Raport Ştiinţific şi Tehnic
Etapa a II-a, an 2015:
„Dezvoltarea Experimentală a Componentelor”

Aceste rezultate au fost obținute prin finanțare in cadrul programului Parteneriate în domenii prioritare, PN II, derulat cu sprijinul MEN – UEFISCDI, proiect nr. PN-II-PT-PCCA-2013-4-1660: **„Sistem Mobil de Asistare Vocala in Reintegrarea Persoanelor cu Afonii Chirurgicale” SWARA**

Denumirea organizaţiei participante in proiect	Acronim organizație	Tip organizaţie	Rolul organizaţiei în proiect (Coordonator/partener)
Universitatea Tehnică din Cluj-Napoca	UTCN	UNI	CO
SC FORTECH SRL	FORTECH	SRL	P1
Universitatea de Medicină şi Farmacie Iuliu Hatieganu	UMF	UNI	P2
Universitatea Babeș-Bolyai	UBB	UNI	P3

Date de identificare proiect

Număr contract:	**Nr. 6 / 2014, PN-II-PT-PCCA-2013-4-1660**
Acronim / titlu:	**SWARA – „Sistem Mobil de Asistare Vocală în Reintegrarea Persoanelor cu Afonii Chirurgicale”**
Titlu raport:	**Raport Științific și Tehnic (Etapa a II-a, 2015)**
Termen:	**Decembrie 2015**
Editor:	**Mircea Giurgiu (Universitatea Tehnică din Cluj-Napoca)**
Adresa de eMail editor:	**Mircea.Giurgiu@com.utcluj.ro**
Autori, in ordine alfabetică:	**Mihaela Dinsoreanu, Camelia Florea, Mircea Giurgiu, Camelia Lemnaru, Silviu Matu, Remus Pop, Rodica Potolea, Bogdan Orza, Radu Soflau, Adriana Stan**
Ofițer de proiect:	**Silvia Geicu**

Rezumat:

Acest document prezintă o sinteză a realizărilor de natură științifică și tehnică obținute în a doua etapă de implementare a proiectului SWARA (perioada Ianuarie – Decembrie 2015). Realizările se referă la:

- dezvoltarea unei versiuni preliminare a sistemului de sinteză text vorbire
- identificarea variabilelor psiho-sociale care trebuie personalizate pentru aplicația asistivă
- dezvoltarea primei versiuni a bazei de date audio-video
- dezvoltarea unui model și a unui sistem preliminar pentru predicție text
- identificarea metodelor și a unei soluții de recunoaștere vizuală a vorbirii
- dezvoltarea unui sistem experimental de sinteză accesibil în Cloud
- diseminarea rezultatelor intermediare.

Activitățile de cercetare desfășurate în etapa a doua de implementare a proiectului (2015) au condus la obținerea rezultatelor asteptate și ele sunt în concordanță cu obiectivele specifice ale etapei. Astfel, rezultatele raportate în acest document și descrise detaliat în cele 11 livrabile aferente perioadei de raportare, pregătesc pentru etapa următoare cadrul de integrare a componentelor în noul sistem de sinteză de înaltă calitate, cu posibilități de creare și adaptare a vocilor sintetice, cu predicția rapidă a textului și accesibil de pe echipamente mobile. De asemenea, acest raport prezinta detalii referitoare la activitatile de management si comunicare, precum si de diseminare a rezultatelor.

Cuprins

1. Activitățile etapei de raportare în contextul general al proiectului

În etapa anterioară (2014) au fost indexate resursele (baze de date de semnal vocal, resurse de text si adnotari de natura lingvistica ale acestora, instrumente software utilizate in procesarea semnalului vocal si a textului aplicate in scopul sintezei din text a semnalului vocal) si s-au elaborat specificațiile funcționale pentru componentele software ale sistemului de sinteza.

Astfel, în etapa de raportare 2015 s-au desfășurat activități de cercetare și dezvoltare experimentală pentru modulele sistemului de sinteză, pentru achiziția unui corpus lărgit de semnal vocal, pentru dezvoltarea unui model de predicție rapidă a textului și pentru dezvoltarea unui sistem de sinteză experimental disponibil online, conform cu specificațiile anterior definite și cu obiectivele care sunt prezentate în secțiunea a doua a acestui raport.

În etapa următoare (2016) va alea loc optimizarea și integrarea componentelor, dezvoltarea metodelor pentru adaptarea și crearea de noi voci sintetice, dezvoltarea integrală a serviciilor web cu accesibilitate de pe echipamente mobile propuse de către partenerul industrial, sesiuni de evaluare finală a sistemului.

2. Gradul de realizare a obiectivelor specifice pentru Etapa a 2-a

Obiectivele specifice ale Etapei a 2-a, „Dezvoltarea experimentală a componentelor", impreuna cu gradul lor de realizare, activitatile si principalele rezultate obtinute în anul 2015 sunt prezentate in lista de mai jos.

Obiectiv2.a: *Dezvoltarea unei versiuni preliminare a sistemului de sinteză text vorbire*

Grad realizare: Obiectiv realizat integral

Rezultate:

- resurse de date audio și de text folosite pentru antrenarea sistemului
- dicționare și lexicoane pentru procesarea hibridă a textului
- componente software funcționale (testate și evaluate în diverse scenarii) pentru: 1) normalizarea textului, 2) restaurarea automată a diacriticelor, 3) transcriere fonetică automată, 4) silabificare și predicția automată a accentului lexical, 5) predicția automată a părților de vorbire
- un modul pentru antrenarea modelelor acustice din datele audio și text
- un modul de sinteză de voce bazat pe sistemul HTS
- un demonstrator online de sinteză text vorbire în variantă preliminară
- 3 articole științifice publicate la conferințe internaționale [6] [7] [8]
- un livrabil (D1.2) cu titlul „Sistem preliminar de sinteză text vorbire", care descrie rezultatele menționate mai sus.

Obiectiv2.b: *Identificarea variabilelor psiho-sociale care trebuie personalizate*

Grad realizare: Obiectiv realizat integral

Rezultate:

- metodologie de identificare și raportare
- rezultate ale evaluării versiunii preliminare cu pacienții
- articole științifice publicate la conferințe internaționale [1] [2] [3] [4] [5]
- un livrabil (D2.2) cu titlul „Raport asupra variabilelor psiho-sociale ce trebuiesc personalizate" și care descrie rezultatele menționate mai sus.

Obiectiv2.c: *Dezvoltarea primei versiuni a bazei de date audio-video*

Grad realizare: Obiectiv realizat integral

Rezultate:

- izolarea fonică a unei locații folosită pentru înregistrări audio-video
- activarea unei metodologii pentru înregistrare și sincronizare date
- înregistrări audio pentru 3 noi vorbitori și video pentru 1 vorbitor
- un livrabil (D3.2a) cu titlul „Baza de date vers1.”, care descrie rezultatele menționate mai sus.

Obiectiv2.d: *Dezvoltarea unui model și a unui sistem preliminar pentru predicție text*

Grad realizare: Obiectiv realizat integral

Rezultate:

- experimente preliminare pentru modelarea limbajului natural și analiza bigramelor și trigramelor pentru predicșia textului
- definirea unui model de predicție de text
- dezvoltarea și implementarea conceptului într-un sistem preliminar de predicție care folosește indexul inversat
- un articol publicat la conferință internațională [6]
- două livrabile: (D4.1) cu titlul „Model de context pentru predicție text” și (D4.2) „Sistem preliminar de predicție text”.

Obiectiv2.e: *Identificarea metodelor de recunoaștere vizuală a vorbirii*

Grad realizare: Obiectiv realizat integral

Rezultate:

- trei tipuri de metode identificate și raportate
- un experiment preliminar folosind înregistrările video de la Obiectivul 2.c pentru implementarea unui experiment de recunoaștere vizuală a vorbirii
- un livrabil: (D4.3a) cu titlul „Raport privind metodele de recunoaștere vizuală a vorbirii”.

Obiectiv2.f: *Dezvoltarea unui sistem experimental de sinteză accesibil în Cloud*

Grad realizare: Obiectiv realizat integral

Rezultate:

- instalarea și testarea infrastructurii hardware în Cloud
- dezvoltarea arhitecturii software: interfața web, serverul HTTP, serverul de aplicație, integrarea motorului de sinteză vocală
- integrarea a 3 noi voci folosind datele colectate la Obiectivul 2.c
- evaluarea online a sistemului experimental în Obiectivul 2.b
- două articole despre tehnologii asistive [4] [9]
- un livrabil: (D6.2) cu titlul „Versiune experimentală a sistemului de sinteză text vorbire în Cloud accesibil de pe mobil”.

Obiectiv2.g: *Diseminarea rezultatelor intermediare*

Grad realizare: Obiectiv realizat integral

Rezultate:

- actualizarea dinamică și monitorizarea cu Google Analytics a site-ului
- planul de diseminare pentru anul 2015
- materiale promoționale: pliant, prezentarea PPT a proiectului, 2 postere articole, pagina web pentru demonstratorul online, pagina web cu mostre de semnal vocal sintetizat pentru diferite voci
- 9 articole prezentate și comunicate la conferințe internaționale
- 4 livrabile referitoare la diseminare: D7.1, D7.2, D7.3, D7.4.

3. Rezultatele etapei și descrierea lor științifică și tehnica

3.1. Sistem complet de sinteză text vorbire în versiunea preliminară

Rezultatele raportate în aceasta secțiune corespund Obiectivului 2.a din lista de obiective specifice Etapei a 2-a, iar ele sunt descrise in extenso in livrabilul (D1.2) cu titlul „Sistem preliminar de sinteză text vorbire”.

3.1.1. Modulul de normalizare a textului

Textul furnizat la intrarea sistemului de sinteză poate să conţină, pe lângă cuvintele obişnuite şi o serie de secvenţe de caractere alfanumerice sau caractere speciale care generează o semnificație specifică din punct de vedere lingvistic. De exemplu: abrevieri, numere, data si ora, acronime, numere de telefon, semne speciale pentru specificarea sumelor de bani, etc. Aceste secvențe se numesc secvențe non-standard NSW - Non Standard Words.

Problema normalizării textului este doar aparent simplă, deoarece ea implică o serie de dificultăți. De exemplu, un aspect relativ banal cum este segmentarea la nivel de frază pe baza semnelor de punctuație, de exemplu punctul, poate induce ambiguități majore deoarece punctul poate să apară atât la sfârșit de frază cât şi în abrevieri (dvs.), acronime (P.N.L.), numere (12.300,56) sau indicaţia că se omite un anumit fragment de text.

Pentru implementarea acestui modul s-a definit o taxonomie a NSW, s-a definit un ansamblu de functionalități, iar pe baza lor s-a implementat o metodă hibridă de normalizare (Fig.1), care are în vedere:

- normalizarea pe bază de reguli pentru secvențele identificate a fi numerice, dată, oră, ani
- normalizarea pe baza unui dicționar organizat sub forma unui lexicon de cuvinte NSW. Exemple: simboluri (<, <=, >, >=,), (, &,], [, {, }, +, -, @, #, “, ‘, $, |, *, etc), orice succesiune de două sau mai multe majuscule este considerată acronim, pentru acronimele care nu se găsesc în dicționar literele lor se vor normaliza individual, literele (a, b, c, etc)

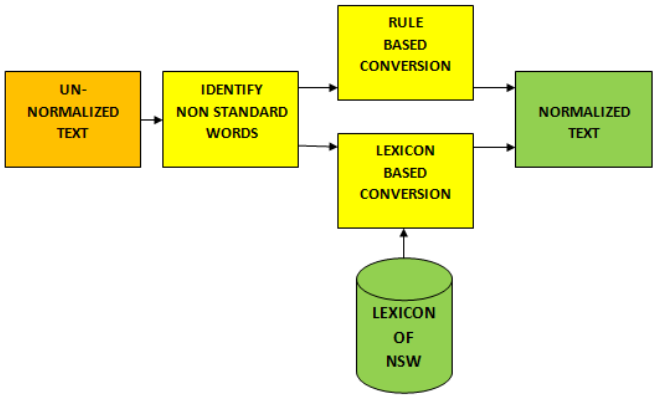

Figura 1. Fluxul de procesări implementat pentru normalizarea textului

Tabel 1. Selecție din rezultatele privind normalizarea textului

Tip NSW	Intrare	Ieșire
Numere mai mici de 10^12	123	o sută douăzeci și trei
Numere mai mari de 10^12	750580558282384	șapte cinci zero cinci opt zero cinci cinci opt doi opt doi trei opt patru
Ora în format hh:mm:ss	11:35:40	ora unsprezece treizeci și cinci de minute și patruzeci de secunde
Data în format m.y	09.1944	septembrie o mie nouă sute patruzeci și patru

3.1.2. Modulul de restaurare automată a diacriticelor

Modulul de restaurare automată a diacriticelor (Fig. 2) este o componentă importantă pentru sistemul de sinteză text-vorbire, deoarece scrierea fără diacritice va produce fie secvențe sonore neinteligibile local, fie ambiguități sintactice și semantice care pot conduce la ne-inteligibilitatea mesajului la nivel global, de propoziție sau frază. Vezi detalii în D1.2.

Pornind de la soluțiile studiate și având in vedere cerințele de viteză de procesare și de memorie impuse pentru aplicația finală, s-au propus câteva metode de restaurare automată a diacriticelor bazate pe algoritmi de invățare supervizată rapizi și cu consum redus de memorie. Pentru selecția acestor metode s-au realizat teste preliminare pe corpusuri de text de dimeniune redusă. Metodele testate au fost: Naïve Bayes, Retele neuronale de tip Multi Layer Perceptron (MLP), Random forest, Logistic Regression, Support Vector Machines de tip multicriterial, Radial Basis Functions, Arbori de decizie folosind algoritmul C4.5 (CART-Classification and Regression Trees, J48), Instance Based Learning (IBL, tool folosit TIMBL).

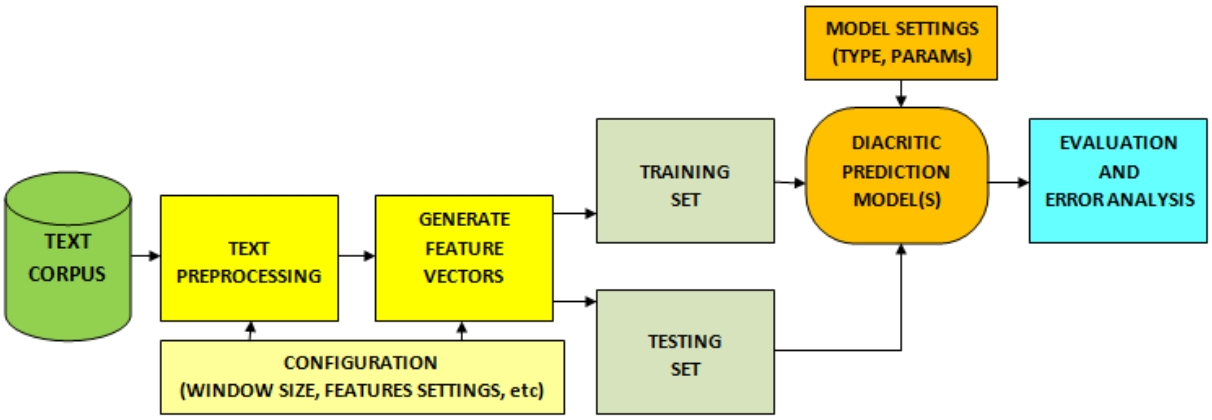

Figura 2. Fluxul de prelucrări pentru restaurarea automată a diacriticelor

Tabelul 2. Exemplu de generarea vectorilor de antrenare pe baza contextului de litere

Vector (context 3 litere), diacritic							Cuvânt proveniență
i,	*n,*	*v,*	*t,*	*a,*	*m,*	***ă***	*-învățăm- (învățământului)*
n,	*v,*	*a,*	*a,*	*m,*	*a,*	***ț***	*-nvățămâ- (învățământului)*
v,	*a,*	*t,*	*m,*	*a,*	*n,*	***ă***	*-vățământ- (învățământului)*
t,	*a,*	*m,*	*n,*	*t,*	*u,*	***â***	*-țământu- (învățământului)*
t,	*<SP>,*	*c,*	*r,*	*u,*	*i,*	***ă***	*-t cărui – (cărui)*
u,	*l,*	*t,*	*t,*	*i,*	*l,*	***ă***	*-ultățil- (facultățile)*
l,	*t,*	*a,*	*i,*	*l,*	*e,*	***ț***	*-ltățile- (facultățile)*

Corpusuri folosite pentru evaluarea soluțiilor propuse:

- **RomParl** – acesta este un corpus colectat de către grupul de cercetare din Universitatea Tehnica din Cluj-Napoca și conține partea de audio și text (21 de ore) a ședințelor plenare din parlamentul Romaniei din perioada 2011-2014.
- **RomLit** – acesta este un corpus de text colectat din revista România Literară din anii 2006 – 2015.
- **RomWiki** – acesta este un corpus colectat automat din paginile Wikipedia în limba română.

Tabelul 3. Rezultate cu e*valuarea modelelor TIMBL aferente diacriticelor în funcție de context*

Diacritic	Dimens vector (context)	Vectori U / W (Unic / Within features)		
		Nr. vectori antrenare	Entropie	Performanța [%]
a – ă	9	28.000	0,76	89,64
	11	40.000	0,81	92,76
a - â	9	22.600	0,22	99,16
	11	31.500	0,27	99,32
a – ă - â	9	21.000	0,47	91,07
	11	41.400	1,00	92,81
i - î	9	25.000	0,42	98,46
	11	34.000	0,52	99,09
s - ș	9	11.200	0,83	95,90
	11	15.600	0,83	96,88
t - ț	9	16.000	0,62	95,80
	11	21.000	0,70	96,80
toate	9	56.000	2,04	94,87
	11	99.900	2,63	96,28

Concluziile acestui experiment sunt următoarele:

- modelele TIMBL create pentru fiecare diacritic în parte au o performanța mai bună decât modelul global. Ordinea de performanță pentru contextul de 5 litere la stânga și 5 la dreapta, cu diacritic inclus în vector este: a – â (99,16%), i – î (99,09%), s - ș (96,88%), t - ț (96,80%), a – ă (92,76%)
- de aici a apărut ideea de a adopta modele de predicție adaptate la diacritic și aplicarea lor incremental, în ordinea performanțelor.

Tabelul 4. Matricea de confuzie folosind la antrenare 109.999 vectori, context 11 litere.

	a	s	t	i	ă	ț	ș	î	â
a	***23.528***	*254*	*249*	*1.671*	*1.137*	*7*	*100*	*79*	*50*
s	*378*	***8.554***	*371*	*243*	*41*	*39*	*68*	*2*	*0*
t	*221*	*344*	***14.019***	*595*	*90*	*179*	*30*	*0*	*0*
i	*1.291*	*204*	*567*	***24.073***	*1.397*	*62*	*27*	*120*	*48*
ă	*806*	*17*	*95*	*1.218*	***6.689***	*2*	*0*	*0*	*49*
ț	*11*	*67*	*197*	*82*	*9*	***2.880***	*38*	*0*	*0*
ș	*33*	*87*	*82*	*48*	*1*	*44*	***3.015***	*2*	*0*
ă	*73*	*0*	*0*	*203*	*0*	*0*	*28*	***2.807***	*0*
â	*22*	*1*	*0*	*41*	*7*	*0*	*0*	*0*	***959***

Tabelul 5. Comparație între performanțele modelelor J48 și TIMBL pentru diferite diacritice

Model pentru predicția perechii …	J 48 cel mai bun	TIMBL	
		(nU), cu diacritic inclus	(U), cu diacritic inclus
	[%]	[%]	[%]
a - ă	*96,04*	*95,27*	92,73
a - â	*99,64*	*99,71*	99,43
i - î	*99,84*	*99,50*	99,29
s - ș	*98,95*	*98,10*	97,20
t - ț	*98,75*	*98,34*	97,09
a – ă - â	*95,10*	*95,06*	92,77
model global (toate)	*98,15*	*97,71*	96,25

În final, ca soluție practică s-a ales pentru sistemul preliminar predicția folosind arbori de decizie. Pentru sistemul final vom cerceta posibilitatea creșterii performanțelor (dacă este posibil fără nici o eroare la predicția diacriticelor) prin aplicarea de metode hibride, prin utilizarea unor dicționare adaptive colectate din interacttiunea efectivă cu utilizatorii, sau alte metode adaptate la contextul și la modelul de limbaj folosit de către utilizatori.

3.1.3. Modulul de transcriere fonetică automată

Pornind de la studiile și rezultatele prezentate în literatura de specialitate și având în vedere aspecte legate de optimizarea performanțelor sistemului de procesare de text din cadrul proiectului, metoda propusă este una hibridă. Aceasta folosește într-o primă fază căutarea în dicționar, iar apoi, doar pentru literele cu valori fonetice polivalente se face o predicție cu ajutorul arborilor de decizie. De exemplu, pentru cuvântul *'casă'* cu o fereastră de predicție de lungime 7, vor rămâne în setul de antrenare, doar următoarele cazuri: *'- - - c a s ă -'* și *'- - c a s ă -'*, literele *s* și *ă* fiind monovalente din punct de vedere fonetic. Arborii de decizie au fost selectați datorită vitezei de procesare și a dimensiuni reduse de stocare. Setul de antrenare este cel din dicționarul NaviRO extins: 138.500 de cuvinte transcrise fonetic conform cu codarea SAMPA pentru limba română și utilizând un set de 31 de foneme.

Într-o primă etapă, acest dicționar este interogat pentru a găsi rapid o transcriere a cuvântului sau a cuvintelor pentru care se face procesarea de text. Pentru a optimiza căutarea, dicționarul este segmentat în sub-secțiuni aferente fiecărei litere de început a cuvântului, iar modul de stocare este cel serializat. Căutarea este de asemenea eficientizată prin utilizarea unor funcții hash specifice dicționarelor din limbajul de programare Python. Cuvintele care nu sunt găsite în dicționar sunt direcționate către arborii de decizie.

Antrenarea arborilor de decizie se face utilizând o fereastră de context în care caracterul central este cel pentru care se face predicția, așa cum s-a prezentat mai sus. Contextul este dependent de cuvânt, ceea ce înseamnă că fereastra de analiza se va limita doar la literele din cuvântul din care face parte litera. Tot ca și optimizare a timpului de predicție, fiecărei litere și fiecărui fonem îi este atribuit un cod numeric întreg pozitiv, fapt ce permite eficientizarea dimensiunii modelelor antrenate stocate, precum și a timpului de procesare. Rezultatele experimentale se referă în mod strict la predicția cu ajutorul arborilor de decizie și doar pentru acele litere ce au valori fonetice multiple: *a, ă, c, e, g, h, i, k, o, q, u, w, x, y.* Arborii de decizie au fost antrenați cu ajutorul modulului Scikit-Learn din Python, opțiunea Decision Tree Classifier. Acuratețea predicției este măsurată prin validare încrucișată cu 10 partiții.

Tabelul 6. Evaluarea transcrierii fonetice automate pentru fonemele polivalente

Lungime fereastră (nr. de litere)	Procent date de antrenare [%]	Acurateţe [%]
1	*25*	*92.48*
	50	*92.38*
	100	*92.35*
3	*25*	*98.35*
	50	*98.35*
	100	*98.37*
5	*25*	*98.91*
	50	*99.15*
	100	***99.30***
7	*25*	*98.72*
	50	*99.05*
	100	*99.28*

De notat că acuratețea măsurată în acest caz se referă la fonemele prezise și nu la cuvinte integral transcrise corect. Aceste rezultate urmează să le obținem în etapa de optimizare a componentelor. Se vor testa și alte metode mai eficiente de predicție, sau combinarea predicției cu modulul de silabificare, predicția părții de vorbire și poziționare a accentului.

3.1.4. Modulul de silabificare și de predicție a accentului

Silabificarea reprezintă procesul de separare a cuvintelor în entitati de dimensiune mai mica, silabe. Pozitionarea accentului se referă la identificarea acelei vocale dintr-un cuvant care este pronunțata mai intens sau pe un ton mai înalt. Identificarea poziției accentului și a despărțirii în silabe sunt sarcini importante in componentele de procesare a textului din sistemele de sinteza a vorbirii, deoarece prin intermediul lor se comandă intonația în propoziție, obținând astfel naturalețea și expresivitatea din vorbire. Realizarea eficientă a acestor sarcini va îmunătăți semnificativ calitatea semnalului sintetizat.

Tabelul 7. Exemple de particularități problematice în poziționare accent și silabificare

	Cuvant	**Parte de vorbire**
Despartire in silabe diferita, aceasi pozitionare a accentului	*i-gnór* *ig-nór*	Verb (prezent)
Parte de vorbire diferita, pozitionare diferita a accentului	*e-chi-pá* *e-chí-pa*	Verb (trecut) Substantiv (singular)
Timp diferit, pozitionare diferita a accentului	*no-ti-fi-că* *no-tí-fi-că*	Verb (trecut) Verb (prezent)
Sens diferit, pozitionare diferita a accentului	*re-gíi* *ré-gii*	Substantiv (plural)
Parte de vorbire diferita, despatire in silabe si pozitionare accent diferite	*bi-blio-gra-fi-á* *bi-bli-o-gra-fí-a*	Verb (prezent) Substantiv (singular)

Figura 3 prezinta arhitectura modulului software de antrenare și predicție automată a silabificării și poziționării accentului.

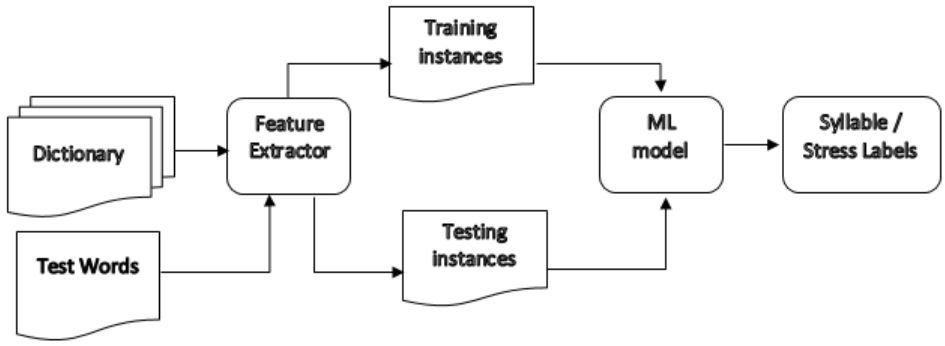

Figura 3. Fluxul de procesari pentru crearea modelelor antrenate supervizat

Pentru crearea sistemului am utilizat cel mai complet dictionar care contine informatii de silabisire și pozitionare a accentului, RoSyllabiDict. Dicționarul conține 525,534 de forme flexionare ale cuvintelor, mai mult de 65.000 de leme, cu variantele lor silabisire și accente plasate.

Într-o primă versiune, vectorul de trasaturi (Feature Extractor) a fost definit într-un proces iterativ, urmarind rezultatele validărilor experimentale.

În a doua versiune a vectorului de trasaturi, am considerat informații legate de pozitionarea accentului ca trăsătura de intrare, i.e. pentru fiecare instanta am adaugat o trasatura binara care sa contina aceasta informatie legata de caracterul curent.

Într-o a treia versiune, am considerat, de asemenea, adăugarea de bigrame. Acest lucru implică faptul că fiecare instanta este caracterizata de grupuri vecine de două litere fiecare, urmând aceeași strategie de cinci vecini pe partea stângă și pe partea dreaptă. Am observat din experimentele noastre că timpul de procesare pentru antrenare si testare creste în mod semnificativ pentru această versiune a vectorului de trasaturi. Mai mult decât atât, având în vedere aceeași dimensiune a seturilor de antrenare si evaluare pentru unigrame și bigrame, cresterea acurateii de clasificare nu este notabila.

Tabelul 8. Exemplu de creare a vectorilor de antrenare

Litera	*Trasaturi*	*Clasa*
î	*0 1 * * * * * î n ţ e l e*	*no*
n	*0 0 * * * * î n ţ e l e g*	*yes*
ţ	*0 0 * * * î n ţ e l e g i*	*no*
e	*0 1 * * î n ţ e l e g i **	*yes*
l	*0 0 * î n ţ e l e g i * **	*no*
e	*1 1 î n ţ e l e g i * * **	*no*
g	*0 0 n ţ e l e g i * * * **	*no*
i	*0 1 ţ e l e g i * * * * **	*yes*

După construirea setului de antrenare, am procedat la selectia clasificatorului si identificarea parametrilor modelului. Ne-am concentrat pe urmatorii algoritmi de clasificare: Support Vector Machines (SVMs), Random Forest (RF), Ada Boost, si Naive Bayes. Pentru a evalua clasificatorii am generat cinci seturi de antrenare aleatoare de cuvinte scrise cu diacritice din intregul dicționar. Numărul de cuvinte din fiecare multime este egal cu 4.300, rezultand intr-un numar de cazuri între 42.434 și 42.737. Aceeasi multime de 860 de cuvinte evaluate cu diacritice a fost utilizata pentru fiecare dintre cele cinci seturi de antrenare pentru a evalua clasificatorii.

Tabelul 9. Analiza comparativa a clasificatorilor RF, SMO, NaiveBayes si Ada Boost

Setul de antrenare	*Clasificator*	*Acuratete de clasificare*	*Precizie*
Set 1	Random Forest	***99,46%***	*99,5%*
	SMO	*96,27 %*	*96,3 %*
	Naive Bayes	*87,03 %*	*87,5 %*
	Ada Boost	*80,92 %*	*81,3 %*
Set 2	Random Forest	***99,00 %***	*99.0%*
	SMO	*96,04%*	*96,1 %*
	Naive Bayes	*87,09 %*	*87,6%*
	Ada Boost	*80,92 %*	*81,3%*
Set 3	Random Forest	***98,93 %***	*98,9%*
	SMO	*96,02 %*	*96,0%*
	Naive Bayes	*87,13 %*	*87,7%*
	Ada Boost	*80,92 %*	*81,3%*
Set 4	Random Forest	*98,93%*	*98,9%*
	SMO	***95,94 %***	*96,0%*
	Naive Bayes	*86,90 %*	*87,4%*
	Ada Boost	*80,85 %*	*81,2%*
Set 5	Random Forest	***98,93 %***	*98,9 %*
	SMO	*96,06 %*	*96,1%*
	Naive Bayes	*87,05 %*	*87,6%*
	Ada Boost	*80,92 %*	*81,3%*

Tabelul 10. Performanța de clasificare obtinuta de RF pentru poziționarea accentului

Test	Nr. instante	Nr. cuvinte	Acuratete (la nivel litera)
Set 1	*58.434*	*13.210*	*96,37 %*
Set 2	*58.626*	*13.209*	*96,37 %*
Set 3	*58.207*	*13.210*	*96,26 %*
Set 4	*58.293*	*13.211*	***96,58 %***
Set 5	*58.288*	*13.210*	*93,53 %*

3.1.5. Modulul de adnotare automată a părții de vorbire

Etichetarea automată și fără erori a părților de vorbire (POS – Part of Speech tagging) este una dintre cele mai dificile probleme ale lingvisticii computaționale. În cercetările noastre am realizat un ansamblu de studii amănunțite asupra procesului de etichetare a părților de vorbire pentru limba română și pentru limba engleză folosind diverse metode, cu scopul de a descoperi care este cea mai potrivită pentru contextul sintezei din text a semnalului vocal.

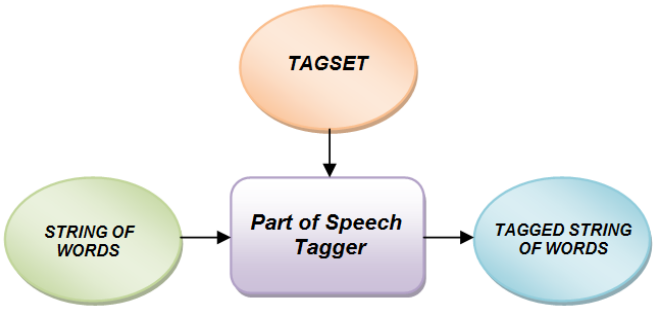

Figura 4. Procesul de etichetare automată a părților de vorbire

Tabelul 11. Structura corpusului de text folosit pentru crearea modelelor

Tip, gen literar	*Politic, stiri*
Nivel segmentare	*Propozitie, cuvant*
Tipuri de tag-uri	*MSD*
Multilingualitate	*Bilingv*
Tip multilingualitate	*Paralel*
Codare caracatere	*UTF-8*
Perechi propozitii	*39.956*
Nr. cuvinte - Romana	*757.550*
Nr. cuvinte - Engleza	*843.832*
Nr. cuvinte - Total	***1.601.382***

Modelul de adnotare este:

Propoziție:	*cuvânt_1*	*cuvânt_2...*	*cuvânt_x*	*cuvânt_y*	*cuvânt_z*
Tag-uri:	*TAG_1*	*TAG_2 ...*	*TAG_x*	*TAG_y*	*TAG_z*

Exemplu:

a	*fost*	*o*	*decizie*	*politică*	*?*
Va--3s	*Vmp--sm*	*Tifsr*	*Ncfsrn*	*Afpfsrn*	*QUEST*

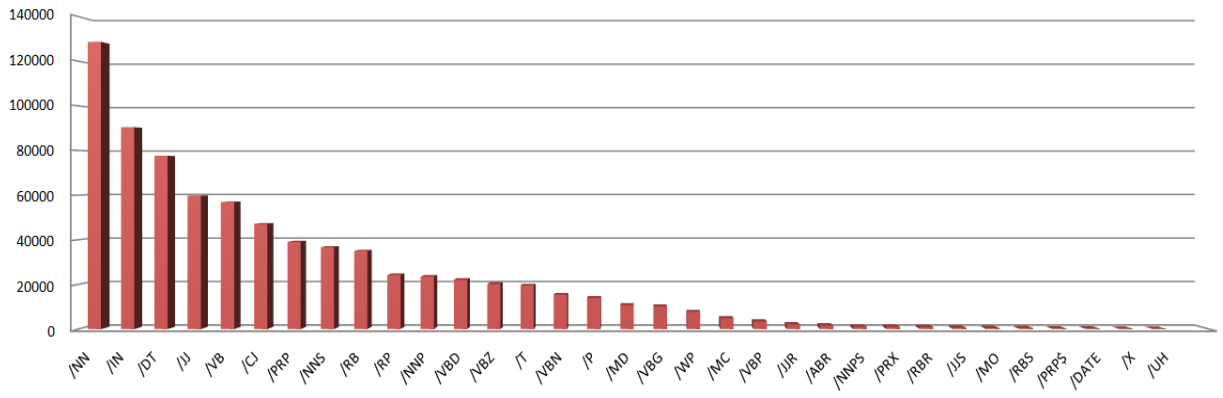

Figura 5. Distribuția tag-urilor pentru corpusul în limba română (de notat: substantivele, 107.101 apariții, au frecvența cea mai mare, urmate de verbe și prepoziții)

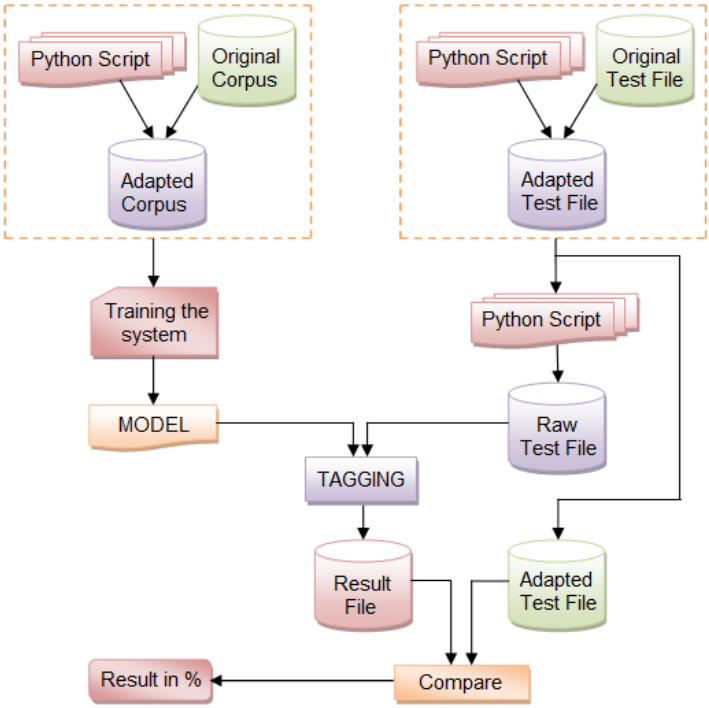

Figura 6. Fluxul de prelucrări pentru adnotarea automată a părților de vorbire (POS)

Rezultatele adnotării automate se realizează, pentru fiecare propoziție în parte, sub forma: *cuvânt / tag,* ca în exemplul de mai jos.

Adnotarea de referință:
*un/T lucru/NN e/VBZ **simplu/JJ** ,/, Adrian_Năstase/NNP nu/RP prea/RB **are/VBZ** concurent/NN serios/JJ în/IN sondaje/NNS și/CJ în/IN partid/NN ./.*

Adnotarea generată de sistemul de predicție automată:
*un/T lucru/NN e/VBZ **simplu/RB** ,/, Adrian_Năstase/NNP nu/RP prea/RB **are/VB** concurent/NN serios/JJ în/IN sondaje/NNS și/CJ în/IN partid/NN ./*

Tabelul 12. Rezultate pentru diferite mărimi ale corpusului (structura entirePOS)

Tool\Training Corpus (in lines)	1.000 linii	5.000 linii	10.000 linii	20.000 linii	30.000 linii	40.000 linii
Kytea - SVM	75.07%	85.06%	87.80%	90.14%	91.74%	**95.87%**
Kytea - LR	75.12%	85.10%	87.80%	90.14%	91.77%	95.80%
HMM Tagger	67.02%	82.84%	85.84%	88.19%	90.19%	93.47%
NLTK - TnT	**75.29%**	**85.50 %**	**88.35%**	**90.27%**	**91.78%**	95.12%

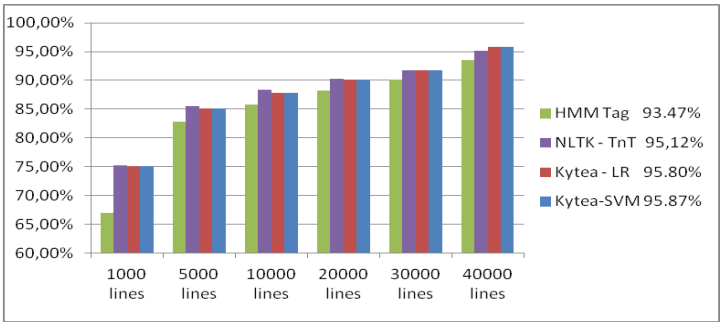

Figure 7. Rezultatele predicției POS pentru diferite mărimi ale corpusului (entirePOS)

Tabelul 13. Rezultate pentru diferite mărimi ale corpusului în limba română (structura onlyPOS)

Tool\Training Corpus (in lines)	1.000 linii	5.000 linii	10.000 linii	20.000 linii	30.000 linii	40.000 linii
Kytea - SVM	76.41%	86.48%	89.30%	91.57%	93.20%	**97.14%**
Kytea - LR	**76.43%**	**86.55%**	89.27%	**91.67%**	**93.25%**	97.08%
HMM Tagger	68.99%	84.69%	87.84%	90.00%	92.09%	95.29%
NLTK - TnT	76.23%	86.47%	**89.43%**	91.54%	93.07%	96.66%

În urma rezultatelor putem spune că cele mai mai bune rezultate pentru limba română s-au obținut folosind Kytea cu SVM. Mult mai multe rezultate sunt prezentate în livrabilul D1.2.

3.1.6. Modulul de sinteză a semnalului vocal

Pentru sistemul de sinteză dezvoltat în cadrul proiectului SWARA algoritmii de generare a semnalului vocal sunt cei bazați pe modele parametrice, de tip probabilist cu lanțuri Markov (HMM). În sistemele parametrice, semnalul este mai întâi parametrizat pentru a obține o reducere a dimensiunii caracteristicilor sale. Această parametrizare este apoi utilizată pentru a antrenare modele probabilistice la nivel de fonem, dependente de context.

Generarea semnalului vocal se va face apoi prin înlănțuirea acestor modele și netezirea traiectoriilor fiecărui set de parametri. Pentru a generaliza modelelor acustice antrenate, lanțurile Markov contextuale sunt grupate în clustere determinate de arbori de decizie binari, a căror splitare se face folosind informația lingvistică contextuală. Astfel că, având un alt text de intrare, nemaiîntâlnit în setul de date de antrenare, deciziile arborilor vor determina accesarea uneia dintre frunzele lor și în acest fel se va putea genera un anumit set de parametri pentru textul respectiv.

Ca și date de antrenare pentru sistemul preliminar, s-au folosit două seturi de înregistrări din baza de date RSS, un set de înregistrări colectate de pe internet și un set înregistrat în cadrul proiectului:

- “Adriana” (RSS): 2.5 ore, la frecvența de eșantionare de 48kHz și având 16bps.
- “Elena” (RSS): 2 ore, la frecvența de eșantionare de 48 kHz și având 16 bps.
- “Victor” (Internet – Cartea Sonoră): 3 ore de înregistrări în studio semi-profesional.
- “Sergiu” (Înregistrare SWARA): 1.5 ore, studio semi-profesional, 48 kHz, 16 bps.

Pe lângă sistemul HTS ce folosește ca și vocoder STRAIGHT, s-a antrenat și o voce ce utilizează doar coeficienții cepstrali și frecvența fundamentală. Această voce are avantajul faptului că necesită un timp mai scurt de generare a semnalului vocal și astfel permite o comunicare în timp real mai bună, însă calitatea este inferioară. Ca și date de antrenare pentru această voce s-a folosit setul „Adriana”. Aceste rezultate au fost diseminate prin lucrarea [8]. Mai multe detalii despre sistemul de sinteză integral, pot fi găsite în livrabilul D6.2.

3.1.7. Demonstratorul online în versiune preliminară

Dupa testarea și validarea individuală a modulelor de procesare a textului și de generare a semnalului vocal descrise mai sus, acestea au fost integrate de către P1 într-o versiune experimentală a unui demonstrator online[1], astfel ca acest demonstrator să poată fi testat de către partenerii P2 și P3. Mostre audio care demonstrează calitatea semnalului sintetizat pentru versiunea preliminară a sistemului de sinteză-text vorbire sunt accesibile in pagina web a proiectului[2].

[1] http://swara.fortech.ro/audio/

[2] *http://speech.utcluj.ro/swara/listeningTest/*

3.2. Identificarea variabilelor psiho-sociale care trebuie personalizate

Inițial s-a realizat o documentare bibliografică ce scoate în evidența că vocea pe care o dobândesc pacienții cu laringectomie ca urmare a utilizării unui sistem de asistare vocală are un rol important în adaptarea psiho-socială a acestora. Este important cum percep pacienții această voce, dar este important și cum o percep interlocutorii În ceea ce privește factorii psiho-sociali care ar putea să afecteze această percepție s-a ținut seama de două perspective.

În primul rând, vocea în sine ar putea să sugereze o serie de caracteristici psihologice sau elemente ce țin de interacțiunea socială. Mai specific, pe lângă conținutul propriu zis al unui mesaj, vocea poate extinde semnificația acestuia prin intonație, pauze, inflexiuni, adică valența emoțională a unui mesaj. Am avut în vedere trei tipuri de valență a mesajului: negativă, pozitivă și neutră. În privința contextului, am abordat două tipuri de contexte: familal, respectiv formal.

În al doilea rând, așa cum au arătat și unele studii, alte variabile ce ar putea să afecteze percepția vocii țin chiar de interlocutor. De exemplu: atitudinea față de o voce sintetică, starea emoțională, experiența în comunicarea cu persoane laringectomizate, vârsta, genul, nivelul de educație. S-a analizat impactul acestor variabile asupra percepției vocilor sintetice în raport cu vocea unui pacient în vederea adaptării corespunzătoare a sistemului de sinteză.

Metodologie. Participanții au fost invitați să participe la o evaluare online și li s-au oferit în mod aleatoriu două seturi de înregistrări: un set cu nouă înregistrări ale unei voci de pacient cu sistem de asistare vocală, respectiv un set cu nouă înregistrări ale aceluiași text, dar generate de sistemul de sinteză vocală în versiunea sa preliminară. Fiecare participant a ascultat în mod aleatoriu una din patru voci sintetice posibile: două cu voce masculină și două cu voce feminină. Cele nouă mesaje transmise de ambele categorii de voci au avut diferite tipuri de valență și nivele de familiaritate. Chestionarul, metodologia aplicată și rezultate în detaliu sunt prezentate în livrabilul D2.2.

Tabelul 8. Caracteristici ale resurselor audio folosite în evaluare

	Voce sintetică			*Voce proteză vocală*		
	Valența mesajului			Valența mesajului		
	pozitivă	negativă	neutră	pozitivă	negativă	neutră
familiar	Poz & Fam	Neg & Fam	N & Fam	Poz & Fam	Neg & Fam	N & Fam
nefamiliar	Poz & Nefam	Neg & Nefam	N & Nefam	Poz & Nefam	Neg & Nefam	N & Nefam
neutru	Poz & N	Neg & N	N & N	Poz & N	Neg & N	N & N

Efecte ale variabilelor demografice. S-a constatat ca variabilele psiho-sociale luate in analiza nu sunt influentate de nivelul de educatie. In schimb, pentru participanții mai în vârstă s-a dedus o asociere semnificativă între voce și nivelul de familiaritate al contextului ($r=0.427$, $p=.015$), respectiv atitudinea față de posibile interacțiuni viitoare cu persoane care au o voce similară cu a pacientului ($r=0.433$, $p=0.13$).

Analiza globală. Pentru aceasta s-a realizat o analiză de varianță mixtă de tip ANOVA (Figura 8.a) având ca factori intra-subiecți tipurile de voce, iar ca factori inter-subiecți genul vocii. S-a identificat un efect semnificativ al tipului de voce, sintetică vs. naturală, de valoare $F(1, 30)=26.582$, $p<0.001$, $\eta 2$parțial$=.47$. Efectul grupului (voce sintetică feminină vs. masculină), respectiv cel de interacțiune (tip voce X grup) au fost nesemnificative ($p>0.05$). Comparațiile mediilor marginale estimate arată scoruri semnificativ mai bune pentru vocile sintetice decât pentru vocea pacientului. În ceea ce privește naturalețea vocii, apare un efect semnificativ intra-subiecți, $F(1, 30)=389.851$, $p<0.001$, $\eta 2$parțial$=0.929$, respectiv efecte nesemnificative pentru grup și interacțiune ($p>0.05$). Pentru celelalte variabile dependente (concordanța cu valența emoțională și nivelul de familiaritate al mesajului, respectiv atitudinea față de interacțiuni viitoare) nu s-a găsit nici un efect semnificativ (toate având $p>0.05$), fapt ce poate sugera că ambele tipuri de voci au fost percepute similar de către participanți.

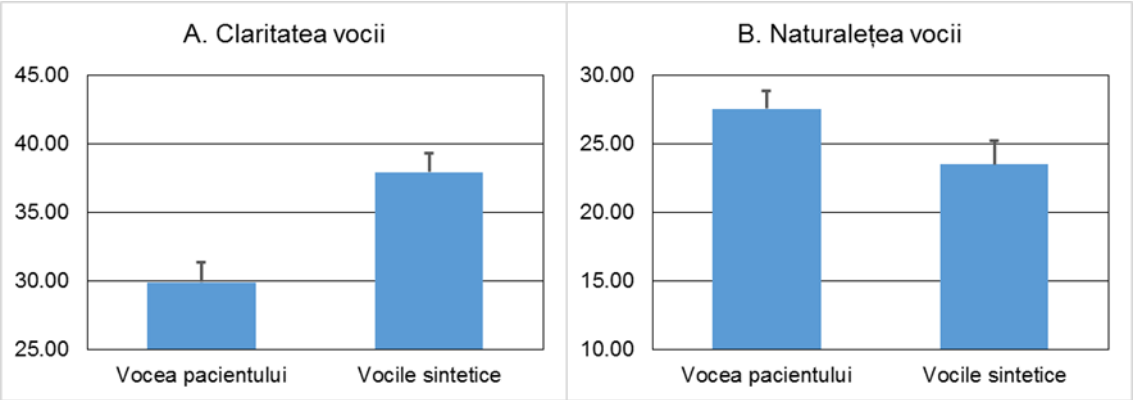

Fig 8.a. Mediile marginale estimate din analiza multifactoriala ANOVA: (A) claritatea percepută a vocii, (B) naturaleţea percepută a vocii (diferenţele sunt semnificative statistic)

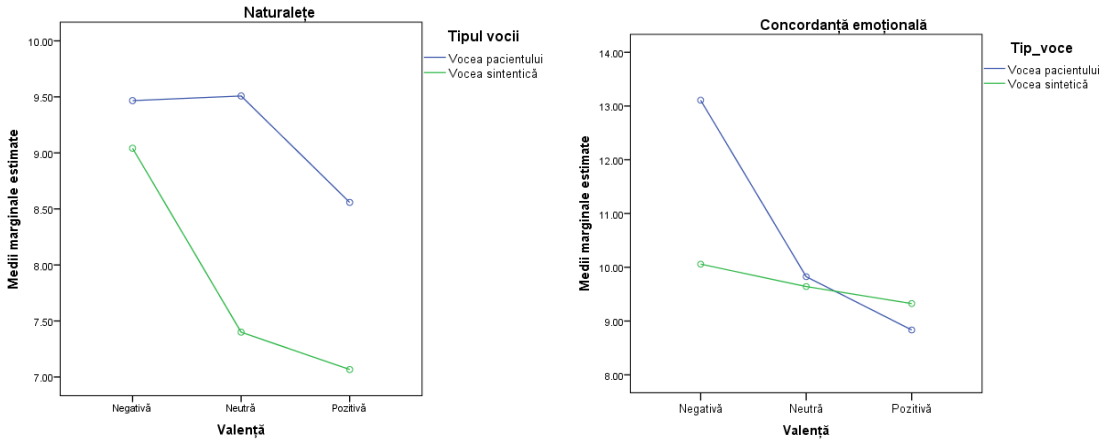

Fig 8.b. Efectul de interacţiune dintre valenţă şi tipul vocii asupra naturaleţii (stanga), respectiv concordanţei emoţionale (dreapta)

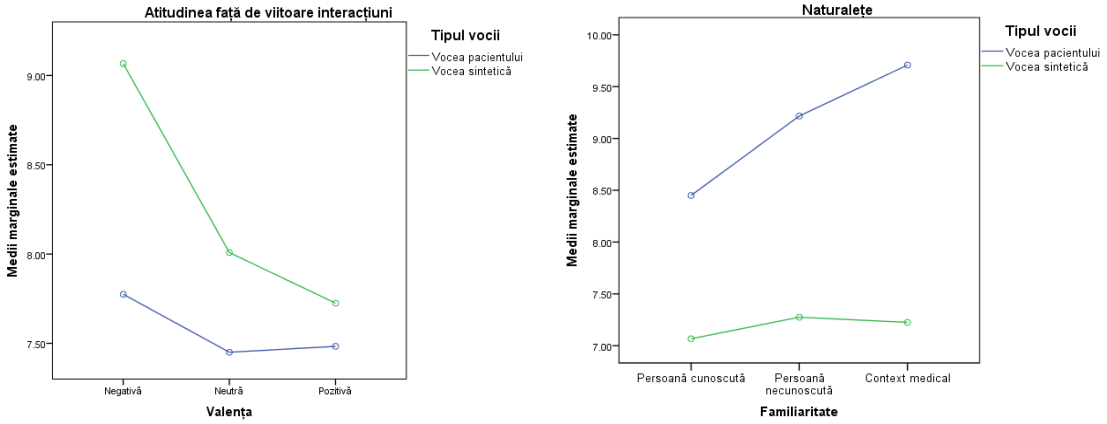

Fig 8.c. Efectul de interacţiune dintre valenţa şi tipul vocii asupra atitudinilor în interacţiunea socială (stânga – scoruri mari indică o atitudine pozitivă, diferenţe apar la valenţa negativă), respectiv dintre familiaritate şi tipul vocii asupra naturaleţii vocii percepute (dreapta – diferenţe semnificative pentru toate contextele şi tipurile de voce)

Figura 8 (a, b, c). Rezultate sintetice privind variabilele care afectează modul de percepţie a vocilor sintetice şi interdependenţa dintre acestea.

Valența emoțională. În ceea ce privește naturalețea vocii în funcție de valența emoțională exprimată s-au identificat câteva efecte semnificative (Figura 8.b). Vocea sintetică este percepută mai clar, indiferent de valența emoțională a mesajului sau de genul vocii sintetice pe care au ascultat-o, $F(1, 30)=26.582$, $p<0.001$, $\eta 2$parțial=.470. De asemenea, s-a identificat un efect de interacțiune între valență și genul vocii sintetice, dar nediferențiat între vocea sintetică și cea a pacientului ($F=2, 60)=3.513$, $p=0.036$, $\eta 2$parțial=0.105. În plus, am obținut un efect semnificativ al interacțiunii dintre toți cei trei factori (tipul vocii, valența și genul vocii sintetice), $F(2, 60)=4.063$, $p=0.042$, $\eta 2$parțial=0.100.

Naturalețea vocii. Tipul vocii a explicat din nou semnificativ varianța naturaleții vocii, $F(1, 30)=5.420$, $p=0.027$, $\eta 2$parțial=.153. Vocea pacientului este percepută ca mai naturală decât cele sintetice. Mesajele care au exprimat o valență negativă au fost percepute mai naturale decât cele care au exprimat valențe neutre sau pozitive. Cele două tipuri de voci sunt similare din perspectiva naturaleții în situația valenței negative, dar sunt percepute diferit (cele sintetice ca fiind mai puțin naturale). În ceea ce privește concordanța emoțională dintre conținutul mesajului și voce, au reieșit semnificative aceleași efecte ca cele pentru naturalețe, dar pattern-urile specifice au fost diferite (Figura 8.b). Există diferențe între cele două tipuri de voci în condițiile valenței negative ($p<0.001$), dar nu și în condițiile celorlalte tipuri de valență ($p>0.05$ pentru ambele comparații) (Figura).

Familiaritatea contextului. Pentru claritatea vocii am identificat un efect semnificativ pentru tipul vocii, $F(1, 30)=26.582$, $p<0.001$, $\eta 2$parțial=0.470, cu scoruri mai mari pentru vocea sintetică. S-a identificat de asemenea un efect semnificativ al familiarității, $F(2, 60)=4.197$, $p=0.020$, $\eta 2$parțial=0.123. Comparațiile perechilor de valori arată o claritate mai redusă pentru mesajele adresate într-un limbaj familiar, prin comparație cu mesajele cu un caracter formal ($p=0.031$), precum și cu cele cu un conținut care face referire la contextul medical (Fig. 8.c).

Starea emoțională. S-a identificat un singur coeficient de corelație semnificativ indicând o relație inversă între distresul (emoții negative) ca dispoziție și evaluarea clarității mesajelor cu valență pozitivă exprimate de vocea pacientului ($r=-0.483$; $p=0.005$). Această corelație indică faptul că participanții cu niveluri mai ridicate de distres au perceput vocea pacientului pentru aceste tipuri de mesaje ca fiind mai puțin clară (Fig.8.c).

Alte variabile psihosociale. În realizarea studiului am planificat să luăm în considerare și o serie de alte variabile psihosociale care ar putea avea un impact asupra percepției vocii pacientului și a celor sintetice. De exemplu: familiarizarea în utilizarea tehnologiilor de asistare vocală, atitudinea personală față de utilizarea tehnologiilor asistive, etc. Rezultatele sunt centralizate în matricea de corelație non-parametrică Spearman.

Rezultatele prezentate într-o maniera mult mai extinsă în livrabilul D2.2 indică faptul că vocea sintetică este percepută ca fiind mai clară, în timp ce vocea generată în baza sistemului de asistare vocală ca fiind mai naturală. Dacă luăm în considerare și valența emoțională, aspectele sunt mai nuanțate. Vocile par să fie similare în privința naturaleței, pentru valențele negative ale mesajelor, dar sunt diferite pentru celelalte tipuri de valențe. Totuși, concordanța emoțională este mai bună pentru valența negativă în cazul pacientului, și dispar diferențele dintre tipurile de voci pentru celelalte condiții ale valenței.

Pe de altă parte, în condițiile unei relații apropiate între interlocutori, vocea sintetică este preferată pentru interacțiuni viitoare. Alte variabile luate în considerare, cum sunt genul participanților, dispoziția și starea lor emoțională, atitudinile lor față de utilizarea tehnologiei în domeniul medical, respectiv utilizarea vocilor sintetice pentru recuperarea vocii persoanelor cu afecțiuni ale laringelui, nu au influențat modul în care vocile sintetice sunt percepute. În studiile viitoare vom ține seama de posibila varietate a vocilor generate prin sistemele de asistare vocală și vom compara vocile sintetice cu un eșantion de astfel de voci, de ambele sexe dacă este posibil.

3.3. Baza de date audio – video, versiunea 1

A fost dezvoltată prima versiune (vers 1) a bazei de date audio-video necesară pentru extinderea numărului de voci disponibile în sistemul de sinteză, pentru adaptarea vocii la vorbitor, precum și evaluarea metodelor de recunoaștere vizuală a vorbirii. În livrabilul D3.2a "Baza de date vers. 1" este descrisă infrastructura funcțională în două locații de înregistrare: una în studioul de Tehnologii Multimedia (pentru partea de înregistrări video cu camere multiple), una în studioul izolat fonic pentru partea de înregistrări audio si video sinconizate.

Datele audio-video în format brut au fost pre-procesate prin activitățile: sincronizarea semnalului audio cu semnalul video, extragerea regiunii de interes din secvențele video, adnotarea semi-automată a semnalului vocal folosind toolkitul dezvoltat de grupul nostru de cercetare, verificarea și evaluarea adnotărilor. Pâna în prezent au fost înregistrați audio 3 noi vorbitori și 1 vorbitor pentru partea de video. Numărul actual de vorbitori noi înregistrați este suficient pentru validarea sistemului preliminar de sinteză pe care l-am expus deja online cu posibilitatea de selectare a vocilor noi înregistrate. De asemenea, au fost realizate cu aceste date o serie de experimente preliminare reușite în ceea ce privește recunoașterea vizuală a vorbirii (vezi livrabilul D3.4a). Structura datelor este realizată astfel încât acestea să fie asociate fiecărui vorbitor, permițând în acest fel asocierea dintre acestea.

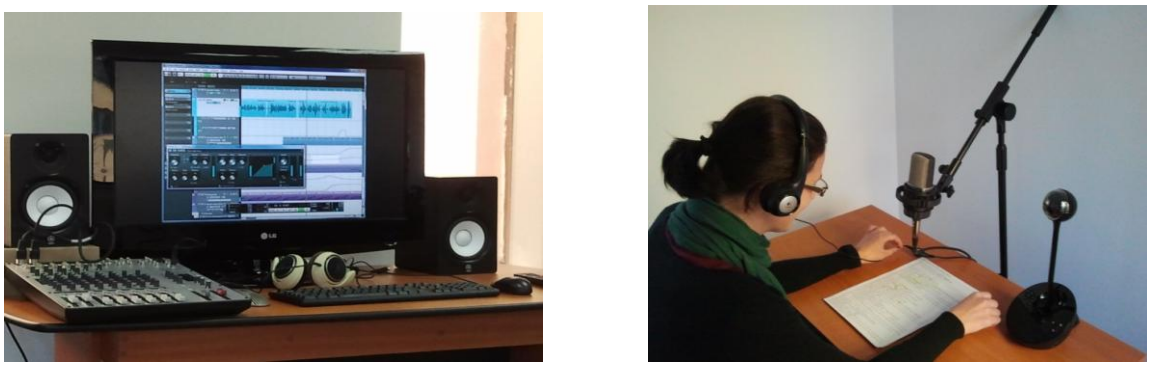

Figura 9. Echipamente pentru înregistrările audio și sesiune de acomodare a vorbitorului

Figura 10. Studioul de înregistrări video și audio

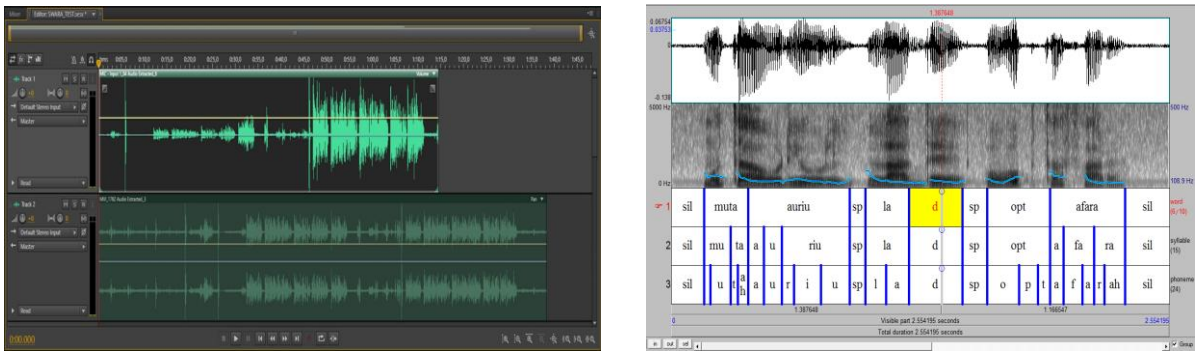

Figura 11. Sincronizarea secvențelor video și audio și adnotarea fonetica a semnalului vocal

3.4. Model de context pentru predicția textului

Există, în principal, două aspecte care au fost considerate la definirea modelului pentru predicția textului (vezi detalii și în livrabil D4.1 „Model de context pentru predicție text”):

1. propunerea de adaptare a metodelor și algoritmilor pentru limba română, pentru a oferi o tehnologie actualizată. Pentru acest scop avem nevoie de două resurse importante: a) resurse sub formă de corpusuri de text în limba română; b) resurse software pentru procesarea datelor text și realizarea predicției textului în funcție de modelul de limbaj.
2. au fost identificate în etapa anterioară (vezi livrabil „D2.1 Raport asupra funcționalităților sistemului și a scenariilor predefinite”) cerințele speciale ale persoanelor cu deficiențe de vorbire raportat la predictia textului intr-un anumit context de utilizare. Pentru acesta aspect am propus să realizăm selectarea unor scenarii sau contexte reprezentative în care predicția poate fi utilizată (ex. comunicare in familie, la medic, în viața de zi cu zi).

Pentru definirea modelului s-a realizat o analiză a stadiului actual în domeniu și s-au reținut o serie de soluții: a) utilizarea indexului inversat pentru predicții la nivel de cuvânt, b) utilizarea arborilor de tip Fuzzy pentru predicții la nivel de frază, c) utilizarea dezambiguării cuvintelor pe bază de dicționare și predicția statistică pe bază de n-grame, d) alte soluții referitoare la contextul utilizatorului și la creșterea vitezei de procesare. Pe baza acestei analize s-a trecut la faza de identificare a posibilelor soluții, prin două etape de teste preliminare.

Într-o primă etapă am utilizat un corpus de text larg (câteva milioane de cuvinte extrase atât din Wikipedia, cât și din trascripturile din Parlamentul României - RomParl) pe baza căruia am generat un model de limbă. În acest model am folosit bigramele și trigramele de cuvinte pentru a realiza predicții pe baza probabilității de apariție a n-gramelor (Tabelul 15). Am constatat ca deși avem un model de limba general, rezultatele predicției rapide a textului sunt modeste din cauză că utilizatorii obișnuiți folosesc de obicei pentru transmiterea mesajelor pe dispozitivele mobile un limbaj mai simplu, cu propoziții scurte și mai puțin elaborate decât modelul generic.

Tabelul 15. Exemple de unigrame, bigrame și trigrame din corpusul de test

	Cuvinte (N=1,2,3)	Log(Probabilitate)	Log(N-gramWeight)
unigrame	*avem*	*-4.3595*	*-0.2116*
	generat	*-4.1834*	*-0.1536*
	avere	*-4.6606*	*-0.0867*
bigrame	*este cinstită*	*-3.7082*	*-0.0491*
	este clar	*-2.8740*	*-0.2252*
	este de	*-1.2477*	*-0.1893*
trigrame	*a face ceva*	*-1.9271*	*indisponibil*
	a face declarații	*-2.2229*	*-*
	a face parte	*-0.9363*	*-*

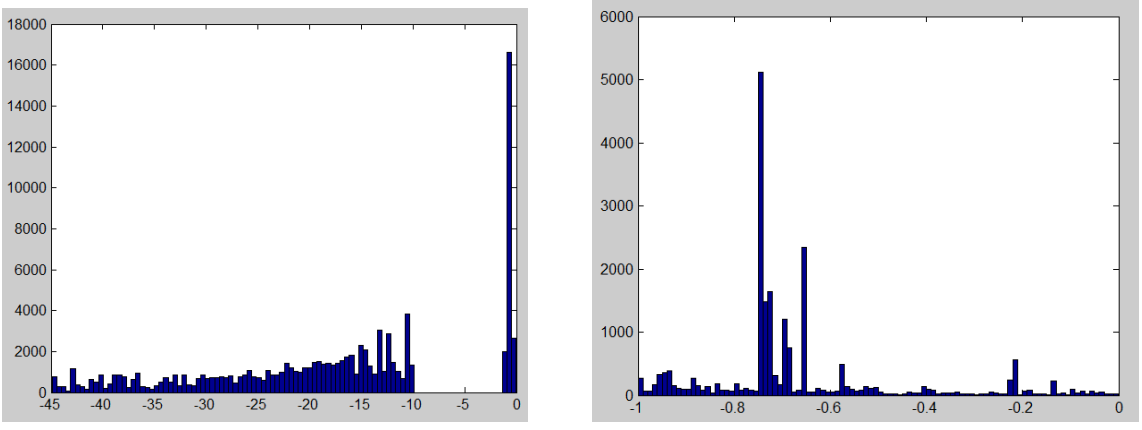

Figura 12. Histograma valorilor probabilităților de apariție a bi-gramelor – stânga, respectiv detaliu pe bigrame cu probabilități mai mari în modul decât pragul Th = -3.

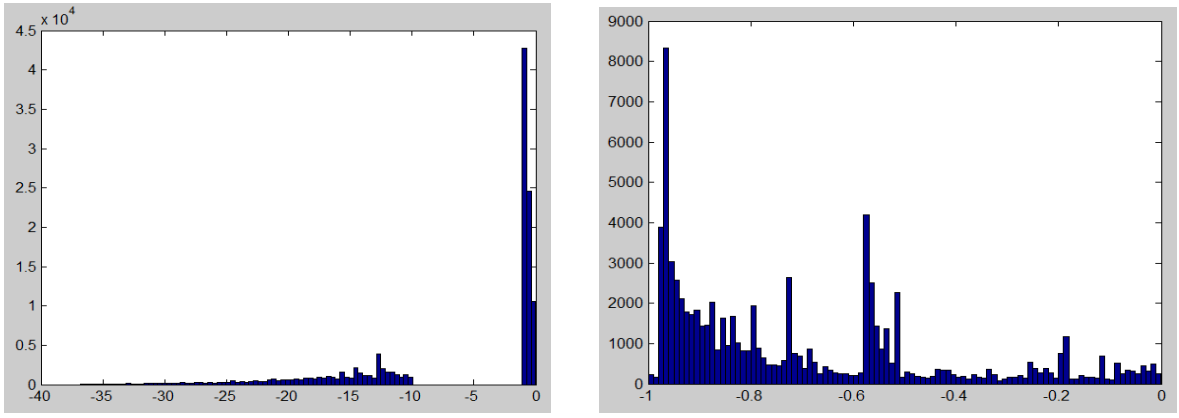

Figura 13. Histograma valorilor probabilităților de apariție a tri-gramelor – stânga, respectiv detaliu pe trigrame cu probabilități mai mari în modul decât Th = -3.

Analiza histogramelor probabilităților n-gramelor pune în evidență: a) distribuția unigramelor arată că există multe cuvinte care apar foarte rar – ca atare, apare ideea de a reține în predicții doar cuvintele care apar mai des; b) distribuția bi-gramelor arată ca în general acestea sunt relativ uniform distribuite, dar există o serie de bigrame cu logaritmul probabilității cuprins între -0,8 și -0,6 care sunt în număr foarte mare – ca atare aceste bigrame le vom considera cu prioritate în predictie; c) distribuția tri-gramelor arată că există o serie de tri-grame cu logaritmul probabilității cuprins între -1,0 și -0,8, respectiv -0,6 și -0,5 care sunt în număr foarte mare – ca atare aceste trigrame le vom considera cu prioritate în modelul de predictie.

Într-o a doua etapă, s-a colectat un corpus de text mult mai mic, specific dialogului pacient – medic. S-a constatat că predicțiile sunt mult mai adecvate cerințelor utilizatorilor, dar totuși modelul de limbaj trebuie antrenat cu un volum mai mare de date. Din acest punct de vedere, pentru livrabilele următoare se are în vedere acest lucru.

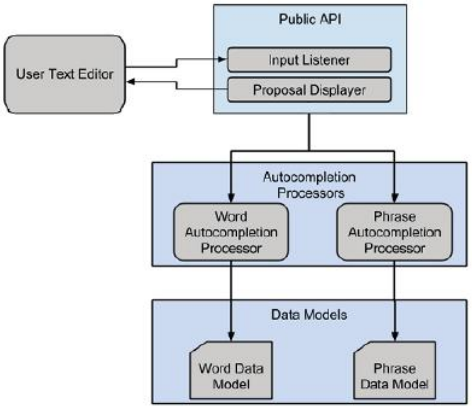

Figura 14. Definirea modelului pentru predicția textului

Pe baza acestor experimente preliminare, conceptul propus se are în vedere trei componente principale (Figura 14), cu functionalități independente și cu canale de comunicare specifice: editorul de text (interfața de intrare cu utilizatorul), procesorul de predicție a textului (interfața de ieșire către utilizator), și modelul de context (modelul folosit pentru generarea predicțiilor).

3.5. Sistem preliminar de predicție a textului

Pe baza modelului de context definit mai sus s-a trecut la cercetarea, dezvoltarea și implementarea practică a unor modele pentru predicția textului. S-a realizat implementarea unui model de predicție a textului bazat pe indexul inversat. Această metodă este independentă de limbă și se bazează pe un index ce asociază cuvintele izolate cu documentele în care acestea se găsesc, urmat de un proces de căutare binară și ordonare a cuvintelor prezise. Sistemul a fost evaluat și validat pe 3 corpusuri de texte:

- **SmallRo:** date extrase din blog-uri și articole de stiri precum și mesaje din reteaua Facebook, toate in limba română. Setul conține 72.000 de cuvinte și are o dimensiune de 3MB.
- **MediumEn:** date extrase din articole de pe Wikipedia precum și documente scrise de utilizator despre produse software, toate în limba engleză. Acest set are 1 milion de cuvinte și dimensiunea de 6MB.
- **BigEn:** date din documente pe diferite teme: fitness, ciclism, jocuri, tehnologii web, articole de uz casmic, telefoane și notebook-uri, chimie, matematica, economie, călătorii, rețete. Setul are 7,4 milioane de cuvinte și ocupă 46MB. Este in limba engleză.

Rezultatele experimentale sunt prezentate în Tabelul 16 și figurile de mai jos.

Tabelul 16. Evaluarea performanțelor folosind fereastra de 3 cuvinte și FL = 4

Data Set	Model	Precision	Recall	Runtime (ms)
SW Products	User Oriented	89%	87%	1
	Simple	71%	61%	1
Facebook Messages	User Oriented	80%	78%	4
	Simple	71%	68%	4
Food recipes	User Oriented	84%	82%	6
	Simple	76%	66%	6

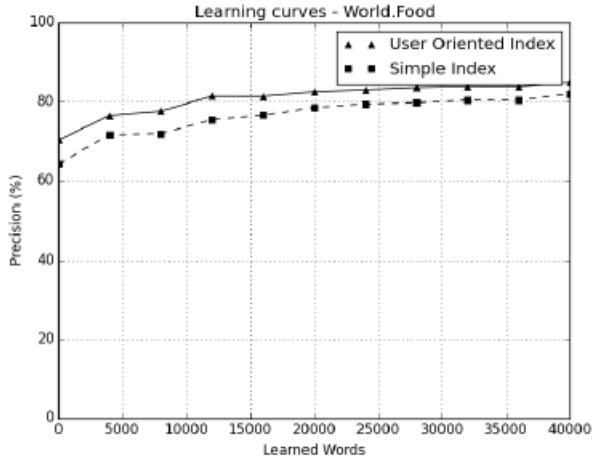

Fig. 15. Precizia pentru Simple vs UserOriented (set de date: BigEn)

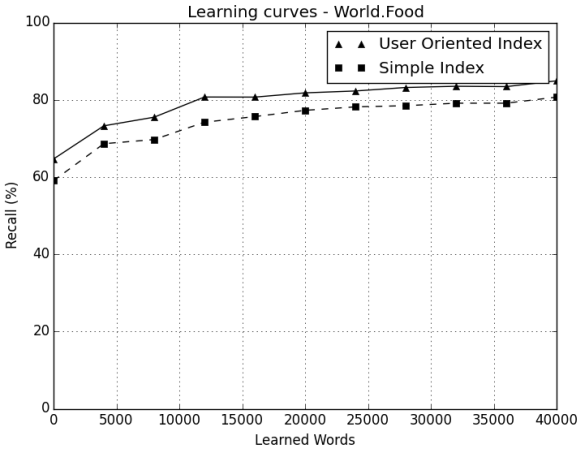

Fig. 16 Recall-ul comparativ pentru BigEn

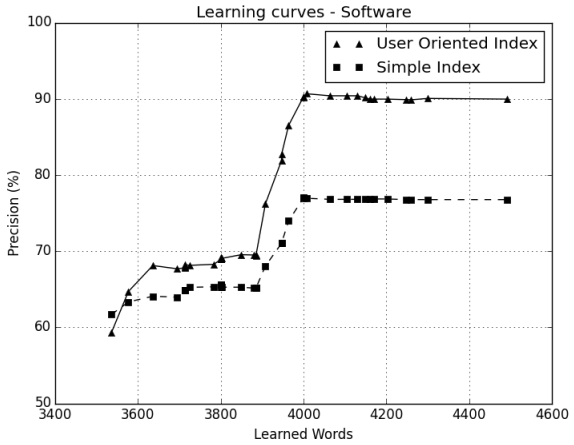

Fig. 17. Precizia pentru Simple vs UserOriented (set de date: MediumEn)

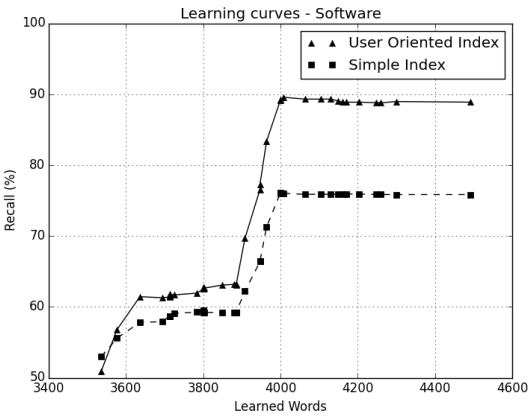

Fig. 18. Recall-ul comparativ pentru MediumEn

Se observă că performața predicției cuvintelor în context de 3 cuvinte este cuprinsă între 71% și 86%, în funcție de corpusul cu care se face testarea. Metoda indexului invers, împreună cu noua strategie de ordonare a cuvintelor, îmbunătățește cu 35% capabilitățle sistemelor tradiționale pentru autocompletarea cuvintelor și conduce în final la o creștere a performanței acestora cu 18%. În plus, prin adoptarea tehnicii de căutare BGBI (Bidirectional Group Boundary Identification) timpul de căutare scade cu 80% față de tehnica de căutare liniară. De asemenea, este demonstrată utilitatea predicției de text la nivel de frază pe echipamente mobile. Aceste rezultate au fost diseminate prin lucrarea [6] și sunt descrise în detaliu în D4.2.

3.6. Metode și experimente preliminare privind recunoașterea vizuală a vorbirii

Atunci când informația audio lipsește din anumite cauze, tehnica de citire a buzelor lipreading (RVV – Recunoașterea Vizuală a Vorbirii) este o alternativă pentru recunoașterea vorbirii. Această tehnică se bazeaza pe interpretarea vizuală a mișcării buzelor, feței și limbii. Descoperirile recente în domeniul procesărilor de imagini, recunoașterii trăsăturilor, dar și de prelucrare a semnalului vocal, au condus la un interes tot mai mare în automatizarea acestei sarcini extrem de dificile de a citi pe buze. RVV a primit o atenție extrem de mare în ultimul deceniu pentru utilizarea ei potențială în aplicații cum ar fi interacțiunea om-calculator, recunoașterea vorbirii cu caracteristici audio-vizuale, recunoașterea limbajului semnelor, dar și supraveghere video. Inițial, s-a realizat o cercetare amănunțită a metodelor de recunoaștere vizuală a vorbirii: metode bazate pe aspect, metode bazate pe transformări de imagini, și metode hibride.

Sistemele bazate pe caracteristici geometrice pornesc de la reprezentarea buzelor prin măsuri geometrice cum ar fi înălțimea sau lățimea, limita buzei exterioare sau interioare, conturul buzelor, forma maxilarelor si obrajilor, deschiderea gurii, cavitatea orala si perimetrul acesteia. Acestea sunt reprezentate prin modele parametrice. Avantajul este că aceste caracteristici extrase sunt de dimensionalitate redusă și invariante la iluminare. Pe de altă parte, aceste modele necesită metode de detecție a caracteristicilor gurii și feței, care in practica sunt destul de greu de aplicat.

Sistemele bazate pe aspect folosesc informațiile nivelelor de gri dintr-o regiune de interes pentru extragerea trăsăturilor. Aceste informații pot fi extrase din imaginea neprelucrată sau după ce aceasta este prelucrată cu ajutorul tehnicilor de procesare de imagini. Vectorii de caracteristici sunt calculați pe baza pixelilor din regiunea de interes. Aceasta abordare include analiza PCA (Principal Component Analysis), bazată pe abordarea “Eigenlips”, unde primii n coeficienți ai tuturor configurațiilor posibile sunt reprezentate ca Eigenlip.

Abordările pe bază de transformări de imagini au în vedere transformări de tip Haar bazate pe varianță, DCT (Discrete Cosine Transform), mașini cu vectori suport pentru detectare de fețe în timp real și filtrare Kalman pentru a realiza detectarea și urmărirea buzelor.

Abordările hibride metodă au în vedere metode de urmărire a buzelor bazate pe combinația între forma, culoarea și mișcarea buzelor. Aceste pot fi integrate într-un model cu stări multiple pentru a reprezenta stările gurii: deschisă, închisă relativ, sau închisă.

Pornind de la acest studiu bibliografic s-a propus experimentarea unui model folosind transformarea DCT (Discrete Cosine Transform) a imaginilor și un sistem de clasificare bazat pe SVM (Support Vector Machines). Datele au fost selectate dintr-un subset al corpusului descris în livrabilul D3.2a „Baza de date vers 1.0” și reprezintă numerele de la „unu” la „zece” pronunțate de câte două ori de un vorbitor masculin, respectiv un vorbitor feminin (Figura 21). Metoda de recunoaștere experimentată preliminar constă din trei etape (Figura 19):

1. **Localizarea buzelor** - Detecția și urmărirea buzelor se poate realiza prin mai multe metode: cu contur activ, cu ajutorul a două semi elipse, combinând culoarea, forma și mișcarea buzelor. În experimentul nostru am ales algoritmul Viola-Jones pentru detecția regiunilor de interes. Regiunea buzelor se detectează doar în primul cadru al secvenței video după care este extrasă regiunea detectată și ajustată de utilizator din fiecare cadru. Algoritmul Viola-Jones folosește clasificatorii Haar pentru a detecta fața respectiv gura și clasificatori în cascadă pentru a mări precizia detectării. Trăsăturile Haar necesare sunt selectate prin algoritmul AdaBoost iar cele nefolositoare sunt rejectate. Această metodă este una dintre cele mai rapide.

2. **Extragerea vectorului cu trăsături** - Extragerea trăsăturilor se realizează prin calcularea coeficienților 3D DCT (Figura 20), ordonarea lor în zig-zag și cuantizarea lor pentru o eficiență ridicată. Ordonarea se face după o funcție hiperboloidă, astfel sunt selectate valorile cele mai semnificative. Se face o limitare a coeficienților stabilind empiric un factor de trunchiere, care în cazul de față este un număr întreg având valoarea egală cu volumul cubului coeficienților împărțit la zece. Trăsăturile sunt salvate într-un fișier binar.
3. **Antrenarea și evaluarea sistemului** - Clasificarea cuvintelor se face prin intermediul SVM (Support Vector Machine). SVM este un algoritm care învață prin intermediul exemplelor să atribuie etichete obiectelor. Pentru a antrena clasificatorul sunt încărcate fișiere care conțin etichetele și trăsăturile.

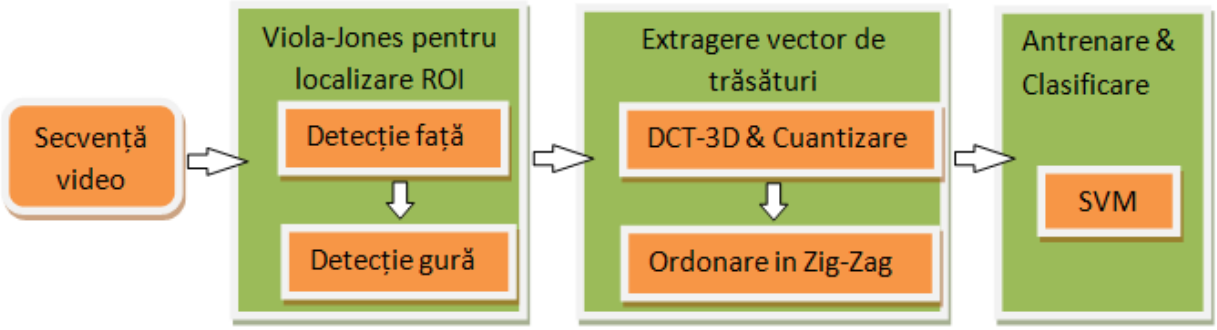

Figura 19. Arhitectura sistemului prototip pentru recunoașterea vizuală a vorbirii

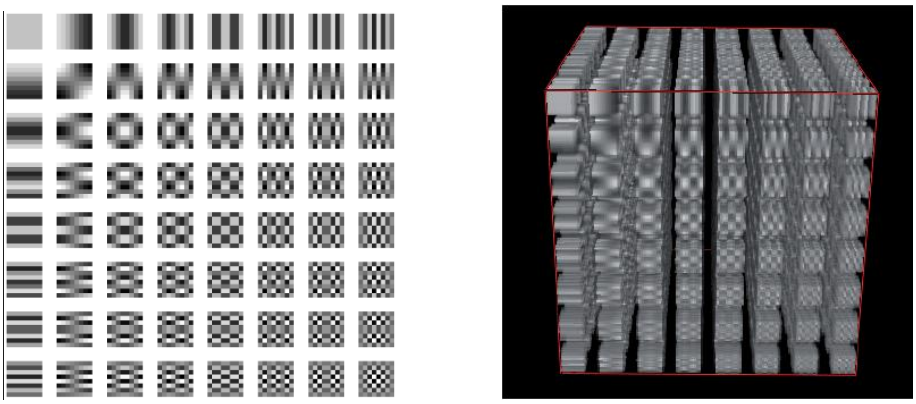

Figura 20. Exemplu cu coeficienții DCT pentru N = 8 (stânga), respectiv 3D-DCT (dreapta)

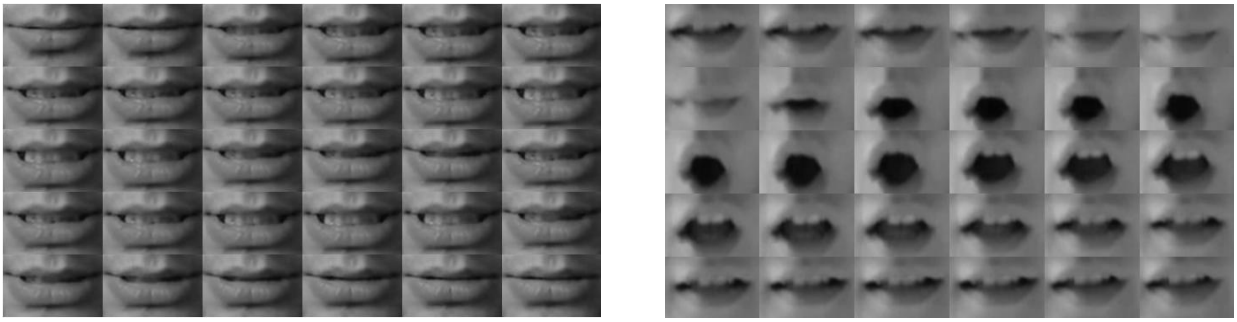

Fig 21. Secvențe cu pronunțiile cifrelor zece, masculin (stînga), respectiv opt, feminin (dreapta)

Sistemul experimental a fost testat pe 20 de secvențe video, iar rata medie de recunoaștere pe acest set este de peste 91%, în funcție de cuvântul pronunțat și de vorbitor. De asemenea, au fost extrase din acest experiment preliminar o serie de concluzii practice privind selectarea zonei de interes, a parametrilor folosiți și a performanței modelelor de clasificare.

Metodele și experimentele raportate sintetic în această secțiune sunt detaliate pe larg în livrabilul D4.3a „Raport privind metodele de recunoaștere vizuală a vorbirii”, anexă la acest raport.

3.7. Versiune experimentală în Cloud a sistemului de sinteză text vorbire accesibilă de pe echipamente mobile

Este implementată și funcțională online[3] versiunea experimentală a sistemului de sinteză text vorbire cu acces din Internet si de pe dispozitive mobile. Sistemul preliminar de sinteză dezvoltat în laborator și descris în D1.2 a fost adaptat pentru functionare pe platforma Cloud. Aceste dezvoltări au la bază specificațiile elaborate în etapa anterioară și descrise în livrabilul D6.1 “Specificațiile sistemului de sinteză în Cloud”. Sistemul experimental de sinteză a vorbirii a fost proiectat folosind modelul client-server și o arhitectură organizată pe mai multe straturi.

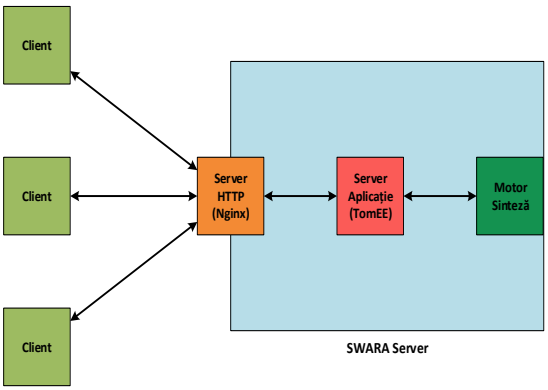

Fig. 22. Modelul client-server al sistemului

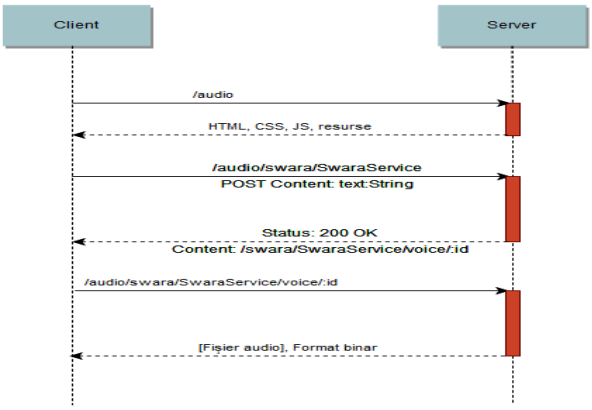

Fig. 23. Diagrama fluxului de procesări

Clientul este reprezentat de interfața web, serverul fiind alcătuit din următoarele componente (Figura 22): serverul HTTP, serverul de aplicație și motorul de sinteză. Aceste componente sunt localizate fizic pe același server în Cloud-ul pe care partnerul P1 îl gestionează. Pentru interfața web s-a propus o soluție care folosește pagini HTML (Hyper Text Markup Language) pentru redarea în browser, respectiv JavaScript, jQuery și AJAX pentru comunicarea cu serverul REST (REpresentational State Transfer).

Pentru serverul HTTP s-au analizat o serie de opțiuni tehnologice, iar pentru infrastructura Cloud folosită s-a ales și justificat instalarea, configurarea și integrarea unui server de tip nginx. Pentru acest server s-au conceput și implementat practic protocolul de comunicare dintre clienți și aplicația web (Figura 23).

Serverul de aplicație este de tip TomEE și mediază legătura dintre client și motorul de sinteză. Pentru motorul de sinteză s-au integrat componentele de procesare de text (Figura 26) dezvoltate de parteneri și s-a găsit soluția tehnică de integrare a acestuia in infrastructura de tip Cloud.

În acest sens, motorul de sinteză (Figura 24) și (Figura 25) este alcătuit dintr-o colecție de scripturi Shell, scripturi Python și programe C care lucrează împreună având ca scop realizarea tuturor procesărilor necesare sintezei audio integrale.

Performanța sistemului experimental a fost testată și evaluată cu succes astfel: test la conexiuni multiple și procesare paralelă a sintezei, teste în vederea optimizării timpilor de execuție, evaluarea securității și vulnerabilității în Internet, evaluarea semnalului sintetizat. Pentru etapa următoare sunt propuse activități de creștere a calitații semnalului sintetizat prin optimizarea performanței modulelor individuale, posibilitatea selectării vocilor, dezvoltarea unor servicii de sinteză in Cloud conform cu planul de business al partenerului industrial.

Alte detalii privind modul de implementare a sistemului sunt disponibile în livrabilul D6.2 „Versiune experimentală a sistemului de sinteza text vorbire în Cloud accesibil pe mobil”, anexă la acest raport. Rezultate au fost diseminate prin lucrarile [8] si [9].

[3] *http://swara.fortech.ro/audio/*

Fig. 24. Interfața de acces în Cloud

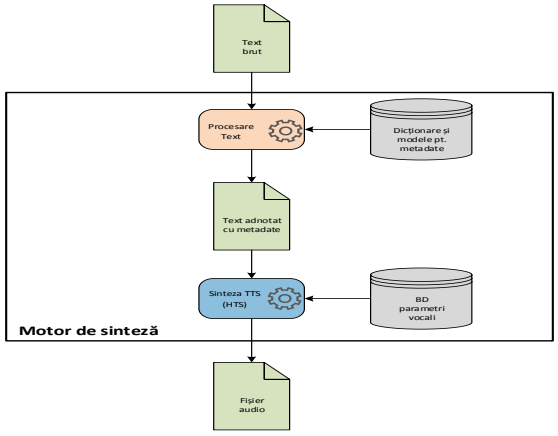

Fig. 25. Motorul de sinteză vocală

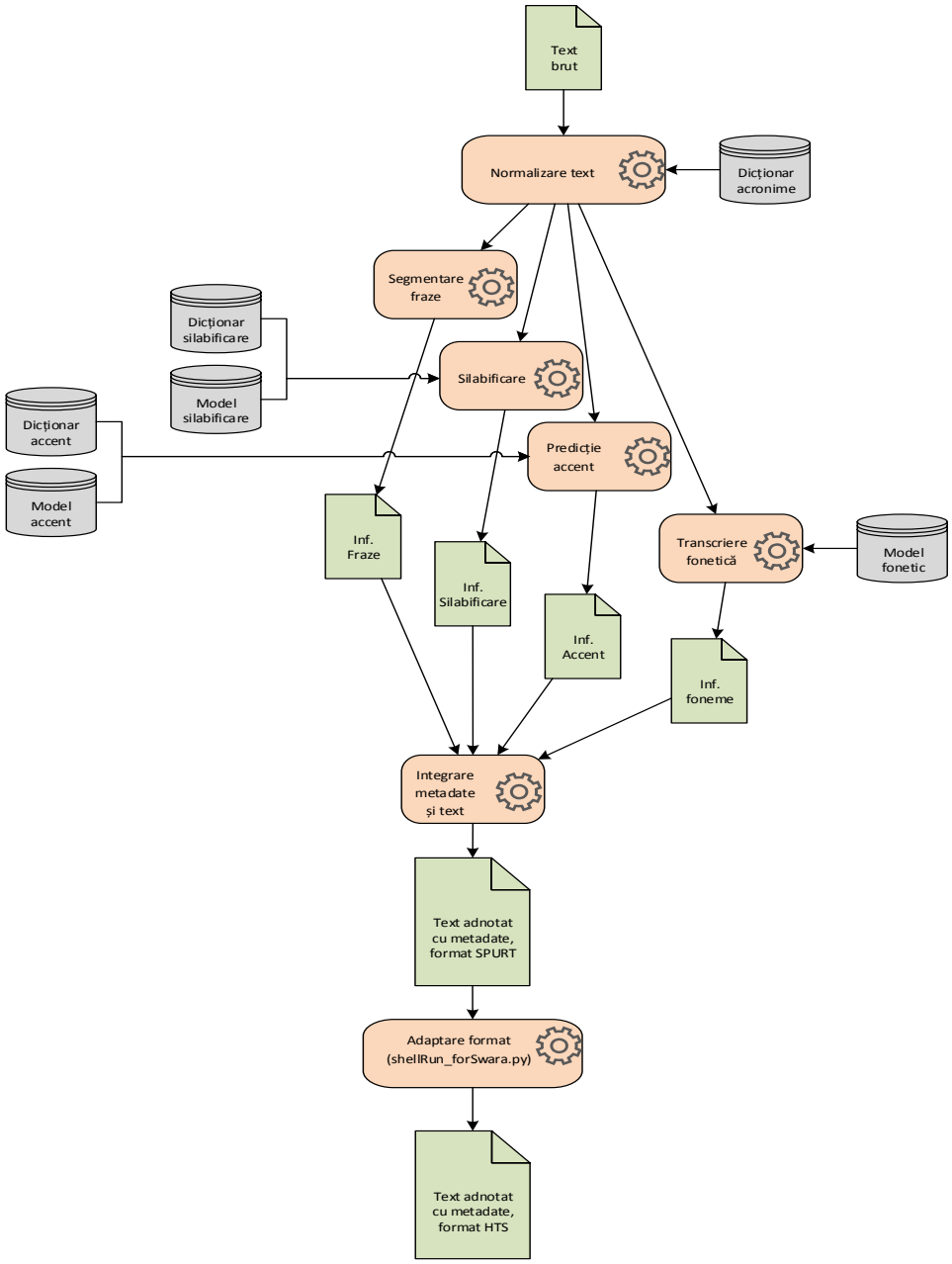

Figura 26. Descrierea structurii interne a componentei de procesare a textului

4. Management si comunicare

Activitățile de management au fost orientate, în special către managementul grupurilor de cercetare constituite în jurul obiectivelor etapei și a interacțiunii dintre acestea. Astfel, întâlniri ale întregului parteneriat au fost realizate în 29.01.2015 (pentru planificarea anuală) și în 08.10.2015 (pentru pregătirea raportării anuale). Grupurile de cercetare au avut întâlniri lunare, cu excepția perioadei din vacanța de vară, iar coordonarea lor s-a facut prin eMail și Skype. În mod special, se remarcă interesul și implicarea partenerului industrial (P1), atât în colaborarea cu coordonatorul, dar și în organizarea reuniunilor de lucru bilaterale între cercetătorii mai tineri.

Din punct de vedere administrativ s-au primit 4 tranșe de avans, s-au întocmit 2 acte aditionale (unul în Martie 2015 pentru reducerea bugetului pe anul 2015, altul în Octombrie 2015 pentru suplimentarea de buget) și s-au derulat la zi activitățile privind achizițiile de materiale pentru izolarea fonică a studioului de înregistrări audio, respectiv dotarea cu echipamentele necesare.

5. Diseminarea rezultatelor

O preocupare continuă a Consorțiului în etapa de raportare a fost implementarea și îndeplinirea cu succes a obiectivelor stabilite în strategia de diseminare a rezultatelor elaborată în cadrul propunerii de proiect. Astfel, adecvat acestei etape de dezvoltare experimentală a componentelor pentru sistemul de sinteză text vorbire pentru asistare vocală s-au identificat canalele de diseminare si s-a acționat pe următoarele direcții: a) actualizarea dinamică a paginii web cu rezultatele obținute incremental in proiect, inclusiv cu secțiuni demonstrative; b) elaborarea unui plan de diseminare pentru anul 2015; c) dezvoltarea de materiale promoționale adecvate etapei; d) publicarea rezultatelor științifice la conferințe internaționale de prestigiu în domeniul proiectului (tehnic, tehnologii medicale asistive).

5.1. Pagina web a proiectului

Pagina web a proiectului[4], are un conținut dinamic, adaptat cu realizările din proiect, astfel ca tot la această adresă se pot accesa mostre demonstrative cu semnale sintetizate, articolele științifice publicate, livrabilele cu caracter public, precum si legături catre serviciile web care se vor dezvolta. Începând cu data de 30.10.2015 a fost instalat și serviciul de monitorizare automată a site-ul, Google Analitics, cu scopul de a evalua modul în care utilizatorii accesează site-ul și ca atare să putem să le adresăm posibile mesaje de colaborare sau de marketing pentru viitoarele servicii comerciale care vor rezulta din acest proiect. La o lună de la instalare (30.11.2015) statisticile sunt: sesiuni: 301, utilizatori: pagini vizualizate: 309, rata de crestere in intervalul de raportare: 98%. Accesul pe țări: neidentificabil: 26,31%, SUA: 21,26%, Rusia: 13,95%, China: 6,31%, Olanda: 4,98%, UK: 4,65%, etc. Modul de structurare și alte informații relevante despre site sunt incluse în livrabilul D7.1 „Pagina web proiect”, anexă la acest raport.

5.2. Planul de diseminare pe anul 2015

Planul de diseminare este un document prin care s-au identificat la începutul anului 2015 (apoi actualizat pe parcurs) posibilitățile de diseminare și publicare de articole la conferințe științifice, prezentări și comunicări ale rezultatelor proiectului la diferite reuniuni cu tematică apropiată de cea a proiectului. Realizarea acestora este în funcție de rezultatele și de resursele pe care consorțiul le are la dispoziție. Conferințe identificate pe profil tehnic: 12, din care s-a participat la 3 cu un număr total de 3 articole. Conferințe identificate pe profil psihologic și medical: 6, din care s-a participat la 3, cu un număr total de 6 articole. Mai multe detalii se găsesc în livrabilul D7.2 „Plan de diseminare”, anexă la acest raport.

5.3. Materiale promoționale

Elaborarea materialelor promoționale a avut în vedere obiective mai mult decât de informare, cât mai ales de conștientizare a tehnologiei propuse, de interacțiune directă cu demonstratorul online, de evaluare a feed back-ului și de implicare a sectorului industrial. Adecvat rezultatelor obținute în această etapă pe direcția dezvoltării la nivel de componente

[4] *http://speech.utcluj.ro/swara*

experimentale ale sistemului asistiv de sinteză vocală s-au realizat următoarele materiale promoționale: a) pliantul proiectului, b) prezentarea PPT a proiectului, c) prezentări PPT și postere ale articolelor susținute la conferințe[5], d) pagina web pentru demonstrarea online a sistemului de sinteză preliminar[6], e) pagină web cu mostre de semnal vocal sintetizat[7]. Detalii se găsesc în livrabilul D7.3 „Materiale promționale", anexă la acest raport.

5.4. Publicații științifice

O parte din rezultatele științifice obținute în etapa de raportare au fost prezentate și publicate la conferințe internaționale de prestigiu, așa cum au fost ele identificate în D7.2 „Plan de diseminare". Alte rezultate sunt în curs de publicare. Pentru vizibilitate directă, publicațiile științifice sunt listate mai jos și incluse în livrabilul D7.4 „Publicații și tutoriale".

[1] M. Chirila, C. Tiple, F.V. Dinescu, SD. Bolboaca, "Clinical investigation of Quality-of-Life data among laryngectomized patients." Proceedings of 49th Annual Scientific Meeting of the European Society for Clinical Investigation, 27-30 May, 2015, Cluj-Napoca, Romania, pg 103-106, Medimond Publishers, Bologna, Italy, ISBN: 978-88-7587-719-4.

[2] Tiple C., Dinescu F. V., Chirila M., Muresan R., Drugan T., Cosgarea M, "Impact on Quality of Life and Voice Handicap in Laryngectomees after Vocal Rehabilitation", In Proc. of the 3rd Congress of European ORL-HNS, 7-11 June, Prague, Czech Republic, 2015

[3] Dinescu F. V., Tiple C., Chirila M., Muresan R., Drugan T., Cosgarea M., "Evaluation of HRQL with EORTC QLQ-C30 and QLQ-H&N35 in Romanian Laryngeal Cancer Patients", Presented at 3rd Congress of European ORL-HNS, Published by : European Archives of Oto-Rhino-Laryngology, Volume 272 / 2015m, ISSN : 0937-4477, Springer Berlin Heidelberg, pp.1-6, 2015.

[4] Matu S., Șoflău R., Cîmpean A., David D., Chirilă M., Ţiple C., Dinescu V., Mureșan R., "What do patients with laryngectomy expect from the next generation of vocal assistive systems? A qualitative and quantitative analysis of users' needs and expected improvements.", In Proc. of the 3rd Congress of European ORL-HNS, 7-11 June, Prague, Czech Republic, 2015

[5] Matu S., Șoflău R., Cîmpean A., David D., Chirilă M., Ţiple C., Dinescu V., Mureșan R., "Psychological Mechanisms Linking Vocal Handicap and Quality of Life in Laryngectomy Patients", In Proc. of the 3rd Congress of European ORL-HNS, 7-11 June, Prague, Czech Republic, 2015

[6] Stefan Prisca, Rodica Potolea, Mihaela Dinsoreanu, "A language independent user adaptable approach for word auto-completion", In Proceedings of the 11th International Conference on Intelligent Computer Communication and Processing (ICCP), ISBN: 978-1-4673-8199-4, pp. 43-49, 3-5 Septemeber, Cluj-Napoca, Romania, 2015.

[7] Diana Balc, Anamaria Beleiu, Rodica Potolea and Camelia Lemnaru, "A learning-based Approach for Romanian Syllabification and Stress Assignment", In Proceedings of the 11th International Conference on Intelligent Computer Communication and Processing (ICCP), ISBN: 978-1-4673-8199-4, pp. 37-42, 3-5 September, Cluj-Napoca, Romania, 2015.

[8] Adriana Stan, Cassia Valentini-Botinhao, Mircea Giurgiu, Simon King, "Phonetic Segmentation of Speech using STEP and t-SNE", In Proceedings of the 8th International Conference on Speech Technology and Human-Computer Dialogue (SpeD), ISBN: 978-1-4673-7559-7, pp.11-16, 14-17 October, Bucuresti, Romania, 2015.

[9] Cristina Tiple, et al, Voice-Related Quality of Life Results in Laryngectomies With Today's Speech Options and Expectations From the Next Generation of Vocal Assistive Technologies, Proc of 5th IEEE Int Conference on eHealth and Bioengineerings, 21-25 Npvember, 2015, Iasi.

6. Concluzii

Activitățile de cercetare desfășurate în etapa a doua de implementare a proiectului (2015) au condus la obținerea rezultatelor asteptate și ele sunt în concordanță cu obiectivele specifice ale etapei. Astfel, rezultatele raportate în acest document și descrise detaliat în cele 11 livrabile aferente perioadei de raportare, pregătesc cadrul de integrare a componentelor în noul sistem de sinteză de înaltă calitate, cu posibilități de creare și adaptare a vocilor sintetice, cu predicția rapidă a textului și accesibil de pe echipamente mobile.

[5] *http://speech.utcluj.ro/swara/publications.html*

[6] *http://swara.fortech.ro/audio*

[7] *http://speech.utcluj.ro/swara/listeningTest/*

7. Referințe la livrabilele aferente etapei a doua, anul 2015 (Anexe la raport)

[1] Livrabil D1.2:	*„Sistem preliminar de sinteză text vorbire”,* Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Decembrie 2015. *Nivel diseminare: Confidential*
[2] Livrabil D2.2:	*„Raport asupra variabilelor psiho-sociale ce trebuiesc personalizate”,* Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Decembrie 2015. *Nivel diseminare: Confidential*
[3]Livrabil D3.2a	*„Baza de date vers.1”,* Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Decembrie 2015. *Nivel diseminare: Confidential*
[4] Livrabil D4.1:	*„Model de context pentru predicție text”,* Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Decembrie 2015. *Nivel diseminare: Confidential*
[5] Livrabil D4.2:	*„Sistem preliminar de predicție text”,* *Nivel diseminare: Confidential*
[6]Livrabil D4.3a	*„Raport privind metodele de recunpaștere vizuală a vorbirii”,* Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Decembrie 2015. *Nivel diseminare: Confidential*
[7] Livrabil D6.2:	*„Versiune experimentală a sistemului de sinteză text vorbire în Cloud, accesibil de pe mobil”,* *Nivel diseminare: Public (http://swara.fortech.ro/audio)*
[8] Livrabil D7.1:	*„Pagina web proiect”,* *Nivel diseminare: Public (http://speech.utcluj.ro/swara/)*
[9] Livrabil D7.2:	*„Plan de diseminare”,* Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Decembrie 2015. *Nivel diseminare: Public*
[10] Livrabil D7.3	*„Materiale promoționale”,* Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Decembrie 2015. *Nivel diseminare: Public(http://speech.utcluj.ro/swara/)*
[11] Livrabil D7.4	*„Publicații și tutoriale”,* Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Decembrie 2015. *Nivel diseminare: (http://speech.utcluj.ro/swara/results.html#publications)*