

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

EXPERIMENTE DE RECUNOAȘTERE 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 își propune să prezinte o serie de experimente realizate de autori în domeniul recunoașterii cifrelor rostite izolat, în cadrul mai larg al sistemelor de selecție a numerotării cu comandă vocală. Prin urmare, recunoașterea se realizează independent de vorbitor, rostirile candidat constituindu-se într-o înregistrare continuă. Performanța urmărită a fost obținerea unui compromis între procentul de recunoaștere și 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 detecție a marginilor de cuvânt destinat delimitării rostirilor candidat în cadrul înregistrării continue, blocul de recunoaștere propriu-zisă a cuvintelor izolate și partea de generare a comenzilor de selecție.

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

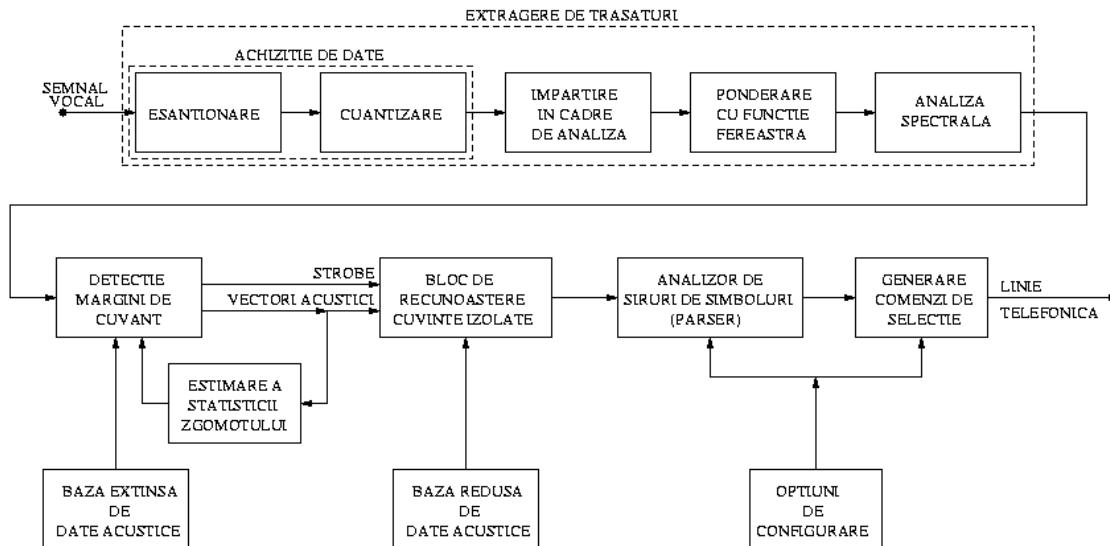


Figura 1 - Schema bloc a unui sistem de selecție comandat prin voce

2. PROBLEMA DETECȚIEI MARGINILOR DE CUVÂNT

O bună localizare a regiunilor ce conțin vorbire, nu numai că reduce timpul de calcul, ci și mărește procentul de recunoaștere. Să presupunem că se compară printr-un

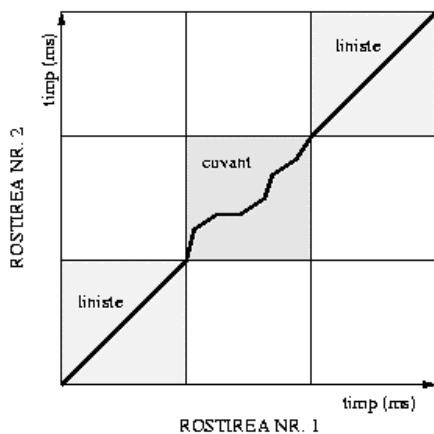


Figura 2 – Alinierea temporală dinamică a două rostiri ce conțin largi regiuni de zgomot

procedeu de aliniere temporală dinamică (DTW) [9] două segmente acustice, fiecare conținând aproximativ două treimi de „liniște” (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 dificultățile majore în decelarea cifrelor ce prezintă rostiri de lungime scurtă („2” și „8”) [6][8].

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

3. BAZĂ DE DATE ACUSTICĂ CU VORBITORI MULTIPLI

62
63
64
65
66
67
68

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

70 Modelele etalon au fost
72 organizate într-o bază de date acustică
74 ce conține rostiri înregistrate de la 14
76 vorbitori (7 bărbați și 7 femei).
78 Fiecare vorbitor a rostit fiecare cifră
80 cu trei viteze diferite și trei contururi

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

Tabelul 1 – Organizarea bazei de date acustice cu vorbitori multipli

81 prozodice, conform tabelului 1 [6]. Prin urmare, baza de date cuprinde 1260 rostiri.

82
83

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

84
85
86

87 S-au efectuat trei seturi de experimente. În primul set s-a utilizat un algoritm DTW,
88 semnalul vocal fiind reprezentat prin zece coeficienți de predicție (LPC-10) [6]. În al

90 doilea set, s-a utilizat tot un
92 algoritm de aliniere temporală,
94 dar cu o parametrizare cepstrală
96 (CEPC-10), respectiv parame-
98 trizare cepstrală normată
100 (SDCEPC-10) a semnalului

Condiții experimentale	Procentul de recunoaștere (%)
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 recunoaștere

101 vocal [3][7]. În ultimul set s-a utilizat un clasificator hibrid compus dintr-o rețea
102 neuronală de tip cuantizare vectorială prin învățare (LVQ) [2][5] și un modul de
103 aliniere temporală [1]. Rezultatele sunt prezentate în tabelul 2 [1][6][7].

5. CONCLUZII

104

105

106 O bază de date acustică care a fost concepută cu atenție poate duce la îmbunătățiri
107 semnificative atât la nivelul procentului de recunoaștere cât și la acela al timpului de
108 calcul. Un algoritm robust de detecție a marginilor de cuvânt este strict necesar
109 pentru mărirea procentului de recunoaștere. Procentul de recunoaștere de aproximativ
110 94% și timpul de calcul de ordinul a 20 de secunde demonstrează eficiența abordării
111 hibride DTW+LVQ și a parametrizării semnalului vocal prin coeficienți cepstrali
112 calculați direct din spectru.

113

114

- 115 [1] Burileanu D., M. Sima, C. Burileanu, V. Croitoru, *A Neural Network-Based*
116 *Speaker-Independent System for Word Recognition in Romanian Language*,
117 *Proceedings of the First Workshop on Text, Speech, Dialogue – TSD'98*, Brno,
118 *Czech Republic*, 1998, pp. 177-1982.
- 119 [2] Fausett L., *Fundamentals of Neural Networks – Architectures, Algorithms, and*
120 *Applications*, Prentice-Hall, Englewood Cliffs, New Jersey, 1994.
- 121 [3] Picone J., *Signal Modeling Techniques in Speech Recognition*, *Proceedings of*
122 *the IEEE*, Vol. 81, No. 9, September 1993, pp. 1215-1247.
- 123 [4] Rabiner L. R., M. R. Sambur, *An Algorithm for Determining the Endpoints of*
124 *Isolated Utterances*, *The Bell System Technical Journal*, Vol. 54, No. 2, 1975,
125 pp. 297-315.
- 126 [5] Rabiner L. R., R. W. Schafer, *Digital Processing of Speech Signals*,
127 Prentice-Hall, Englewood Cliffs, New Jersey, 1978.
- 128 [6] Sima M., V. Croitoru, D. Burileanu, *Performance Analysis on Speech*
129 *Recognition using Neural Networks*, *Proceedings of the International*
130 *Conference on Development and Application Systems*, Suceava, Romania,
131 1998, pp. 259-266.
- 132 [7] Sima M., D. Burileanu, V. Croitoru, C. Burileanu, *The Application of Neural*
133 *Network Paradigms in Speech Recognition for a Romanian Voice Dialing*
134 *System*, *Proceedings of Communications'98*, Bucharest, Romania, 1998,
135 pp. 233-238.
- 136 [8] Sima M., D. Burileanu, C. Burileanu, V. Croitoru, *Full-Custom Software for*
137 *Start/End Point Detection of Isolated-Spoken Words*, *Proceedings of the*
138 *12th International Conference on Control Systems and Computer Science*,
139 Vol. II, Bucharest, Romania, 1999, pp. 19-24.
- 140 [9] Tubach J. P., Éditeur principal, *La parole et son traitement automatique*,
141 Masson, Paris, 1989.