

RECONNAISSANCE ANALYTIQUE MULTILOCUTEURS DE MOTS ISOLES*

L.C. SAUTER, D. GROSSETETE-FOURNOL

LABORATOIRES DE MARCOUSSIS
Centre de Recherches de la CGE
Route de Nozay
91460 MARCOUSSIS, FRANCE

Les systèmes de reconnaissance multilocuteurs de mots isolés utilisent en général plusieurs exemples de chaque mot du vocabulaire prononcés par un nombre assez grand de locuteurs différents. L'apprentissage d'un vocabulaire représente alors un travail considérable. Pour cette raison, nous étudions un système de reconnaissance analytique de mots isolés indépendant du locuteur. Chaque mot du vocabulaire pourra être représenté par un treillis de segments de référence et la reconnaissance d'un mot sera le choix du treillis minimalisant une certaine distance. Chaque segment sera représenté par un certain nombre d'empreintes extraites d'élocutions de plusieurs locuteurs. Différentes approches sont possibles pour chercher le mot le plus proche, selon qu'on effectue ou non une segmentation préalable, selon le caractère ascendant ou descendant de l'algorithme de recherche de l'optimum.

1. Sans segmentation

- a) Reconstitution d'empreintes globales : les différentes variantes de chaque segment peuvent être concatenées pour générer un grand nombre de variantes de chaque mot. Il faudra, selon une technique à déterminer (classification automatique), choisir un nombre convenable de références pour chaque mot. L'algorithme de reconnaissance sera alors identique à celui de l'approche globale. Dans [1], Rosenberg décrit une version monolocuteur d'un tel système qui donne de bons résultats.

*Cette étude a été financée en partie par un contrat ESPRIT.

b) Une autre approche pourrait utiliser un algorithme de reconnaissance de mots enchaînés, en remplaçant les mots par des segments, et en utilisant le lexique en guise de syntaxe. Des exemples de reconnaissance de mots enchaînés sont donnés dans [2, 3, 4]. Seul [3] aborde le problème de références multiples pour chaque mot.

2. Avec segmentation

Le mot inconnu est segmenté avant la reconnaissance. La segmentation devant être fiable, nous avons choisi de préférence une détection de noyaux vocaliques (segmentation en syllabes). Chaque segment peut ensuite être identifié soit globalement (références globales ou obtenues en concaténant des segments plus petits) soit analytiquement (chercher la meilleure séquence de sous-segments correspondant à un segment valide).

3. Apprentissage

Les segments de base seront des demi-syllabes ainsi que quelques phonèmes en position finale. Ces sous-segments doivent être choisis avec l'idée de pouvoir reconstituer n'importe quel mot du vocabulaire. Un programme interactif a permis de commencer la création d'un premier corpus (20 locuteurs, 5 élocutions par locuteur, 100 segments).

Ce premier corpus doit nous permettre d'évaluer les différentes approches décrites ci-dessus. Les performances sur un vocabulaire de 35 mots seront comparées à celles obtenues par l'approche globale (des résultats seront disponibles pour les journées d'étude). Ce corpus sera progressivement complété (100 locuteurs, un millier de segments).

[1] E. Rosenberg, L.A. Rabiner, J.G. Wilpon et D. Kahn, "Demisyllable-Based Isolated Word Recognition System", IEEE Trans. on Acoust. Speech and Sign. Proc. vol ASSP-31, n°3, juin 1983.

[2] L. Sauter, "RAPACE, un système de reconnaissance analytique de la parole continue", 4ème congrès Reconnaissance des Formes et Intelligence Artificielle, Paris, janvier 1984.

[3] C.S. Myers et S.E. Levinson, "Speaker Independent Connected Word Recognition Using a Syntax-Directed Dynamic Programming Procedure", IEEE Trans. on Acoust. Speech and Sign. Proc., vol ASSP-30 n°4, août 1982.

[4] J.S. Bridle, M.D. Brown et R.M. Chamberlain, "An Algorithm for Connected Word Recognition", JSRU Research Report n°1010, oct. 1981.