



(12) **CERERE DE BREVET DE INVENȚIE**

(21) Nr. cerere: **a 2018 00512**

(22) Data de depozit: **06/07/2018**

(66) Prioritate internă:

**07/07/2017 RO a 2017 00463**  
**06/07/2017 RO a 2017 00458**

(41) Data publicării cererii:

**29/03/2019** BOPI nr. **3/2019**

(71) Solicitant:

• **SOFTWIN S.R.L., STR. MĂGURICEA**  
**NR. 20, BL. 7P, AP. 7, SECTOR 1,**  
**BUCUREȘTI, B, RO**

(72) Inventatori:

• **DIACONESCU ȘTEFAN STELIAN,**  
**STR. TEIUL DOAMNEI, NR.9, BL.32, SC.A,**  
**ET.2, AP.15, S2, BUCUREȘTI, B, RO;**  
• **RIZEA MONICA, CALEA FERENTARI**  
**NR.11, BL.82, SC.3, ET.5, AP.121,**  
**SECTOR 5, BUCUREȘTI, B, RO;**  
• **CRISTESCU MIHAELA,**  
**STR.NICOLAE TITULESCU NR.111,**  
**GĂEȘTI, DB, RO;**  
• **DOROBANȚU LIVIU CĂTĂLIN, BD.UNIRII,**  
**BL.B17, ET.6, AP.25, BUZĂU, BZ, RO;**  
• **DINESCU ADRIAN,**  
**STR. LIVIU REBREANU NR. 27, BL. M12,**  
**SC. 2, ET. 1, AP. 27, SECTOR 3,**  
**BUCUREȘTI, B, RO;**

• **MINCA ANDREI, STR.MĂRȚIȘOR NR.80,**  
**ET.3, AP.12, SECTOR 4, BUCUREȘTI, B,**  
**RO;**

• **FULEA ȘTEFAN, STR.POSTĂVARUL**  
**NR.21, BL.10, SC.1, AP.4, SECTOR 3,**  
**BUCUREȘTI, B, RO;**

• **RUSU MIRCEA SORIN,**  
**INTRAREA BĂRSEI NR.8, BL.G6, SC.1,**  
**ET.5, AP.32, SECTOR 3, BUCUREȘTI, B,**  
**RO;**

• **BURILEANU CORNELIU, BL. LACUL TEI**  
**NR. 75 BL. 16 SC. A AP. 3 SECTOR 2,**  
**BUCUREȘTI, B, RO;**

• **CUCU HORIA, ALEEA POLITEHNICII**  
**NR. 6, BL. 3, SC. 4, ET. 2, AP. 42,**  
**SECTOR 6, BUCUREȘTI, B, RO;**

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

• **RĂDULESCU ANA MONICA,**  
**STR.BARBU LĂUTARU, NR.2, BL.21, SC.A,**  
**ET.3, AP.13, SECTOR 1, BUCUREȘTI, B,**  
**RO**

(74) Mandatar:

**ROMINVENT S.A.,**  
**STR. ERMIL PANGRATTI NR.35,**  
**SECTOR 1, BUCUREȘTI**

(54) **METODĂ ȘI SISTEM DE COMPRESIE A SEMNALULUI  
SONOR, CU UTILIZĂRI ÎN RECUNOAȘTEREA  
CARACTERISTICILOR DIN PLÂNSETUL BEBELUȘILOR  
SAU RECUNOAȘTEREA AUTOMATĂ A VORBIRII BAZATĂ  
PE DETECȚIA ȘI ANALIZA SILABELOR**

(57) Rezumat:

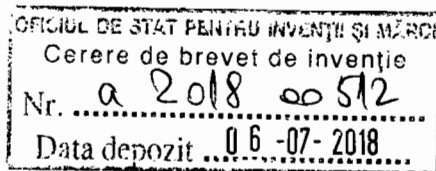
Invenția se referă la o metodă de prelucrare a unui semnal acustic, reprezentat printr-o secvență de eșantioane, care cuprinde: identificarea de mulțimi având ca elemente cel puțin trei invarianți similari, unde un invariant este definit de două sau mai multe eșantioane consecutive, iar invarianții similari au același număr de eșantioane, aceeași amplitudine medie, definită ca o funcție de amplitudinile eșantioanelor din care este format invariantul respectiv, în limitele unei marje de toleranță, același tip de bază, definit ca o funcție care depinde de poziția relativă a eșantioanelor din invariantul respectiv, și invarianții se află la o distanță egală, în limitele unei marje de toleranță, unui față de ceilalți, reprezentată, de exemplu, ca număr de eșantioane dintre respectivi invarianți, și păstrarea unui

număr predefinit de elemente în fiecare mulțime de invarianți similari, restul elementelor fiind eliminate prin invalidarea eșantioanelor care le constituie, pentru a obține un semnal comprimat, constituit din eșantioanele rămase după eliminarea eșantioanelor invalidate. Invenția se referă și la un sistem și la o metodă de recunoaștere a vorbirii, precum și la un sistem și la o metodă de recunoaștere automată a caracteristicilor din plânsetul nou-născuților, în care semnalul sonor captat este prelucrat folosind această metodă de prelucrare a semnalului acustic.

Revendicări: 20  
Figuri: 16

Cu începere de la data publicării cererii de brevet, cererea asigură, în mod provizoriu, solicitantului, protecția conferită potrivit dispozițiilor art.32 din Legea nr.64/1991, cu excepția cazurilor în care cererea de brevet de invenție a fost respinsă, retrasă sau considerată ca fiind retrasă. Întinderea protecției conferite de cererea de brevet de invenție este determinată de revendicările conținute în cererea publicată în conformitate cu art.23 alin.(1) - (3).





Invenția se referă la o metodă și sistem de compresie a semnalului sonor cu numeroase aplicații practice, cum ar fi recunoașterea caracteristicilor din plânsetul bebelușilor sau recunoașterea automată a vorbirii bazată pe detecția și analiza silabelor în acord cu modelul de limbă.

Invenția de față își propune să furnizeze o metodă de compresie a semnalului acustic (sonor), în care semnalul obținut după prelucrare este reprezentat de o cantitate de date redusă semnificativ în comparație cu semnalul inițial. În același timp, metoda de compresie conform prezentei invenții asigură păstrarea acelor date necesare și suficiente pentru clasificarea și/sau compararea caracteristicilor asociate semnalului sonor.

Astfel, într-un prim aspect al invenției, este furnizată o metodă de prelucrare a unui semnal acustic achiziționat. Semnalul acustic (unda sonoră), un semnal continuu, este achiziționat prin convertirea acestuia într-o secvență de eșantioane (semnale discrete). Metoda nouă de compresie a unui semnal acustic, reprezentat printr-o secvență de eșantioane, cuprinde etapele:

a) identificarea de mulțimi având ca elemente cel puțin 3 invarianți similari, unde un invariant este definit de două sau mai multe eșantioane consecutive iar invarianții similari au:

- același număr de eșantioane;
- aceeași amplitudine medie, definită ca o funcție de amplitudinile eșantioanelor din care este format invariantul respectiv, în limitele unei marje de toleranță;
- același tip de bază, definit ca o funcție care depinde de poziția relativă a eșantioanelor din invariantul respectiv și
- se află la o distanță egală, în limitele unei marje de toleranță, unii față de ceilalți, reprezentată de exemplu ca număr de eșantioane dintre numiții invarianți, și

b) păstrarea unui număr predefinit de elemente în fiecare mulțime de invarianți similari, restul de elemente fiind eliminate prin invalidarea eșantioanelor care le constituie;

pentru a obține un semnal comprimat constituit din eșantioanele rămase după eliminarea eșantioanelor invalidate.

Metoda de compresie conform invenției folosește algoritmul VTS conform căruia semnalul acustic, reprezentat printr-o secvență de eșantioane la intrarea în etapa de compresie, este supus unui proces de invalidare a eșantioanelor considerate redundante, pentru a obține un semnal comprimat reprezentat numai de eșantioanele rămase valide. Algoritmul VTS folosit în prezenta invenție este avantajos prin aceea că reduce considerabil cantitatea de date ce trebuie prelucrate ulterior, reducând timpul de procesare a semnalului inițial. În același timp, modul inovativ de determinare a eșantioanelor redundante este esențial pentru asigurarea păstrării acelor date necesare și suficiente pentru clasificarea caracteristicilor asociate semnalului vocal pe baza unui semnal comprimat.

Pentru stabilirea eșantioanelor de invalidat, se formează invarianți din două sau mai multe eșantioane consecutive și se compară între ei pentru identificarea invarianților similari. Prin invarianți similari se înțeleg invarianții care au următoarele proprietăți comune:

- sunt formați din același număr de eșantioane, dar diferă prin cel puțin un eșantion;

- au aceeași amplitudine medie, definită ca o funcție de amplitudinile eșantioanelor din care este format invariantul respectiv, de exemplu ca media aritmetică a amplitudinilor eșantioanelor din care este format invariantul. Amplitudinea medie este considerată aceeași în limitele unei marje de toleranță.
- au același tip de bază, definit ca o funcție care depinde de poziția relativă a eșantioanelor din invariantul respectiv. Prin tip de bază se înțelege în prezenta invenție o mărime extrasă dintr-un semnal sau dintr-un ansamblu de semnale și care nu depinde de anumite caracteristici ale aceluși semnal. Mai preferabil, prin tip de bază se înțelege o mărime extrasă dintr-un semnal sau dintr-un ansamblu de semnale care nu depinde de amplitudine ca valoare absolută ci de variația de amplitudine și care nu depinde de timp ca durată ci analizează timpul ca secvență.
- se află la o distanță egală, în limitele unei marje de toleranță, unii față de ceilalți. Considerând un semnal care variază în timp, eșantionat într-o secvență de eșantioane, cu o anumită frecvență, atunci fiecare eșantion este situat la o anumită poziție în timp în semnalul respectiv. Astfel, distanța între invarianti poate fi exprimată în mai multe moduri, de exemplu în funcție de timp, ca număr de invarianti sau ca număr de eșantioane dintre numiții invarianti, sau mai exact între primul eșantion din fiecare invariant. Considerând o distanță  $D$  între primul și al doilea invariant, în cazul în care se găsește, în urma analizei făcute, un al treilea invariant având aceleași proprietăți ca cele menționate mai sus: număr de eșantioane, amplitudine medie și același tip de bază, aflat la aceeași distanță  $D$ , cu o marjă de eroare, față de cel de-al doilea invariant, atunci cei trei invarianti vor forma o mulțime de invarianti similari cu trei elemente. Procedura se repetă pentru al patrulea invariant și următorii, iar în cazul în care se regăsesc la aceeași distanță  $D$ , se vor adăuga ca elemente în mulțimea menționată.

Se creează astfel mai multe mulțimi de invarianti similari, fiecare mulțime conținând invarianti cu proprietăți comune așa cum au fost menționate mai sus.

În fiecare mulțime de invarianti similari se va păstra numai un număr de elemente sub un număr predefinit. De preferat, numărul de elemente care se păstrează este doi. Restul de elemente sunt invalidate, și anume eșantioanele care formează invariantii similari corespunzători elementelor suplimentare sunt trecuți într-o stare invalidă.

Semnalul comprimat va fi reprezentat de o succesiune de eșantioane rămase valide în urma etapei de comprimare.

Într-un exemplu de realizare preferat, pentru a identifica invariantii similari, se face o analiză în două etape: se selectează invariantii care au următoarele proprietăți la fel: amplitudine medie, în limitele unei marje de toleranță, tip de bază, poziția într-un grup de secvențe repetitive și se verifică distanța dintre ei. Considerând o undă eșantionată, care conține secvențe cu frecvențe aproximativ constante, fiecare frecvență având un anumit număr de perioade, numim secvență repetitivă o perioadă asociată unei frecvențe. Fiecare astfel de secvență repetitivă va avea o anumită dimensiune, reprezentată de numărul de eșantioane din care este formată. Se poate genera astfel un grup de secvențe repetitive, fiecare reprezentând o altă frecvență posibilă a unei unde. Fie grupul de secvențe repetitive constituit dintr-un șir de indecși reprezentați, de exemplu, prin valori numerice crescătoare, fiecărei secvențe repetitive îi este asociat un index de început și un număr de indecși subsecvenți, corespunzători dimensiunii secvenței. Numim poziția în grupul de secvențe repetitive indexul sau indecșii din grupul menționat care poate fi asociat(ți) unui invariant. Într-un exemplu preferat, se stabilește poziția

invariantului într-un grup de secvențe repetitive în funcție de poziția invariantului respectiv în semnalul de analizat, de exemplu pe baza numărului de eșantioane dintre începutul semnalului și invariantul menționat.

O astfel de metodă de compresie permite prelucrarea și analiza rapidă și precisă a diferite semnale sonore și poate fi folosită pentru eficientizarea unor metode și sisteme de recunoaștere și/sau prelucrare a diferite tipuri de semnale acustice, cum sunt plânsetul bebelușilor sau vorbirea.

Astfel, este un alt aspect al prezentei invenții acela de a furniza o metodă rapidă, precisă și eficientă și un sistem de recunoaștere automată a caracteristicilor din plânsetul bebelușilor utilizând metoda de compresie descrisă anterior.

Plânsul copilului, în special la nou-născuți și până în jurul vârstei de 3 luni, este principalul mod de a comunica al acestora. Prin plâns, bebelușii pot exprima diverse stări cum ar fi foame, sete, durere, somnolență, frică, surpriză, furie, stare de frig sau cald, etc.

Există în literatura de specialitate metode de analiză a plânsetului bebelușului prin procesarea semnalului vocal.

În cererea de brevet internațională WO2014036263A1 se prezintă o metodă de analiză a plânsetului nou-născuților prin estimarea frecvenței fundamentale a plânsetului și calculul coeficienților cepstrali (metode de predicție liniară) - în scopul detectării automate a stărilor patologice, sau a stărilor de dezvoltare întârziată.

În cererea de brevet US nr.2003/0033094A1 sunt prezentate metode de analiză EMD (Empirical Mode Decomposition) a semnalelor acustice focalizate pe exploatarea transformatei Hilbert-Huang (HHT). Semnalele provin din surse biologice, ex. de aplicare: sunetul produs de sânge în sistemul circulator, recunoașterea vorbirii sau din surse mecanice ex de aplicare: polizor. Cererea menționată nu propune metode privind detecția automată a stărilor exprimate în plânsetele nou-născuților.

De asemenea, în studiul "Assessment of pain expression in infant cry signals using Empirical Mode Decomposition", Methods InfMed Biosignal Interpretation, Katholieke Universiteit Leuven, 2010, B. Mijović, Silva, B.R.H. Van den Bergh, K. Allegaert, J.M. Aerts, D. Berckmans, S. Van Huffel se arată că procesarea prin EMD a semnalului vocal preluat de la bebeluș este o modalitate eficientă de analiză a plânsetului de durere provocată prin prelevare de sânge, dar concluziile vizează doar estimarea corelației medii dintre variațiile frecvenței fundamentale și anvelopa de contur a semnalului în scopul evaluării gradului de maturizare a sistemului nervos central al bebelușului.

Procesarea EMD (Empirical Mode Decomposition) extrage dintr-un semnal inițial, preprocesat ca semnal de medie nulă (eliminarea componentei continue), setul de funcții esențiale și necesare transmiterii mesajului fonic denumite IMF-uri (Intrinsic Mode Functions). Fiecare IMF ce este extras din semnal, are prin definiție două proprietăți fundamentale:

- numărul de extreme și numărul de treceri prin zero diferă cel mult cu valoarea 1;
- media valorii anvelopelor definite de maximele locale și de minimele locale este 0.

Spre deosebire de funcțiile armonice ale unei decompoziții bazate pe serii Fourier, IMF-urile pot prezenta variații atât în frecvență cât și în amplitudine în timp. În procesul de descompunere, primul IMF conține cele mai înalte frecvențe asociabile semnalului inițial. Fiecare nou IMF procesat/generat conține componente de frecvență din ce în ce mai joasă. Amplitudinile instantanee și frecvențele unghiulare

asociate fiecarui IMF depind de amplitudinea și de faza numărului complex compus de IMF și de transformata Hilbert-Huang (HHT) astfel: partea reală a numărului complex este IMF-ul iar partea imaginară este transformata HHT a respectivului IMF. Frecvența instantanee asociată IMF-ului reprezintă derivata fazei locale, contribuind astfel la formarea tiparului.

Procesarea este repetată pentru fiecare IMF posibil de a fi extras pentru a satisface rafinamentul analizei din perspectiva frecvență/timp asupra semnalului original. Calculul HHT este în principiu o convoluție a unui IMF,  $x(t)$ . Scopul convoluției este de a evidenția proprietățile locale ale semnalului  $x(t)$ . Această descriere locală prezervă structura amplitudinii și frecvenței semnalului  $x(t)$ .

Rolul EMD este de a descompune semnalul complet, într-un set de IMF-uri, extrăgându-l pe fiecare printr-un proces de cernere, denumit "sifting" și de diferențiere față de starea anterioară procesării. Procesul de "sifting" și de diferențiere este repetat până când componenta reziduală  $r(t)$  nu mai conține informație relevantă scopului, respectiv frecvența semnalului  $r(t)$  tinde la valoarea 0 Hz, practic: Freq.  $r(t) < 1/2$  Hz.

Este un scop al prezentei invenții acela de a furniza o metodă și un sistem de recunoaștere automată a caracteristicilor din plânsetul bebelușilor rapidă, precisă și eficace. Acest scop este rezolvat de o metodă de recunoaștere automată a caracteristicilor din plânsetul bebelușilor care cuprinde etapele:

- a) captarea unui semnal vocal de tip plânset de la un subiect și eșantionarea și segmentarea acestuia în segmente de plânset, fiecare segment de plânset fiind definit de un moment de început și de un moment de sfârșit și fiind caracterizat de una sau mai multe frecvențe și amplitudini ale semnalului vocal;
- b) prelucrarea segmentului de plânset din etapa a), prin descompunerea acestuia folosind metoda EMD în funcții IMF și determinarea a cel puțin unei funcții **IMF relevante** prin analiza spectrului de putere al acestora, pentru a obține un segment de plânset recompus prin fuziunea sumativă a numitei cel puțin o funcție **IMF relevantă**;
- c) compresia fiecărui segment de plânset recompus din etapa b), reprezentat printr-o secvență de eșantioane, prin metoda conform revendicării 1;
- d) conversia fiecărui segment de plânset de la punctul c), reprezentat printr-o secvență de eșantioane ale semnalului comprimat, într-un șir de date;
- e) compararea șirului de date al segmentului de plâns obținut în etapa d) cu o bază de date de referință care cuprinde șiruri de date asociate fiecărei caracteristici din plânsetul numitului subiect de recunoscut, pentru a obține **gradul de similaritate S** dintre segmentul de plâns și fiecare caracteristică menționată de recunoscut

În contextul prezentei invenții, termenii "plâns" sau "plânset" sunt folosiți cu același sens.

Caracteristica din plânsetul unui utilizator poate fi în contextul prezentei invenții o nevoie exprimată de un nou-născut, preferabil cu vârsta între 0 și 3 luni, cum ar fi una din următoarele nevoi fundamentale: foame, eructație, disconfort, colici, oboseală, durere. În mod preferat, prin "subiect" se înțelegeun nou-născut sau un bebeluș cu vârsta până la 0-3 luni.

În **etapa a)** are loc captarea unui semnal de tip plânset de la un subiect care a fost eșantionat într-o secvență de eșantioane la o anumită frecvență și segmentat în segmente de plânset. Segmentarea semnalului de tip plânset are ca scop separarea porțiunilor de semnal care reprezintă plâns de alte porțiuni care reprezintă altceva, precum zgomot de fundal, liniște, etc. Preferabil, segmentarea se poate

realiza bazându-ne pe energie, și anume se pleacă de la presupunerea că porțiunile de plâns au energie mai mare decât restul semnalului. Totuși, nici în porțiunile în care nu este plâns, energia nu este nulă. De aceea alegerea pragului de energie pentru determinarea momentului de început și de sfârșit al plânsului va fi un compromis între admiterea zgomotului în analiză (alegere mai liberală care corespunde unui prag mai scăzut) și tăierea extremelor plânsului (alegere mai conservativă care corespunde unui prag mai ridicat). Momentul de început și de sfârșit ale segmentului de plânset se determină prin trecerea peste și respectiv sub un prag energetic. Pragul folosit în această metodă, denumită metoda simplă, este determinat empiric, prin variația lui și estimarea acurateții în fiecare caz.

În mod alternativ, momentul de început al segmentului de plânset se poate determina prin trecerea succesivă peste cel puțin două praguri energetice, ca fiind momentul în care este atins pragul cu energia cea mai scăzută iar momentul de sfârșit de plânset îl constituie momentul în care semnalul scade sub pragul menționat cu energia cea mai scăzută. În cadrul prezentei dezvoltări, această metodă a fost denumită metoda cu întoarcere. Folosind metoda cu întoarcere, nu mai apar problemele de compromis detaliate pentru metoda simplă prezentate mai sus. Semnalul de plâns, ca orice alt sunet uman, începe cu energie mică, energia crescând în timp. Astfel, primele ferestre au energie mai mică decât cea a zgomotului de ambianță. Aceste ferestre sunt tăiate cu metoda simplă. Folosind metoda cu întoarcere, se evită tăierea excesivă din semnalul de plâns care avea loc în metoda simplă.

Etapa b)

Fiecare segment de plânset este prelucrat pentru a obține informațiile folosite ca date de intrare în etapa de comparare.

Pentru acest scop se poate folosi metoda EMD (Empirical Mode Decomposition), ca instrument de analiză timp-frecvență, dovedit eficient, pentru procese/semnale neliniare și netaționare, instrument al cărui principiu algoritmic "data driven" este guvernat de natura datelor analizate "grounded data". EMD este o metodă recent propusă în domeniul analizei biosemnalelor.

Scopul folosirii metodei EMD menționate este de a prelucra segmentul de plânset din etapa a), prin descompunerea acestuia în funcții IMF (Intrinsic Mode Functions) și determinarea funcțiilor IMF relevante prin analiza spectrului de putere al acestora. Prin termenul "funcții IMF relevante" se înțelege acele funcții IMF care sunt esențiale și necesare transmiterii mesajului asociat caracteristicii, de exemplu nevoia sau starea exprimată în plânset care se diferențiază de restul componentelor IMF ale plânsetului, inerente transmiterii fonice a acestuia, cum ar fi:

- purtătoarea fonică – aceasta este simultan: dominant energetică, cu vagă relevanță în ceea ce privește mesajul (nevoia exprimată) și prezintă cele mai înalte frecvențe,
- artefacte asociate exprimării fonice – generate de motricitatea tractului vocal: inspirație, convulsii, etc. Acestea nu sunt dominante energetic și prezintă cele mai joase frecvențe.

Prin "fuziunea sumativă a numitei cel puțin o funcție **IMF relevantă**" se înțelege că în situația în care este de interes o singură funcție IMF relevantă, rezultatul fuziunii sumative este cea unică funcție IMF relevantă iar în cazul în care sunt de interes două sau mai multe funcții IMF relevante, acestea se fuzionează sumativ.

Într-un exemplu de realizare, funcțiile IMF relevante îndeplinesc următoarele condiții de relevanță:

- sunt de ordin mai mare sau egal cu 2, în contextul eşantionării semnalului inițial cu o frecvență de 8000 Hz,
- depășesc un prag de energie determinat iterativ-experimental, exprimat procentual față de energia semnalului nedescompus prin EMD, al magnitudinii spectrului de putere exprimat în domeniul frecvență, prin analiza FFT atât pentru fiecare IMF, cât și pentru semnalul nedescompus în IMF-uri. Într-un exemplu preferat, numitul prag de energie poate fi mai mare de 15%.

Odată stabilite funcțiile IMF relevante, se obține segmentul de plânsset recompus prin fuziunea sumativă a acestora care este preluat în etapa de compresie c).

**Etapa de compresie c)** are ca scop identificarea și reducerea secvențelor repetitive dintr-un segment de plânsset, obținându-se în final o formă redusă a segmentului inițial, care păstrează caracteristici reprezentative pentru numitul segment, necesare comparării și clasificării. Metoda de compresie conform invenției descrisă mai sus folosește algoritmul VTS conform căruia segmentul de plânsset, reprezentat printr-o secvență de eşantioane la intrarea în etapa de compresie, este supus unui proces de invalidare a eşantioanelor considerate redundante, pentru a obține un segment de plânsset comprimat reprezentat numai de eşantioanele rămase valide. Algoritmul VTS folosit în prezenta invenție este avantajos prin aceea că reduce considerabil cantitatea de date ce trebuie prelucrate în etapa de conversie d) și comparare e), reducând timpul de procesare a semnalului inițial. În același timp, modul inovativ de determinare a eşantioanelor redundante este esențial pentru asigurarea păstrării acelor date necesare și suficiente pentru clasificarea caracteristicii sau nevoii asociate semnalului vocal pe baza unui semnal comprimat.

Fiecare segment de plânsset de la punctul c), reprezentat printr-o secvență de eşantioane ale semnalului comprimat, este ulterior convertit în etapa d) într-un format utilizabil în procesul de clasificare efectuând pașii:

1. Conversia plânssetului în invarianți
2. Compresia și ponderarea secvențelor de invarianți

Conform prezentei invenții, segmentul de plânsset de analizat este un vector de eşantioane reprezentate sub forma unor numere întregi pozitive. Acest vector este o reprezentare numerică a unei forme de undă. Prin analiza secvențelor de eşantioane ale unui plânsset se determină invarianții ce compun aceste componente. Preferabil, dacă se consideră un segment de o anumită lungime  $L$ , determinarea invarianților se face analizând grupe de câte  $n$  eşantioane consecutive pornind cu fiecare eşantion al segmentului respectiv, cu excepția ultimelor  $L - ENT(L / n) * n$  puncte, unde  $ENT(L / n)$  reprezintă partea întreagă a lui  $L / n$ .

Numim invariant extins un invariant format din cel puțin două eşantioane consecutive, care are asociate informațiile:

- tipul de bază al invariantului, ca o funcție care nu depinde de amplitudinea sau frecvența segmentului de plânsset studiat dar depinde de poziția relativă a eşantioanelor din invariantul respectiv. Modul de determinare a unui tip de bază este similar celui prezentat în etapa c) de compresie. Tipul de bază referit aici poate fi determinat în același mod sau diferit de tipul de bază din etapa c) de compresie, fie pornind de la un număr diferit de eşantioane pentru fiecare invariant, fie variind codurile asociate fiecărui tip.

- amplitudinea de referință a invariantului, ca funcție de amplitudinile eșantioanelor din care este format invariantul respectiv. Într-un exemplu preferat, amplitudinea de referință este amplitudinea primului eșantion care definește invariantul.

- momentul de referință al invariantului, ca funcție ce caracterizează temporal invariantul. Într-un exemplu preferat, momentul de referință este momentul de apariție a primului eșantion care definește invariantul, relativ la începutul unde.

- un parametru, denumit cod extins sau tip extins, care definește relația dintre amplitudinile de referință a doi invarianti consecutivi, având același tip de bază. Această caracteristică permite o descriere a segmentului de plânsset mai apropiată de forma unei unde.

Într-un exemplu preferat, se compară amplitudinea de referință a fiecărui invariant cu amplitudinea de referință a invariantului imediat anterior de același tip de bază, rezultând parametrul menționat sub forma unei valori numerice  $b$ ,  $b+m$  sau  $b+2m$ , corespunzător situației în care amplitudinea de referință a invariantului este mai mică, egală sau respectiv mai mare decât amplitudinea de referință a invariantului anterior menționat. Prin această operație, fiecare componentă este reprezentată ca o succesiune de coduri care au valori între 0 și  $3 * m - 1$ . Se observă că nu orice secvență de invarianti extinși consecutivi este posibilă. De exemplu, dacă trei invarianti consecutivi au același cod de bază, de exemplu 0, nu este posibil ca al doilea dintre aceștia trei să aibă amplitudinea de referință mai mare decât primul și al treilea mai mică decât al doilea.

Fiecare invariant extins are asociată o pondere, care este stabilită în funcție de proprietățile invariantilor extinși consecutivi având același tip de bază. Într-un exemplu preferat, numitele proprietăți pot fi lungimea formei de undă a invariantului curent, frecvența tipului de bază pe tot semnalul, număr invariantilor consecutivi cu același tip de bază sau tip extins. Într-un exemplu mai preferat, numitele proprietăți se referă la amplitudinile de referință și momentele de referință ale invariantilor extinși consecutivi având același tip de bază, așa cum au fost definite anterior.

Ponderea menționată se obține într-un exemplu de realizare preferat, prin comprimarea și ponderarea secvențelor de invarianti. Comprimarea secvenței de invarianti constă în păstrarea doar a unui singur invariant de un anumit tip dintr-o secvență de invarianti de acel tip. În prezenta metodă, prin ponderare se înțelege atașarea la fiecare invariant rămas a unei ponderi sau cost care depinde de mai multe elemente, după cum se va vedea în continuare.

O metodă preferată de compresie și ponderare cuprinde următorii pași:

a) determinarea unei tabele de tronsoane de invarianti extinși. Fiecare intrare în tabela de tronsoane corespunde unei secvențe de unul sau mai mulți invarianti consecutivi având același tip de bază și conține:

- tipul extins al tronsonului care este tipul extins al invariantilor care compun tronsonul;
- amplitudinea de referință a tronsonului, ca funcție de amplitudinile de referință ale invariantilor extinși din care este compus tronsonul. Mai preferabil, amplitudinea de referință a tronsonului este suma amplitudinilor de referință ale invariantilor din care este compus tronsonul;
- momentul de referință al tronsonului, ca funcție de momentele de referință ale invariantilor extinși din care este compus tronsonul. Mai preferabil, momentul de referință este suma momentelor de referință ale invariantilor care compun tronsonul.



b) sortarea tabelii de tronsoane de la punctul a) folosind drept cheie tipurile de bază extrase din tipurile extinse ale tronsoanelor.

c) împărțirea tabelii de tronsoane în subtronsoane având același cod de bază.

d) determinarea, pentru fiecare subtrason, a lungimii subtrasonului în număr de intrări în tabela de tronsoane.

e) calcularea mediei amplitudinilor de referință ale elementelor subtrasonului pentru fiecare subtrason se obține media, și anume suma amplitudinilor de referință ale elementelor subtrasonului împărțită la lungimea subtrasonului;

f) calcularea mediei momentelor de referință ale elementelor subtrasonului pentru fiecare subtrason, ca suma momentelor de referință ale elementelor subtrasonului împărțită la lungimea subtrasonului;

g) înlocuirea amplitudinii de referință a fiecărui element al subtrasonului cu media amplitudinilor de referință ale elementelor subtrasonului.

h) înlocuirea momentului de referință al fiecărui element al subtrasonului cu media momentelor de referință ale elementelor subtrasonului.

i) sortarea tabelii de tronsoane în ordinea inițială. În acest moment fiecare element al tabelii de tronsoane va conține o amplitudine de referință modificată și un moment de referință modificat.

j) generarea unei noi secvențe de invarianți sub forma unor dubleți (tip de bază, cost) ce conțin, pentru fiecare intrare în tabela de tronsoane:

- tipul extins al invariantului, egal cu tipul extins al elementului de trason

- ponderea sau costul egal cu suma momentului de referință și amplitudinii de referință a elementului de trason

k) ajustarea ponderilor invarianților din secvența obținută conform unei curbe de ajustare. Preferabil, curba menționată este definită în funcție de lungimea L, în număr de invarianți, a segmentului de plânsset, astfel:

- primii  $L / 4$  din invarianți vor avea ponderea înmulțită cu 0.5

- următorii  $L / 2$  din invarianți vor avea ponderea înmulțită cu 1

- restul invarianților vor avea ponderea înmulțită cu 1.5.

În mod alternativ, se pot exprima IMF-urile relevante obținute în etapa b) sub forma unor seturi de simboluri care sunt introduse ca date de intrare în etapa de comparare e), cu sau fără aplicarea metodei de compresie detaliate în etapa c).

Metoda se realizează prin translatarea domeniilor de frecvență relative prin vecinătate din cadrul unui aceluiași IMF relevant într-un set de simboluri, în care comutarea într-un nou simbol se face astfel:

- asocierea fiecărui domeniu a unui simbol, la atingerea unui prag stabilit de limitele minie și maxime a trei domenii de frecvență, punctul central al fiecăruia din cele trei domenii fiind stabilit de cele mai mari trei maxime locale ale întregii funcții spectrale de putere, obținute prin FFT și filtrată prin "moving average" într-o fereastră de 40 de eșantioane cu un pas de deplasare egal cu 1, unde

- la situarea valorii funcției spectrale de putere sub un prag procentual de minim ( $P_{PS}$ ), din media puterii spectrale a respectivului IMF, determinat experimental, de exemplu  $P_{PS} = 10\%$ , codul pentru respectivele zone de frecvență devine P (pauză),

- la conjunctia a două domenii de frecvență învecinate, din funcția spectrală de putere, aflate peste pragul  $P_{PS}$

iar ulterior fiecare simbol fiind translatat în domeniul timp prin transpunere corespunzătoare cu frecvența relativă, pentru fiecare semiperioadă a funcției IMF procesate,

iar în situația în care amplitudinea maximă a unei semiperioade este mai mică decât un prag procentual, determinat experimental, de exemplu 3%, din valoarea amplitudinii medii a semnalului IMF curent procesat, atunci simbolul pentru intervalul de timp asociat semiperioadei respective comută în P.

Preferabil se folosesc patru astfel de simboluri alfabetice: A,B,C,P.

Avantajul este că în acest mod fiecare semnal IMF relevant va fi exprimat autoreferențial, indiferent de caracteristicile tonale individuale ale fiecărui subiect emitent. Condiția necesară pentru coerența comparărilor ce vor urma în etapele următoare este aceea că frecvența de eșantionare a tuturor semnalelor, inclusiv ale celor din baza de referință, trebuie să fie aceeași.

Fiecare segment de plânset este reprezentat printr-o succesiune (un șir) de invarianți sau, în mod alternativ, de simboluri, obținute în etapa precedentă, aceste informații fiind preluate pentru comparare cu o bază de date de referință cu specimene cunoscute reprezentând caracteristicile sau nevoile sub formă de șiruri de invarianți sau de simboluri, procesate în același mod ca cel descris anterior. Preferabil, baza de date de referință menționată conține mai multe specimene care exprimă fiecare caracteristică sau nevoie, constituite pe baza semnalului vocal preluat din mai multe surse. În urma comparației, se determină gradul de similaritate dintre segmentul de plânset analizat și fiecare caracteristică sau nevoie de recunoscut.

Într-un exemplu de realizare preferat, pentru **compararea** informațiilor, se folosește metoda SRA1 și SRA2 descrisă în brevetul de invenție RO 121497, inclusă aici prin referință.

În urma comparației, pentru fiecare segment de plânset de la intrare rezultă un set de scoruri de similaritate relativ la speciimenele din baza de referință, câte un scor de similaritate pentru fiecare specimen din baza de referință. Acest set de scoruri de similaritate față de speciimene referință va fi transformat, de exemplu folosind un algoritm de clasificare-decizie cum ar fi kNN - k Nearest Neighbours, într-un grup (tuplu) de scoruri de similaritate, câte un scor pentru fiecare nevoie/caracteristică din sistem.

În mod avantajos, deoarece procedura de comparare SRA1 descrisă anterior suportă mai multe componente ca și intrare, folosind metoda alternativă, se pot folosi ca date de intrare mai multe seturi de simboluri, corespunzătoare mai multor IMF-uri relevante asociate unui singur segment de plânset. Analiza în paralel a mai multor astfel de IMF-uri ale unui aceluiași plânset, în cadrul SRA, conduce la fenomenul benefic de fuziune între caracteristicile specifice fiecărui IMF.

Astfel caracteristicile specifice unei nevoi sunt evidențiate mai clar prin efectul produs de fuziunea IMF-urilor relevante. În cazul folosirii șirului de invarianți, fuziunea are loc prin însumarea IMF-urilor, anterior compresiei, respectiv în cazul folosirii șirului de simboluri, fuziunea între IMF-uri are loc prin tratarea acestora ca multiple componente de intrare pentru modulul de comparare SRA, comprimate în șiruri de simboluri.

Prin fuziunea IMF-urilor, procentul de identificare corectă a nevoii a crescut cu câteva procente. Acest efect competitiv-cumulativ-cooperant, aplicat în cazul acestei invenții între IMF-urile aceluiași semnal/fenomen, este susținut de teoria Fuziunii Sensoriale (Sensor Fusion).

Modulul de clasificare descris anterior oferă informații necesare clasificării la nivel de segment de plânset. Semnalul preluat de la subiect însă este sub forma unui episod de plânsete, format din mai multe segmente de plânset. Se obțin, deci, vectori pentru gradul de similaritate pentru fiecare astfel de segment în parte. În mod preferat, aceste rezultate se combină conform etapei f) în vederea obținerii unui rezultat mai exact cu privire la caracteristica sau nevoia asociată episodului de plâns.

În mod preferat, decizia cu privire la caracteristica sau nevoia dominantă este luată pe baza cotelor de relevanță pentru fiecare caracteristică sau nevoie existentă.

Într-un alt aspect al invenției, este furnizat un calculator sau rețea de calculatoare care cuprinde mijloace pentru executarea metodei de recunoaștere automată a caracteristicilor din plânsetul bebelușilor descrise mai sus. Prin calculator se înțelege orice ansamblu de componente hardware și componente software ce permite coordonarea și controlul executării operațiilor prin intermediul programelor de calculator.

Într-un alt aspect al invenției, este furnizat un suport non-tranzitoriu care poate fi citit de calculator, care conține instrucțiuni executabile de calculator, care atunci când sunt executate de către procesorul calculatorului efectuează metoda descrisă mai sus. Prin suport non-tranzitoriu se înțelege orice mediu care poate fi citit de calculator și care poate stoca date, inclusiv programe de calculator.

Avantajele prezentei invenții constau în furnizarea unei metode și a unui sistem de recunoaștere a caracteristicilor din plânsetul unui subiect, preferabil un nou-născut, cu un grad de precizie ridicat, care poate să prezinte un rezultat pe baza comparării cu mai multe caracteristici sau nevoi, rapid și eficient.

Totodată, un alt aspect al prezentei invenții este acela de a furniza o metodă și un sistem de analiză acustică și lingvistică a semnalului vocal în vederea **recunoașterii automate a vorbirii** utilizând metoda de compresie conform revendicării 1, care să fie rapidă, precisă și eficace.

Recunoașterea vorbirii reprezintă una din sarcinile majore ale procesării limbajului natural (NLP – natural language processing), prin care un computer realizează transformarea semnalului acustic de vorbire în text. Sistemele de recunoaștere a vorbirii (ASR – automatic speech recognition) cunoscute folosesc metode care cuprind, în general, etapele:

- achiziția de semnale acustice de la unul sau mai mulți vorbitori, sub formă de vorbire continuă sau de unități (sunete, silabe, cuvinte) izolate, opțional cu achiziționarea mai multor variații ale fiecărui specimen de semnal sonor (de exemplu, variații de pronunție, de accent, variații date de factori de mediu cum ar fi reverberații, zgomote de fond etc.);
- modelarea acustică și/sau lingvistică a semnalelor acustice achiziționate în etapa precedentă, cu crearea unor baze de date de specimene de semnal vocal (de exemplu, baze de date de cuvinte, de silabe);
- conversia specimenelor de semnal vocal din baza de date obținută conform etapei precedente, prin extragerea din undele sonore a unor mărimi caracteristice (vectori de trăsături acustice) pentru fiecare specimen de semnal vocal și crearea unei baze de date de mărimi caracteristice asociată bazei de date de segmente de semnal vocal;
- achiziția semnalului vocal de recunoscut, compararea acestuia cu bazele de date create conform etapelor anterioare și asocierea acestuia cu unul sau mai multe specimene de semnal vocal;
- alegerea semnalului vocal asociat optim.

Prin modelare acustică se înțelege, în contextul acestor metode cunoscute din stadiul tehnicii (de exemplu EP1569201, EP1642234) analiza semnalelor acustice în vederea detecției sau marcării în acestea de specimene de semnal vocal reprezentând unități de vorbire, care pot fi sunete (*phones*, *phonemes*), grupuri de sunete (*diphones*, *triphones*, *silabe*, *grupuri de silabe*), cuvinte sau grupuri de cuvinte ale unui vocabular (de exemplu, vocabularul unei anumite limbi).

Modelarea lingvistică presupune, de asemenea în contextul acestor metode cunoscute din stadiul tehnicii, analiza segmentelor de semnal vocal și asocierea unui set de reguli de succesiune a acestora, pe baza regulilor gramaticale ale limbii respective. Modelarea lingvistică se poate face direct printr-un model de tip gramatical, dar, cel mai adesea și mai eficient, se face pe baza unui model statistic, care stabilește ce unitate de vorbire poate urma într-un context dat. Cel mai cunoscut astfel de model lingvistic statistic este modelul n-gram, care prezice elementul următor dintr-o succesiune de n elemente (de exemplu silaba sau cuvântul următor) pe baza frecvențelor relative ale elementelor din succesiune. Pentru o succesiune de două elemente se aplică un digram, pentru trei elemente un trigram șamd.

Un alt scop al prezentei invenții este acela de a furniza o metodă și un sistem de recunoaștere automată a vorbirii rapidă, precisă și eficientă. Acest scop este rezolvat de o metodă de recunoaștere a vorbirii care cuprinde etapele:

**a) achiziție și segmentare:** achiziția de semnale acustice și segmentarea acestora în secvențe de segmente de semnal vocal sau voce, fiecare segment de semnal vocal sau voce fiind definit de un moment de început și de un moment de sfârșit, determinate prin identificarea zonelor cu amplitudine sau energie sau presiune joasă și suficient de lungi;

**b) compresie:** compresia segmentelor de semnal vocal sau voce obținute în etapa precedentă conform metodei din revendicarea 1;

**c) conversie:** conversia fiecărui segment comprimat de semnal vocal sau voce obținut în etapa precedentă în câte un șir de invarianți extinși, rezultând pentru fiecare *secvență de segmente de semnal vocal sau voce* achiziționată câte un *șir de invarianți extinși corespunzător*;

**d) detecția silabelor:** interpretarea fiecărui șir de invarianți extinși obținut în etapa precedentă prin:

d1) parsarea fiecărui șir de invarianți extinși ;

d2) consumarea fiecărui șir de invarianți extinși parsat, ca întreg sau ca mai multe variante de secvență de sub-șiruri de invarianți extinși și marcarea fiecărui sub-șir cu una sau mai multe opțiuni de secvență de silabe, astfel încât fiecare sub-șir de invarianți extinși:

- este marcat parțial cu zgomot pe un sub-sub-șir de la început, de dimensiune determinată experimental în cazul în care nu s-a reușit detecția de silabe, caz în care se reia analiza etapei d) începând cu sfârșitul zonei de marcaj cu zgomot;

- este marcat cu una sau mai multe opțiuni de secvență de silabe în urma unei analize de comparație cu șiruri de invarianți extinși dintr-o bază de referință de șiruri de invarianți extinși asociată unei baze de referință de secvențe de silabe și

- are o dimensiune determinată prin maximizarea unui scor de asemănare a mai multor căutări în jurul dimensiunii medii a șirurilor de invarianți extinși de referință cu care este comparat;

d3) atribuirea de către un clasificator a unui *scor de asemnănare* pentru fiecare asociere și de invarianți extinși – opțiune de secvență de silabe;

**e) analiza silabelor:** analiza fiecărei variante de secvență de silabe obținută în etapa precedentă și asocierea acesteia cu una sau mai multe variante de secvență de cuvinte prin:

- e1) consumarea fiecărei variante de secvență de silabe ca sub-secvențe de silabe, marcarea lor cu variante de cuvinte, și validarea sub-secvențelor de silabe care au fost marcate ca variante de cuvinte;
- e2) verificarea ortografică a sub-secvențelor de silabe ce nu trec de validare și marcarea acestora cu sugestii de variante de cuvinte, pentru fiecare sugestie de variantă de cuvânt calculându-se un *cost de distanță de cuvânt sugestie*, care arată cât de mare este asemnănarea/diferența sub-secvenței de silabe față de cuvântul sugerat;
- e3) atribuirea unui *scor de analiză* pentru fiecare asociere variantă secvență de silabe – variantă de secvență de cuvinte;

**f) alegerea cuvintelor:** alegerea variantei optime de secvență de cuvinte recunoscute prin:

- f1) atribuirea unui *scor de recunoaștere* fiecărei variante de secvență de cuvinte obținute în etapa precedentă prin combinarea scorurilor de asemnănare ale șirurilor de invarianți extinși corespunzători cu scorurile de analiză ale variantelor de secvență de silabe corespunzătoare;
- f2) alegerea variantei optime de secvență de cuvinte pe baza scorului de recunoaștere atribuit.

Invenția de față îmbunătățește performanțele pentru recunoașterea vorbirii prin procesarea semnalului vocal la nivel de silabă. Abordarea propusă se bazează pe detecția și analiza silabelor făcând uz de o bază audio reprezentativă pentru silabele limbii țintă dar și de o ierarhie a silabelor construită peste un vocabular cu acoperire largă.

Într-un prim pas, semnalul acustic de la intrare este supus *analizei acustice* în vederea detecției silabelor. La începutul analizei acustice semnalul acustic este supus unei transformări de *segmentare* ce are ca rezultat stabilirea zonelor potențiale cu semnal vocal. Altfel spus, analiza de segmentare echivalează cu identificarea zonelor de liniște ce depășesc o anumită lungime fixă sau relativă. Segmentele de semnal vocal identificate sunt supuse apoi unei transformări de *compresie* de semnal, folosind metoda de compresie conform invenției, în care se urmărește reducerea variabilă în timp a zonelor de semnal succesive privite ca forme de undă cvasi-asemănătoare. Ulterior, fiecare segment de semnal vocal comprimat este supus unor transformări de *conversie* în șiruri de invarianți extinși sau caracteristici de formă de undă. Fiecare șir de invarianți extinși este trecut apoi printr-un *interpretor* ce consumă șirul în vederea maximizării unui scor de asemnănare prin *comparație* cu șirurile de invarianți extinși ale silabelor sau ale secvențelor de silabe din baza de referință. Interpretorul, în urma comparațiilor de șiruri, preia scorurile de asemnănare și operează o analiză de *clasificare* determinând cel mai bun candidat pentru silabă sau secvența de silabe detectată. Interpretorul de șir de invarianți extinși realizează consumarea șirului și în funcție de ultimele silabe detectate care dictează candidații de silabă următoare pentru detecție (și implicit șirurile de invarianți extinși din baza de referință folosite în comparații) în conformitate cu următoarele structuri de asocieri de silabe specifice cuvintelor (extrase

peste vocabularul limbii), în această ordine: *ierarhia de silabe*, *secvențele de câte două silabe*, *lista de silabe*.

Într-un al doilea pas, o secvență de silabe este supusă *analizei lingvistice* și astfel, prin analiza silabelor se poate deduce secvența de cuvinte corespunzătoare semnalului vocal. Analiza lingvistică începe cu validarea de subsecvențe de silabe din secvența de silabe folosind o ierarhie de silabe. Se vor genera astfel una sau mai multe variante de cuvinte pentru fiecare subsecvență validată. Subsecvențele de silabe ce nu trec validarea prin ierarhia de silabe vor fi supuse unei transformări de ajustare de tip verificare ortografică.

Prin **achiziție de semnale acustice** se înțelege captarea unui semnal acustic (de exemplu cu un microfon) și prelucrarea acestuia pentru a putea fi utilizat pe calculator.

Prin **segmente de semnal vocal sau voce** se înțelege un enunț vorbit mărginit de liniște (de exemplu, unul sau mai multe cuvinte, una sau mai multe propoziții, una sau mai multe fraze rostite împreună).

Conform invenției, prin **silabă** se înțelege orice sunet sau grup de sunete succesive care alcătuiesc un cuvânt sau o parte dintr-un cuvânt. În mod particular, prin silabă se înțelege o unitate sonoră a cuvântului, formată dintr-unul sau mai multe foneme, cu o culme de sonoritate și de expirație și despărțită de celelalte prin depresiuni de sonoritate și de expirație.

Metoda de recunoaștere a vorbirii conform invenției se folosește de **baze de referință** specifice de cunoștințe lingvistice și referințe acustice. Într-un exemplu de realizare preferat, se pot folosi elementele de resurse și instrumente specifice sistemului **GRAALAN** (GRAMmar Abstract LANguage), care reprezintă un sistem complex de procesare a limbajului natural care face uz de un limbaj specializat de reprezentare a cunoștințelor privind o limbă naturală, precum și corespondențele între două limbi, ceea ce permite unui lingvist o descriere eficientă a cunoștințelor privind o limbă naturală și a corespondențelor dintre două limbi. Descrieri mai detaliate ale acestui sistem se regăsesc, de exemplu, în documentele (1) S. Diaconescu: Complex Natural Language Processing System Architecture, in Corneliu Burileanu, Horia-Nicolai Teodorescu (Eds.), *Advances in Spoken Language Technology*, The Publishing House of the Romanian Academy, Bucharest 2007, pp. 228-240; (2) Diaconescu, M. Mateescu, A. Minca, G. Masei, B. Paun: System for Managing a Complex Lexicon Comprising Multiword Expressions and Multiword Expression Templates (US 8,762,131 B1); (3) Diaconescu, I. Dumitrascu, C. Ingineru, O. Bulibasa, M. Rizea, B. Paun: System and methods for Natural Language Processing Including Morphological Analysis, Lemmatizing, Spell Checking, and Grammar Checking (US 8,762,130 B1); (4) S. Diaconescu, C. Ingineru, F. Codiriasu, M. Rizea, O. Bulibasa: General System for Normal and Phonetic Inflection - SpeD 2009 Conference on Speech Technology and Human-Computer Dialogue, 18-21 june, Constanta; (5) Șt. S. Diaconescu, M. M. Rizea, F. C. Codiriasu, M. Ionescu, M. Rădulescu, A. Mincă, Șt. Fulea: Methods for Automatic Generation of GRAALAN-based Phonetic Databases - in the Proceedings of the 8th Conference on Speech Technology and Human-Computer Dialogue (SpeD2015), Bucharest, 2015, pp. 135-142, ISBN: 978-1-4673-7560-3; (6) Șt. S. Diaconescu, M. M. Rizea, M. Ionescu, A. Mincă, Liviu Dorobanțu, Șt. Fulea, M. Rădulescu, Horia Cucu, Dragoș Burileanu: Building a Representative Audio Base of Syllables for Romanian - in the Proceedings of the 9th Conference on Speech Technology and Human-Computer Dialogue (SpeD2017), Bucharest, 2017, încorporate aici prin referință.

Sistemul GRAALAN folosește cunoștințe lingvistice specializate, reguli și instrumente specializate de silabisire și flexiune pe baza cărora se pot genera în mod automat baze de cunoștințe de o largă acoperire a limbii, care cuprind, printre altele, formele sintetice ale limbii (flexiunile mono-cuvânt) dar și formele analitice (flexiunile multi-cuvânt), toate acestea fiind însoțite și de caracteristicile fonetice, morfologice și de silabă. Existența unui vocabular de largă acoperire a limbii, în formă silabisită, permite crearea unei liste reprezentative a silabelor precum și a unei baze audio reprezentative a silabelor. Suplimentar, existența unui vocabular de largă acoperire a limbii permite crearea unei ierarhii de silabe ca un arbore de secvențe de silabe de lungimi variabile. Silaba ca structură acustică împachetează mai multă informație față de un fonem sau grupuri de foneme (difoneme, trifoneme) și în același timp prezintă o complexitate scăzută în alcătuirea unui dicționar de referință (zeci de foneme vs. mii de silabe vs. zeci de mii de morfeme vs. milioane de cuvinte), contribuind astfel la îmbunătățirea recunoașterii unui sistem de recunoaștere a vorbirii atât din punct de vedere al acoperirii cât și al preciziei; silabele sunt privite ca entități rezistente la variații temporale (atunci când vorbirea își schimbă ritmul fonemele sunt afectate în mod diferit în timp ce silabele păstrează proporții relative). Așadar, folosirea unei baze audio reprezentativă de silabe construită peste un vocabular cu acoperire largă contribuie la îmbunătățirea recunoașterii vorbirii atât în acoperire cât și în precizie, iar modelarea lingvistică bazată pe reguli și cunoștințe lingvistice asigură un tratament controlabil și în detaliu al limbii contribuind astfel la creșterea indicatorilor de acoperire și precizie ai recunoașterii.

Construcția și încărcarea bazelor de date specifice cu ajutorul sistemului GRAALAN se realizează cu:

- extragerea de cunoștințe lingvistice și funcții de proofing specifice recunoașterii de vorbire din bazele GRAALAN de cunoștințe lingvistice generale și anume:
  - baza de referință de secvențe de silabe ( *n-grams* de silabe)
    - setul (lista) de silabe specific vocabularului ( *unigrami* de silabe )
    - setul de secvențe de două silabe specifice vocabularului ( *digrami* de silabe)
    - ierarhia de silabe specifică vocabularului ( *n-grams* de silabe de lungime variabilă comprimate sub formă de automat finit determinist minimizat (arbore de tip *trie* )
  - informații adiacente silabelor
    - dicționare-asocieri de transformare text normal de silabă în text fonetic de silabă și invers
  - verificator ortografic (peste vocabularul limbii)
- extragerea de referințe acustice pentru silabe din baza audio reprezentativă a silabelor:
  - unul sau mai multe șiruri de invarianți extinși pentru fiecare silabă.

**Etapa a)** a metodei de recunoaștere a vorbirii (**etapa de achiziție și segmentare**) presupune achiziția și segmentarea semnalului acustic ce are ca rezultat stabilirea zonelor potențiale cu semnal vocal din acest semnal. Altfel spus, analiza de segmentare echivalează cu identificarea zonelor de liniște ce depășesc o anumită lungime fixă sau relativă prin identificarea zonelor cu amplitudine joasă și suficient de lungi.

**Segmentarea** unui semnal de vorbire achiziționat are loc de preferință după cum urmează. Semnalul de vorbire este eșantionat într-o secvență de eșantioane la o anumită frecvență și segmentat în segmente de vorbire. Segmentarea semnalului de vorbire are ca scop separarea porțiunilor de semnal care reprezintă vorbire de alte porțiuni care reprezintă altceva, precum zgomot de fundal, liniște, etc.

Preferabil, segmentarea se poate realiza bazându-ne pe energie, și anume se pleacă de la presupunerea că porțiunile de vorbire au energie mai mare decât restul semnalului.

Segmentul astfel obținut este supus în etapa b) compresiei conform invenției, așa cum a fost descrisă prin algoritmul VTS.

**În etapa b) (etapa de compresie)** are loc compresia segmentelor de semnal vocal sau voce obținute în etapa precedentă prin metoda de compresie a unui semnal sonor conform invenției (descrisă mai sus). Prin această metodă se obține în final o formă redusă a segmentului inițial, care păstrează caracteristici reprezentative pentru numitul segment. Metoda de compresie conform invenției folosește algoritmul VTS, care este avantajos prin aceea că reduce considerabil cantitatea de date ce trebuie prelucrate în etapele următoare.

În **etapa c)** a metodei de recunoaștere a vorbirii (**etapa de conversie**), fiecare segment de semnal vocal sau voce din etapa b), reprezentat printr-o secvență de eşantioane ale semnalului comprimat, este convertit într-un șir de invarianți extinși. Numim invariant extins un invariant format din cel puțin două eşantioane consecutive, care are asociate informațiile:

- tipul de bază al invariantului, ca o funcție care nu depinde de amplitudinea sau frecvența semnalului vocal studiat dar depinde de poziția relativă a eşantioanelor din invariantul respectiv. Modul de determinare a unui tip de bază este similar celui prezentat în etapa c) de compresie. Tipul de bază referit aici poate fi determinat în același mod sau diferit de tipul de bază din etapa c) de compresie, fie pornind de la un număr diferit de eşantioane pentru fiecare invariant, fie variind codurile asociate fiecărui tip.

- amplitudinea de referință a invariantului, ca funcție de amplitudinile eşantioanelor din care este format invariantul respectiv. Într-un exemplu preferat, amplitudinea de referință este amplitudinea primului eşantion care definește invariantul.

- momentul de referință al invariantului, ca funcție ce caracterizează temporal invariantul. Într-un exemplu preferat, momentul de referință este momentul de apariție a primului eşantion care definește invariantul, relativ la începutul unde.

- un parametru, denumit cod extins sau tip extins, care definește relația dintre amplitudinile de referință a doi invarianți consecutivi, având același tip de bază. Această caracteristică permite o descriere a semnalului mai apropiată de forma unei unde.

Într-un exemplu preferat, se compară amplitudinea de referință a fiecărui invariant cu amplitudinea de referință a invariantului imediat anterior de același tip de bază, rezultând parametrul menționat sub forma unei valori numerice  $b$ ,  $b+m$  sau  $b+2m$ , corespunzător situației în care amplitudinea de referință a invariantului este mai mică, egală sau respectiv mai mare decât amplitudinea de referință a invariantului anterior menționat. Prin această operație, fiecare componentă este reprezentată ca o succesiune de coduri care au valori între 0 și  $3 * m - 1$ . Se observă că nu orice secvență de invarianți extinși consecutivi este posibilă. De exemplu, dacă trei invarianți consecutivi au același cod de bază, de exemplu 0, nu este posibil ca al doilea dintre aceștia trei să aibă amplitudinea de referință mai mare decât primul și al treilea mai mică decât al doilea.

Fiecare invariant extins are asociată o pondere, care este stabilită în funcție de proprietățile invarianților extinși consecutivi având același tip de bază. Într-un exemplu preferat, numitele proprietăți pot fi lungimea formei de undă a invariantului curent, frecvența tipului de bază pe tot semnalul, număr



invariantilor consecutivi cu același tip de bază sau tip extins. Într-un exemplu mai preferat, numitele proprietăți se referă la amplitudinile de referință și momentele de referință ale invariantilor extinși consecutivi având același tip de bază, așa cum au fost definite anterior.

În **etapa d)** a metodei de recunoaștere a vorbirii (**etapa de detecție a silabelor**), interpretarea fiecărui șir de invarianti extinși este realizată de către un Interpretor, care este responsabil cu consumarea șirului de invarianti extinși în vederea maximizării unui scor de asemănare prin comparație cu șirurile de invarianti extinși ale silabelor dintr-o bază de referință; în urma acestei comparații interpretorul, pe baza scorurilor de asemănare, operează o analiză de clasificare, determinând cel mai bun candidat pentru o silabă detectată sau o secvență de silabă detectată.

Într-un exemplu de realizare preferat, Interpretorul operează prin utilizarea următoarelor resurse și componente:

– **o bază de referință de șiruri de invarianti extinși:**

Pornind de la o bază audio reprezentativă a silabelor unei limbi, se construiește un set de referințe acustice pentru fiecare silabă sau **și anumite** secvențe de silabe, sub forma unui set de șiruri de invarianti extinși. Pentru aceasta, într-un exemplu de realizare preferat, fiecare înregistrare audio de silabă sau **și anumite** secvențe de silabe este supusă operațiilor de: filtrare, compresie (conform metodei de compresie a semnalului descrisă mai sus) și conversie în șir de invarianti extinși (conform metodei de conversie descrise mai sus). De preferință, fiecare silabă sau **și anumite** secvențe de silabe va avea asociate cel puțin 5 șiruri de invarianti extinși provenite în urma procesării înregistrărilor audio de la cel puțin 5 vorbitori; acestea reprezintă șirurile de invarianti extinși de referință care se folosesc în calcularea scorurilor de asemănare prin compararea cu șirurile de invarianti extinși de intrare precum și în a determina cât să se consume din șirul de invarianti extinși de intrare;

– **o bază de referință de secvențe de silabe;**

Secvențele de silabe sunt folosite de către Interpretor pentru a determina, pe măsură ce consumă șirul de invarianti extinși, silabele candidat următoare pentru detecție și pentru a accesa șirurile de invarianti extinși de referință asociate acestora. Într-un exemplu de realizare preferat, baza de referință de secvențe de silabe cuprinde următoarele structuri de asocieri de silabe extrase peste vocabularul limbii, în această ordine: ierarhia de silabe specifică vocabularului menționată, lista de secvențe de câte două silabe (digrami de silabe) și lista de silabe specifică vocabularului (unigrami de silabe). Aceste trei baze de secvențe de silabe sunt interogate exact în această ordine pentru a reduce spațiul căutărilor. Așadar, confruntarea variantelor de secvență de silabe deja detectate se realizează doar cu ierarhia de silabe iar în caz de eșec se realizează doar cu secvențele de câte două silabe iar în caz de eșec se realizează cu lista de silabe; eșecul este definit atât ca lipsa unui candidat de silabă următoare cât și ca obținerea, în urma comparării, a unui scor sub un prag de asemănare. Așadar, folosirea unei baze audio reprezentativă de silabe construită peste un vocabular cu acoperire largă contribuie la îmbunătățirea recunoașterii vorbirii atât în acoperire cât și în precizie, iar modelarea lingvistică bazată pe reguli și cunoștințe lingvistice asigură un tratament controlabil și în detaliu al limbii contribuind astfel la creșterea indicatorilor de acoperire și precizie ai recunoașterii.

**Ierarhia de silabe** menționată are de preferință o structură tip arbore *trie* de silabe (un automat finit determinist DFA – deterministic finite automaton), și este construită din secvențele de silabe de lungime variabilă specifice cuvintelor din vocabular. Pentru optimizare, această structură se poate minimiza pe sufixe (partea dreaptă a secvențelor) și/sau pe prefixe (partea stângă a secvențelor). Ierarhia de silabe este folosită și în etapa de analiză a variantelor de secvențe de silabe în sensul în care consumă secvența de silabe detectate având astfel rol de validare prin marcarea cuvintelor.

După cum s-a menționat, folosirea de către interpretor a unei baze de referință de secvențe de silabe cum este cea conform sistemului GRAALAN, este în mod special avantajoasă, având ca rezultat creșterea indicatorilor de acoperire și precizie ai recunoașterii, întrucât silaba ca structură acustică împachetează mai multă informație față de un fonem sau grupuri de foneme (difoneme, trifoneme) și în același timp prezintă o complexitate scăzută în alcătuirea unui dicționar de referință (zeci de foneme vs. mii de silabe vs. zeci de mii de morfeme vs. milioane de cuvinte), contribuind astfel la îmbunătățirea recunoașterii unui sistem de recunoaștere a vorbirii atât din punct de vedere al acoperirii cât și al preciziei; în plus, silabele sunt privite ca entități rezistente la variații temporale (atunci când vorbirea își schimbă ritmul fonemele sunt afectate în mod diferit în timp ce silabele păstrează proporții relative).

– **o componentă de comparare**

Această componentă realizează compararea șirurilor de invarianți extinși cu șirurile de invarianți extinși de referință; în urma comparării această componentă atribuie scoruri de asemănare. Fiecare scor de asemănare este specific unui șir de invarianți extinși de referință și are asociat și un identificator de silabă;

– **un clasificator de silabe**

Clasificatorul primește lista de scoruri de asemănare de la componenta de comparare și calculează, utilizând de exemplu un algoritm de tip kNN (*k nearest neighbours*), scoruri generale de asemănare cu fiecare silabă, pe baza cărora determină cel mai bun candidat de silabă.

Într-un exemplu de realizare preferat, etapa d) de detecție de silabe se desfășoară în următorii pași:

1) Interpretorul consumă și marchează, în șirul de invarianți extinși, una sau mai multe variante de prim șir de invarianți extinși și pentru fiecare dintre aceste variante asociază una sau mai multe variante de primă silabă sau primă secvență de silabe astfel: numitul prim șir de invarianți extinși este comparat cu șirurile de invarianți extinși de referință (mai întâi cu ierarhia de silabe, și, în caz de respingere, digramii de silabe, și, în caz de respingere, cu unigramii de silabe) și primește un scor de asemănare pe baza căruia este clasificat de către interpretor (prin componenta sa clasificator) cu o prioritate asociată. Acest pas se realizează în următoarele sub-etape:

1a) Silabele sau secvențele de silabe candidat pentru detecție se stabilesc ca listă de silabe sau secvențe de silabe aflate la început în ierarhia de silabe.

1b0) Dacă lista de candidate este respinsă de clasificator, atunci noile candidate se stabilesc ca setul de digrami de silabe mai puțin silabele sau secvențele de silabe încercate deja

1b) Dacă clasificatorul respinge candidatele de la 1b0), atunci noile candidate se stabilesc ca lista tuturor silabelor (unigrami de silabe) mai puțin silabele încercate deja.

1c) Dacă clasificatorul respinge orice candidat atunci șirul de invarianți extinși este marcat parțial ca zgomot și se revine la pasul 1a),

1d) Ciclul se reia începând cu sfârșitul marcajului de zgomot până când clasificatorul acceptă una sau mai multe **variante de primă silabă** sau primă secvență de silabe cărora le stabilește, pe baza scorurilor de asemănare, câte o prioritate asociată.

2) Interpretorul continuă să consume și să marcheze, în șirul de invarianți extinși și în funcție de silabele detectate deja dar și de șirurile de invarianți extinși de referință ai acestora, una sau mai multe **variante de sub-șir de invarianți extinși următoare** și pentru fiecare dintre aceste variante se asociază una sau mai multe **variante de silabă sau secvență de silabe următoare**. Fiecare variantă de silabă sau secvență de silabe are o prioritate asociată dedusă de către interpretor (prin componenta sa clasificator) din scorul de asemănare în urma comparațiilor șirurilor de invarianți extinși de referință cu șirul de invarianți extinși curent. Analiza se reia în mod similar până la consumarea întregului șir de invarianți extinși. Acest pas se realizează în următoarele subetape:

2a) Silabele sau secvențele de silabe candidat pentru detecție se stabilesc confruntând **ultimele variante de silabă sau variante de secvențe de silabe acceptate** cu ierarhia de silabe;

2b) Dacă lista de candidate este nulă sau clasificatorul le respinge, atunci noile candidate se stabilesc ca silabele obținute confruntând ultimele variante de silabă acceptate cu setul de secvență de două silabe (digrami de silabe) și din rezultat se elimină silabele încercate deja;

2c) Dacă noua listă de candidate este nulă sau clasificatorul le respinge, silabele candidat se stabilesc ca lista tuturor silabelor (unigrami de silabe) mai puțin silabele încercate deja;

2d) Dacă clasificatorul respinge orice candidat atunci sub-șirul de invarianți extinși este marcat parțial ca zgomot și se revine la pasul 1a);

2e) Dacă clasificatorul acceptă una sau mai multe variante de silabă sau secvență de silabă următoare, le stabilește, pe baza scorurilor de asemănare, câte o prioritate asociată și se revine la pasul 2a);

2f) Ciclul se reia în mod similar până la consumarea întregului șir de invarianți extinși.

**Etapa e)** a metodei de recunoaștere a vorbirii (**etapa de analiză a silabelor**) este etapa de transformare a variantelor de secvențe de silabe în una sau mai multe variante de secvență de cuvinte. Într-un exemplu de realizare preferat, în etapa e) a metodei de recunoaștere a vorbirii (etapa de analiză a silabelor), un analizor de silabe primește la intrare o variantă de secvență de silabe rezultată din etapa precedentă, pe care o consumă ca sub-secvețe de silabe și marchează fiecare sub-secvență de silabe cu variante de cuvinte până la terminarea secvenței de silabe. Marcarea sub-secvențelor de silabe ca variante de cuvinte se realizează prin confruntarea sub-secvențelor de silabe cu ierarhia de silabe menționată și, dacă nu se obțin rezultate, cu un verificator ortografic peste sub-secvența curentă. Orice sub-secvență de silabe satisfăcută de un drum în ierarhia de silabe va fi marcată ca o variantă de cuvânt. În caz de insucces se reia confruntarea cu ierarhia de silabe deplasând începutul sub-secvenței de silabe cu o silabă. Analiza continuă pe mai multe căi, câte una pentru fiecare marcaj de cuvânt, până când varianta de secvență de silabe este consumată complet. Variantele de sub-secvențe de silabe care nu au fost marcate drept cuvinte nu sunt validate. Variantele de sub-secvențe de silabe nevalidate sunt analizate de un verificator ortografic, care va oferi una sau mai multe sugestii de corecție (sugestii de

variante de cuvinte), pentru fiecare variantă de cuvânt calculându-se un *cost de distanță de cuvânt sugestie*, care arată cât de mare este asemănarea/diferența sub-secvenței de silabe față de cuvântul sugerat. Etapa se încheie cu atribuirea de către un clasificator a unui *scor de analiză* pentru fiecare asociere variantă secvență de silabe – variantă de secvență de cuvinte.

În etapa f) de alegere a variantei optime de secvență de cuvinte recunoscute (**etapa de alegere a cuvintelor**) se alege, pe baza scorurilor de asemănare și de atribuire acordate în etapele precedente, varianta optimă de secvență de cuvinte care va fi secvența de cuvinte recunoscută de metoda de recunoaștere a vorbirii. Pentru fiecare variantă de secvență de cuvinte obținută în etapa e) se atribuie un scor de recunoaștere, care se calculează prin combinarea scorurilor de asemănare ale șirurilor de invarianți corespunzători obținute în etapa d) cu scorurile de analiză ale variantelor de secvență de silabe corespunzătoare obținute în etapa e).

Într-un exemplu de realizare preferat, listele de scoruri de asemănare și atribuire corespunzătoare variantelor de secvențe de cuvinte sunt transmise unui clasificator, care calculează scoruri generale de asemănare cu fiecare silabă, numite scoruri de clasificare, iar scorul de recunoaștere se calculează prin combinarea următoarelor scoruri:

- funcție a tuturor scorurilor de clasificare de silabă detectată ( Syllable Recognition Rate - **indicatorul SRR** )
- funcție a tuturor scorurilor de clasificare de silabă detectată pe prima și ultima poziție în cuvintele obținute fără verificare ortografică ( Syllable Word Ends Recognition Rate - **indicatorul SweRR** )
- funcție a costurilor de distanță de cuvânt sugestie obținute prin verificare ortografică ( Word Spelling Cost - **indicatorul wsc** ).

De exemplu, variantele de secvență de cuvinte se pot ordona descrescător după indicatorul *SRR* și în caz de nedeterminare sub un anumit prag se pot departaja descrescător folosind indicatorul *SweRR* urmând ca în ultima instanță departajarea să se facă folosind indicatorul *wsc* în mod crescător.

Într-un alt aspect al invenției este furnizat un sistem de recunoaștere a vorbirii care cuprinde mijloace pentru executarea unei metode de prelucrare a semnalelor acustice sau de recunoaștere a vorbirii. Prin sistem de recunoaștere a vorbirii se înțelege orice ansamblu de componente hardware și componente software ce permite coordonarea și controlul executării operațiilor prin intermediul programelor de calculator.

Preferabil, **sistemul de recunoaștere** a vorbirii cuprinde cel puțin un computer configurat să cuprindă sau să fie conectat operabil la:

- mijloace pentru achiziționarea de semnale acustice;
- mijloace pentru segmentarea fiecărui semnal acustic achiziționat în câte o secvență de segmente de semnal vocal sau voce;
- mijloace pentru compresia semnalului sonor segmentat folosind metoda de compresie conform invenției menționată anterior;
- mijloace pentru conversia fiecărei secvențe de segmente de semnal sonor comprimate într-un șir de invarianți extinși ;

- un interpretor al șirului de invariante extinse capabil să realizeze parsarea fiecărui șir de invariante extinse și asocierea șirurilor de invariante extinse cu una sau mai multe variante de silabe sau secvență de silabe, folosind: o bază de referință de șiruri de invariante extinse asociată unei baze de referință de silabe ale unei limbi, mijloace de comparare a șirurilor de invariante extinse de interpretat cu cele din bazele de referință, precum și un clasificator capabil să atribuie scoruri de asemănare pentru fiecare variantă de semnal sonor asociată;
- un analizor de silabe capabil să asocieze fiecare variantă de secvență de silabe generată de interpretor cu una sau mai multe variante de secvențe de cuvinte folosind numita bază de referință de silabe și un verificator ortografic și să aleagă varianta optimă de secvență de cuvinte .

Într-un alt aspect al invenției este furnizat un suport non-tranzitoriu care poate fi citit de calculator, care conține instrucțiuni executabile de calculator, care atunci când sunt executate de către procesorul calculatorului efectuează o metoda de prelucrare a semnalelor acustice așa cum au fost prezentate în prezenta invenție. Prin suport non-tranzitoriu se înțelege orice mediu care poate fi citit de calculator și care poate stoca date, inclusiv programe de calculator.

Se dau în continuare mai multe exemple de realizare ale metodelor și sistemelor conform invenției, exemple care nu sunt limitative și care sunt în legătură și cu figurile 1 – 14 care reprezintă:

Fig. 1 Schema generală a metodei de recunoaștere automată a caracteristicilor din plânsetul bebelușilor

Fig. 2 Schema modului de clasificare a segmentelor de plânset

Fig. 3 Principiul de funcționare a modului de segmentare a semnalului de plânset conform unui exemplu de realizare folosind două praguri energetice

Fig. 4 Grafic ce reprezintă modul de determinare a unui element din tabela \_pAmplitudeTable într-un exemplu preferat de metodă de compresie

Fig. 5 Reprezentare schematică a modului de realizare a unui exemplu preferat de metodă de compresie

Fig. 6 Diagramă reprezentând modul de definire a tipului de bază pentru un invariant format din 3 eșantioane

Fig. 7 Diagramă reprezentând modul de comparare a unui segment de plânset cu una sau mai multe nevoi, evidențiind gradul de similaritate cu fiecare nevoie

Fig. 8 Schema de funcționare StypeFulea a criteriului de agregare a mai multor segmente de plânset conform unui exemplu de realizare

Fig. 9 Schema de funcționare StypeMinca a criteriului de agregare a mai multor segmente de plânset conform unui al doilea exemplu de realizare

Fig. 10 Schema de funcționare StypeRusu a criteriului de agregare a mai multor segmente de plânset conform unui al treilea exemplu de realizare

Fig 11: diagramă generală a metodei și sistemului de analiză acustică și lingvistică a semnalului vocal în vederea recunoașterii vorbirii;

Fig. 12: reprezentare schematică a principiului GRAALAN pentru recunoașterea vorbirii;

Fig. 13: reprezentarea schematică a metodei de funcționare a unui detector de silabe al Sistemului de recunoaștere a vorbirii;

Fig. 14: reprezentarea schematică a metodei de funcționare a unui analizor de silabe al Sistemului de recunoaștere a vorbirii

Fig. 15: diagramă reprezentând algoritmul de derulare a unei metode de compresie preferate conform prezentei invenții

Fig. 16: diagramă reprezentând algoritmul de derulare a etapei de determinare a invarianților similari in metoda preferată de compresie prezentată în Fig 15

O metodă preferată de derulare a unei metode de compresie a unui semnal sonor este descrisă în conformitate cu Fig. 15. Semnalul acustic este eșantionat într-o secvență de eșantioane. Două sau mai multe eșantioane consecutive, de exemplu trei eșantioane consecutive, formează un invariant, astfel încât în etapa S100 semnalul acustic este reprezentat de o secvență S de invarianți de dimensiuni (număr de eșantioane) egale. De exemplu, primele trei eșantioane consecutive formează primul invariant, iar al doilea, al treilea și al patrulea eșantion formează al doilea invariant, și se continuă în acest fel până se obține toată secvența S. În etapa S200 se setează ca invariant curent, primul invariant din secvența de invarianți S. În etapa S300 se caută din secvența S toate mulțimile M care conțin cel puțin trei invarianți similari care sfârșesc cu invariantul curent setat.

Un exemplu de algoritm de determinare a mulțimilor M este redat în Fig. 16. În pasul S301 se setează mulțimea MList, inițial fără niciun element, și anume egală cu mulțimea vidă. Apoi, se stabilește perioada curentă, reprezentând distanța dintre doi invarianți de comparat. Perioada curentă poate fi reprezentată ca număr de eșantioane între doi invarianți de comparat, de exemplu între primul eșantion din fiecare dintre cei doi invarianți sau poate fi reprezentată de numărul de invarianți dintre invarianții de comparat. În pasul S302, perioada curentă se setează ca având valoarea 1.

În pasul S303 se setează ca invariant INVc, invariantul curent preluat din pasul S200 sau S700 așa cum se observă în Fig. 15.

În pasul S304, se stabilește mulțimea M, inițial fără niciun element, și anume egală cu mulțimea vidă.

În pasul S305 se adaugă invariantul INVc în mulțimea M și se trece la pasul S306.

În pasul S306 se caută invariantul INVnext în secvența S, la stânga lui INVc și situat la distanța P' față de INVc. Distanța P' este reprezentată de perioada curentă P, cu o marjă de eroare acceptată, de exemplu  $\pm 1$  unitate. De asemenea, distanța P' poate să nu aibă nicio marjă de eroare acceptată, caz în care va fi egală cu P.

În același mod, invariantul INVnext se poate căuta la dreapta lui INVc și la distanța P', obținându-se în final același rezultat.

În pasul S307, se analizează invariantul INVnext găsit și se verifică dacă are aceeași amplitudine medie și tip de bază ca invariantul curent INVc. În acest exemplu de realizare, amplitudinea medie este dată de media aritmetică a amplitudinilor eșantioanelor din care este format invariantul respectiv, plus sau minus o marjă de eroare acceptată. Într-un exemplu mai preferat, marja de eroare se raportează de fiecare dată la primul element al mulțimii M, reprezentat de invariantul curent inițial. Tipul de bază se determină așa cum este explicat mai jos.

Dacă se determină că INVnext are aceeași amplitudine medie și tip de bază se trece la etapa S308, altfel se trece la etapa S309.

În pasul S308, dacă se determină că INVnext are aceeași amplitudine medie și tip de bază ca invariantul INVc, aceștia sunt considerați invarianți similari. Se setează INVnext ca INVc și se adaugă în mulțimea M, alături de celelalte invarianți similari, urmând ca evaluarea următorului invariant să se facă în comparație cu noul invariant curent INVc.

Dacă se determină că INVnext nu are aceeași amplitudine medie, și anume amplitudinea medie este sub o marjă de toleranță comparativ cu cea a invariantului INVc sau dacă se determină că INVnext nu are același tip de bază cu cel al invariantului INVc, se trece la pasul S309 unde se verifică dimensiunea mulțimii formate M.

Dacă mulțimea M are cel puțin 3 elemente, se trece la etapa S311 iar dacă mulțimea M are mai puțin de 3 elemente, se trece la etapa S310.

În etapa S311, se adaugă mulțimea M în mulțimea MList, după care se trece la etapa S310.

În etapa S310, se incrementează perioada curentă P. În etapa S312, se verifică dacă noua perioadă P este mai mică sau egală cu o perioadă maximă PMax presetată. În cazul în care P este mai mică sau egală cu PMax, se reia etapa S303 cu noua perioadă curentă, altfel se trece la etapa S313.

În etapa S313, se returnează mulțimea MList conținând toate mulțimile M cu cel puțin 3 elemente reprezentate de invarianți similari, pentru a fi prelucrată în algoritmul de derulare a metodei de compresie, și anume în etapa S400 din Fig. 15.

În etapa S400 se verifică dacă există elemente M în mulțimea MList. Dacă DA, se trece la pasul S500, altfel se trece la pasul S600.

În etapa S500, pentru fiecare mulțime M din MList, se determină dacă numărul de elemente **noInvS**, și anume de invarianți similari, este mai mare față de un număr predeterminat K. Într-un exemplu preferat, K poate fi 2 sau 3. În cazul în care numărul de elemente este mai mare decât K, se păstrează primii K invarianți din listă și se marchează pentru eliminare eșantioanele celorlalți invarianți nepăstrați, și anume ultimelor **noInvS-K** elemente din mulțimea M.

În etapa S600, se verifică dacă mai sunt invarianți în secvența S. Dacă DA, se trece la etapa S700, altfel se trece la etapa S800.

În etapa S700, se preia următorul invariant din secvența S și se setează ca invariant curent, urmând să se reia procesul de la etapa S300 cu noul invariant curent.

În etapa S800, se elimină eșantioanele marcate pentru eliminare în etapa S500.

Ca urmare, semnalul sonor prelucrat va fi reprezentat de o nouă secvență de eșantioane necesare și suficiente pentru clasificarea și/sau compararea caracteristicilor asociate semnalului sonor.

În Fig. 1 se observă schema generală a metodei de recunoaștere automată a caracteristicilor din plânsetul unui utilizator, de exemplu se analizează plânsetul unui nou-născut sau bebeluș, pentru a determina cărei nevoi fundamentale îi poate fi asociat. Într-o primă etapă, are loc segmentarea 100 a unui episod de plânset din care rezultă un număr k de segmente de plânset care vor fi ulterior supuse separat metodei de clasificare 200, pentru a obține gradul de asemănare sau similaritate S, pentru fiecare segment, cu fiecare dintre nevoile fundamentale. Datele astfel obținute sunt ulterior prelucrate în etapa de agregare 300 a rezultatelor pentru a identifica în final nevoia fundamentală asociată plânsetului analizat.

Fiecare segment de plânsset 1 până la k va fi supus metodei de clasificare 200, unde metoda 200(1), în urma căreia se obține gradul de similaritate sau asemănare  $S^1$ , corespunde clasificării segmentului 1 și se continuă până la segmentul k, clasificat prin metoda 200(k), unde se obține gradul de similaritate  $S^k$ .

Pentru a decide momentul de început și de final al fiecărui segment de plânsset s-au folosit două metode: metoda simplă sau metoda cu întoarcere. Metoda simplă constă în determinarea momentului de început și de sfârșit ale segmentului de plânsset pr prin trecerea peste și respectiv sub un prag energetic  $E_{\text{prag}}$ .

În mod alternativ, se poate folosi metoda cu întoarcere. Principiul de funcționare al metodei cu întoarcere este prezentat în **Fig. 3**. Plecând de la începutul semnalului, se caută momentul în care energia depășește pragul de jos  $E_{\text{low}}$  și se reține această valoare și momentul respectiv. Dacă următoarele valori ale energiei scad sub  $E_{\text{low}}$ , atunci valorile reținute se șterg și se caută un alt început de plâns. În schimb, dacă în următoarele valori se depășește și pragul de sus  $E_{\text{high}}$ , atunci momentul în care s-a trecut pragul de jos se consideră momentul de început al plânsului. Pentru estimarea momentului de sfârșit al segmentului de plâns se procedează la fel dar analiza se face parcurgând semnalul invers, plecând de la sfârșitul lui. Pragurile  $E_{\text{high}}$  și  $E_{\text{low}}$  folosite pentru sfârșitul plânsului rămân aceleași. Energia semnalului se calculează după împărțirea semnalului în ferestre, de exemplu de 20ms, aplicând formula:

$$\text{energie} = \sum_{n=1}^N x^2(n),$$

unde N este lungimea ferestrei.

Metoda de clasificare 200 presupune o secvență de etape prezentate în detaliu în **Fig. 2**. Fiecare segment de plânsset obținut în etapa de segmentare 100 este prelucrat pentru a obține informațiile folosite ca date de intrare în etapa de comparare 240 și clasificare 250. Pentru acest scop, se pornește cu extragerea dintr-un segment de plânsset a funcțiilor IMF relevante folosind metoda EMD 210. Conform metodei EMD se execută următorii pași, considerând semnalul eșantionat  $y(t)$ :

1. Se determină locația tuturor maximelor  $y_{\text{max}}(t)$  și minimelor  $y_{\text{min}}(t)$   $y(t)$ .
2. Se aproximează prin interpolare polinomială curba "spline", sau prin orice altă metodă de interpolare aleasă ca fiind caracteristică fenomenului analizat, între seturile de valori  $y_{\text{max}}(t)$ ,  $y_{\text{min}}(t)$ .
3. Se calculează media curbei de interpolare între fiecare pereche  $m(t) = (y_{\text{max}}(t) + y_{\text{min}}(t))/2$ .
4. Se diferențiază media,  $m(t)$ . Fie  $d(t) = y(t) - m(t)$ .
5. Se verifică dacă  $d(t)$  este un IMF, și anume îndeplinirea condițiilor de relevanță:
  - numărul de extreme și numărul de treceri prin zero diferă cel mult cu valoarea 1,
  - media valorii anvelopelor definite de maximele locale și de minimele locale este 0.
6. Dacă  $d(t)$  întrunește criteriul de definire a unui IMF, fie  $c_i(t) = d(t)$  și se incrementează  $i$  cu 1. Se extrage reziduala  $r(t) = y(t) - d(t)$ . Dacă  $d(t)$  nu îndeplinește criteriile, se modifică/deplasează fereastra de analiză.
7. Semnalul  $c_i(t)$  este  $\text{IMF}_i$ ; Semnalul  $r(t)$  este reziduala din care se va extrage  $\text{IMF}_{i+1}$ .

Se repetă pașii 1 -5, substituind pentru fiecare repetare  $y(t) = r(t)$ .

Prin însumarea tuturor celor N componente obținute va rezulta semnalul original:

$$y(t) = \sum_{i=1}^N c_i(t) + r_N(t)$$



Separarea funcțiilor IMF relevante se realizează prin analiza și alegerea, pe bază de praguri de energie, determinate iterativ-experimental, a amplitudinii spectrului de putere exprimat în frecvența prin analiza transformatei Fourier rapide (FFT) pentru fiecare IMF, prin referențiere la totalitatea spectrelor FFT ale tuturor IMF-urilor unui aceluiași segment de plâns. Alegerea IMF-urilor relevante se face începând cu cel de ordin 2, care este de regulă cel mai relevant, în contextul eşantionării semnalului inițial cu o frecvență de 8000 Hz, în ordine descrescătoare a magnitudinii spectrului de putere al fiecărui IMF.

Odată stabilite funcțiile IMF relevante, se obține segmentul de plâns recompus prin fuziunea sumativă a numitorilor funcțiilor IMF relevante care este preluat în etapa de compresie .

În **figura 4** este reprezentat schematic modul de funcționare a unei metode preferate de compresie a unui semnal sonor conform invenției. Astfel, secvențele de sunet cu perioadele care sunt aproximativ la fel, vor avea aproximativ aceleași eşantioane, și anume eşantioane cu aceeași valoare a amplitudinii medii și tipului de bază, în limita unor marje de eroare acceptate, situate la distanțe aproximativ egale. Considerând două eşantioane notate cu  $j$  și  $k$ , se va face o analiză a eşantioanelor care au aceeași amplitudine, din punct de vedere al distanței între ele. Analiza se va face pentru amplitudinea medie  $am_j$ ,  $am_k$  a invariantului care se poate construi pornind de la fiecare eşantion. Într-un exemplu de realizare, amplitudinea medie a unui invariant este media obținută din amplitudinile eşantioanelor cu care se calculează invariantul.

Se consideră următoarele date de intrare în analiză:

- a) `_inputBuffer` - Tablou unidimensional cu eşantioanele unde care se prelucrează.
- b) `_pOutputBuffer` - Tabloul unidimensional cu eşantioanele obținute în rezultatul prelucrării.
- c) `_inputBufferIndex` - Index de baleiere a eşantioanelor din tabloul de intrare `_inputBuffer`.
- d) `_pAmplitudeTable` - Tablou tridimensional, în care cele trei dimensiuni sunt:
  - `_amplitudeLevels`: Numărul de nivele de amplitudine ce se va lua în considerare. Se consideră că, pentru problema dată, un număr de 256 de nivele de amplitudine per eşantion este satisfăcător (reprezentare pe 1 octet, eventual cu semn). În principiu ar putea fi 2 octeți deci maximum 65535 nivele de amplitudine.
  - `_invariantAlphabet`: Numărul de tipuri de invarianti (altfel spus numărul de simboluri din alfabetul de invarianti). Sunt diverse valori pe care le va putea lua: 3, 9, 13, 59, etc.
  - `_periodSituation`: Numărul de poziții luate în considerare pentru identificarea începutului unei secvențe de dimensiunea `_invariantConsecutivePoints` începând de la `_inputBufferIndex`. (Aceasta determină, în ultimă instanță un fel de fereastră de analiză a eşantioanelor din tabloul de intrare care conține eşantioanele de analizat).

Un element din tabloul `_pAmplitudeTable` conține 3 informații care se referă la un anumit tip de invariant întâlnit pe o anumită poziție din cele pe care le poate avea un eşantion în cadrul unei ferestre de analiză și având un anumit nivel de amplitudine:

- Momentele de timp `_sampleTime1` și `_sampleTime2` (exprimate în număr de eşantioane) la care s-a întâlnit penultimul și respectiv ultimul eşantion cu caracteristicile corespunzătoare celei curente din `_pAmplitudeTable`. Fiind exprimate în număr de eşantioane, sunt totodată indecși în lista de eşantioane `_inputBuffer`.

- Numărul de eşantioane de tipul respectiv (având caracteristicile corespunzătoare celei curente din `_pAmplitudeTable`) întâlnite în secvența curentă de analiză. Poate avea valorile:
  - 0: Nu s-a întâlnit niciun eşantion corespunzător acestei situații. În acest caz `_sampleTime1` și `_sampleTime2` sunt nuli.
  - 1: S-a întâlnit un eşantion corespunzător acestei situații. În acest caz `_sampleTime1` este indexul acelui eşantion și `_sampleTime2` este nul.
  - 2: S-au întâlnit două eşantioane corespunzătoare acestei situații. În acest caz `_sampleTime1` este indexul primului eşantion și `_sampleTime2` este indexul celui de al doilea.
  - >2: S-au întâlnit mai mult de două eşantioane. În acest caz `_sampleTime1` este indexul penultimului eşantion și `_sampleTime2` este indexul ultimului eşantion.
- e) `_pGroupTable` - Un tablou de „grupuri” va indica, pentru fiecare frecvență  $i$  care se va lua în considerare din totalul de `_groupNumber` grupuri, și anume numărul de frecvențe luate în considerare:
  - $\_entryIndex = i * (i + 1) / 2$
  - $\_entryLength = ((i + 1) * (i + 2) / 2) - (i * (i + 1) / 2)$

Semnificația acestor valori este următoarea: considerând că  $n$  este numărul de eşantioane care descriu o perioadă corespunzătoare unei frecvențe oarecare  $f_n$ . Dacă unda pe care o analizăm conține numai frecvența  $f_n$ , atunci, indiferent unde ar începe o perioadă, fiecare al  $n$ -lea eşantion va fi aproximativ același (va avea aceeași amplitudine). Un eşantion  $k$  al unei analize, va cădea în poziția  $k\%n$  (restul împărțirii lui  $k$  prin  $n$ ) a unei perioade. Dacă vom păstra evidența a  $n$  eşantioane, vom putea, deci verifica faptul că fiecare al  $n$ -lea eşantion are amplitudinea constantă și se găsește la o „distanță” de  $n$  eşantioane de eşantionul corespunzător din perioada anterioară. Dacă vom considera că unda analizată poate avea cel mult  $m$  frecvențe care ne interesează (iar acestea se descriu prin respectiv 1, 2, 3, ...,  $m$  eşantioane), atunci va trebui să păstrăm evidența a  $1 + 2 + 3 + \dots + m = \sum_{i=1}^m i = (1 + m) * m / 2$  eşantioane. Numim grup de eşantioane, setul de eşantioane corespunzător unei frecvențe. În setul de celule care vor memora informații despre aceste  $(1 + m) * m / 2$  eşantioane, celulele corespunzătoare unui grup încep la poziția `_entryIndex+1` și se termină la poziția `_entryIndex+_entryLength`. Ca urmare, poziția corespunzătoare unui eşantion  $k$  din tabloul de intrare `_inputBuffer` va fi  $k\%\_entryLength + \_entryIndex$ . În felul acesta se calculează coordonata `_periodSituation` din tabelul tridimensional `_pAmplitudeTable`.

Preferabil, intrările în `_groupTable` corespund unor frecvențe pentru care perioadele sunt exprimate prin numere întregi (o unitate corespunzând la o valoare reală anumită) care sunt prime între ele. Dacă se pornește de la valoarea 2 atunci perioadele pot fi de exemplu, o secvență de numere prime: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, etc.
- f) `_invariantConsecutivePoints` - Numărul de eşantioane consecutive care se iau în considerare atunci când se calculează un invariant.
- g) `sampleType` - Numărul de ordine al invariantului curent în tabela de invarianți conform alfabetului de invarianți cu care se lucrează.
- h) `index` - Indexul intrării de analiză din `_pAmplitudeTable`.
- i) `entry` - Intrarea curentă din `_pAmplitudeTable` (de la poziția `index`)

- j) `_maximumKeptPeriod` - Numărul maxim de perioade asemănătoare consecutive ce se păstrează în semnalul de ieșire
- k) `Amplitude` - Tipul unei valori de amplitudine (având în general 1 sau 2 octeți).
- l) `averageAmplitude` - Amplitudinea medie a unui invariant (media amplitudinilor eșantioanelor care compun invariantul)
- m) `currentAmplitude` - Valoarea amplitudinii medii a invariantului curent exprimată în niveluri de amplitudine. De exemplu:  

$$\text{currentAmplitude} = \text{averageAmplitude} * \text{_amplitudeLevels} / 2^{\text{sizeof(AMPLITUDE)*8}};$$
- n) `_maxBuffer` - Dimensiune tabloului de intrare `_inputBuffer`.
- o) `_pValidationTable` - Tablou de validare a eșantioanelor de intrare. Preferabil, prelucrarea semnalului de intrare presupune o analiză a fiecărui eșantion din tabloul `_inputBuffer` ca fiind necesar sau nu.

Desfășurarea algoritmului VTS este următoarea:

- execută o serie de filtrări pe eșantioanele din `_inputBuffer` de tip `SMOOTH_FILTER` – filtrare de netezire, `FFT_FILTER` – filtrare Fast Fourier, `SLOPE_FILTER` – filtrare de pantă, etc.
- execută o analiză a secvenței de eșantioane ale semnalului de intrare. Analiza are ca rezultat actualizarea tabloului de validare `_pValidationTable`.
- generează eșantioanele semnalului de ieșire pe baza semnalului de intrare și a tabelii de validare `_pValidationTable`.
- pentru toate eșantioanele din tabloul de intrare `_inputBuffer`, identificate prin indexul curent `_inputBufferIndex`, delimitat de `_maxBuffer`: calculează invariantul care pleacă din indexul curent în tabloul de intrare `_inputBuffer` și care:
  - va avea `_invariantConsecutivePoints` eșantioane, corespunzătoare alfabetului de invariante ales
  - va avea numărul `sampleType` de ordine în lista de simboluri a alfabetului de invariante, unde `sampleType` servește ca prima coordonată pentru tabloul tridimensional de analiză
- calculează amplitudinea medie `averageAmplitude` a acestui invariant, împărțind suma amplitudinilor celor `_invariantConsecutivePoints` amplitudini la `_invariantConsecutivePoints`;
- Calculează amplitudinea invariantului curent în nivele de amplitudine:  

$$\text{currentAmplitude} = \text{averageAmplitude} * \text{_amplitudeLevels} / 2^{\text{sizeof(AMPLITUDE)*8}};$$
 și folosirea `currentAmplitude` ca a doua coordonată în tabloul tridimensional de analiză;

Se execută următoarea buclă după toate grupurile:

- pentru toate grupurile `j` existente în `_groupTable`, determină grupul curent al perioadelor:  
`currentGroup = (_inputBufferIndex % _groupTable[j]._entryLength) + _groupTable[j]._entryIndex;`  
 unde `currentGroup` servește ca a treia coordonată în tabloul tridimensional de analiză

Se determină celula din tabloul tridimensional de analiză:

`entry = _pAmplitudeTable[sampleType][currentRoundedAmplitude][currentGroup];`

Se procesează celula curentă a tabloului de amplitudini: `processAmplitudeEntry(entry)` astfel:

- dacă adică este prima oară când apare un eșantion de tipul curent, eșantionul curent va fi păstrat în ieșire

- altfel, dacă a mai apărut un eșantion de tipul curent, eșantionul va fi păstrat în ieșire
- altfel dacă au apărut cel puțin două eșantioane de tipul curent, dacă intervalul de timp între penultimul și ultimul eșantion marcate în entry este aproximativ egal cu intervalul de timp între eșantionul curent și ultimul eșantion marcat în entry, atunci mută eșantionul ultim din `_currentAmplitudeEntry` în eșantionul penultim din entry și pune eșantionul curent ca eșantion ultim în entry:

Dacă `entry._sampleNumber >= _maximumKeptPeriod` (adică noul eșantion ar trebui eventual șters din eșantioanele de ieșire deoarece el se află în secvență după cele `_maximumKeptPeriod` perioade), pentru un număr `i` de eșantioane consecutive specificat de `_invariantConsecutivePoints`, marchează eșantioanele ca fiind nevalide:

```
_pValidationTable[_inputBufferIndex + i] = fals;
```

Altfel, fiindcă nu au apărut atâtea perioade asemănătoare consecutive câte se dorește a fi păstrate, deci eșantionul curent ar trebui eventual păstrat, dacă el se află la un interval corespunzător de ultimul eșantion de acel tip marcat în entry:

```
entry._sampleNumber = entry._sampleNumber + 1;
```

Altfel, adică eșantionul curent nu este la intervalul corespunzător de față de ultimul eșantion marcat în entry, se consideră că o nouă secvență trebuie analizată, ultima secvență analizată fiind „incompletă”, adică având mai multe perioade consecutive asemănătoare dar mai puține decât `_maximumKeptPeriod`.

Se marchează eșantionul curent ca primul eșantion dintr-o nouă secvență.

Modul de determinare a unui element din tabela `_pAmplitudeTable` este reprezentat în

Într-un exemplu de realizare preferat, operațiile făcute cu amplitudini și momente de timp se fac cu anumite aproximații. Astfel:

- operațiile privind amplitudinile trebuie făcute ținând cont de posibilitatea ca o amplitudine  $x$  să se găsească oriunde între o valoare `_amplitudeApproximationLow` și `_amplitudeApproximationHigh`. Toate valorile de amplitudine ce se găsesc între  $x - \_amplitudeApproximationLow$  și  $x + \_amplitudeApproximationHigh$  se consideră echivalente, deci trebuie făcute încercări pentru toate valorile între aceste limite, cu un pas unitar.

- în mod analog, operațiile privind momentele de timp trebuie facute ținând cont de posibilitatea ca la un moment de timp  $y$  să se găsească oriunde între o valoare `_timeApproximationLow` și `_timeApproximationHigh`. Toate valorile de timp ce se găsesc între  $y - \_timeApproximationLow$  și  $y + \_timeApproximationHigh$  se consideră echivalente, deci trebuie făcute încercări pentru toate valorile între aceste limite, cu un pas unitar.

După ce s-a stabilit numărul de perioade corespunzător fiecărei frecvențe, are loc etapa de reducere a numărului de perioade, asociate fiecărei frecvențe, determinate în etapa precedentă, dacă numărul acestora este mai mare decât un număr prestabilit, la numitul număr prestabilit, așa cum reiese din **fig. 5**. De exemplu, numărul prestabilit poate fi 2.

Pentru a obține datele de intrare pentru etapa de comparare **240 din metoda de recunoaștere automată a caracteristicilor din plânsetul bebelușilor**, se poate folosi metoda de conversie a segmentului de sunet într-un șir de invarianți, parcurgând etapele

- Conversia plânsetului în invarianți
- Compresia și ponderarea secvențelor de invarianți

Fiecare invariant extins este caracterizat de:

- Tipul de bază al invariantului  $inv_i$ .
- Amplitudinea de referință  $ar_i$  a invariantului, ca amplitudinea primului eșantion din cele  $n$  peste care se definește invariantul.
- Momentul de referință  $tr_i$  al invariantului ca momentul de apariție al primului eșantion (din cele  $n$  peste care se definește invariantul), relativ la începutul segmentului de plânsset.

Tipul de bază poate fi determinat în felul următor: cu  $n$  puncte se definesc  $n - 1$  segmente de dreaptă. În cazul în care  $n = 3$ , se formează  $p_0, p_1$  pantele celor două segmente de dreaptă. Se pot defini următoarele  $m = 13$  tipuri de invarianți, după cum reiese și din Fig. 6 cărora le sunt asociate coduri  $cod$  între 0 și  $m - 1$ :

$$cod_0 = 0 : p_0 > 0, p_1 > p_0$$

$$cod_1 = 1 : p_0 > 0, p_1 = p_0$$

$$cod_2 = 2 : p_0 > 0, p_1 < p_0, p_1 > 0$$

$$cod_3 = 3 : p_0 > 0, p_1 = 0$$

$$cod_4 = 4 : p_0 > 0, p_1 < 0$$

$$cod_5 = 5 : p_0 = 0, p_1 > p_0$$

$$cod_6 = 6 : p_0 = 0, p_1 = p_0$$

$$cod_7 = 7 : p_0 = 0, p_1 < p_0$$

$$cod_8 = 8 : p_0 < 0, p_1 > 0$$

$$cod_9 = 9 : p_0 < 0, p_1 = 0$$

$$cod_{10} = 10 : p_0 < 0, p_1 < 0, p_1 > p_0$$

$$cod_{11} = 11 : p_0 < 0, p_1 = p_0$$

$$cod_{12} = 12 : p_0 < 0, p_1 < p_0$$

Se observă că tipul de bază nu depinde de frecvența și amplitudinea semnalului. Dacă semnalul ar fi amplificat de două ori sau s-ar micșora frecvența de două ori, s-ar obține aceleași secvențe de invarianți.

Se observă de asemenea că nu orice secvență de invarianți este posibilă. De exemplu, după un invariant de tip 0 pot veni doar invarianți de tip 0, 1, 2, 3, 4.

Pentru a face o descriere care să fie mai apropiată de forma unei unde, se compară amplitudinea de referință a fiecărui invariant  $ar_i$  al undei cu amplitudinea de referință  $ar_k$  a invariantului imediat anterior de același tip de bază adică  $tip(ar_i) = tip(ar_k) = b_i$ . Dacă nu există invariant anterior de același tip se consideră că  $ar_k = ar_i$ . Apar următoarele 3 situații:

- a)  $ar_i < ar_k$  În acest caz se va acorda invariantului  $i$  codul  $b_i$
- b)  $ar_i = ar_k$  În acest caz se va acorda invariantului  $i$  codul  $b_i + m$
- c)  $ar_i > ar_k$  În acest caz se va acorda invariantului  $i$  codul  $b_i + 2 * m$

Prin această operație fiecare componentă este reprezentată ca o succesiune de coduri denumite coduri extinse sau tipuri extinse și care au valori între 0 și  $3 * m - 1$ .

Următoarea etapă în prelucrarea undei reprezentate acum prin secvențe de invarianți codificați cu coduri extinse constă în comprimarea și ponderarea secvențelor de invarianți urmând pașii

a) se determină o tabelă de tronsoane de invarianți. Fiecare intrare în tabela de tronsoane corespunde la o secvență de unu sau mai mulți invarianți consecutivi având același tip de bază și conține:

- tipul extins al tronsonului care este tipul extins al invarianților care compun tronsonul.
- amplitudinea de referință a tronsonului ca fiind suma amplitudinilor de referință ale invarianților din care este compus tronsonul.
- Momentul de referință al tronsonului ca fiind suma momentelor de referință ale invarianților care compun tronsonul.

b) Se sortează tabela de tronsoane folosind drept cheie tipurile de bază extrase din tipurile extinse ale tronsoanelor.

c) Se împarte tablea de tronsoane în subtronsoane având același cod de bază.

d) Pentru fiecare subtronson se determină lungimea subtronsonului în număr de intrări în tabela de tronsoane.

e) Pentru fiecare subtronson se obține media amplitudinilor de referință ale elementelor subtronsonului (suma amplitudinilor de referință ale elementelor subtronsonului împărțită la lungimea subtronsonului).

f) Pentru fiecare subtronson se obține media momentelor de referință ale elementelor subtronsonului (suma momentelor de referință ale elementelor subtronsonului împărțită la lungimea subtronsonului).

g) Se înlocuiește amplitudinea de referință a fiecărui element al subtronsonului cu media amplitudinilor de referință ale elementelor subtronsonului.

h) Se înlocuiește momentul de referință al fiecărui element al subtronsonului cu media momentelor de referință ale elementelor subtronsonului.

i) Se sortează tabela de tronsoane în ordinea inițială. În acest moment fiecare element al tablei de tronsoane va conține o amplitudine de referință modificată și un moment de referință modificat.

j) Se generează o nouă secvență de invarianți sub forma unor dubleți ( $inv_i, cost_i$ ) ce conțin, pentru fiecare intrare în tabela de tronsoane:

- Tipul extins al invariantului  $inv_i$  (egal cu tipul extins al elementului de tronson).
- Ponderea (costul)  $cost_i$ ; egal cu suma momentului de referință și amplitudinii de referință a elementului de tronson.

k) Se ajustează ponderile invarianților din secvența obținută conform unei curbe (funcții) de ajustare. Dacă lungimea (în număr de invarianți) a unei componente în acest moment este  $L$ , atunci vom considera că această curbă este definită astfel:

- Primii  $L / 4$  din invarianți vor avea ponderea înmulțită cu 0.5.
- Următorii  $L / 2$  din invarianți vor avea ponderea înmulțită cu 1.
- Restul invarianților vor avea ponderea înmulțită cu 1.5.

Se obțin astfel ponderile asociate fiecărui invariant din șirul de invarianți ce compun segmentul de plânset.

În mod alternativ, datele de intrare în metoda de comparare **240** sunt sub forma unui șir de simboluri, așa cum este exemplificat în continuare. Această metodă se realizează prin translatarea a trei domenii de frecvență, relative prin vecinatate, din cadrul unui același IMF relevant obținut în etapa **210**, într-un șir de simboluri (alfabetice de exemplu: A,B,C). În acest mod fiecare IMF relevant va fi exprimat autoreferențial. Analiza și stabilirea vecinătăților în frecvență ce determină comutarea simbolului curent în

altul se face pe bază de praguri stabilite experimental pe spectrul de putere (obținut prin FFT) al respectivului IMF. Domeniile de frecvență aflate sub prag capătă simbolul P (pauză). Cele patru simboluri A,B,C,P, astfel determinate se translatează în domeniul timp, prin transpunere corespunzătoare cu frecvența relativă, pentru fiecare semiperioadă a semnalului IMF procesat. Procesările pasului alternativ se realizează astfel:

- pentru estimarea celor trei domenii de frecvență  $f(A)$ ,  $f(B)$ ,  $f(C)$ , ce pot conține simbolurile A, respectiv B, respectiv C, se consideră primele trei maxime globale ale întregii funcții spectrale de putere, obținute prin FFT asupra unui IMF relevant. Se aproximează maximele ca fiind puncte relativ centrale ale celor trei domenii. Funcția spectrală de putere este anterior filtrată prin "moving average" într-o fereastră de 40 de eșantioane cu un pas de deplasare egal cu 1. Conjunctia a două domenii de frecvență învecinate este determinată de minimul local dintre două maxime consecutive dintre cele trei anterior estimate.

- se asociază provizoriu fiecărui domeniu  $f(X)$  un simbolul X corespunzător, începând cu A și apoi se procedează astfel: la situarea valorii locale a funcției spectrale de putere sub un prag procentual de minim ( $P_{PS}$ ), din media puterii spectrale a respectivului IMF, determinat experimental, de exemplu  $P_{PS} = 10\%$ , codul pentru respectivele zone de frecvență comută în P ( pauză).

- la conjunctia a două domenii de frecvență învecinate, care este determinată de minimul dintre două maxime consecutive dintre cele trei anterior decise, din funcția spectrală de putere, aflate peste pragul  $P_{PS}$ , pot exista doar comutări din A în B sau din B în C.

Fiecare simbol este ulterior translatat în domeniul timp prin transpunere corespunzătoare cu frecvența relativă, pentru fiecare semiperioadă a funcției IMF procesate. În final, după translatarea în domeniul timp, în situația în care amplitudinea maximă a unei semiperioade este mai mică decât un prag procentual  $P_a$ , determinat experimental, de exemplu 3%, din valoarea amplitudinii medii a semnalului IMF curent procesat, atunci simbolul pentru intervalul de timp asociat semiperioadei respective comută în P dacă nu este deja P.

Exemplificativ șirul, exprimat în domeniul frecvență poate arăta astfel: P P P P P A A A A A A A A A A A P P A A B B B B P P P P P C C C C C C C C C C P P, cu lungime fixă pentru oricare IMF, iar exprimat în domeniul timp șirul poate avea aliura:

P C C C C C P C C C C C P C C C C P P P P P P P P A A A A A A B B B B B B A A P A A A B B B B B B P P P P P C C A A P P C C A A A P ...  
....

având o lungime corespondentă duratei segmentului de plânset și în care fiecare simbol este corespondent și descrie o semiperioadă a semnalului IMF procesat.

Astfel, rezultatul este alcătuit dintr-un șir de simboluri poziționate autoreferențial din perspectiva variației atât a frecvenței relative cât și a amplitudinii relative din respectiva funcției intrinseci (IMF).

Datele de intrare folosite în metoda de comparare **240** sunt fie sub forma unui șir de invarianți sau, în mod alternativ, un șir de simboluri, al cărui mod de obținere a fost detaliat anterior.

Într-un exemplu preferat de realizare, metoda SRA1 de comparare a segmentului de plânset cu baza de date de referință se aplică după cum urmează: se pornește de la calculul distanței dintre două șiruri de invarianți folosind distanța Levenshtein, așa cum este descris în Christian Charras, Thierry Lecroq: Sequence comparison, LIR (Laboratoire d'Informatique de Rouen) et ABISS (Atelier Biologie

Informatique Statistique Socio-linguistique) Faculté des Sciences et des Techniques Université de Rouen 76821 Mont-Saint-Aignan Cedex France) cu următoarele precizări:

- a) tipurile de simboluri care se compară conform distanței Levenshtein sunt codurile extinse de invarianți sau setul de simboluri
- b) costurile simbolurilor sunt ponderile (costurile) invarianților în cazul folosirii codurilor extinse de invarianți și respectiv costurile simbolurilor se înlocuiesc cu valoarea 1 în cazul folosirii ca date de intrare a setului de simboluri.
- c) atunci când, conform distanței Levenshtein, se compară două simboluri și acestea sunt identice ca tip, costul rezultat este egal cu modulul diferenței costurilor celor două simboluri.
- d) atunci când, conform distanței Levenshtein, se compară două simboluri și acestea sunt diferite ca tip, atunci costul rezultat este după cum urmează:
  - i. Dacă este vorba de o ștergere, atunci costul rezultat este costul simbolului șters
  - ii. Dacă este vorba de o inserție, atunci costul rezultat este costul simbolului inserat
  - iii. Dacă este vorba despre o substituție, atunci costul rezultat este suma costurilor celor două simboluri
- e) în final, dacă rezultatul (distanța Levenshtein) este  $D$ , atunci distanța luată în considerare (normată)  $d$  va fi:

$$d = 1 - \frac{D}{\sum_i cost_i + \sum_j cost_j}$$

unde  $cost_i$  și  $cost_j$  reprezintă costurile invarianților celor două componente.

Prin compararea celor  $m$  componente ale șirului de invarianți asociat segmentului de plânsset cu respectiv cele  $m$  componente ale șirului de invarianți corespunzător fiecărui specimen din baza de referință rezultă  $m$  distanțe. Combinarea celor  $m$  distanțe se poate face în mai multe feluri, de exemplu, considerând că distanța finală  $d_{SRA1}$  este media celor  $m$  distanțe:

$$d_{SRA1}(Sa, Sb) = \frac{\sum_{i=1}^m d_i}{m}$$

În urma comparației, pentru fiecare segment de plânsset de la intrare rezultă un set de scoruri de similaritate relativ la speciemenele din baza de referință, câte un scor de similaritate pentru fiecare specimen din baza de referință. Acest set de scoruri de similaritate față de speciemen referință va fi transformat, într-un grup (tuplu) de scoruri de similaritate, câte un scor pentru fiecare nevoie/clasă din sistem.

În cazul în care se folosesc mai multe seturi de simboluri pentru același segment de plânsset, care corespund mai multor IMF-uri relevante, comparația se face de  $n$  ori, unde  $n$  este minimum dintre numărul de IMF-uri relevante asociate comparației curente.

Într-un exemplu de realizare, ca algoritm de clasificare-decizie se folosește algoritmul kNN (k-Nearest Neighbours) (cei mai apropiați  $k$  vecini) și o derivație a acestuia kNNperClass (cei mai apropiați  $k$  vecini per clasă/nevoie). Pot apărea următoarele cazuri particulare:



- $k = 0$ , se consideră toate speciemenle din referință
- $k = 1$ , se consideră cel mai bun specimen din referință (ca scor) sau cel mai bun specimen din fiecare clasă/nevoie în cazul derivației kNNperClass

Așa cum reiese din Fig. 7, se definește  $s_{ij}$  ca fiind gradul de similaritate dintre un segment de plânset de intrare și un specimen plânset din baza de referință etichetat cu nevoia  $i$  și având indexul  $j$ .

Se consideră că se vor identifica  $m$  nevoi ale nou-născutului.

#### Descrierea metodei (varianta A):

- a) Se calculează vectorul cu similarități:

$$V = [s_{11}, s_{12}, \dots, s_{1N1}, s_{21}, s_{22}, \dots, s_{2N2}, \dots, s_{m1}, s_{m2}, \dots, s_{mNm}]$$

- b) Se ordonează descrescător vectorul  $V$  și se obține  $V' = [v_1, v_2, \dots, v_k]$ , având cele mai mari  $k$  valori în vectorul  $V$ . Se definește mulțimea:

$$X_i = \{s_{ij} \mid s_{ij} \in V'\}$$

ca fiind mulțimea tuturor gradelor de similaritate calculate între plânsetul de intrare și plânsetele din baza de referință etichetate cu nevoia  $i$ , și care se regăsesc în vectorul  $V'$ .

- c) Se calculează scorul de similiaritate general pentru nevoia  $i$  astfel:

$$S_i = \frac{\sum s_{ij}}{\text{card}(X_i)}, \text{ unde } s_{ij} \in X_i$$

#### Descrierea metodei (varianta B):

În mod alternativ se poate aplica metoda:

- a) Se calculează vectorul cu similarități:

$$V = [s_{11}, s_{12}, \dots, s_{1N1}, s_{21}, s_{22}, \dots, s_{2N2}, \dots, s_{m1}, s_{m2}, \dots, s_{mNm}]$$

- b) Se ordonează descrescător gradele de similaritate pentru nevoia  $i$ :  $s_{ij}$  din vectorul  $V$  și se aleg primele  $k$  valori, obținându-se:  $V_i = [v_1, v_2, \dots, v_k]$ .

- c) Se calculează scorul de similiaritate general pentru nevoia  $i$  astfel:

$$S_i = \frac{\sum v}{\text{card}(V_i)}, \text{ unde } v \in V_i$$

Modulul de similaritate va returna un vector cu scorurile de similaritate generale pentru fiecare nevoie:

$$S = [S_1, S_2, \dots, S_m]$$

Vectorul  $S$  pentru fiecare segment de plânset este folosit ulterior în etapa de agregare a rezultatelor de la mai multe sau de la toate segmentele de plânset care constituie episodul de plânset.

Pentru un episod de plânsete se definesc:

- $L_T$  = sum a lungimilor unităților de plânset care fac parte din episod
- $u$  = numărul de unități de plânsete dintr-un episod de plânsete

Pentru fiecare unitate de plânset avem asociate următoarele informații:

- $L_i$  = lungimea unității de plânset cu indexul  $i$  (în număr de eșantioane)
- un vector cu scorurile de similaritate pentru fiecare nevoie  $S^j = [S_1^j, S_2^j, \dots, S_m^j]$ , asociat unității de plânset  $j$

Au fost definite două criterii de departajare bazate pe vot:

a) **Vot simplu**

Pentru fiecare clasă  $i$  (unde  $i = \overline{1, m}$ ,  $m$  = numărul de nevoi), se calculează relevanța asociată clasei:

$$R_i = \sum_{j=1}^u S_i^j$$

b) **Vot ponderat bazat pe lungime**

Pentru fiecare clasă  $i$  (unde  $i = \overline{1, m}$ ,  $m$  = numărul de nevoi), se calculează relevanța asociată clasei:

$$R_i = \sum_{j=1}^u S_i^j \cdot \frac{L_j}{L_T}$$

Fiind mai multe criterii de departajare, au fost definite mai multe strategii de aplicare ale acestora în vederea obținerii unui rezultat final. Rezultatul poate fi concludent, prin determinarea unei nevoi dominante, neconcludent, sau respins. Un rezultat neconcludent mandatează trierea după un nou criteriu.

Un prim exemplu de realizare este prezentat în **Fig. 8**, în care criteriile sunt aplicate doar câte unul, pentru a obține câte un rezultat. Este notabil că, în această strategie de agregare, rezultatele din aplicarea unui nou criteriu sunt utilizate doar pentru a confirma dominanța clasei cu relevanță maximă din criteriul precedent. Dacă din aplicarea unui nou criteriu reiese dominantă o altă clasă decât cea din criteriul precedent, rezultatul este unul de respingere.

Un al doilea exemplu de realizare este prezentat în **Fig. 9**, în care criteriile de clasificare sunt aplicate preliminar în grup. Relevanțele claselor astfel obținute servesc ca criterii de diferențiere într-o ordonare particularizată. În această ordonare, sunt comparate relevanțele claselor obținute cu unul și același criteriu, diferența dintre ele necesitând a depăși un prag predefinit. Nedepășirea acestui prag în compararea relevanțelor după un criteriu mandatează compararea relevanțelor aceluiași clase obținute însă cu un următorul criteriu. Clasa dominantă este pronunțată cea din fruntea listei ordonate, dacă comparația mai sus amintită o diferențiază de următoarea clasă din listă după cel puțin un criteriu de clasificare, altfel răspunsul este unul de respingere.

Un al treilea exemplu de realizare este prezentat în **Fig. 10**, în care criteriile de clasificare sunt aplicate preliminar în grup. Calculul relevanțelor este parametrizabil, formula implicită este de a însuma pentru fiecare clasă relevanțele calculate conform diferitor criterii. Relevanța clasei dominante trebuie să depășească un prag prestabilit pentru un rezultat concludent al întregii operații de agregare.

Într-un exemplu de realizare preferat, se pot agrega vectorii obținuți pentru același segment de plânset, în urma parcurgerii etapei de comparare 240 și clasificare 250 în primul rând pentru șirul de invarianți asociați segmentului respectiv și în al doilea rând pentru setul sau seturile de simboluri, pentru a obține un procent crescut al gradului de recunoaștere a caracteristicii sau nevoii asociată semnalului inițial.

În **figura 11** este reprezentat schematic un Sistem de recunoaștere a vorbirii conform invenției. Se observă că semnalul vocal achiziționat este transmis unui detector de silabe, care, folosind baze audio (baze de șiruri de invarianți extinși de referință) și n-grams de silabe (baze de secvențe de silabe de referință), realizează detecția unor variante de secvențe de silabe, care sunt apoi analizate de către un

analizor de silabe cu unelte de verificare a limbajului care folosesc baze de cunoștințe lingvistice. În urma acestei analize, sistemul generează variante de text pentru semnalul vocal achiziționat.

În **figura 12** este reprezentat modul cum este aplicat sistemul GRAALAN în metoda de recunoaștere a vorbirii. Se observă că Sistemul de recunoaștere a vorbirii realizează recunoașterea unui semnal vorbit prin accesarea unor baze de date construite conform principiilor sistemului GRAALAN. Astfel, sistemul GRAALAN folosește o bază de cunoștințe lingvistice specializate, care cuprinde reguli și instrumente specializate de silabisire și flexiune. Folsind această bază de cunoștințe lingvistice se construiește rețeaua de silabe peste vocabularul unei limbi (adică baza de secvențe de silabe organizate sub forma ierarhiei de silabe, a digramilor și unigramilor de silabe). Tot pe baza acestor cunoștințe lingvistice se realizează extracția de silabe ale limbii și se alcătuiește lista de silabe sau și lista de secvențe de silabe, atât în alfabet normal cât și în alfabet fonetic. Față de această listă și pe baza unor înregistrări audio, se realizează baza audio de referință a silabelor. Prin prelucrarea înregistrărilor audio din această bază de referință a silabelor, se creează o bază de caracteristici acustice de referință ale silabelor (cum este baza de șiruri de invarianți extinși de referință). Totodată, sistemul GRAALAN cuprinde un verificator care face analiza ortografică peste vocabularul limbii precum și instrumente de modelare lingvistică (analiză semantică și sintactică pe bază de reguli și cunoștințe lingvistice). Sistemul de recunoaștere a vorbirii va putea astfel să analizeze înregistrarea audio de recunoscut față de toate aceste instrumente ale sistemului GRAALAN, ceea ce va duce la generarea unui text recunoscut cu o recunoaștere îmbunătățită atât ca acoperire cât și ca precizie.

În **figura 13** este redată schematic etapa de detecție a silabelor a metodei de recunoaștere a vorbirii conform invenției. Semnalul vocal trece mai întâi printr-o etapă de segmentare, prin care este împărțit în segmente de semnal vocal. Fiecare astfel de segment de semnal vocal trece apoi printr-o etapă de compresie (conform metodei de compresie descrise mai sus). Segmentele de semnal vocal astfel comprimate trec apoi printr-o etapă de conversie, în urma căreia sunt convertite în șiruri de invarianți extinși. Șirurile de invarianți extinși vor fi apoi interpretate de către un interpretor prin comparație cu o bază de referință de șiruri de invarianți extinși și cu o bază de referință de secvențe de silabe (n-grams de silabe), cu ajutorul unui clasificator de silabe. În urma acestei interpretări și comparări, se obțin la sfârșitul etapei de detecție de silabe una sau mai multe variante de secvențe de silabe pentru semnalul vocal de analizat.

În **figura 14** este redată schematic etapa de analiză de silabe a metodei de recunoaștere a vorbirii conform invenției. Variantele de secvențe de silabe rezultate la sfârșitul etapei de detecție a silabelor sunt analizate de către un analizor de silabe, folosind baza de referință de secvențe de silabe (n-grams de silabe) precum și un verificator ortografic cum este cel al sistemului GRAALAN. În urma acestei analize, sunt generate una sau mai multe variante de text.

## REVEDICĂRI

1. Metodă de prelucrare a unui semnal acustic, reprezentat printr-o secvență de eşantioane, care cuprinde:

a) identificarea de mulțimi având ca elemente cel puțin 3 invarianți similari, unde un invariant este definit de două sau mai multe eşantioane consecutive iar invarianții similari au:

- același număr de eşantioane;
- aceeași amplitudine medie, definită ca o funcție de amplitudinile eşantioanelor din care este format invariantul respectiv, în limitele unei marje de toleranță;
- același tip de bază, definit ca o funcție care depinde de poziția relativă a eşantioanelor din invariantul respectiv și
- se află la o distanță egală, în limitele unei marje de toleranță, unii față de ceilalți, reprezentată de exemplu ca număr de eşantioane dintre numiții invarianți, și

b) păstrarea unui număr predefinit de elemente în fiecare mulțime de invarianți similari, restul de elemente fiind eliminate prin invalidarea eşantioanelor care le constituie;

pentru a obține un semnal comprimat constituit din eşantioanele rămase după eliminarea eşantioanelor invalidate.

2. Metodă de recunoaștere a vorbirii care cuprinde etapele:

**a) achiziție și segmentare:** achiziția de semnale acustice și segmentarea acestora în secvențe de segmente de semnal vocal sau voce, fiecare segment de semnal vocal sau voce fiind definit de un moment de început și de un moment de sfârșit, determinate prin identificarea zonelor cu amplitudine sau energie sau presiune joasă și suficient de lungi;

**b) compresie:** compresia segmentelor de semnal vocal sau voce obținute în etapa precedentă conform metodei din revendicarea 1;

**c) conversie:** conversia fiecărui segment comprimat de semnal vocal sau voce obținut în etapa precedentă în câte un șir de invarianți extinși, rezultând pentru fiecare *secvență de segmente de semnal vocal sau voce* achiziționată câte un *șir de invarianți extinși* corespunzător;

**d) detecția silabelor:** interpretarea fiecărui șir de invarianți extinși obținut în etapa precedentă prin:

d1) parsarea fiecărui șir de invarianți extinși ;

d2) consumarea fiecărui șir de invarianți extinși parsat, ca întreg sau ca mai multe variante de secvență de sub-șiruri de invarianți extinși și marcarea fiecărui sub-șir cu una sau mai multe opțiuni de secvență de silabe, astfel încât fiecare sub-șir de invarianți extinși:

- este marcat parțial cu zgomot pe un sub-sub-șir de la început, de dimensiune determinată experimental în cazul în care nu s-a reușit detecția de silabe, caz în care se reia analiza etapei d) începând cu sfârșitul zonei de marcaj cu zgomot;
  - este marcat cu una sau mai multe opțiuni de secvență de silabe în urma unei analize de comparație cu șiruri de invarianți extinși dintr-o bază de referință de șiruri de invarianți extinși asociată unei baze de referință de secvențe de silabe și
  - are o dimensiune determinată prin maximizarea unui scor de asemănare a mai multor căutări în jurul dimensiunii medii a șirurilor de invarianți extinși de referință cu care este comparat;
- d3) atribuirea de către un clasificator a unui *scor de asemănare* pentru fiecare asociere șir de invarianți extinși – opțiune de secvență de silabe;

**e) analiza silabelor:** analiza fiecărei variante de secvență de silabe obținută în etapa precedentă și asocierea acesteia cu una sau mai multe variante de secvență de cuvinte prin:

- e1) consumarea fiecărei variante de secvență de silabe ca sub-secvențe de silabe, marcarea lor cu variante de cuvinte, și validarea sub-secvențelor de silabe care au fost marcate ca variante de cuvinte;
- e2) verificarea ortografică a sub-secvențelor de silabe ce nu trec de validare și marcarea acestora cu sugestii de variante de cuvinte, pentru fiecare sugestie de variantă de cuvânt calculându-se un *cost de distanță de cuvânt sugestie*, care arată cât de mare este asemănarea/diferența sub-secvenței de silabe față de cuvântul sugerat;
- e3) atribuirea unui *scor de analiză* pentru fiecare asociere variantă secvență de silabe – variantă de secvență de cuvinte;

**f) alegerea cuvintelor:** alegerea variantei optime de secvență de cuvinte recunoscute prin:

- f1) atribuirea unui *scor de recunoaștere* fiecărei variante de secvență de cuvinte obținute în etapa precedentă prin combinarea scorurilor de asemănare ale șirurilor de invarianți extinși corespunzători cu scorurile de analiză ale variantelor de secvență de silabe corespunzătoare;
- f2) alegerea variantei optime de secvență de cuvinte pe baza scorului de recunoaștere atribuit.

**3. Metodă conform revendicării 2** în care invariantul extins  $IE_i$  este definit de cel puțin două eșantioane consecutive ale semnalului comprimat iar  $i$  reprezintă poziția invariantului în șirul de invarianți extinși, fiecare invariant extins  $IE_i$  fiind exprimat printr-un **cod extins** și având asociată o **pondere**, codul extins menționat al invariantului extins  $IE_i$ , cuprinzând elementele:

- $inv_i$ , care reprezintă tipul de bază al invariantului  $IE_i$ , ca o funcție care nu depinde de amplitudinea sau frecvența segmentului de semnal vocal sau voce studiat dar depinde de poziția relativă a eșantioanelor semnalului comprimat din invariantul respectiv;

- $ar_i$ , care reprezintă amplitudinea de referință a invariantului  $IE_i$ , ca funcție de amplitudinile eșantioanelor semnalului comprimat din care este format invariantul respectiv;

-  $tr_i$ , reprezentând momentul de referință al invariantului  $i$ , ca funcție ce caracterizează temporal invariantul  $IE_i$ ;

-  $ar_{i-h}$ , reprezentând un parametru care definește relația dintre amplitudinile de referință  $ar_i$  și  $ar_h$  ale invariantului extins  $IE_i$  și respectiv invariantului extins  $IE_h$  imediat anterior, având același tip de bază  $inv_i=inv_h$ ;

iar ponderea asociată invariantului  $IE_i$  fiind stabilită în funcție de proprietățile invariantilor extinși consecutivi având același tip de bază.

4. Metodă conform oricărei revendicării de la 2 la 3, în care în etapa d2) alegerea șirurilor de invarianti extinși de referință pentru compararea cu sub-șirul de invarianti extinși curent a fi marcat, se realizează dintr-un set de șiruri de invarianti extinși asociat unei liste de silabe sau de secvență de silabe candidat, iar această listă de candidați este determinată prin confruntarea variantelor de secvență de silabe deja detectate cu următoarele asocieri de silabe, specifice cuvintelor, construite peste vocabularul limbii: o ierarhie de silabe, secvențe de câte două silabe, secvențe de câte o silabă (lista de silabe).

5. Metodă conform oricăreia dintre revendicările 3 sau 4 în care compararea variantelor de secvență de silabe deja detectate se realizează doar cu ierarhia de silabe iar în caz de eșec se realizează doar cu secvențele de câte două silabe iar în caz de eșec se realizează cu lista de silabe, unde eșecul este definit atât ca lipsa unui candidat de silabă următoare cât și ca obținerea, în urma comparării, a unui scor sub un prag de asemănare.

6. Metodă conform revendicării 5 în care etapa d) de detecție a silabelor cuprinde pașii:

- i) consumarea și marcarea în șirul de invarianti extinși, a uneia sau mai multor de **prim** sub-șir de invarianti extinși și pentru fiecare dintre aceste variante asocierea uneia sau mai multor opțiuni de **primă silabă** sau primă secvență de silabe prin compararea numitului prim sub-șir de invarianti extinși cu șirurile de invarianti extinși de referință asociate opțiunilor;
- ii) stabilirea silabelor sau secvențelor de silabe candidat pentru detecție ca listă de silabe aflate la început în ierarhia de silabe;
- iii) dacă sunt respinse candidatele stabilite în etapa ii), stabilirea noilor candidate ca prima silabă din setul de digrami de silabe sau ca setul de digrami de silabe mai puțin silabele sau secvențele de silabe încercate deja
- iv) dacă sunt respinse noile candidatele stabilite în etapa iii), stabilirea noilor candidate ca lista tuturor silabelor (unigrami de silabe) mai puțin silabele încercate deja;
- v) dacă sunt respinse candidatele stabilite în etapa iv), marcarea sub-șirului de invarianti extinși parțial ca zgomot pe un sub-sub-șir de la început, de dimensiune determinată experimental, și revenirea la pasul i),
- vi) reluarea ciclului i)-v) până când sunt acceptate una sau mai multe opțiuni de primă silabă sau primă secvență de silabe cărora li se stabilește, pe baza scorurilor de asemănare, câte o prioritate asociată;
- vii) continuarea procedurii de consumare și marcarea, în șirul de invarianti extinși și în funcție de silabele detectate, una sau mai multe variante de sub-șir de invarianti extinși **următoare** și asocierea pentru

fiecare dintre aceste variante a uneia sau mai multor opțiuni de **silabă sau secvență de silabe următoare**;

viii) stabilirea silabelor sau secvențelor de silabe candidat următoare pentru detecție confruntând ultimele variantele de silabe sau secvențe de silabe acceptate deja cu ierarhia de silabe;

ix) dacă lista de candidate este nulă sau candidatele sunt respinse, stabilirea silabelor candidat confruntând ultimele variante de silabă acceptate cu setul de secvență de două silabe (digrami de silabe) și din rezultat se elimină silabele încercate deja.

x) dacă noua listă de candidate este nulă sau candidatele sunt respinse, stabilirea silabelor candidat confruntând ultimele variantele de silabă acceptate cu lista tuturor silabelor (unigrami de silabe) mai puțin silabele încercate deja;

xi) dacă sunt respinse și ultimele candidate, atunci sub-șirul de invarianți extinși este marcat parțial ca zgomot pe un sub-sub-șir de la început, de dimensiune determinată experimental, și se revine la pasul i);

xii) dacă una sau mai multe opțiuni de silabă sau secvență de silabe următoare este acceptată, acestora li se stabilește, pe baza scorurilor de asemănare, câte o prioritate asociată și se revine la pasul vi);

xiii) ciclul se reia în mod similar până la consumarea întregului șir de invarianți extinși.

**7. Metodă conform oricăreia dintre revendicările 2 - 6 în care în etapa e1) validarea sub-secvențelor de silabe se realizează prin confruntarea lor cu ierarhia de silabe definită la revendicarea 7.**

**8. Metodă conform oricăreia dintre revendicările 2 - 7 în care în etapa f) scorul de recunoaștere se calculează pe baza unor scoruri obținute din numitele liste de scoruri de asemănare și de atribuire prin calculul unor scoruri generale de asemănare cu fiecare silabă, numite scoruri de clasificare, iar scorul de recunoaștere se calculează prin combinarea următoarelor scoruri:**

- funcție a tuturor scorurilor de clasificare de silabă detectată,

- funcție a tuturor scorurilor de clasificare de silabă detectată pe prima și ultima poziție în cuvintele obținute fără verificare ortografică,

- funcție a costurilor de distanță de cuvânt sugestie obținute prin verificare ortografică.

**9. Sistem care cuprinde mijloace pentru executarea unei metode conform oricărei revendicări de la 1 la 8.**

**10. Suport non-tranzitoriu care poate fi citit de calculator, care conține instrucțiuni executabile de calculator, care atunci când sunt executate de către procesorul calculatorului efectuează metoda conform oricărei revendicări de la 1 la 9.**

**11. Metodă, executată într-un sistem de calcul, de recunoaștere a caracteristicilor din plânsetul unui subiect, care cuprinde pașii:**

a) captarea unui semnal vocal de tip plânsset de la un subiect și eșantionarea și segmentarea acestuia în segmente de plânsset, fiecare segment de plânsset fiind definit de un moment de început și de un moment de sfârșit și fiind caracterizat de una sau mai multe frecvențe și amplitudini ale semnalului vocal;

b) prelucrarea segmentului de plânsset din etapa a), prin descompunerea acestuia folosind metoda EMD în funcții IMF și determinarea a cel puțin unei funcții **IMF relevante** prin analiza spectrului de putere al acestora, pentru a obține un segment de plânsset recompus prin fuziunea sumativă a numitei cel puțin o funcție **IMF relevantă**;

c) compresia fiecărui segment de plânsset recompus din etapa b) utilizând metoda de prelucrare a unui semnal acustic conform revendicării 1, în care semnalul acustic este reprezentat de segmentul de plânsset recompus din etapa b), pentru a obține un segment de plânsset comprimat constituit din eșantioanele rămase după eliminarea eșantioanelor invalidate.

d) conversia fiecărui segment de plânsset de la punctul c), reprezentat printr-o secvență de eșantioane ale semnalului comprimat, într-un șir de date.

e) compararea șirului de date al segmentului de plânsset obținut în etapa d) cu o bază de date de referință care cuprinde șiruri de date asociate fiecărei caracteristici din plânssetul numitului subiect de recunoscut, pentru a obține **gradul de similaritate S** dintre segmentul de plânsset și fiecare caracteristică menționată de recunoscut

**12.** Metodă conform revendicării 11 în care în etapa a) momentul de început și de sfârșit ale segmentului de plânsset se determină prin trecerea peste și respectiv sub un prag energetic.

**13.** Metodă conform revendicării 11 sau 12 în care în etapa a) momentul de început al segmentului de plânsset se determină prin trecerea succesivă peste cel puțin două praguri energetice, ca fiind momentul în care este atins pragul cu energia cea mai scăzută iar momentul de sfârșit de plânsset îl constituie momentul în care scade sub pragul cu energia cea mai scăzută.

**14.** Metodă conform oricărei revendicări de la 11 la 13 în care în etapa b) funcțiile IMF relevante se obțin folosind metoda EMD, prin extragerea din fiecare segment de plânsset menționat, preprocesat ca semnal de medie nulă, a funcțiilor IMF care îndeplinesc condițiile de relevanță:

- sunt de ordin mai mare sau egal cu 2, în contextul eșantionării semnalului inițial cu o frecvență de 8000 Hz,
- depășesc un prag de energie determinat iterativ-experimental, exprimat procentual față de energia semnalului nedescompus prin EMD, al magnitudinii spectrului de putere exprimat în domeniul frecvență, prin analiza FFT atât pentru fiecare IMF, cât și pentru semnalul nedescompus în IMF-uri, numitele IMF-uri relevante fiind ulterior fuzionate sumativ pentru a obține segmentul de plânsset recompus.



15. Metodă conform oricărei revendicări de la 11 la 14 în care șirul de date obținut în etapa d) este un șir de invarianți extinși  $IE_i$ , unde cel puțin două eșantioane consecutive ale semnalului comprimat definesc un invariant extins iar  $i$  reprezintă poziția invariantului în șirul de invarianți, fiecare invariant extins  $IE_i$  fiind exprimat printr-un **cod extins** și având asociată o **pondere**, codul extins menționat al invariantului extins  $IE_i$ , cuprinzând elementele:

-  $inv_i$ , care reprezintă tipul de bază al invariantului  $IE_i$ , ca o funcție care nu depinde de amplitudinea sau frecvența segmentului de plânsset studiat dar depinde de poziția relativă a eșantioanelor semnalului comprimat din invariantul respectiv

-  $ar_i$ , care reprezintă amplitudinea de referință a invariantului  $IE_i$ , ca funcție de amplitudinile eșantioanelor semnalului comprimat din care este format invariantul respectiv

-  $tr_i$ , reprezentând momentul de referință al invariantului  $i$ , ca funcție ce caracterizează temporal invariantul  $IE_i$

-  $ar_{i-h}$ , reprezentând un parametru care definește relația dintre amplitudinile de referință  $ar_i$  și  $ar_h$  ale invariantului extins  $IE_i$  și respectiv invariantului extins  $IE_h$  imediat anterior, având același tip de bază  $inv_i = inv_h$

iar ponderea asociată invariantului  $IE_i$  fiind stabilită în funcție de proprietățile invariantilor extinși consecutivi având același tip de bază.

16. Metodă conform oricărei revendicări de la 11 la 14 în care șirul de date obținut în etapa d) este un șir de simboluri obținut prin procesarea funcțiilor IMF relevante obținute în etapa b) înainte de fuziunea sumativă a acestora, prin translatarea domeniilor de frecvență relative prin vecinătate din cadrul unui aceluiași IMF relevant într-un set de simboluri, în care comutarea într-un nou simbol se face astfel:

- asocierea fiecărui domeniu a unui simbol, la atingerea unui prag stabilit de limitele minie și maxime a trei domenii de frecvență, punctul central al fiecăruia din cele trei domenii fiind stabilit de cele mai mari trei maxime locale ale întregii funcții spectrale de putere, obținute prin FFT și filtrată prin "moving average" într-o fereastră de 40 de eșantioane cu un pas de deplasare egal cu 1, unde

- la situarea valorii funcției spectrale de putere sub un prag procentual de minim ( $P_{PS}$ ), din media puterii spectrale a respectivului IMF, determinat experimental, de exemplu  $P_{PS} = 10\%$ , codul pentru respectivele zone de frecvență devine P (pauză),

- la conjunctia a două domenii de frecvență învecinate, din funcția spectrală de putere, aflate peste pragul  $P_{PS}$

iar ulterior fiecare simbol fiind translatat în domeniul timp prin transpunere corespunzătoare cu frecvența relativă, pentru fiecare semiperioadă a funcției IMF procesate,

iar în situația în care amplitudinea maximă a unei semiperioade este mai mică decât un prag procentual, determinat experimental, de exemplu 3%, din valoarea amplitudinii medii a semnalului IMF curent procesat, atunci simbolul pentru intervalul de timp asociat semiperioadei respective comută în P.

17. Metodă conform oricărei revendicări 11-16 care mai cuprinde etapa f) de agregare a rezultatelor de la segmentele de plânsset care alcătuiesc semnalul vocal captat în etapa a) prin asocierea fiecărui segment de plânsset a unui vector cu toate gradele de similaritate dintre segmentul de plânsset respectiv și fiecare

caracteristică de recunoscut determinate în etapa e) și prelucrarea vectorilor obținuți pentru a asocia o caracteristică semnal vocal de tip plânsset.

**18.** Metodă conform oricărei revendicări 11-17 în care caracteristica menționată este o nevoie fundamentală a unui bebeluș selectată din lista: foame, eructație, disconfort, colici, oboseală, durere.

**19.** Calculator sau rețea de calculatoare care cuprinde mijloace pentru executarea unei metode conform oricărei revendicări de la 1 la 18.

**20.** Suport non-tranzitoriu care poate fi citit de calculator, care conține instrucțiuni executabile de calculator, care atunci când sunt executate de către procesorul calculatorului efectuează metoda conform oricărei revendicări de la 1 la 18.

Figuri

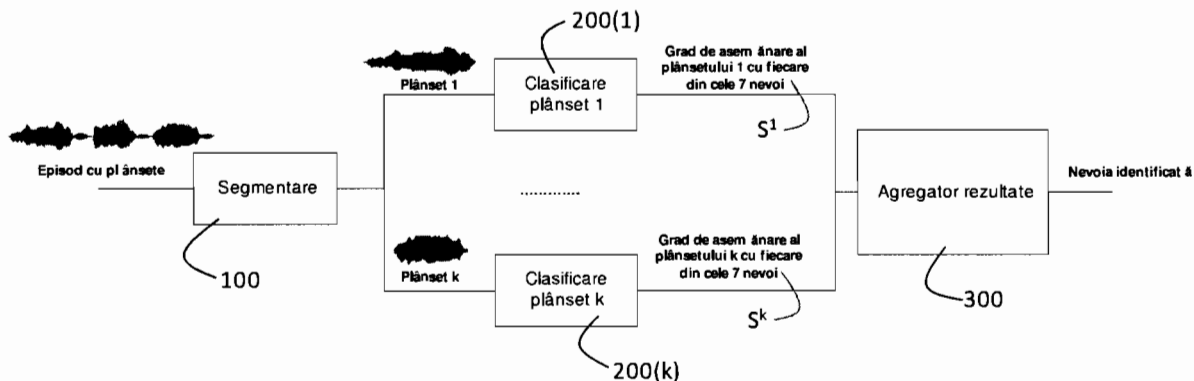


Fig 1

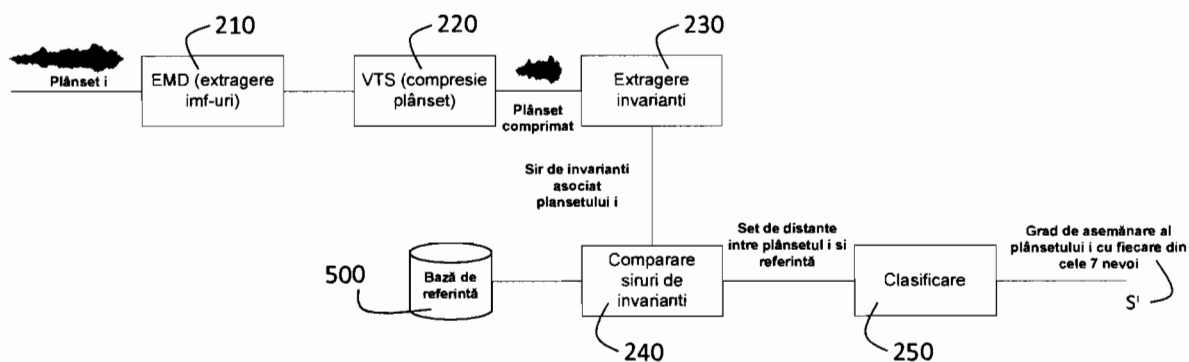


Fig 2

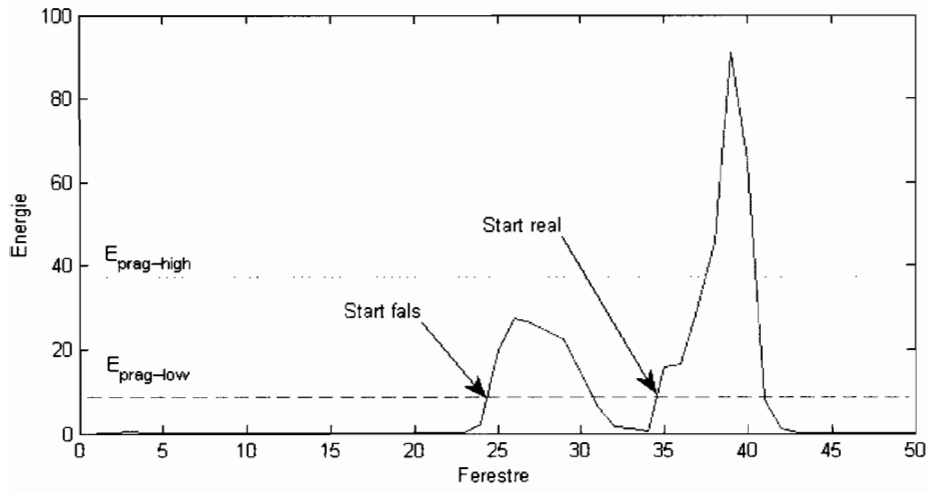


Fig 3

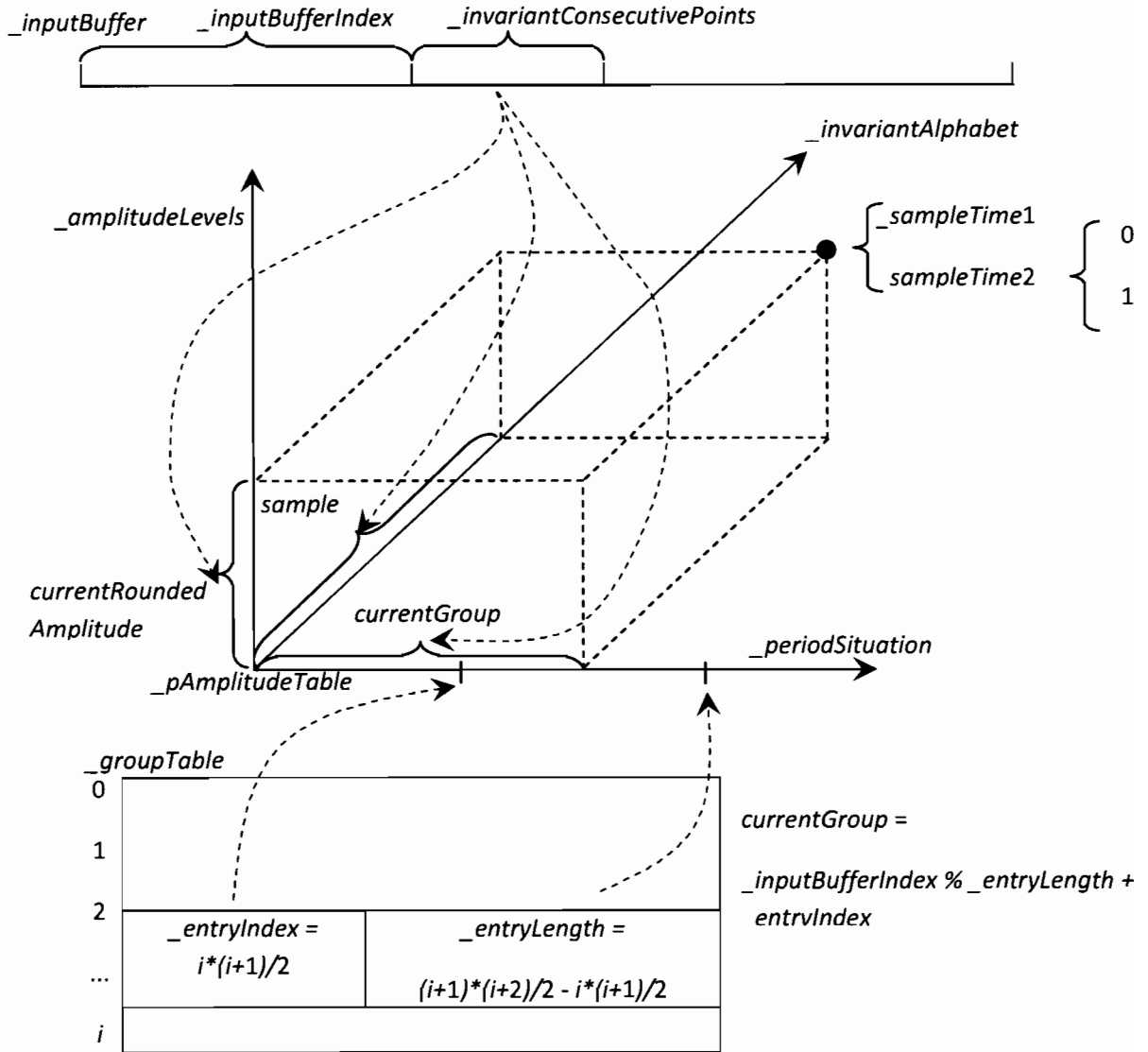


Fig 4

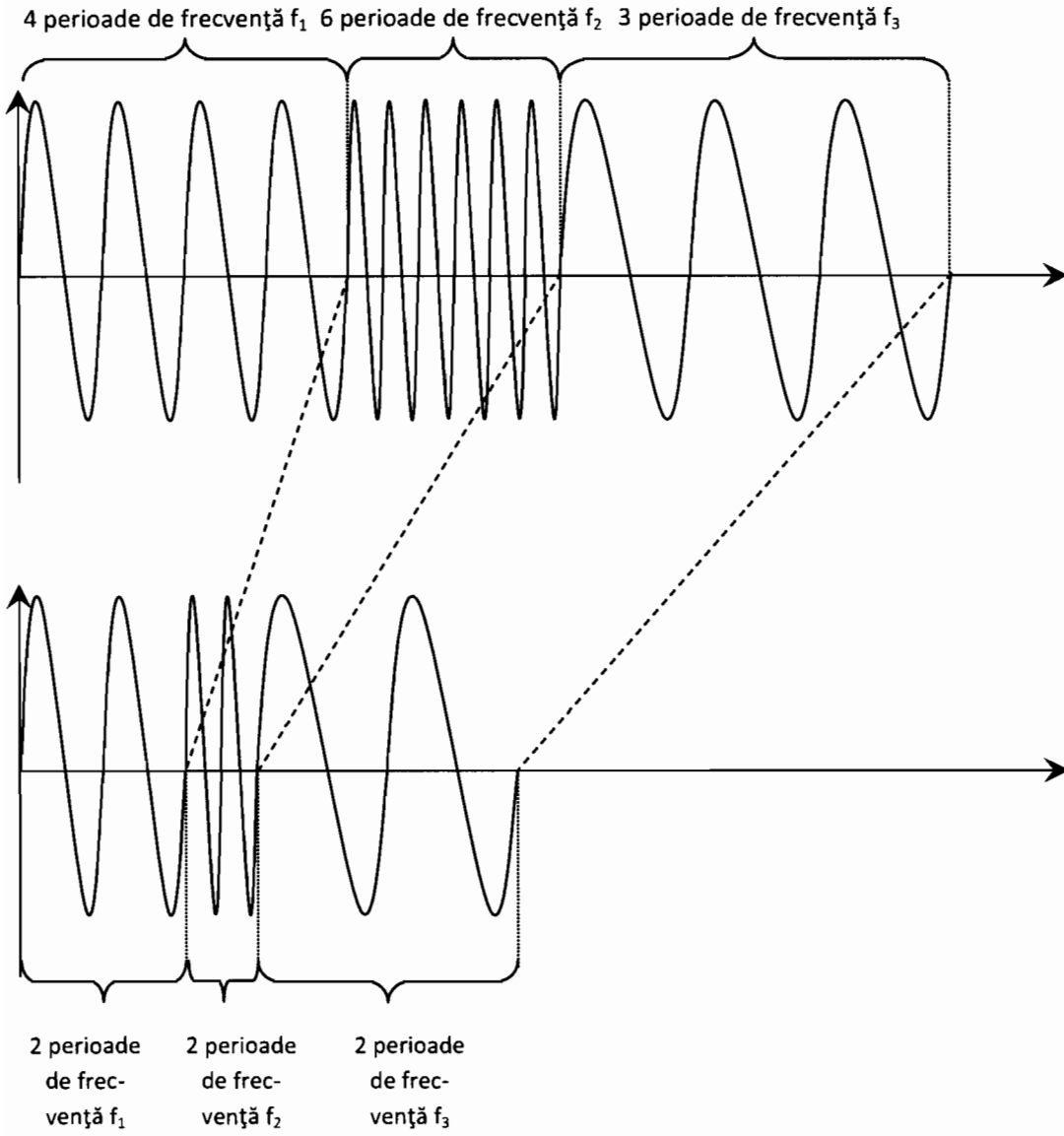


Fig 5

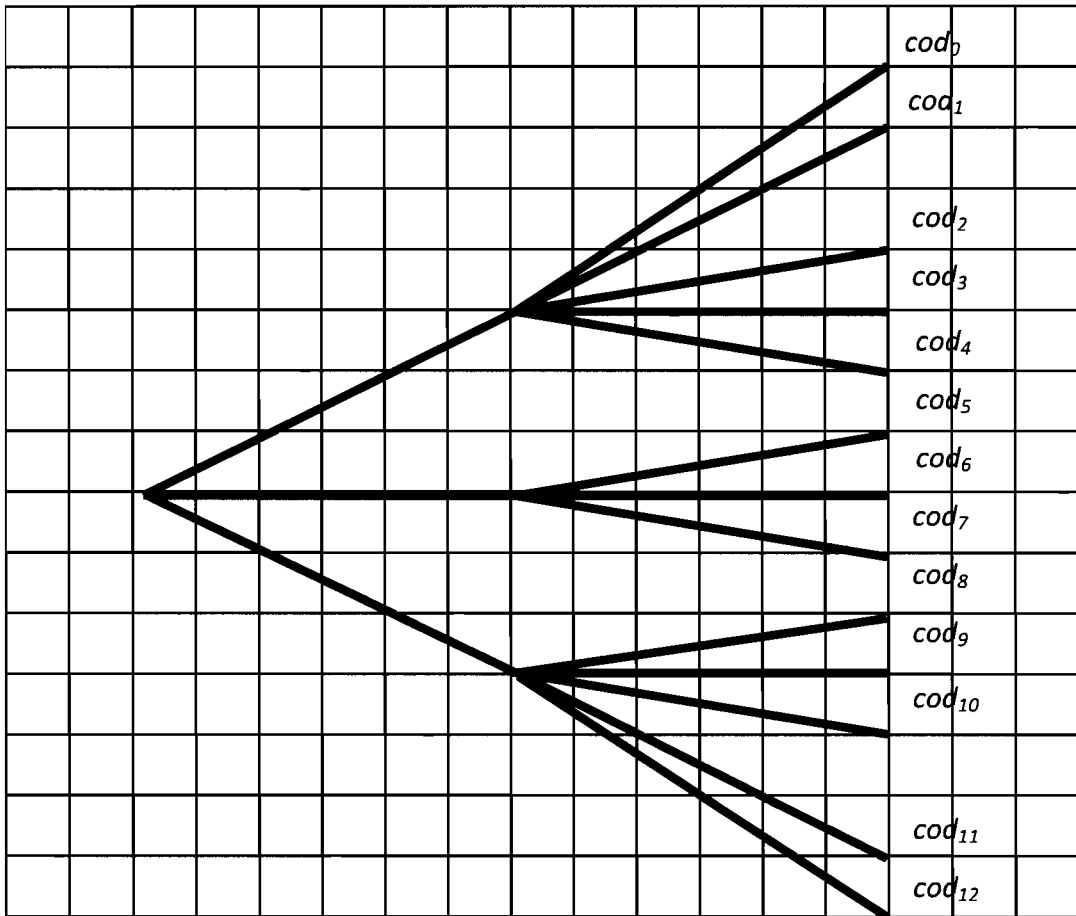


Fig. 6

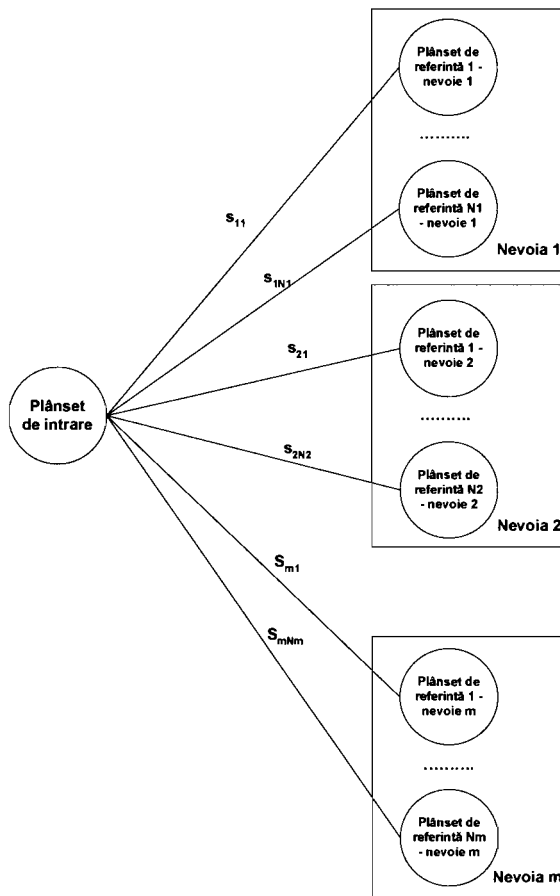


Fig. 7



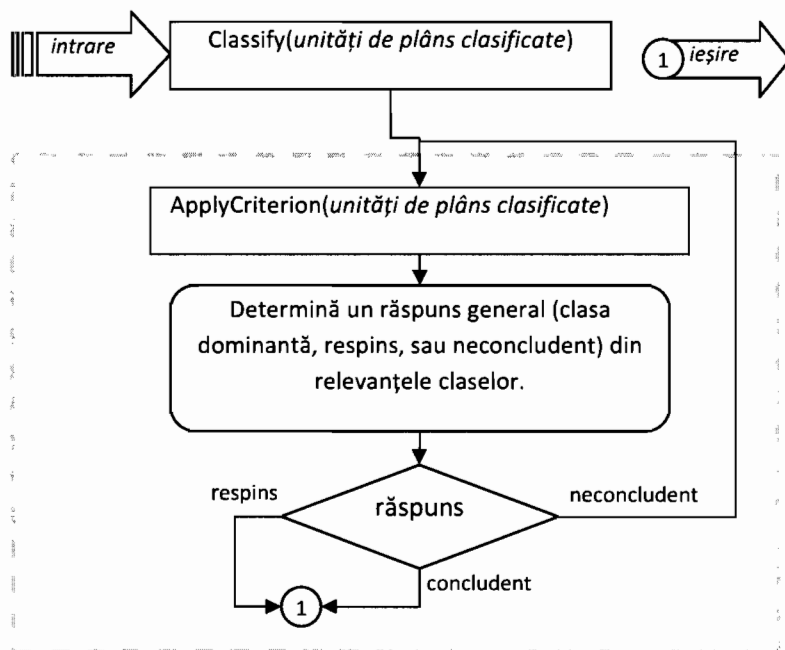


Fig 8

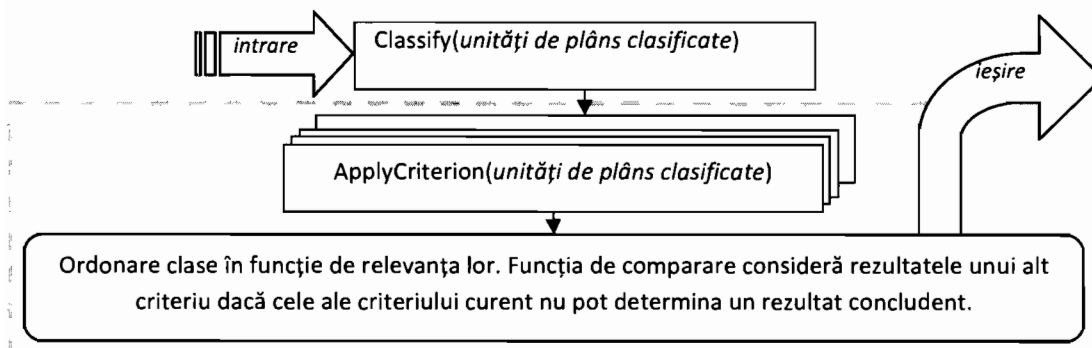


Fig 9

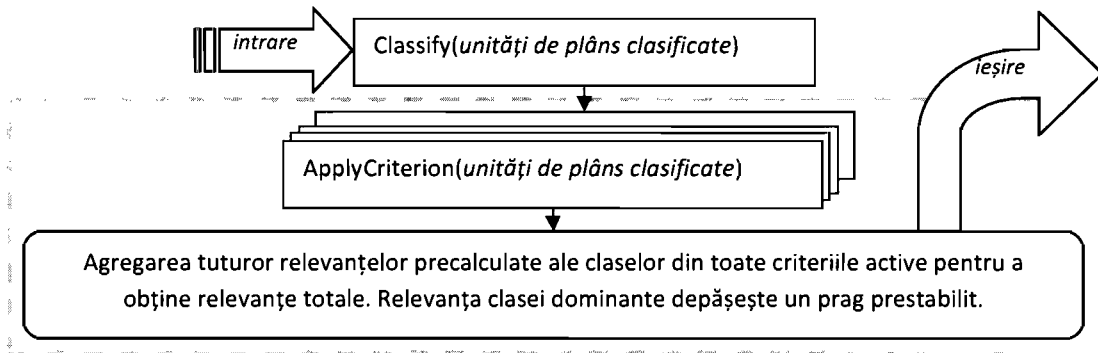


Fig 10

1/5

**Semnal vocal:**

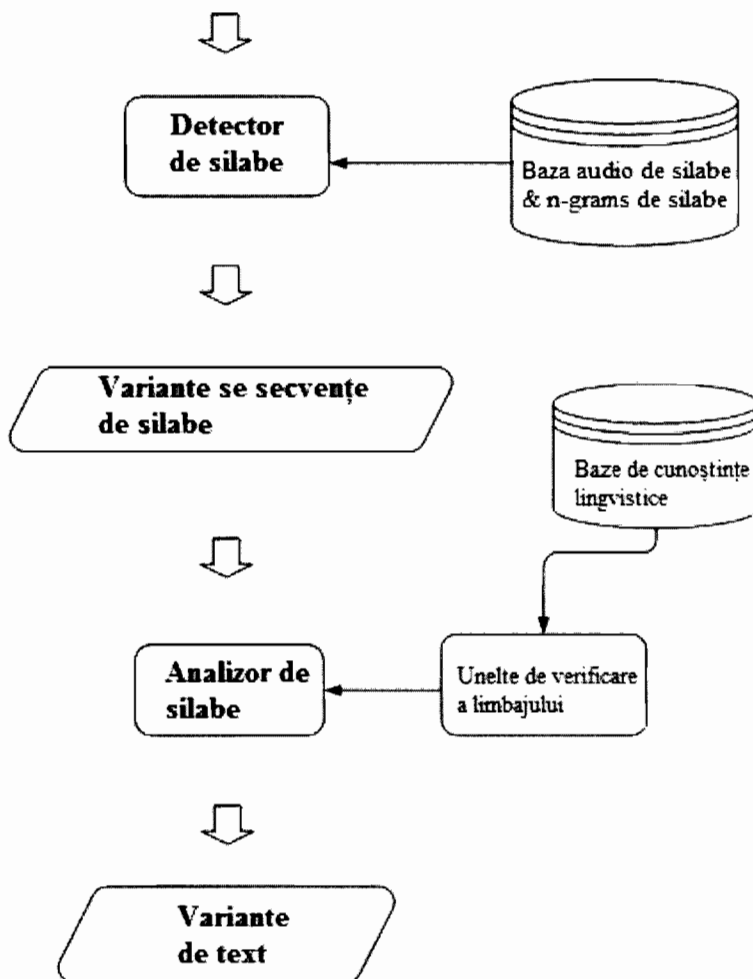


Fig. 11

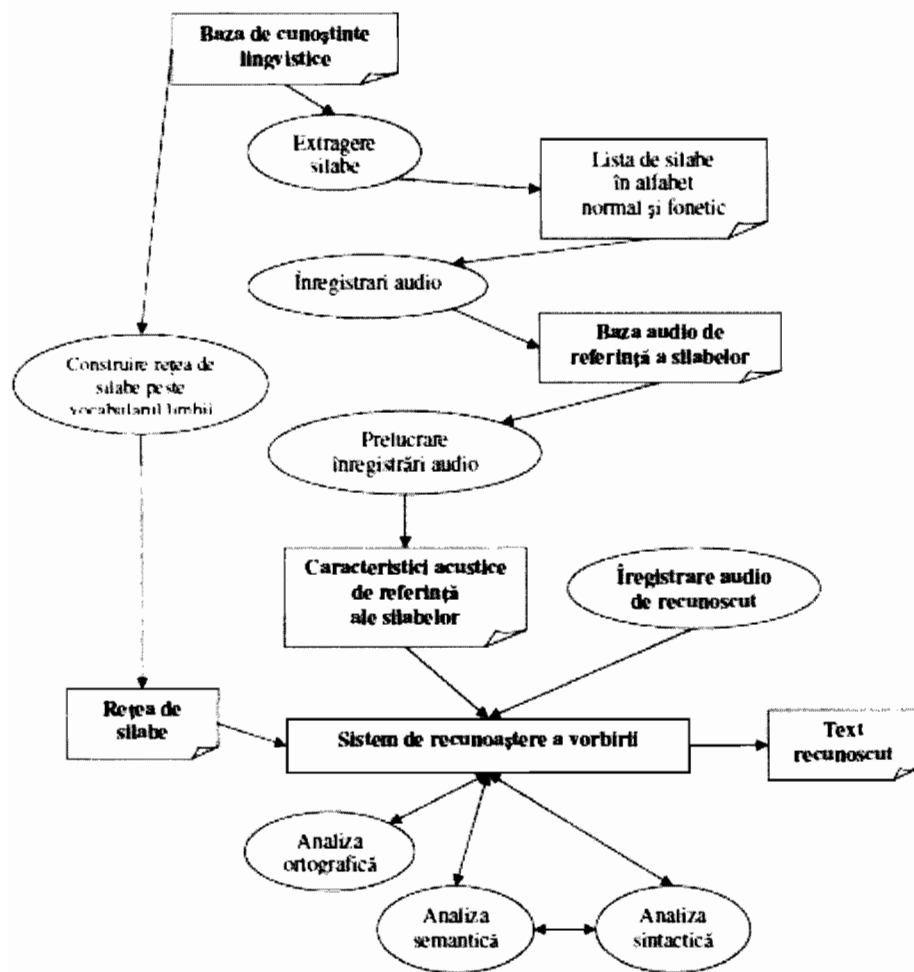


Fig. 12

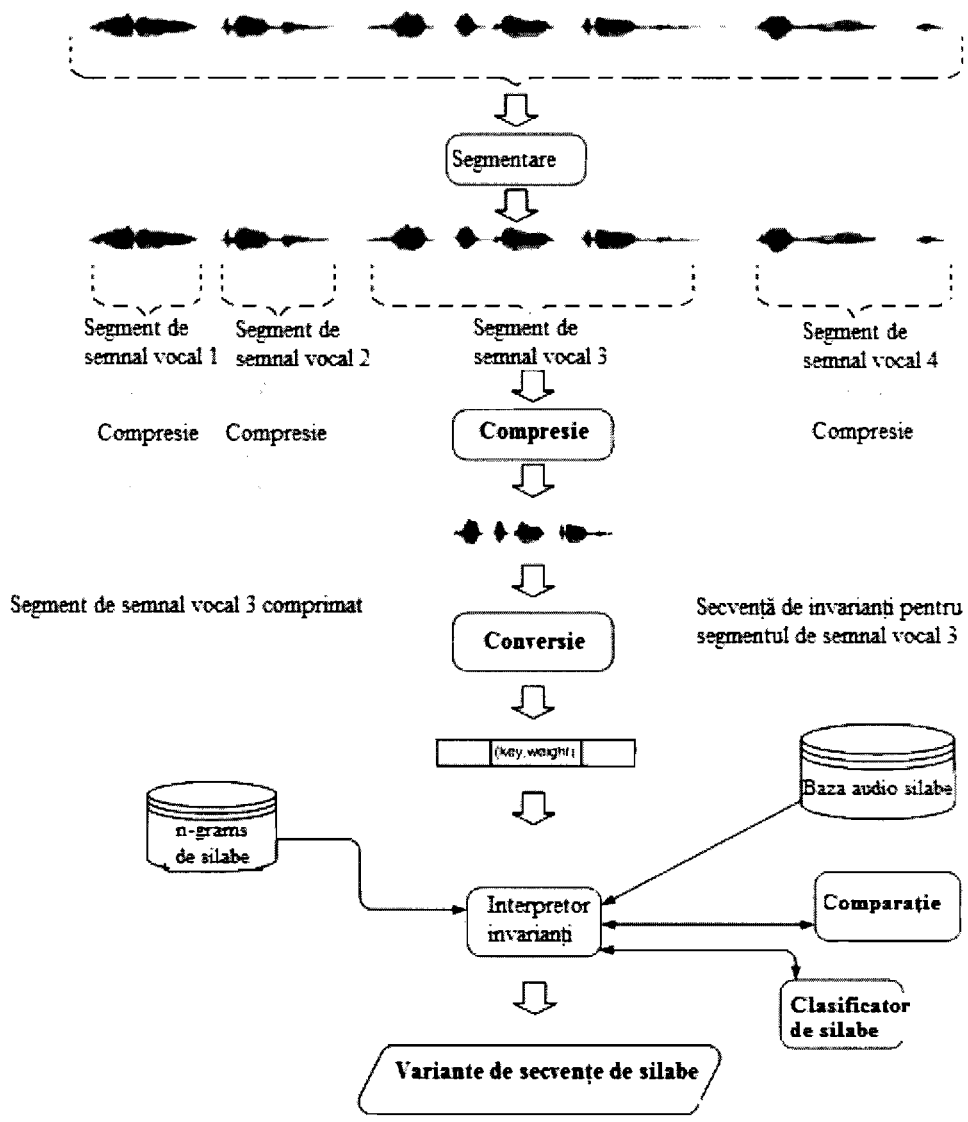


Fig. 13

1/2

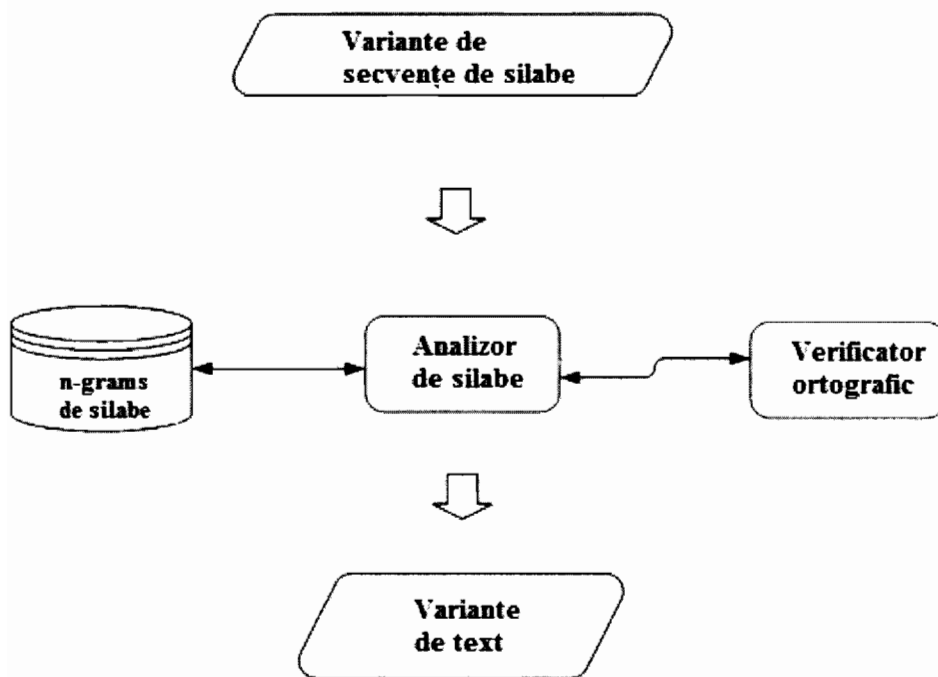


Fig.14

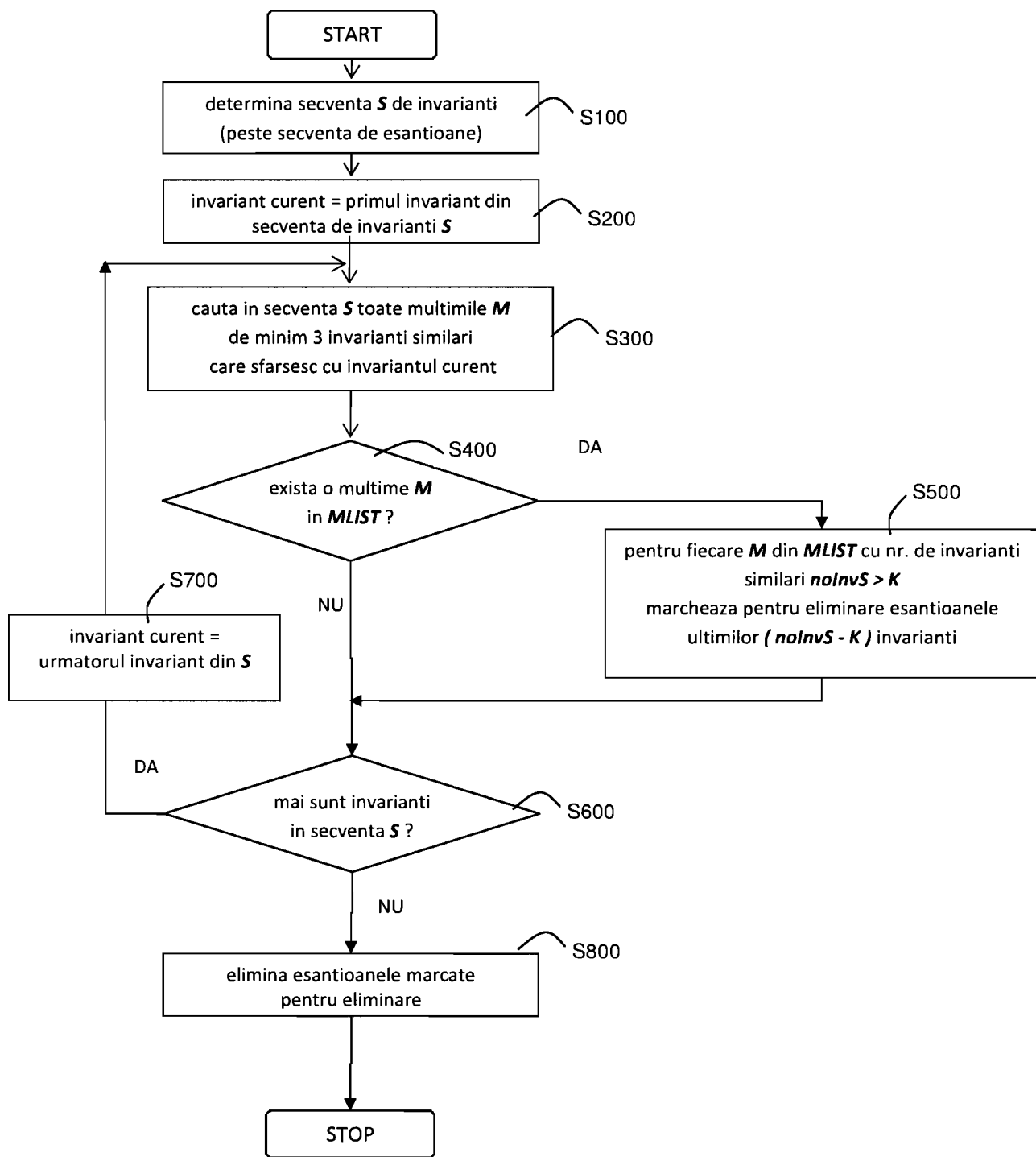


Fig. 15

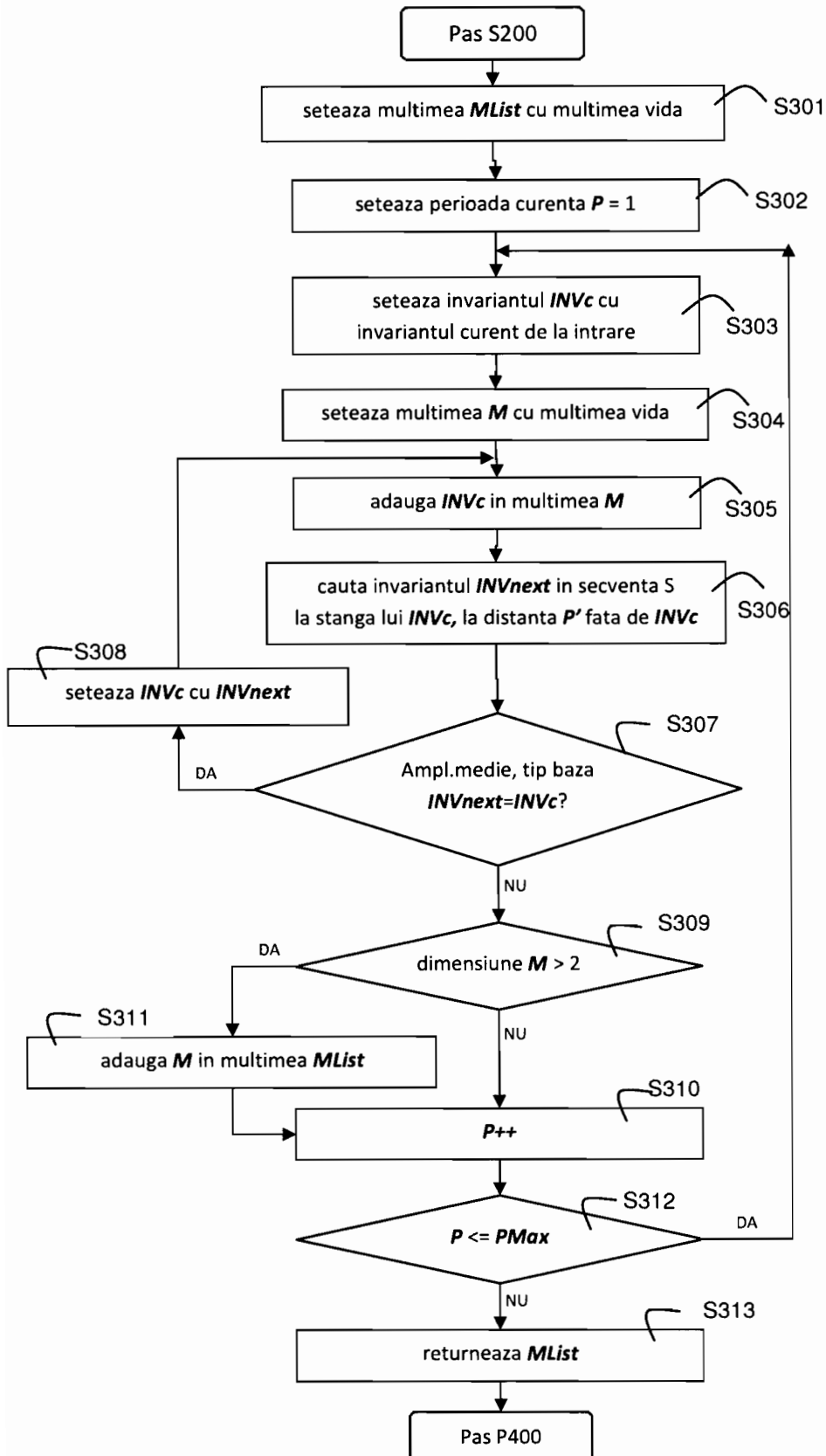


Fig. 16