

UN SYSTEME DE TRAITEMENT DE CONNAISSANCES POUR LE DECODAGE ACOUSTICO-PHONETIQUE

H. Meloni, R. Bulot

G.I.A., Faculté de Luminy, 70 routé Léon Lachamp 13288 Marseille Cedex 09 France

I - INTRODUCTION

Les travaux que nous avons accomplis dans le cadre de la reconnaissance de la parole (Meloni 1982, 1984, 1985) séparaient assez nettement les traitements numériques, exécutés dans un langage algorithmique classique, des traitements de données symboliques effectués en PROLOG. Cette dichotomie artificielle interdisait l'interaction et l'optimisation contextuelles des deux processus. Afin de permettre une coopération simple entre toutes les sources de connaissances, nous proposons, sous la forme d'un ensemble de prédicats du langage PROLOG II (Colmérauer 1984), un environnement souple et efficace pour l'acquisition, la manipulation, l'évaluation, la représentation et le traitement d'informations acoustiques, phonétiques et linguistiques. Les outils développés permettent, de manière interactive, de produire diverses paramétrisations du signal, de décrire et reconnaître des formes simples, de définir et identifier des événements, des propriétés, des indices et des traits acoustico-phonétiques, de coder ces informations et les stratégies qui les utilisent et de structurer l'ensemble des résultats produits sous la forme d'un treillis d'unités valuées. Nous illustrons les possibilités nouvelles de cet environnement en présentant quelques particularités d'un système de décodage acoustico-phonétique réalisé entièrement sous PROLOG II.

II - PARAMETRISATION DU SIGNAL

Le but visé à ce stade du traitement est de caractériser de manière précise et peu coûteuse une portion de signal au moyen d'une suite de vecteurs de paramètres. Les limites de la zone codée, la nature des attributs retenus ainsi que leurs conditions d'évaluation sont déterminées en examinant des connaissances de niveau acoustique, phonétique ou phonologique.

II - 1 - Prédicats évaluables pour la paramétrisation

Nous disposons de 2 prédicats évaluables qui effectuent le calcul des paramètres et leur chargement dans une mémoire accessible à d'autres fonctions réalisant des opérations numériques complexes. Chaque vecteur, produit à intervalles réguliers de 10 ms, est constitué d'une vingtaine d'attributs temporels et spectraux (répartition spectrale de l'énergie, densité des passages par zéro, position, amplitude, émergence et largeur des pics, etc.). Les spectres lissés sont obtenus à partir des coefficients cepstraux ou de LPC dont un ensemble de variables définit les conditions de calcul (portion de signal traitée, méthode utilisée, nombre de coefficients, préemphasis, rayon, etc.).

II - 2 - Utilisations des prédicats de paramétrisation

Les prédicats de paramétrisation du signal ont été employés, dans la phase d'acquisition des connaissances acoustiques, pour évaluer les conditions optimales du codage des sons correspondant aux diverses phases de phonèmes segmentés semi-automatiquement dans un ensemble de 130 phrases prononcées par 2 locuteurs.

La stratégie du système de décodage conduit à une paramétrisation globale d'un énoncé au moyen des 14 pre-

miers coefficients de LPC, mais ces attributs sont localement recalculés lorsque certaines règles de niveau quelconque exigent d'autres conditions d'évaluation. C'est le cas notamment pour l'identification des traits des voyelles nasales, le traitement des explosions d'occlusives ou de portions sourdes du signal.

III - RECONNAISSANCE DES FORMES

Les outils proposés dans ce cadre ont pour objectif la modélisation et la symbolisation des évolutions temporelles de certains groupes de paramètres.

III - 1 - Prédicats évaluables de reconnaissance de formes

Sur un intervalle de temps, ils définissent des fonctions simples d'un paramètre telles que la mesure de ses extrema, le calcul de sa moyenne, la caractérisation de son instabilité. Ils déterminent également des fonctions complexes d'un ensemble d'attributs comme la recherche de formants, la valuation de la continuité ou de la monotonie d'un phénomène. Enfin, ils désignent et identifient des schémas de formes parmi lesquels les modèles de collines ou de vallées sont les plus utilisés. Des variables permettent de préciser les contours d'une forme ; ainsi, la définition d'un type de colline, pour un paramètre quelconque, sera donnée par sa largeur et son minimale, son émergence limite à gauche et à droite, le seuil maximum de déviation acceptable ainsi que le seuil de bruit au dessous duquel le paramètre n'est pas significatif.

III - 2 - Utilisations des prédicats de R.F.

A partir de schémas de formes simples, appliqués pour l'ensemble des paramètres sur les phrases de référence, nous avons sélectionné quelques dizaines de formes représentatives de portions de sons déterminées. Ces éléments (collines, vallées, zones monotones ou stables, etc.) concernent, souvent de plusieurs manières, la plupart des attributs temporellement continus.

Dans le système de décodage, les formes retenues définissent des événements acoustiques et phonétiques et constituent des repères pour guider le processus de reconnaissance vers les phénomènes les plus saillants.

IV - NIVEAUX SYMBOLIQUES DU SYSTEME

Les connaissances acoustico-phonétiques formelles d'un même type sont regroupées en niveaux pour leur présentation, mais chaque règle peut être sollicitée indépendamment de sa classe.

IV - 1 - Prédicats prédéfinis de l'environnement

Ces outils contribuent à rendre plus naturelle l'expression de la connaissance et définissent pour l'essentiel les fonctions suivantes :

- relations temporelles entre des unités du treillis de résultats (coïncidence, intersection, succession, union, adjacence, etc.),
- démonstrations particulières d'une liste de prédicats pour la gestion du contrôle et la visualisation des parcours (effacement déterministe ou complet, vérification de l'existence d'une ou de plusieurs solutions, saturation des effacements, impression de traces, etc.),
- opérations logiques sur des listes de prédicats (conjonction, disjonction, négation, implication, etc.),
- opérations arithmétiques diverses acceptant des fonctions en paramètres,
- manipulations complexes sur les arbres.

IV - 2 - Événements acoustiques et phonétiques

Les événements acoustiques sont définis par regroupement de formes au moyen des prédicats qui décrivent des relations temporelles entre les éléments de base. Les unités engendrées ne reçoivent pas d'interprétation phonétique ; elles mettent en évidence la conjonction de propriétés acoustiques du signal et caractérisent généralement des segments infra-phonémiques.

Les événements phonétiques, identifiés à partir des événements acoustiques, des formes et des relations, constituent des unités que l'on peut associer directement à des phases spécifiques de phonèmes et de transitions (constriction, occlusion, explosion, etc.) ou à des regroupements de segments acoustiquement proches. Des règles contextuelles réunissent ensuite ces éléments pour désigner les limites des phonèmes ou décomposent certains d'entre eux à partir de critères plus fins pour séparer certaines voyelles des consonnes vocaliques qui les entourent.

Les quelques dizaines de clauses qui définissent ces connaissances opèrent dans des contextes souvent très différents suivant qu'il s'agisse d'événements "évidents" ou de segments tributaires de l'identification préalable de l'environnement. Ces règles sont indépendantes du locuteur, elles opèrent une partition peu ambiguë d'un énoncé en macro-classes pseudo-phonétiques. L'exemple ci-dessous décrit et évalue un type particulier d'événement vocalique :

```
evenement-voc(<voc(5),z>) ->
forme(<colline1-er0,z>)
inferieur(5,longueur(z))
voise(z)
ou (coincidence-sur(z,colline1-ebf) ,
coincidence-sur(z,colline1-ap1) ) ;
```

IV - 3 - Traits pseudo-phonétiques

Chaque événement pseudo-phonétique est caractérisé par un faisceau de traits hiérarchisés dont chacun est défini par un ensemble de clauses. Les règles qui représentent ces connaissances utilisent de nombreuses informations contextuelles sur la nature, les paramètres, les propriétés, les indices ou les traits des sons adjacents. Cette étape de la reconnaissance fonctionne comme un filtre phonétique limitant le nombre des phonèmes candidats. Le choix des solutions les plus vraisemblables résulte de l'évaluation d'un score à partir de paramètres sélectionnés et ajustés en fonction de caractéristiques des segments contigus.

L'acquisition et l'évaluation de ces règles sont effectuées de manière interactive sur les phrases de référence. Une étude statistique des paramètres ou de certaines fonctions de plusieurs d'entre eux permet de désigner les attributs les plus discriminants pour la détermination d'un trait d'un phonème dans un environnement précis. Les règles qui en résultent sont immédiatement testées sur l'ensemble des situations où elles sont susceptibles de s'appliquer. La clause suivante décrit et évalue un indice du trait *grave* pour les occlusives sourdes :

```
acuite-occ-sourde(z,grave(2)) ->
inferieur(cgh(z),3200)
inferieur(afmedian(z),ebf(z))
inferieur(afhaul(z),ebf(z))
si-alors(inferieur(moins(fois(2,ebf(z)),10),
plus(afmedian(z),afhaul(z))),
inferieur(afhaul(z),afmedian(z)))
si-alors(inferieur(300,fbas(z)),
inferieur(afbas(z),plus(6,ebf(z)))) ;
```

V - TREILLIS DES RESULTATS

Certains résultats, considérés comme définitivement acquis au cours du décodage d'un énoncé, sont conservés dans un treillis constitué de clauses PROLOG. Cette structure est rendue souple et efficace au moyen d'un ensemble de prédicats qui permet de réaliser des opérations telles que l'ajout et la suppression d'unités en un point quelconque, des parcours multiples, le repérage des zones libres, des accès diversifiés à une unité (par la position de ses bornes, par son type ou ses caractéristiques, etc.), la récupération des arguments et des limites d'un élément, etc.

Dans la phase de décodage acoustico-phonétique, la stratégie conduit à exécuter non séquentiellement les étapes suivantes :

- calcul des paramètres sur l'ensemble de l'énoncé,
- reconnaissances des formes constituant les noyaux nécessaires à la définition d'événements sûrs,
- identification et mémorisation dans le treillis des événements acoustiques évidents,
- recherche dans les zones libres des événements secondaires et transitoires,
- regroupement des segments étiquetés pour produire et ajouter des événements consonantiques,
- affinement de la segmentation des noyaux vocaliques pour déterminer, quelquefois de manière ambiguë, des événements vocaliques qui vont enrichir le treillis,
- identification des traits pseudo-phonétiques puis ajout des phonèmes les plus vraisemblables après filtrage et calcul du score,
- interprétation des zones non reconnues au moyen de l'ensemble des unités du treillis.

VI - CONCLUSION

La brièveté de l'exposé ne donne qu'une image imparfaite de la puissance potentielle de l'environnement proposé. Son utilisation pour les tâches d'apprentissage et d'acquisition des connaissances acoustico-phonétiques nous a permis de constituer très rapidement un important ensemble de règles dont certaines demeurent perfectibles. Le traitement de ces connaissances fournit des résultats bien supérieurs par certains aspects à ceux que nous obtenions au moyen de techniques classiques. La durée du processus de décodage demeure tout à fait raisonnable sur un mini-calculateur, et nous pouvons envisager de réaliser sous PROLOG un système complet de reconnaissance automatique de la parole.

BIBLIOGRAPHIE

- Colmérauer A., Kanoui H., Van Caneghem M.
PROLOG : Bases théoriques et Développements actuels ; TSI, vol. 2, n° 4, juin-juillet 1983, pp 271-311
- Méloni H.
Etude et réalisation d'un système de reconnaissance automatique de la parole continue ; Thèse de Doctorat d'Etat, Université d'Aix-Marseille II, Faculté de Luminy, février 1982
- Méloni H.
Traitement des Contraintes Linguistiques en Reconnaissance de la Parole ; TSI, vol. 2, n° 5, septembre-octobre 1983, pp 349-363
- Méloni H., Gispert J., Guizol J.
Traitement de connaissances pour l'identification analytique de Mots dans le discours continu ; congrès AF'CEI Informatique 5^e Génération, Paris 5-7 mars 1985, pp 339-350