertes par IROISE, le module de reconnaissance acoustico-phonétique du système de reconnaissance de la parole KEAL, actuellement implanté en langage C sous UTS (IBM 3083).

En outre, il le complète en lui adjoignant de nouveaux modules de reconnaissance écrits par d'autres experts.

Le module de reconnaissance phonétique "KEAL-SERAC" commence par lire les échantillons spectraux ; les données acoustiques sont fournies par les analyses spectrales numériques effectuées toutes les 13,3 ms, par un vocodeur à 14 canaux ; il en extrait les paramètres acoustiques. Le paramètre le plus important, pour les programmes que nous implantons, est le vecteur des énergies, appelé "en", qui est constitué de la valeur de l'énergie dans chacun des 14 canaux du vocodeur pour un échantillon temporel donné.

Le module actuel de reconnaissance effectue l'étiquetage phonétique des échantillons, la segmentation en syllabes et en noyaux vocaliques, la reconnaissance des macro-classes... Nous intervenons après le module de détection des noyaux vocaliques.

2. RECONNAISSANCE DES VOYELLES

2.1. Le programme de reconnaissance des voyelles

La recherche des indices a été menée de deux façons : (1) par la méthode d'essai-erreur, (2) au moyen d'une analyse factorielle des correspondances. Le corpus comprenait 300 mots de forme CVCV enregistrés deux fois par 5 locuteurs masculins d'une moyenne d'âge de 25 ans.

La reconnaissance s'effectue selon un mode binaire dans une arborescence où tous les indices sont hiérarchisés : au total, 37 indices représentant essentiellement les traits Ouvert/Fermé, Aigu/Grave, Bémolisé/Non bémolisé, Nasal/Non nasal, Périphérique/Non périphérique. Ils sont fondés sur les variations d'énergie dans le spectre et calculés soit sur la partie centrale de la voyelle, soit sur plusieurs parties de celles-ci, dans le cas des voyelles nasales par exemple, caractérisées par la présence de deux segments distincts, un segment oral et un segmen nasal localisé dans les deux derniers tiers de la voyelle. En général, les voyelles sont tronquées de 20 % aux bornes afin d'éviter de prendre en compte les parties transitoires et de ne conserver que la partie stable de la voyelle.

Les indices portent, par pure commodité, l'étiquette d'un trait, mais il n'existe pas de relation biunivoque entre indices et traits. Les voyelles ne sont pas reconnues par traits mais par configurations d'indices.

2.2. Détermination des indices dans SERAC

Nous avons créé un nouvel objet, "phone-voy", qui représente la voyelle.

Chaque indice est un attribut de "phone-voy" et prend la valeur "vrai", "faux" ("inconnu" au lancement du programme). L'indice est détecté sur une partie déterminée de la voyelle ; les limites temporelles sont également des attributs de "phone-voy".

On teste si l'indice, ou plutôt son attribut, est "vrai" par un problème particulier qui porte le nom de l'indice testé. Nous avons regroupé dans un même fichier tous les problèmes qui testent la valeur d'indices d'un même trait.

Cinq fichiers sont créés pour : (1) les indices d'acuité, (2) les indices de bémolisation, (3) les indices d'ouverture, (4) les indices de nasalité, (5) les indices périphériques.

Dans la plupart des cas les indices sont détectés par des règles simples. Le langage Lisp peut coexister avec IROISE pour les règles plus complexes qui exigent en particulier des itérations.

2.3. Algorithme de reconnaissance des voyelles

Après l'évaluation de "n10", dernier indice testé, l'ensemble des attributs des indices positifs possèdent la valeur "vrai". Les autres sont implicitement "faux". On peut alors déclencher l'algorithme de reconnaissance.

Chaque règle de celui-ci correspond à un des chemins de l'arbre défini dans le par. 2.1. La première règle représente le chemin qui est défini uniquement par des indices positifs : la partie "contexte" de la règle est écrite de la manière suivante :

si (phone-voy ?pv (indic1 vrai) (indic2 vrai)...)

La deuxième règle reprend dans sa partie "contexte" la liste de la première en éliminant le dernier indice de celle-ci. On sait alors, si la règle est appliquée, que cet indice est négatif puisque la règle précédente n'a pas été déclenchée - les règles sont incompatibles entre elles - et on modifie la valeur de l'indice. Par cette méthode, seuls les indices effectivement testés lors de la reconnaissance d'une voyelle déterminée auront la valeur "faux", les autres resteront à "inconnu".

On réitère le processus jusqu'à ce que l'algorithme soit tout entier écrit sous forme de règles. On peut, de cette façon, créer un seul problème qui traite l'ensemble de l'algorithme.

2.4. Les indices fondamentaux d'antériorité (acuité) et d'ouverture (compacité)

Le premier axe de l'analyse factorielle des correspondances représente le trait Aigu/Grave, l'étude des corrélations entre les groupes de voyelles et les canaux permet de diviser le spectre en deux bandes fréquentielles : 650 à 1600 Hz, 1600 à 3400 Hz, à partir desquelles est calculé le principal indice d'acuité appelé AIGU1. L'indice recherche et compare les maxima spectraux dans chacune des bandes ainsi délimitées. En effet, l'énergie caractéristique des voyelles graves [u,o,o] est située dans la bande 650-1600 Hz tandis que celle des voyelles aiguës est située au-delà de cette bande. Il semble que l'énergie caractéristique de [a] soit située à la frontière mais cette voyelle peut apparaître dans l'une ou l'autre classe en fonction du contexte. [a] se comporte comme les voyelles vélaires sauf au contact des consonnes vélo-palatales [k-g] derrière lesquelles apparaît sa variante aiguë. La reconnaissance de la voyelle [a] dans la classe des /+ouvertes, +aiguës/ est accompagnée de la spécification du contexte vélo-palatal qui se trouve vérifiée dans 90 % des cas.

Le trait d'ouverture n'est pas représenté par le deuxième axe, plus complexe, mais l'examen du spectre des voyelles fermées met en évidence l'existence d'une zone d'anti-résonance dans les canaux 3 à 4 (650-1050 Hz) qui disparaît au fur et à mesure que F1 s'élève et que la voyelle devient plus ouverte. D'où l'indice d'ouverture "ouv1" :

si $EK1_2$ EK3 + EK4 alors -ouv1

Ki : ième canal. EKi : énergie dans le ième canal.

Les seules voyelles dont la classification pose un problème par cet indice sont [ ] et [u]. Pour [u], cet échec s'explique par la présence du F2 dans les canaux 3 et 4 qui réduit l'importance relative de F1.

Un second indice est alors proposé pour forcer [u] dans la classe des voyelles fermées. Cet indice, appelé "ouv4" permet de mettre en relief la proéminence du F1 :

si (EK1 - EK2) (EK3 + EK4) - EK1 alors -ouv4

Pour certains locuteurs, la voyelle [ ] est produite comme une voyelle fermée, en conséquence sa reconnaissance est prévue dans les deux classes vocaliques, ouvertes et fermées.

Le taux de reconnaissance des voyelles par l'ensemble des indices sur le corpus défini dans le parag. 2.1 est évalué à 86 % : 1 candidat est présenté dans 60 % des cas, 2 candidats dans 40 % des cas.

2.5. Reconnaissance des macroclasses

Le module de reconnaissance des traits vocaliques pour la reconnaissance des consonnes occlusives (A. Bonneau 1984), vise à regrouper les voyelles en quatre grandes classes vocaliques, selon les traits "ouvert-fermé" et "aigu-grave" :

a) Voyelles aiguës : /i,e, ,y/
   Voyelles graves : /u,o, , , , /
   [a, , oe, ø] peuvent, selon le contexte dans lequel ils apparaissent, appartenir à l'une ou l'autre classe.

b) Voyelles fermées : /i,u,y,(e),(ø),(o)/
   Voyelles ouvertes : / , , ,a, ,( ),( ),(oe)
   Il est difficile de déterminer a priori si, dans une syllabe donnée, en particulier les syllabes atones, le locuteur a prononcé la variante ouverte ou fermée de [e,] ; [o,] ; [ø,oe], c'est pourquoi ces phonèmes, mis entre parenthèses ci-dessus, ne sont pas pris en compte dans les % de reconnaissance selon l'indice d'ouverture.

3. RESULTATS ET CONCLUSIONS

SERAC-IROISE est écrit en Lisp dans le dialecte COMMON LISP, sur VAX 11/780, sous VMS. Le corpus choisi pour l'évaluation est constitué de 139 nombres (de 0 à 999) prononcés par 6 nouveaux locuteurs masculins. L'application des règles de reconnaissance à ces nouveaux locuteurs permet de tester dans quelle mesure les indices utilisés sont indépendants du locuteur :

Le pourcentage de reconnaissance pour les voyelles, pour ce nouveau corpus s'établit comme suit :

* 72 % de reconnaissance pour les nombres prononcés en parole continue. Le fort pourcentage des voyelles nasales [ ] dans les nombres est responsable de la chute du taux de reconnaissance ; cette voyelle, en effet, est la moins bien identifiée du système français. Sans la présence de cette voyelle, le taux de reconnaissance s'élèverait à 86 %. Une ou deux réponses (deux dans 37 % des cas) sont données pour l'identification de la voyelle à reconnaître. Le pourcentage de reconnaissance pour les traits aigu-grave et ouvert-fermé est de 97 %.

REFERENCES

1. Rossi, M., Nishinuma, Y., Mercier, G. (1983), "Indices acoustiques multilocuteurs et indépendants du contexte pour la reconnaissance automatique de la parole", Speech communication, North-Holland, pp. 215-217.

2. Bonneau, A. (1984), "Indices de reconnaissance des consonnes occlusives sourdes du français, en vue d'une application à la reconnaissance automatique de la parole", Thèse de doctorat de troisème cyle, Aix-en-Provence.

3. Mercier, M., Gilloux, M., Tarridec, C., Vaissière, J. (1984), "From Keal to Serac : a new rule-based expert system for speech recognition", Nato Advance Studies Institute, Bonas, Gers, France.