

UN COPROCESSEUR DE PROGRAMMATION DYNAMIQUE POUR LA RECONNAISSANCE DE LA PAROLE CONTINUE

G. Quénot, A. Reichart, J.L. Gauvain, J.J. Gangolf
LIMSI-CNRS - B.P. 30 - 91406 ORSAY CEDEX - France

J.M. Frailong
Ecole Normale Supérieure - 45, rue d'Ulm - 75005 PARIS - France

Les algorithmes de programmation dynamique utilisés pour la reconnaissance de la parole fournissent d'excellentes solutions aux problèmes d'alignement temporel et de segmentation. Cependant, ces algorithmes nécessitent une quantité de calculs, proportionnelle au carré du nombre d'évènements retenus pour représenter les entités à reconnaître, qui ne peut être aisément supportée en temps réel par un microprocesseur standard. Un co-processeur de programmation dynamique capable de supporter la plupart des algorithmes de reconnaissance a été développé au LIMSI.

Cette machine a été spécialement étudiée pour traiter rapidement la partie coûteuse en temps des calculs de comparaison dynamique, c'est à dire : les calculs de distances entre deux spectres et les calculs d'équations locales. Sa structure très générale lui permet d'enchaîner automatiquement les calculs d'équations locales pour une référence ou pour une fenêtre dans une référence, de passer automatiquement à la référence suivante et d'effectuer un prétraitement de haut niveau ne transmettant au processeur maître que les résultats de la reconnaissance.

Le processeur maître n'a qu'à charger les références dans son co-processeur, lui envoyer les spectres de la phrase à reconnaître au fur et à mesure qu'ils arrivent et à traiter les résultats que lui renvoie celui-ci.

Le co-processeur gère pour chaque référence une matrice phrase à reconnaître-référence dans laquelle il effectue les calculs de programmation dynamique en progressant d'une colonne à chaque arrivée d'un spectre de la phrase test. Cette solution favorise l'aspect temps réel.

Ce co-processeur comprend une unité de calcul générale, une unité spécialisée pour le calcul de distance, une autre pour les calculs d'adresse, une mémoire de données de 64 K mots de 16 bits, une mémoire de programme de 1 K mot de 32 bits, une unité de contrôle qui gère l'ensemble et une interface multibus. Toutes ces unités fonctionnent en parallèle, le temps de cycle est de 200 ns.

L'équation locale et la structure des données sont programmables, il faut de 5 à 6 μ s pour effectuer à la fois un calcul de distance avec 16 paramètres et un calcul d'équation locale, ce qui permet de faire de la reconnaissance de mots enchaînés avec un temps de réponse quasi-nul sur des vocabulaires de 150 à 200 références acoustiques, ce qui correspond à la capacité actuelle de la mémoire.