EXPERIMENTE DE RECUNOAŞTERE A CIFRELOR ROSTITE IZOLAT

Mihai SIMA* Victor CROITORU*

This paper describes several experiments of speaker independent isolated word recognition in Romanian. The structure of a system for isolated digit recognition in the framework of voice-dialing applications is presented. The necessity of a start/end point detection is then discussed. The recognition rates are then presented along with details on the experiments. The paper ends with several conclusions.

1. INTRODUCERE

Lucrarea îşi propune să prezinte o serie de experimente realizate de autori în domeniul recunoaşterii cifrelor rostite izolat, în cadrul mai larg al sistemelor de selecţie a numerotării cu comandă vocală. Prin urmare, recunoaşterea se realizează independent de vorbitor, rostirile candidat constituindu-se într-o înregistrare continuă. Performanţa urmărită a fost obținerea unui compromis între procentul de recunoaştere şi timpul de prelucrare.

Schema bloc a unui astfel de sistem este prezentată în figura 1. Se remarcă un bloc de extragere de trăsături, unul de detecţie a marginilor de cuvânt destinat delimitării rostirilor candidat în cadrul înregistrării continue, blocul de recunoaştere propriu-zisă a cuvintelor izolate şi partea de generare a comenzilor de selecţie.

* Universitatea «Politehnica» din Bucureşti, Facultatea de Electronică şi Telecomunicaţii,
tel: (01) 410-64-45, fax: (01) 411-11-87, e-mail: {sima,croitoru}@ADComm.pub.ro

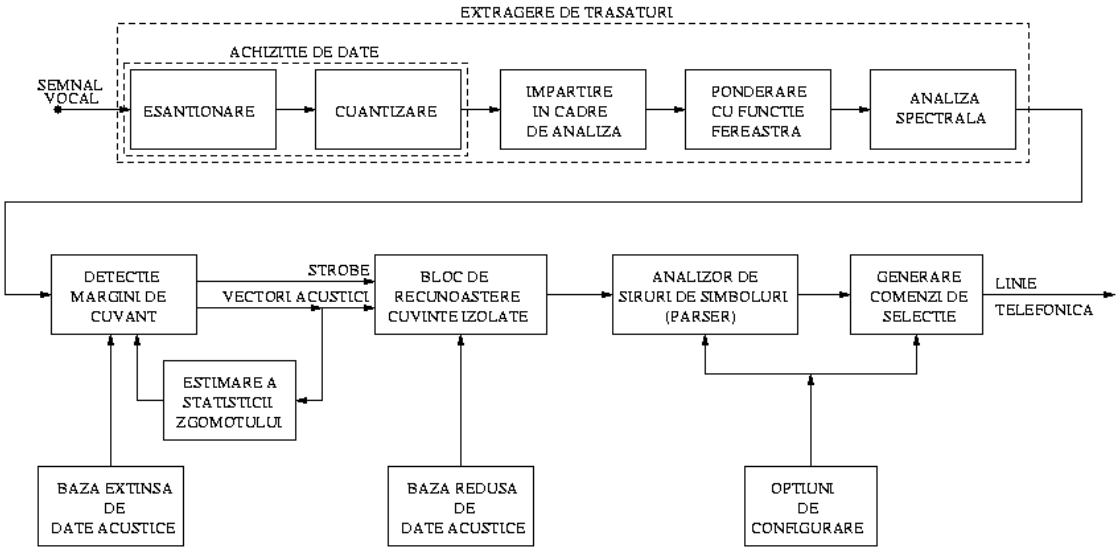

Figura 1 - Schema bloc a unui sistem de selecţie comandat prin voce

2. PROBLEMA DETECŢIEI MARGINILOR DE CUVÂNT

O bună localizare a regiunilor ce conţin vorbire, nu numai că reduce timpul de calcul, ci şi măreşte procentul de recunoaştere. Să presupunem că se compară printr-un procedeu de aliniere temporală dinamică (DTW) [9] două segmente acustice, fiecare conţinând aproximativ două treimi de „linişte” (fig. 2). Regiunile marcate drept zgomot sunt susceptibile să producă erori, deoarece rostirile pot fi considerate drept egale de-a lungul acestor regiuni. Aceasta este una din dificultăţile majore în decelarea cifrelor ce prezintă rostiri de lungime scurtă („2” şi „8”) [6][8].

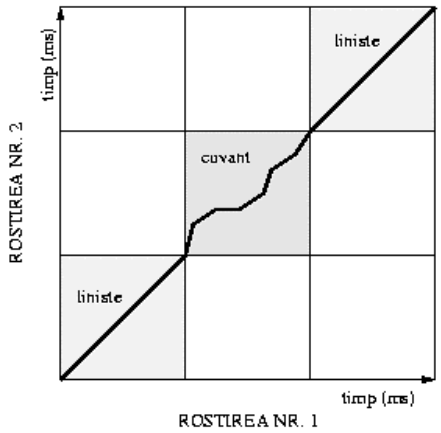

Figura 2 – Alinierea temporală dinamică a două rostiri ce conţin largi regiuni de zgomot

Un algoritm de detecţie a marginilor de cuvânt, care se bazează pe două clase de parametri, una cu semnificaţie de energie şi alta de frecvenţă, este descris în [8]. Acesta extrage cuvintele rostite izolat dintr-o înregistrare continuă, dacă distanţa dintre cuvinte nu este mai mică decât 300 ms.

3. BAZĂ DE DATE ACUSTICĂ CU VORBITORI MULTIPLI

Aşa cum s-a precizat, una din modalităţile de a clasifica o rostire ce corespunde unui cuvânt rostit izolat este compararea acesteia cu o serie de modele etalon şi de a determina care este „cea mai bună potrivire”. Clasificarea se aborda *în sens clasic*, folosind, de exemplu, un algoritm de aliniere temporală dinamică, sau *în sens conexionist*, urmând, de exemplu, procedura descrisă în [1].

Modelele etalon au fost organizate într-o bază de date acustică ce conţine rostiri înregistrate de la 14 vorbitori (7 bărbaţi şi 7 femei). Fiecare vorbitor a rostit fiecare cifră cu trei viteze diferite şi trei contururi prozodice, conform tabelului 1 [6]. Prin urmare, baza de date cuprinde 1260 rostiri.

Conturul prozodic	Viteza rostirii	Tipul vocii
ascendent	mică	masculin
orizontal	medie	feminin
descendent	mare	

Tabelul 1 – Organizarea bazei de date acustice cu vorbitori multipli

4. EXPERIMENTE DE RECUNOAŞTERE INDEPENDENTĂ DE VORBITOR A CIFRELOR ROSTITE IZOLAT

S-au efectuat trei seturi de experimente. În primul set s-a utilizat un algoritm DTW, semnalul vocal fiind reprezentat prin zece coeficienţi de predicţie (LPC-10) [6]. În al doilea set, s-a utilizat tot un algoritm de aliniere temporală, dar cu o parametrizare cepstrală (CEPC-10), respectiv parametrizare cepstrală normată (SDCEPC-10) a semnalului vocal [3][7]. În ultimul set s-a utilizat un clasificator hibrid compus dintr-o reţea neuronală de tip cuantizare vectorială prin învăţare (LVQ) [2][5] şi un modul de aliniere temporală [1]. Rezultatele sunt prezentate în tabelul 2 [1][6][7].

Condiţii experimentale	Procentul de recunoaştere (%)
DTW / LPC-10	81,0
DTW / CEPC-10	88,3
DTW / SDCEPC-10	90,2
DTW+LVQ / CEPC-10	93,9

Tabelul 2 – Procentele de recunoaştere

5. CONCLUZII

O bază de date acustică care a fost concepută cu atenţie poate duce la îmbunătăţiri semnificative atât la nivelul procentului de recunoaştere cât şi la acela al timpului de calcul. Un algoritm robust de detecţie a marginilor de cuvânt este strict necesar pentru mărirea procentului de recunoaştere. Procentul de recunoaştere de aproximativ 94% şi timpul de calcul de ordinul a 20 de secunde demonstrează eficienţa abordării hibride DTW+LVQ şi a parametrizării semnalului vocal prin coeficienţi cepstrali calculaţi direct din spectru.

[1] Burileanu D., M. Sima, C. Burileanu, V. Croitoru, *A Neural Network-Based Speaker-Independent System for Word Recognition in Romanian Language*, Proceedings of the First Workshop on Text, Speech, Dialogue – TSD'98, Brno, Czech Republic, 1998, pp. 177-1982.

[2] Fausett L., Fundamentals of Neural Networks – Architectures, Algorithms, and Applications, Prentice-Hall, Englewood Cliffs, New Jersey, 1994.

[3] Picone J., *Signal Modeling Techniques in Speech Recognition*, Proceedings of the IEEE, Vol. 81, No. 9, September 1993, pp. 1215-1247.

[4] Rabiner L. R., M. R. Sambur, An Algorithm for Determining the Endpoints of Isolated Utterances, The Bell System Technical Journal, Vol. 54, No. 2, 1975, pp. 297-315.

[5] Rabiner L. R., R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[6] Sima M., V. Croitoru, D. Burileanu, *Performance Analysis on Speech Recognition using Neural Networks*, Proceedings of the International Conference on Development and Application Systems, Suceava, Romania, 1998, pp. 259-266.

[7] Sima M., D. Burileanu, V. Croitoru, C. Burileanu, The Application of Neural Network Paradigms in Speech Recognition for a Romanian Voice Dialing System, Proceedings of Communications'98, Bucharest, Romania, 1998, pp. 233-238.

[8] Sima M., D. Burileanu, C. Burileanu, V. Croitoru, Full-Custom Software for Start/End Point Detection of Isolated-Spoken Words, Proceedings of the 12th International Conference on Control Systems and Computer Science, Vol. II, Bucharest, Romania, 1999, pp. 19-24.

[9] Tubach J. P., Éditeur principal, *La parole et son traitement automatique*, Masson, Paris, 1989.