A STUDY ON CLASSIFICATION FOR THE VOWEL SIGNAL ACCORDING WITH EMOTIONAL STATE BASED ON RECURENT FUZZY C-MEANS ALGORITHM

Marius Zbancioc, Monica Feraru
"Institute of Computer Science" Romanian Academy of Iasi
zmarius@etti.tuiasi.ro

Abstract. *In this paper we used the recurent fuzzy C-means algorithm in order to compare the centers of the clusters with the statistic parameters obtained from the input data sets. Based on the experimental results, the FCM recurent is better than the clasical algorithms FCM and K-means. The results are encouraging and justify further research with new features vectors for the automatically emotion recognition.*

Cuvinte-cheie: *algoritm C-means, semnal vocal, stare emoțională*

I. Introducere

Starea emoțională sau fizică a unei ființe umane este cunoscută ca şi aspect emoțional a vorbirii şi poate fi inclusă şi în aşa-numitele aspecte paralingvistice. Deşi starea emoțională nu modifică conținutul lingvistic, este un factor important în comunicarea umană, deoarece ne oferă informații de feedback în mai multe aplicații. Recunoaşterea emoțiilor cu ajutorul unui computer nu este o idee nouă. Primele investigații s-au desfăşurat pe la mijlocul anilor 1980 folosind statistic proprietățile acustice anumite caracteristici [1], [2]. Zece ani mai târziu, evoluția de arhitecturi noi de calculatoare a dus la punerea în aplicare a unor algoritmi complecşi de recunoaştere a emoțiilor.

În momentul actual cercetarea este direcționată spre găsirea unor combinații de clasificatori care măresc eficiența clasificărilor în aplicațiile din viața reală. De exemplu, în proiectele "Prozodie pentru sistemele de dialog" şi "SmartKom", sistemele de rezervare a biletelor sunt dezvoltate astfel încât să fie capabile să recunoască starea de disconfort sau frustrare a unui utilizator, iar răspunsul să fie cel corespunzător [3], [4]. Situații similare se găsesc de asemenea în cadrul aplicațiillor de call center [5], [6].

În viitor, cercetarea din domeniul expresivității emotive va beneficia de disponibilitatea continuă la scară largă de colecții emoționale de date de vorbire, şi se va concentra pe îmbunătățirea modelelor teoretice legate de comunicarea emoțiilor [7]. Într-adevăr, pe de o parte, colecții mari de date care includ o varietate de enunțuri vorbitor în mai multe stări emotionale sunt necesare, în scopul de a evalua fidel performanța algoritmilor de recunoaştere emoțională în vorbire. Colecțiile deja disponibile date constau doar din câteva enunțuri, şi, prin urmare, este dificil să se demonstreze rezultate de încredere în recunoaşterea de emoție. Colecții mari de date care includ o varietate de pronunții ale unui vorbitor cu mai multe stări emoționale sunt necesare, în scopul de a evalua cât mai bine performanța algoritmilor de recunoaştere a stărilor emoționale în timpul vorbirii. Bazele de date disponibile constau doar din câteva rostiri şi prin urmare, este dificil să se demonstreze ca rezultatele obținute sunt concludente în procesul de recunoaştere a emoțiilor în vorbire.

Au fost testate mai multe metode de recunoaştere automată a emoțiilor [8], [9]. De exemplu, Dellaert, et al. au folosit probabilitatea maximă Bayes, regresia Kernel, şi metoda K-NN [10], în timp ce Roy şi Pentland au utilizat metodă de discriminare liniară Fisher [11]. În studiul propus în această lucrare am folosit o variantă de algoritm Fuzzy C-means cu recurență, care s-a dovedit mai performant decât algoritmii clasici K-means şi FCM.

II. Corpusul SROL

Studiul s-a realizat pe singurul corpus emoțional adnotat pentru limba română, disponibil gratuit pe internet, care face parte din proiectul Sunetele Limbii Române (SRoL) [12]. Baza de date SRoL a obținut diploma de excelență şi Medalia de Aur în cadrul Salonului International INVENTICA 2009. Ambasada Franței din România face referire la baza de date SRoL, ca fiind o arhivă cu multiple aplicații în educație, în medicină, în sociolinguistică etc, precum şi învățarea pronunției corecte a limbii române.

Corpusul SRoL conține peste 1500 de înregistrări distincte, disponibile în diverse formate de precizie şi codare. Fişierele sunt grupate în sunete de bază (fişiere de vocale, consoane, diftongi, triftongi, hiatusuri şi sunete specifice - ce, ci, che), în scurte propoziții sau segmente de fraze, cu încărcătură emoțională diferită, şi în propoziții referitoare la aspectele fonetice (în conexiune cu aspectele semantice) în frazele cu subiect dublu în Limba Română. Emoțiile studiate sunt starea de bucurie, de tristețe, de furie şi tonul neutru. Propozițiile înregistrate sunt: "Cine a făcut asta", "Vine mama", "Aseară", "Ai venit iar la mine", "Omul meu îl lucră", "Îți vei câştiga locul dorit", "Oricum îți poți câştiga locul dorit". Vorbitorii au pronunțat fiecare frază de minim trei ori; ei sunt persoane sănătoase din zona Moldovei (Iaşi, Bacău, Vaslui) care prezintă un minim de expresivitate emoțională. Vorbitorii au completat şi nişte chestionare referitoare la starea lor de sănătate. Înregistrarile au fost realizate conform protocolului de înregistrare [12].

Corpusul include şi o secțiune dedicată instrumentelor software de procesare a semnalului vocal. Acestea permit preprocesarea semnalului cu operații de filtrare a semnalului (filtre trece sus 70Hz pentru a limita banda de căutare a lui F0, respectiv trece jos pentru a limita banda formanților, filtre de mediere pentru eliminarea zgomotului), operații de segmentare în vederea delimitării semnalului vocalic de cel consonantic. Pentru detecția frecvenței fundamentale F0 s-au implementat 4 metode: metoda cepstrală, metoda autocorelației, metoda produsului spectrelor armonice HPS şi metoda funcției diferență AMDF. Rezultatele celor patru metode de extragere a lui F0 sunt aplicate unor metode de corecție şi sunt ponderate în funcție de performanțele fiecărui extractor în stabilirea rezultatului final.

Detecția formanților F1-F4 este realizată prin concatenarea fuzzy a unor „spectre netezite" obținute prin filtrarea cepstrului. Pentru fiecare fonem sunt extrase valorile statistice privind durata şi valorile mediei, a deviației standard, Skewness, Kurtosis, a medianei pentru valorile lui F0 şi a formanților F1-F4.

Din rezultatele statistice obținute se pot calcula distribuția normală multivariată, histograme bidimensionale ale variațiilor formanților (uzual în spațiu F1-F2) şi calculul matricii de confuzie pentru vocale în diverse contexte emoționale. Metodologia a fost stabilită de profesor H.N. Teodorescu.

III. Algoritmul Fuzzy C-means Recurent

Algoritmii de clusterizare grupează datele dintr-un set de intrare pe baza similarității acestora, cei mai cunoscuți fiind cei ierarhici (aglomerativi şi de divizare) şi algoritmii de partiționare. Din clasa algoritmilor de partiționare fac parte EM (Expectation Maximization) bazați pe modele probabiliste, QT (quality threshold), algoritmii bazați pe teoria grafurilor şi K-means bazat pe calculul distanței Euclidiene si a erorii pătratice. Din algoritmul K-means clasic au fost dezvoltați ulterior o serie de alți algoritmi ce includ algoritmi genetici FGKA (Fast Genetic K-means Algorithm) sau tehnici fuzzy cum ar fi Fuzzy C-means (FCM).

Pentru clusterizarea datelor am utilizat o variantă îmbunătățită a algoritmului clasic FCM care asociază un coeficient de încredere fiecărui centru asociat unui cluster/partiție fuzzy, coeficient ce se

modifică recurent la fiecare nouă iterație a algoritmului. Algoritmul FCMR (fuzzy c-means recurent) rezolvă unul din marile dezavantaje ale algoritmului clasic, legat de faptul că învățărea este nesupervizată şi că nu se cunosc apriori numărul real de clustere din setul de date de intrare. În această situație centrele clusterelor găsite de FCM nu converg către centrele reale, fapt ce este compensat în noul algoritm FCMR prin introducerea coeficienților de încredere recurenți.

In [13-15] Teodorescu propunea un model de sistemelor fuzzy recurente având ca justificare faptul că oamenii au tendința să îşi ajusteze raționamentele pe baza rezultatelor preliminare. Astfel metacunoştințelor (regulilor) ce conduc la obținerea unor rezultate mai slabe ar trebui să li se asocieze coeficienți de încredere mai mici în comparație cu cele care permit obținerea unor rezultate mai bune. Deoarece sistemele fuzzy recurente – model Teodorescu reprezintă o generalizare a sistemelor clasice existente (Mamdani, Sugeno etc.), orice sistem fuzzy clasic (algoritm fuzzy) din orice domeniu de analiză, poate fi transformat într-un model fuzzy cu recurență la nivelul gradelor de încredere în metacunoştințe/reguli [16].

Datele de intrare $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ sunt vectori de n trăsături, distanța între doi vectori fiind calculată pe baza distanței Euclidiene, dar se pot folosi orice fel de distanțe: Mahalanobis, Pearson, Manhattan, Chebyshev, Lee, Hamming etc. Se va nota cu N dimensiunea setului de date $X = \{\mathbf{x}_k\}$, $k = \overline{1, N}$. Algoritmul FCM determină distanța de la centrul fiecărui cluster c_j la fiecare vector de trăsături $\mathbf{x}_k$, şi îi asociază un grad de apartenență $m_{j,k} : R \to [0,1]$, unde valoarea de 1 indică apartenența totală şi valoarea de 0 non-apartenența. Notăm cu C numărul de clustere/partiții şi cu U matricea de dimensiune CxN ce păstrează toate gradele de apartenență.

Pseudocodul algoritmului FCMR este următorul:

1. Faza de inițializare cu numere aleatoare a matricei $U^{(0)}$ şi normalizarea acesteia, stabilirea numărului de clustere $2 \le C \le N$, a erorii de convergență a algoritmului e şi a exponentului m aplicat gradelor de apartenență. În plus este introdusă o matrice a coeficienților de încredere recurenți $\{CF_{jk}^{(0)} = 1\}$, t=0;
2. Se calculează mulțimea C a centrilor partițiilor $\{\mathbf{c}_j^*\}$, folosind o relație de calcul cu coeficienții de încredere recurenți. Se determină distanțele $\{d_{jk}^*\}$ de la fiecare element al setului de date X la fiecare centru;
3. Pe baza distanțelor de calculează funcția obiectiv J* şi valoarea gradelor de apartenență la momentul următor de timp $U^{(t+1)}$;
4. În functie de noile valori din matricea $U^{(t+1)}$ se reactualizează coeficienții recurenți la momentul t+1;
5. Dacă diferențele funcției obiectiv $\left| J^{(t+1)} - J^{(t)} \right|$ sunt peste limita de convergență (eroarea impusă e) şi nu s-a depăşit numărul maxim de iterații se revine la Pas2.

Prin simulare s-a constatat că performanțele algoritmului FCMR propus pot fi îmbunătățite dacă aplicarea coeficienților de încredere nu se face încă de la început, ci cu „întârziere" după ce a avut loc o deplasare „grosieră" a centrilor clusterelor.

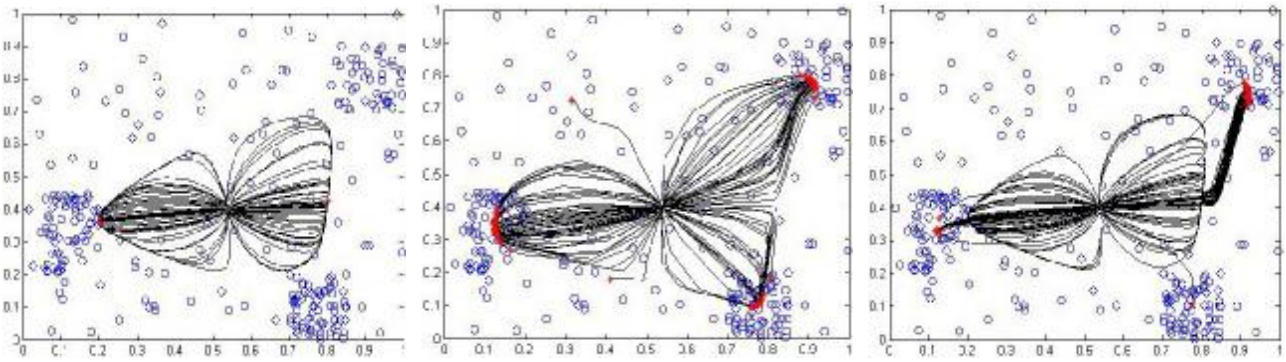

Fig.1 *a) FCM clasic* *b)FCM recurent* *c)FCMR aplicat cu "întârziere"*

În figura 1 se poate observa cum pentru situația când avem un număr real de *M*=3 clustere şi doar *C*=2 clustere specificate algoritmului de partiționare, cum algoritmul clasic FCM poziționează un centru între două clustere, algoritmul FCMR atinge toate 3 centrele clusterelor şi algoritmul FCMR aplicat cu „întârziere" este mai robust găsind în mod frecvent doar 2 clustere. Simulările au fost făcute rulând algoritmii de 50 ori.

Astfel algoritmul FCMR rezolvă parțial o altă critică adusă algoritmilor clasici K-means şi FCM şi anume că aceştia pot converge spre un maxim local şi nu global şi că valoarea finală a centrelor clusterelor depinde de valorile alese inițial pentru centrii partițiilor/clusterelor.

IV. Rezultate şi discuții

Scopul cercetăriilor a fost de a vedea dacă un algoritm de clusterizare bazat pe învățare nesupervizată poate partiționa spațiul caracteristicilor extrase din vocalele limbii române furnizând centri clusterelor cât mai apropiat de valorile medii determinate statistic.

Pe baza fişierelor de adnotare se delimitează din corpusul emoțional SROL poziția de start şi poziția finală a tuturor vocalelor limbii române. Aplicația software implementată extrage valorile mediei, dispersiei, medianei, Skewness, Kurtosis pentru frecvența fundamentală F0 şi formanții F1-F4. Sunt păstrate şi informații privind emoția exprimată, sexul şi numărul de identificare asociat vorbitorului. Mai nou au fost introduşi ca parametri valoarea de jitter pentru variațiile lui F0 şi de shimmer pentru fluctuațiile în amplitudine, care să ajute la identificarea emoției.

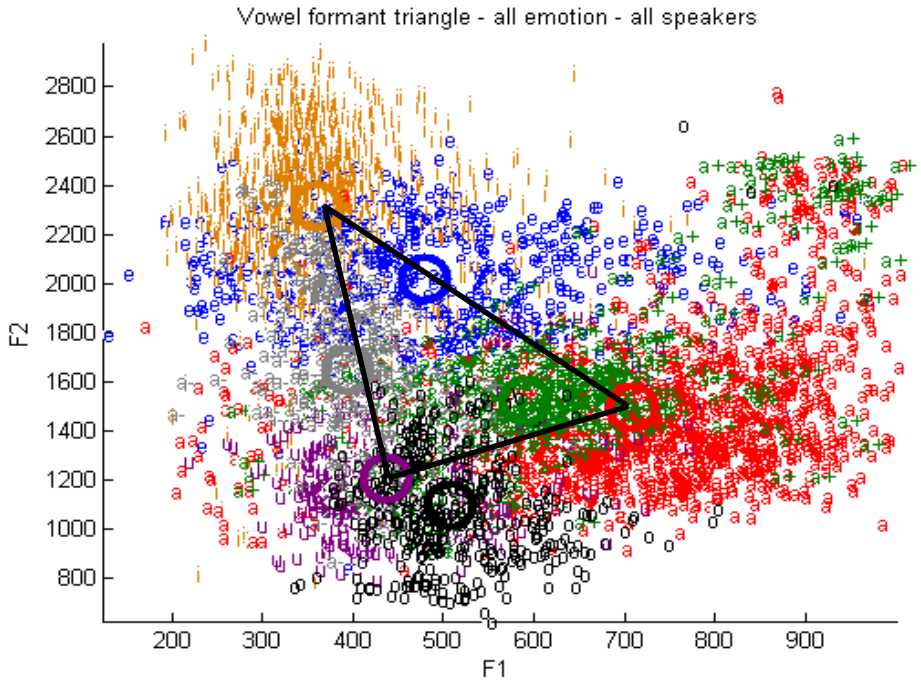

Fig.2 *Reprezentarea valorilor fornanților F1, F2 pentru toate vocalele extrase din corpusul emoțional SROL*

S-au comparat pozițiile clusterelor furnizate de algoritmul FCMR pentru un număr de C=7 clustere (câte o partiție pentru fiecare din cele 7 vocale ale limbii române 'a','e','i','o','u','ă','î') cu valorile obținute statistic în urma unor cercetări aterioare şi raportate în [17][18]. În figura 2 s-au reprezentat în spațiul formanților F1, F2 triunghiul vocalelor determinat pentru toate vocalele şi toți vorbitorii din corpusul emoțional SROL.

În fig. 3 se poate observa cum partiționează algoritmul FCM recurent spațiul datelor de intrare reprezentat prin vectori de trăsături în componența cărora sunt incluşi formanții inferiori F1 şi F2. Repetând numărul de simulări s-a putut observa că algoritmul FCMR găseşte centrii clusterelor în mod frecvent în anumite regiuni. S-au comparat aceste zone cu elipsele de încadrare determinate statistic şi se observă că acestea aproximează destul de bine centrii clusterelor. Diferențe apar în zona vocalei 'a' unde cele două clustere pot fi înglobate într-unul singur şi în zona vocalei 'e' unde algoritmul nu reuşeşte să poziționeze un centru..

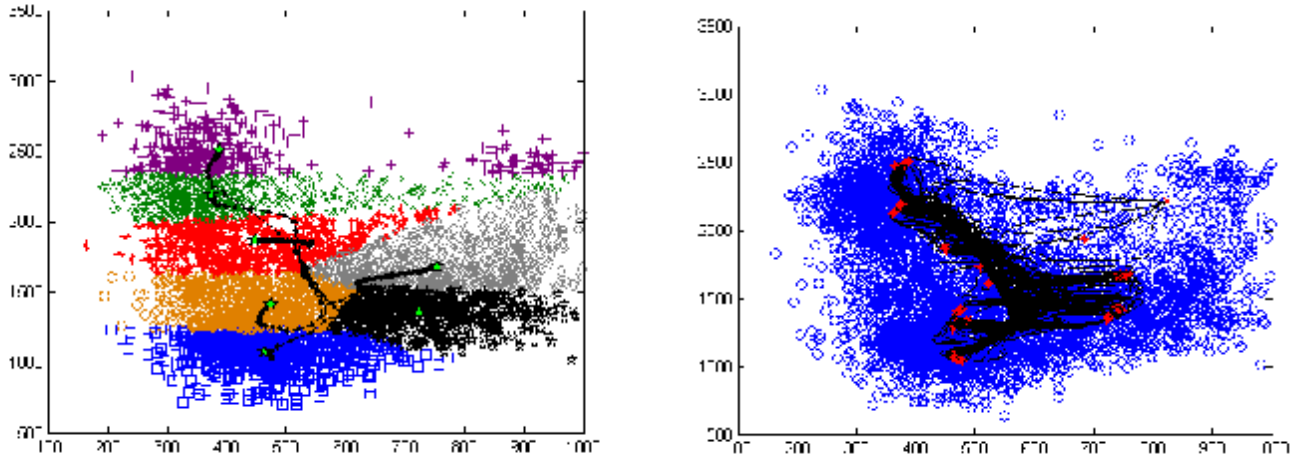

Fig.3 *a) Partiții FCMR* (*C=7*) *pentru o singură rulare, respectiv b) pentru 50 de rulări*

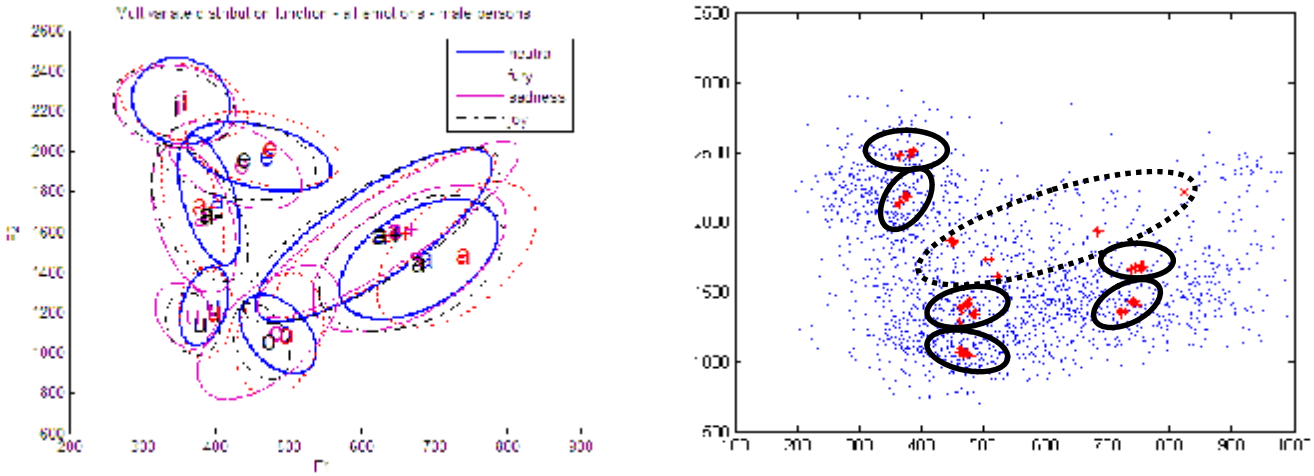

Fig.4 *a) Elipse de încadrare determinate statistic, b) Migrarea clusterelor desterminate de FCM* (*C=7*)

V. Concluzii şi direcții viitoare

S-a verificat experimental că algoritmul FCM recurent este suficient de robust pentru a găsi centrii clusterelor apropiați ca poziționare de valorile statistice ale parametrilor din setul de intare. Prin simulare s-a verificat că introducerea coeficienților recurenți conduce la rezultate mai bune decât algoritmii clasici FCM şi K-means.

Pe baza rezultatelor obținute apare necesitatea introducerii unor algoritmi semisupervizați cu fixarea centrilor clusterilor în puncte considerate a fi reprezenative pentru recunoaşterea unei vocale. Se asemnea prin învățare semisupervizată se pot defini conexiuni între două puncte învecinate de genul "puternic" legate / "slab" legate care să forțeze apartenența la un cluster.

VI. Referințe

1. Van Bezooijen R., The Characteristics and Recognizability of Vocal Expression of Emotions. Foris, Drodrecht, The Netherlands, 1984.
2. Tolkmitt F.J., Scherer K.R., Effect of experimentally induced stress on vocal parameters. J. Exp. Psychol. [Hum.Percept.] 12 (3), 1986., pp. 302–313.
3. Ang J., Dhillon R., Krupski A., Shriberg E., Stolcke A., Prosody-based automatic detection of annoyance and frustration in human–computer dialog. In: Proc. Internat. Conf. on Spoken Language Processing (ICSLP '02), Vol. 3, 2002, pp. 2037–2040.
4. Schiel F., Steininger S., Turk U., The Smartkom multimodal corpus at BAS. In: Proc. Language Resources and Evaluation (LREC '02).
5. Petrushin V.A., Emotion in speech recognition and application to call centers. In: Proc. Artificial Neural Networks in Engineering (ANNIE '99), Vol. 1, pp. 7–10.
6. Lee C.M., Narayanan S.S., Toward detecting emotions in spoken dialogs. IEEE Trans. Speech Audio Process. 13 (2), 2005, pp. 293–303.
7. Scherer K.R., Vocal communication of emotion: a review of research paradigms. Speech Comm. 40, 2003, pp. 227–256.
8. Petrushin V., Emotion in Speech: Recognition and Application to Call Centers, Artificial Neu. Net. In Engr. (ANNIE '99), pp. 7-10.
9. Batliner A., Fischer K., Huber R., Spilker J., Noth E., Desperately Seeking Emotions: Actors, Wizards, and Human Beings, Proceedings of the ISCA Workshop on Speech and Emotion
10. Dellaert F., Polzin T., Waibel A., Recognizing Emotion in Speech, ICSLP'96 Conference Proceedings
11. Roy D., Pentland A., Automatic Spoken Affect Analysis and Classification, In the Proceedings of the International Conference on Automatic Face and Gesture Recognition, Killington, VT. 1996.
12. Proiectul Sunetele Limbii Române - SRoL, coordonator H.N. Teodorescu, http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.html
13. Teodorescu, H.N (2004), Recurrent Rules-Based Fuzzy Decision-Making and Control, WSAS Conference, Udine, Italy
14. Teodorescu H.N., Fuzzy systems with recurrent rules in population and medical models, Proceeding MATH'08 Proceedings of the American Conference on Applied Mathematics World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA 2008, ISBN: 978-960-6766-47-3, pp. 343-349
15. Teodorescu H.N., Fuzzy Systems with Recurrent Rules. A new type of fuzzy systems and applications, Intelligent Systems, Selected papers from ECIT 2004, pag 157-166, Editors: H.N.Teodorescu, Iaşi, Ed. Performantica, ISBN 973-7994-85-X.

16. Zbancioc M. (2005a), "Recurrent Fuzzy Rules (Teodorescu's Fuzzy Systems) in Economic Process Modelling", 15th International Conference on Control Systems and Computer Science, CSCS-15, 25-27 May, 2005, Bucureşti, România
17. H.N. Teodorescu, M. Zbancioc, M. Feraru, "The analysis of the vowel triangle variation for Romanian language depending on emotional states", ISSCS Conference, Iasi, Romania 30June-1Jul.2011, ISBN 978-1-4577-0201-3, pp. 331-334
18. H.N. Teodorescu, M. Zbancioc, M. Feraru, "Statistical characteristics of the formants of the Romanian vowels in emotional states", International Conference on Speech Technology and Human-Computer Dialogue SPeD 2011, 18-21 may 2011 Brasov, Romania, ISBN 978-1-4577-0439-0, pp. 13-22, IEEE publication, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5940725