



(12)

BREVET DE INVENȚIE

(21) Nr. cerere: **a 2014 00347**

(22) Data de depozit: **07/05/2014**

(45) Data publicării mențiunii acordării brevetului: **28/02/2019** BOPI nr. **2/2019**

(41) Data publicării cererii:
29/01/2016 BOPI nr. **1/2016**

(73) Titular:

- **BUZO ANDI**,
BD. GENERAL VASILE MILEA NR. 8,
BL. B2, SC. 3, ET. 4, AP.57, SECTOR 6,
BUCUREȘTI, B, RO;
- **CUCU HORIA**, ALEEA POLITEHNICII
NR. 6, BL. 3, SC. 4, ET. 2, AP. 42,
SECTOR 6, BUCUREȘTI, B, RO;
- **PETRICĂ LUCIAN**, INTRAREA VÂSLEI
NR. 1, BL. PM63, SC. 2, ET. 5, AP. 77,
SECTOR 3, BUCUREȘTI, B, RO;
- **BURILEANU DRAGOȘ**,
STR. JOHANNES KEPLER NR. 2 BL. 2,
SC. 2, AP. 61, SECTOR 2, BUCUREȘTI, B,
RO

(72) Inventatori:

- **BUZO ANDI**,
BD. GENERAL VASILE MILEA NR. 8,
BL. B2, SC. 3, ET. 4, AP.57, SECTOR 6,
BUCUREȘTI, B, RO;

- **CUCU HORIA**, ALEEA POLITEHNICII
NR. 6, BL. 3, SC. 4, ET. 2, AP. 42,
SECTOR 6, BUCUREȘTI, B, RO;
- **PETRICĂ LUCIAN**, INTRAREA VÂSLEI
NR. 1, BL. PM63, SC. 2, ET. 5, AP. 77,
SECTOR 3, BUCUREȘTI, B, RO;
- **BURILEANU DRAGOȘ**,
STR. JOHANNES KEPLER NR. 2, BL. 2,
SC. 2, AP. 61, SECTOR 2, BUCUREȘTI, B,
RO

(74) Mandatar:

CABINET D.NICOLAESCU, STR.TURDA,
NR.102, BL.30A, ET.7, AP.28, BUCUREȘTI

(56) Documente din stadiul tehnicii:

US 8612224 B2; US 8554563 B2;
US 8433567 B2; US 7473838 B2

(54) **METODĂ ȘI SISTEM PENTRU DIARIZARE ÎN TIMP REAL
A SEMNALELOR AUDIO, UTILIZATE
PENTRU RECUNOAȘTEREA AUTOMATĂ A VORBIRII
ȘI A VORBITORULUI**



1 Invenția aparține domeniului sistemelor de procesare a semnalului audio pentru
recunoașterea automată a vorbirii și identificarea automată a vorbitorului.

3 Un sistem de recunoaștere automată a vorbirii are ca scop transcrierea unui semnal
5 audio, de cele mai multe ori înregistrarea unei conversații sau a unui monolog. Prin
recunoaștere se obține un fișier text care conține cuvintele rostite. În cazul în care semnalul
7 audio conține segmente de muzică, zgomot, efecte speciale, sau în cazul în care proprietățile
semnalului audio se schimbă în timp, rezultatele procesului de recunoaștere sunt afectate
9 în mod negativ, prin introducerea de erori. De exemplu, sistemul de recunoaștere va încerca
transcrierea unui semnal muzical, rezultând text fără sens gramatical. Probleme similare sunt
11 întâlnite și în cazul recunoașterii vorbitorului, care nu este posibilă atunci când înregistrarea
audio conține muzică sau zgomot, sau când vorbirea înregistrată conține semnal vocal de
la mai mulți vorbitori.

13 Diarizarea este procesul prin care semnalul audio este analizat și segmentat, astfel
încât fiecare segment are proprietăți uniforme din punct de vedere audio, și conține un singur
15 fel de semnal, care poate fi liniște, muzică, vorbire sau alte tipuri de semnal audio.
Recunoașterea se poate face apoi doar pe segmentele de vorbire.

17 Segmentele de vorbire identificate pot fi suplimentar procesate în cadrul procesului
de diarizare pentru separarea vorbitorilor, folosind proprietăți audio cum ar fi frecvența
19 fundamentală a semnalului vocal. În urma acestui pas, fiecare segment de vorbire aparține
unui singur vorbitor, și poate fi folosit pentru identificarea vorbitorului respectiv.

21 În cele ce urmează vom enunța terminologia folosită în descrierea prezentei invenții,
cu mențiunea că se folosește terminologia consacrată specifică domeniului, în limba engleză:

23 - Frame - Fereastră audio - un număr de eșantioane ale semnalului audio digital,
reprezentând în mod obișnuit un interval de timp fix, de ordinul milisecundelor (10...20 ms);

25 - Speech Features - Vector de caracteristici audio - coeficienții cepstrali de frecvență
(MFCC) și alte măsuri ale semnalului dintr-o fereastră audio, folosite pentru procesul de
27 diarizare. Semnalul audio este reprezentat în procesul de diarizare ca o înșiruire de vectori
de caracteristici consecutivi;

29 - Segment - Segment - un număr de vectori de caracteristici audio consecutivi, care
au proprietăți similare;

31 - Cluster - Grup - un număr de segmente consecutive, care au proprietăți similare,
de exemplu, aparțin aceluiași vorbitor;

33 - Speaker model - Model audio - un model de mixtură Gaussiană (GMM) care
aproximează caracteristicile vorbitorului. Un model de vorbitor poate fi antrenat folosind
35 vectorii de caracteristici audio ai unui grup. Un GMM poate, de asemenea, să aproximeze
clase mai largi de semnal audio, cum ar fi muzică, liniște, vorbire, voce de bărbat, voce de
37 femeie și altele.

 Stadiul cunoscut al tehnicii, în ceea ce privește sistemele de diarizare automată, este
39 analizat în **S. Galliano, G. Gravier, L. Chaubard, "The ester 2 evaluation campaign for
the rich transcription of French radio broadcasts," In Proc. Interspeech, pp. 2583-2586,
41 2009.** Unul dintre sistemele cele mai performante este dezvoltat de laboratoarele LIUM [S.
Meignier, T. Merlin, "LIUM SpkDiarization: An Open Source Toolkit For Diarization",
43 **in Proc. CMU SPUD Workshop, 2010].**

 Acesta procesează semnalul vocal în următoarele etape:

45 1. Extragerea vectorilor de caracteristici audio din semnalul audio.

 2. Segmentare.

47 3. Agregarea segmentelor în grupuri.

 4. Antrenarea modelelor audio.

5. Re-segmentarea Viterbi, folosind toate modelele identificate.	1
6. Identificarea segmentelor care conțin vorbire.	
7. Identificarea bărbat/femeie, folosind modele pre-antrenate.	3
8. Agregarea finală în grupuri a segmentelor care aparțin aceluiași vorbitor.	
Așa cum se prezintă sistemele de diarizare cunoscute, dezavantajele lor sunt	5
inabilitatea de a segmenta semnalul audio pe măsură ce acesta este primit de către sistemul	
de diarizare. Sistemele cunoscute de diarizare nu pot determina caracteristicile segmentului	7
(vorbitor, zgomot de fundal etc.) dacă nu au la dispoziție toate eşantioanele segmentului,	
astfel încât să poată antrena modelele audio necesare. Mai mult, sistemul de diarizare nu	9
poate determina granița între două segmente dacă nu are la dispoziție toate eşantioanele	
pentru ambele segmente. În cazul ideal, diarizarea se face pe tot semnalul vocal, eliminând	11
problemele enunțate, dar această metodă introduce întârzieri mari, inacceptabile pentru	
unele aplicații.	13
De exemplu, rezultatele diarizării sunt utile procesului de recunoaștere a vorbirii prin	
filtrarea vorbire/liniște sau a informației despre vorbitor (pentru adaptarea modelului acustic),	15
și este de dorit ca recunoașterea să aibă loc abia după terminarea diarizării. În multe	
aplicații, de exemplu, cele care implică fluxuri audio foarte lungi, sau care necesită răspuns	17
în timp real al sistemului de recunoaștere a vorbirii, diarizarea pe tot semnalul vocal nu este	
o soluție viabilă, și este necesară o soluție de diarizare în timp real, care să segmenteze	19
semnalul audio pe măsură ce acesta este primit.	
Problema tehnică pe care invenția de față își propune să o rezolve este tocmai	21
această diarizare în timp real a semnalelor vocale, pentru recunoașterea automată a vorbirii	
și a vorbitorului.	23
Invenția se referă la o metodă de diarizare în timp real a semnalelor vocale, prin	
identificarea și marcarea de segmente din fluxul audio vocal ce aparțin aceluiași vorbitor sau	25
aceleiași clase audio, metoda cuprinzând etapele de citire a unui număr de vectori de	
caracteristici audio dintr-o memorie tampon, segmentare a fluxului de vectori de caracteristici	27
menționați, pe baza unor modele statistice stocate într-o istorie de diarizare și într-o memorie	
statică de modele pre-antrenate, actualizare a istoriei de diarizare prin asocierea fiecăruia	29
dintre segmentele audio rezultate la etapa anterioară cu unul dintre modelele de vorbire	
existente și, în cazul în care asocierea nu există, crearea de noi modele statistice pentru	31
acele segmente, și adăugarea lor la istoria de diarizare, precum și eliminarea modelelor	
vechi din istoria de diarizare menționată, în conformitate cu o metodă de curățare a istoriei	33
de diarizare bazată pe reguli, în final trimițându-se către utilizator vectorii audio corespun-	
zători. Invenția se mai referă și la un sistem de diarizare în timp real, pentru implementarea	35
metodei, sistem care cuprinde un grup de memorii format din memorie pentru segmente,	
memorie pentru grupuri de segmente și memorie pentru modele GMM ce stochează istoria	37
de diarizare, două memorii tampon pentru vectori de caracteristici audio, și o memorie	
internă pentru vectori audio care stochează caracteristicile audio înainte de segmentare, o	39
memorie de modele statistice de vorbitor, un automat finit de control, pentru gestionarea	
istoriei de diarizare, și un automat finit de control al diarizării, toate acestea putând fi	41
implementate ca programe software executabile pe un calculator sau circuite integrate, care	
mențin istoria de diarizare și comunică și cu servicii externe de extragere a vectorilor de	43
caracteristici audio, recunoaștere automată a vorbirii, actualizare a modelelor GMM și	
diarizare preliminară.	45
Metoda de diarizare în timp real a semnalelor vocale se realizează folosind atât	
modele statice pre-antrenate de vorbitor, cât și o istorie de segmente și modele de vorbitor	47
create dinamic, istoria fiind gestionată periodic prin actualizarea modelelor de vorbitor și prin	
eliminarea modelelor, folosind o funcție de cost bazată pe vechimea modelului și ponderea	49
sa în istoria de diarizare, atunci când istoria de diarizare depășește o dimensiune	
prestabilită.	51

1 Invenția prezentată are multiple avantaje față de stadiul tehnicii:
2 - metoda propusă, folosind istoria de diarizare, permite execuția în timp real a
3 procesului de diarizare, prin menținerea unui număr relativ mic de modele de vorbitor, și
4 actualizarea acestora pe măsură ce rezultatele diarizării sunt produse. Efortul computațional
5 pentru recunoașterea vorbitorului prezent în fiecare segment de vorbire este proporțional cu
6 numărul de modele de vorbitor avute în vedere, prin urmare, reducerea numărului de modele
7 reduce efortul computațional;
8 - metoda propusă ocupă o cantitate mai mică de resurse de memorie, prin
9 mecanismul de gestiune care elimină modelele de vorbitor, împreună cu segmentele
10 asociate, folosind o funcție de cost ce ține cont de vechimea modelului și ponderea sa în
11 istoria de diarizare. Datorită faptului că necesarul de memorie pentru metoda de diarizare
12 propusă este fix, indiferent de lungimea fluxului audio diarizat și numărul de vorbitori din
13 acest flux, iar dimensiunea efectivă poate fi setată arbitrar de mic, metoda propusă se
14 pretează în special sistemelor încorporate și implementărilor folosind resurse limitate.
15 Se prezintă în continuare, în detaliu, principiile și realizarea invenției, în legătură și
16 cu fig. 1...3, ce reprezintă:
17 - fig. 1, metoda de diarizare propusă, ce presupune folosirea unei istorii de diarizare,
18 conținând modele dinamice de vorbitor, și modele statice de vorbitor, pentru segmentarea
19 unui flux de caracteristici de vorbire;
20 - fig. 2, metoda de gestiune a istoriei de diarizare, prin care segmente noi sunt
21 adăugate și, atunci când istoria depășește o dimensiune dată, sunt eliminate segmente și
22 modele de vorbitor;
23 - fig. 3, sistem de diarizare construit pe baza metodei propuse, constând din memorii
24 pentru istoria de diarizare, și componente pentru gestiunea acestei istorii.
25 Invenția se referă la o metodă pentru separarea unui semnal audio în segmente
26 omogene din punctul de vedere al proprietăților audio (diarizare), incluzând separarea
27 segmentelor de vorbire de segmentele audio de liniște, și separarea segmentelor de vorbire
28 în funcție de vorbitor. Metoda propusă are la bază faptul că segmentarea se face fără a
29 aștepta primirea întregului fișier audio, iar segmentele rezultate dintr-o anumită porțiune a
30 fluxului audio sunt calculate și livrate utilizatorului în timp real, cu o întârziere fixă, relativ la
31 fluxul audio.
32 Metoda propusă pentru diarizare în timp real este prezentată în fig. 1. Metoda se
33 bazează pe citirea, la fiecare T secunde, a unui număr N de vectori de caracteristici audio
34 dintr-o memorie tampon **101**. Vectorii sunt segmentați folosind un serviciu extern de
35 segmentare **102**. Modelele audio folosite pentru segmentare sunt stocate în istoria de
36 diarizare **103** sau în memoria statică de modele pre-antrenate **104**. Istoria **103** acoperă
37 ultimele S secunde de semnal audio, și conține atât segmentele, cât și grupurile de
38 segmente împreună cu modelele audio asociate grupurilor. Memoria statică **104** de modele
39 pre-antrenate conține modele pre-calculate pentru vorbitori considerați a fi importanți, a căror
40 recunoaștere este esențială (persoane publice cunoscute, VIP).
41 Segmentele rezultate din diarizarea de la pasul curent sunt adăugate la istoria de
42 diarizare **103**, și modelele vechi sunt eliminate din istoria **103** conform unei metode **105** de
43 gestiune a istoriei de diarizare. Actualizarea modelelor de vorbitor din istoria **103**, folosind
44 informația audio de la pasul curent, este realizată de un serviciu extern. Vectorii audio
45 corespunzători tuturor segmentelor, în afară de ultimul, sunt apoi transmiși către utilizator.
46 Vectorii audio din ultimul segment sunt păstrați și folosiți ca parte a următorului set de N
47 vectori de caracteristici audio ce vor fi procesați.

Metoda propusă pentru gestiunea istoriei de diarizare este ilustrată în fig. 2. Istoria de diarizare este compusă din trei memorii distincte:	1
- memoria pentru segmente (MS) 201 , ce conține caracteristicile de vorbire corespunzătoare segmentelor identificate anterior prin diarizare;	3
- memoria pentru grupuri de segmente (MG) 202 , ce conține grupurile identificate anterior prin diarizare, și	5
- memoria pentru modele GMM (MM) 203 , ce conține modelele de mixtură gaussiană, calculate pentru fiecare grup de segmente în parte.	7
În urma diarizării unei ferestre audio, segmentele sunt adăugate istoriei de diarizare 103 . Se încearcă asocierea fiecărui segment cu unul dintre modelele de vorbitor existente, la pasul 204 . Dacă asocierea există, se actualizează grupurile de segmente la pasul 205 , și se actualizează modelele de vorbitor ale grupurilor respective la pasul 206 . Dacă asocierea nu există, se creează un grup nou la care segmentul este adăugat, și se generează modelul de vorbitor pentru grupul nou creat. Atât noul grup, cât și modelul său de vorbitor sunt adăugate istoriei de diarizare.	9
Dacă istoria de diarizare depășește o dimensiune D prestabilită de către utilizator, se execută o procedură de curățare a acesteia: se verifică dacă există grupuri în care toate segmentele au vechime mai mare de S secunde, unde S este o valoare specificată de utilizator. Aceste grupuri sunt eliminate din istoria de diarizare la pasul 207 .	11
Dacă pasul anterior nu a dus la scăderea dimensiunii istoriei de diarizare sub dimensiunea D, se calculează o funcție de cost la pasul 208 pentru fiecare grup/model, în felul următor:	13
- se calculează o valoare de cost CV a vechimii modelului, proporțională cu numărul de secunde de la ultima actualizare a modelului;	15
- se calculează o valoare de cost CD a dimensiunii grupului asociat modelului, invers proporțională cu numărul de segmente conținute de grupul respectiv;	17
- se calculează o valoare de cost total CT prin mediere ponderată a CV și CD , cu ponderi alese de utilizator.	19
Modelele sunt ordonate în funcție de valoarea de cost total CT asociată fiecăruia, în lista 209 . În mod repetitiv, modelul cu valoarea cea mai mare de cost este eliminat din istoria de diarizare, împreună cu segmentele și grupul asociate modelului, până când dimensiunea istoriei de diarizare scade sub pragul D.	21
Sistemul propus pentru diarizare în timp real a unui flux audio este ilustrat în fig. 3. Istoria de diarizare este conținută în memorii RAM. Gestiunea istoriei de diarizare, incluzând scrierea și citirea memoriilor 201 , 202 , și 203 care formează istoria de diarizare, este realizată de un automat finit de control 301 , ce poate fi implementat ca un circuit sau ca un program executat pe un microcontroler. Sistemul include o memorie RAM suplimentară pentru vectori audio, 302 , care conține vectori de caracteristici audio citați dintr-o memorie tampon 303 , atât timp cât este necesar pentru diarizare, vectori care apoi sunt scriși în memoria tampon 304 . Întregul proces este controlat de un automat finit de control, ce poate fi implementat ca un circuit sau ca un program executat pe un microcontroler sau un microprocesor.	23
	25
	27
	29
	31
	33
	35
	37
	39
	41

Revendicări

1. Metodă de diarizare în timp real a semnalelor vocale, prin identificarea și marcarea de segmente din fluxul audio vocal ce aparțin aceluiași vorbitor sau aceleiași clase audio, metoda cuprinzând etapele de:

- citire a unui număr de vectori de caracteristici audio (101) dintr-o memorie tampon;
- segmentare (102) a fluxului de vectori de caracteristici menționați pe baza unor modele statistice stocate într-o istorie de diarizare (103) și într-o memorie statică de modele pre-antrenate (104);

- actualizare a istoriei de diarizare (105) prin:
 - (i) asocierea fiecăruia dintre segmentele audio rezultate la etapa anterioară cu unul dintre modelele de vorbire existente, iar în cazul în care asocierea nu există,
 - (ii) crearea de noi modele statistice pentru acele segmente, și adăugarea lor la istoria de diarizare;

- (iii) eliminarea modelelor vechi din istoria de diarizare menționată, în conformitate cu o metodă de curățare a istoriei de diarizare bazată pe reguli;

- transmitere către utilizator a vectorilor audio corespunzători (106).

2. Metodă de diarizare, conform revendicării 1, **caracterizată prin aceea că** etapa de curățare a istoriei de diarizare cuprinde eliminarea tuturor modelelor statistice pentru vorbitorii care nu au mai apărut în fluxul audio de cel puțin S secunde.

3. Metodă de diarizare, conform revendicării 1, **caracterizată prin aceea că** etapa de curățare a istoriei de diarizare cuprinde calcularea unei funcții de cost (CT) pentru modelele statistice dinamice de vorbitor, și eliminarea iterativă a câte unui model statistic din istoria de diarizare, alegându-se de fiecare dată modelul care are costul cel mai mare, până când dimensiunea istoriei de diarizare scade sub un prag D.

4. Metodă de diarizare, conform revendicării 3, folosită pentru eliminarea selectivă a modelelor statistice de vorbitor din istoria de diarizare, metoda fiind **caracterizată prin aceea că** funcția de cost (CT) reprezintă o medie ponderată între costul (CV) dat de vechimea modelului în istoria de diarizare, unde costul (CV) este direct proporțional cu vechimea modelului, și costul (CD) dat de ponderea segmentelor asociate modelului în istoria de diarizare, unde costul (CD) este invers proporțional cu această pondere.

5. Sistem de diarizare în timp real, ce implementează metoda conform revendicării 1, **caracterizat prin aceea că** sistemul cuprinde:

- un grup de memorii:
 - (i) memorie pentru segmente (201);
 - (ii) memorie pentru grupuri de segmente (202);
 - (iii) memorie pentru modele GMM (203) care stochează istoria de diarizare (103);

- două memorii tampon pentru vectori de caracteristici audio (303 și 304), și o memorie internă pentru vectori audio (302) care stochează caracteristicile audio înainte de segmentare (102);

- o memorie de modele statistice de vorbitor (104);

- un automat finit de control, pentru gestionare istoriei de diarizare (301), și

- un automat finit de control al diarizării (305),

care pot fi implementate ca programe software executabile pe un calculator sau circuite integrate, care mențin istoria de diarizare, și comunică și cu servicii externe de extragere a vectorilor de caracteristici audio (EV), recunoaștere automată a vorbirii (ASR), actualizare a modelelor GMM (SA) și diarizare preliminară (SD).

(51) Int.Cl.

G10L 15/08 (2006.01),

G10L 15/04 (2006.01)

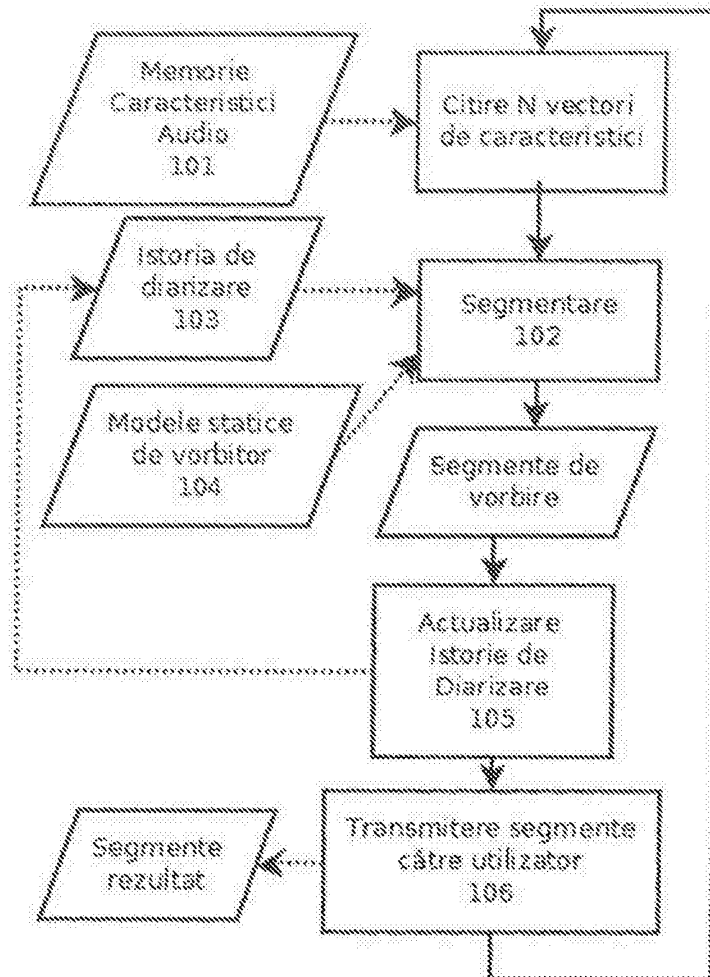


Fig. 1

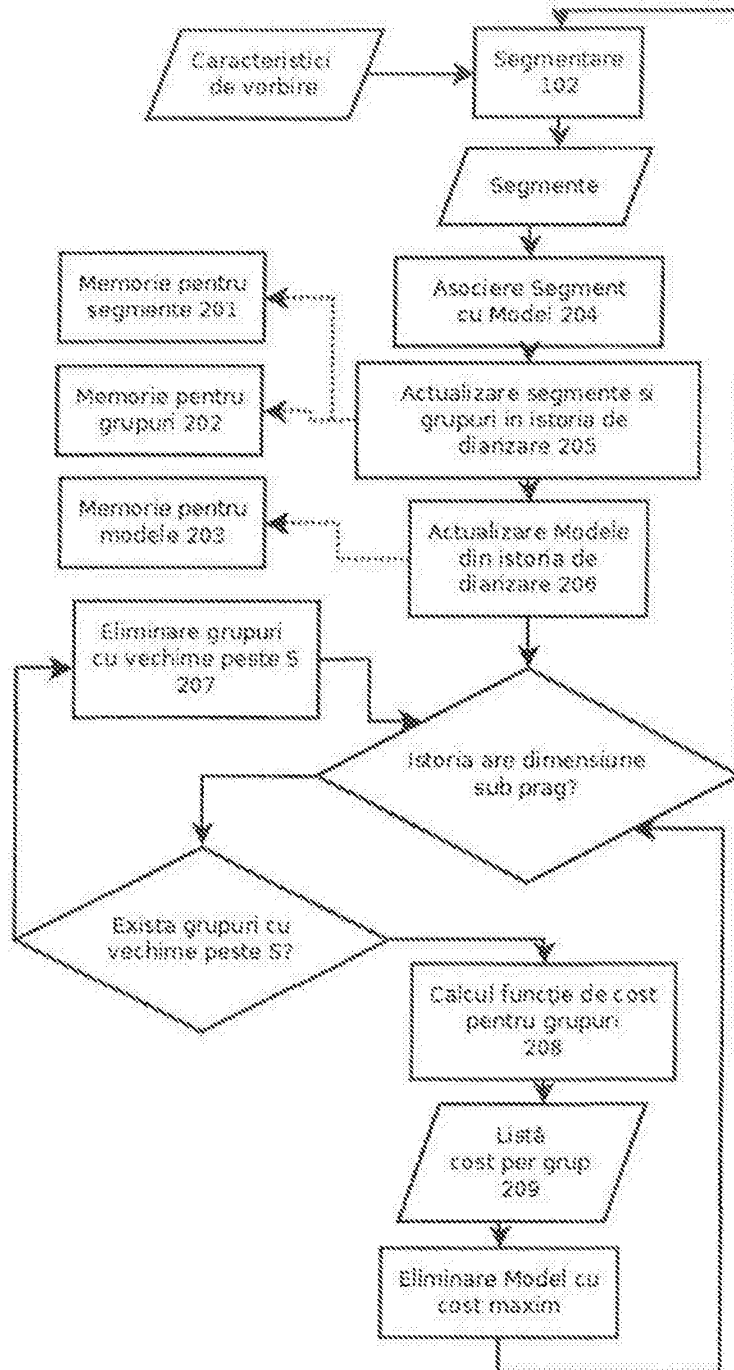


Fig. 2

(51) Int.Cl.

G10L 15/08 (2006.01),

G10L 15/04 (2006.01)

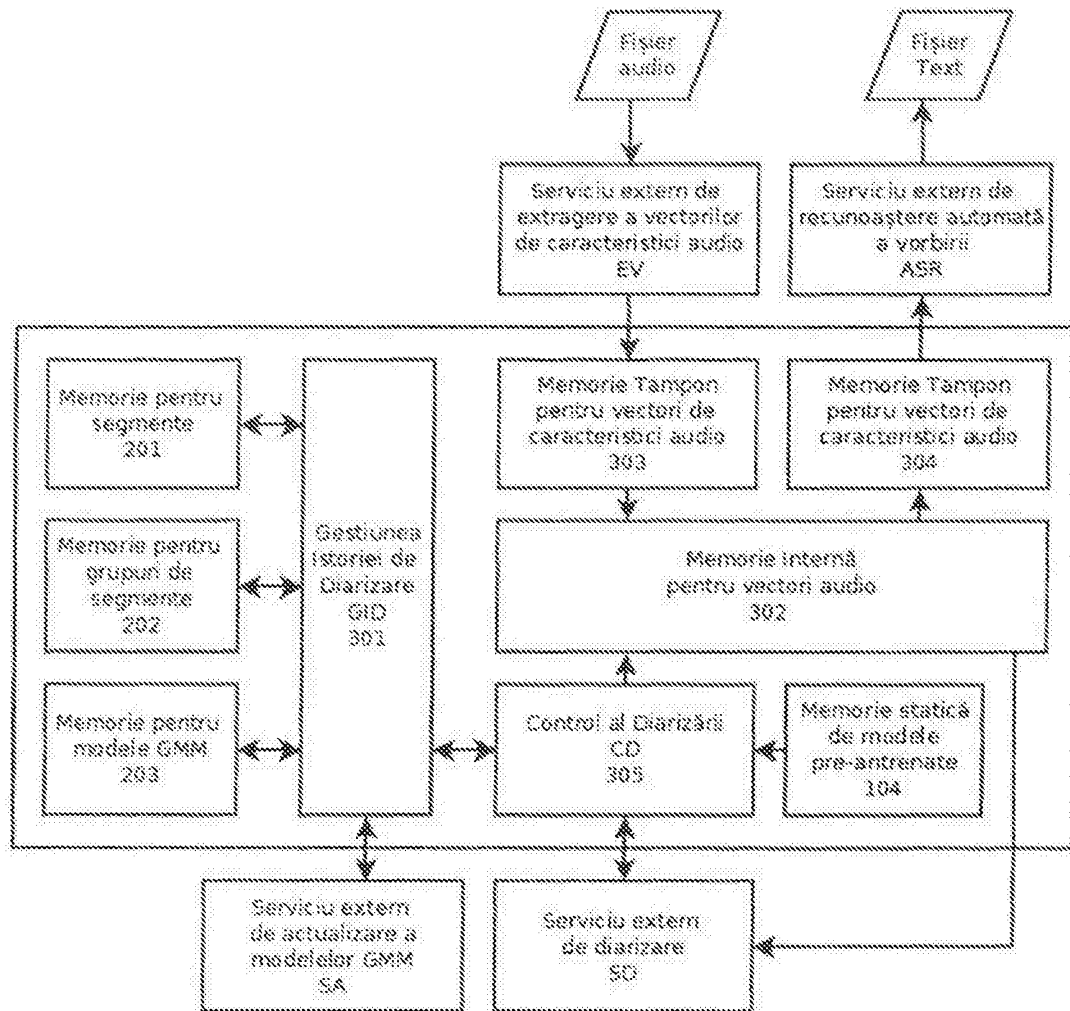


Fig. 3

