

RAPORT FINAL DE ACTIVITATE

Aceste rezultate au fost obținute prin finanțare în 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 Vocală în Reintegrarea Persoanelor cu Afonii Chirurgicale”
SWARA

© 2017 – SWARA

Acest document este proprietatea organizațiilor participante în proiect și nu poate fi reprodus, distribuit sau diseminat către terți, fără acordul prealabil al autorilor.

Denumirea organizației participante în proiect	Acronim organizație	Tip organizație	Rolul organizației în proiect (Coordonator/partener)
Universitatea Tehnică din Cluj-Napoca	UTCN	UNI	CO
SC FORTECH SRL	FORTECH	SRL	P1
Universitatea de Medicină și Farmacie Iuliu Hațieganu	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 Final de Activitate
Termen:	30 Septembrie 2017
Editor:	Mircea Giurgiu (Universitatea Tehnică din Cluj-Napoca)
Adresa de eMail editor:	Mircea.Giurgiu@com.utcluj.ro
Autori, în ordine alfabetică:	Magdalena Chirilă, Daniel David, Vera Dinescu, Mircea Giurgiu, Silviu Matu, Rodica Mureșan, Remus Pop, Radu Șoflău, Adriana Stan, Alexandra Tomiță, Cristina Țiple
Ofițer de proiect:	Andreea Matei

Rezumat:

Acest document prezintă o sinteză a realizărilor de natură științifică și tehnică obținute în proiectul SWARA (2014 – 2017). Activitățile de cercetare desfășurate au condus la obținerea rezultatelor așteptate și ele sunt în concordanță cu obiectivele specifice ale proiectului.

Aceste obiective sunt detaliate distribuit ca și obiective specifice pe fiecare etapă de implementare.

Urmează prezentarea modului de implementare a obiectivelor, indicând principalele realizări de natură științifică și tehnică, respectiv rezultatele obținute.

Impactul acestor rezultate este prezentat din mai multe perspective: vizibilitatea online a rezultatelor, materialele promoționale dezvoltate, modul de implicare a studenților și doctoranzilor în proiect, articolele științifice elaborate și publicate la conferințe internaționale, demonstratoarele online, acordul privind drepturile de proprietate intelectuală, planul de exploatare și valorizare a rezultatelor proiectului, propuneri de colaborarea din partea mediului economic în vederea comercializării rezultatelor, detalierea celui mai semnificativ rezultat.

Secțiunea de concluzii prezintă în sinteză cele mai importante rezultate obținute în fiecare etapă de implementare.

În final sunt prezentate lista celor 21 de articole publicate, respectiv lista celor 29 de rapoarte livrabile elaborate pe durata proiectului.

Cuprins

1. Obiectivele prevăzute ale proiectului	4
1.1. Obiectivul I: Sistem de sinteză text-vorbire personalizat.....	4
1.2. Obiectivul II: Tehnoredactare rapidă și eficientă	5
1.3. Obiectivul III: Tehnologie portabilă	5
2. Realizarea obiectivelor și gradul de atingere a rezultatelor estimate	5
2.1. Realizarea obiectivelor aferente Etapei 1, „Resurse și specificații”	5
2.2. Realizarea obiectivelor aferente Etapei 2 „Dezvoltarea experimentală a componentelor”	7
2.3. Realizarea obiectivelor aferente Etapei 3 „Integrarea componentelor”	10
2.4. Realizarea obiectivelor aferente Etapei 4 „Adaptare sistem și integrarea lui în serviciul de sinteză în Cloud”	12
3. Impactul rezultatelor	14
3.1. Vizibilitatea online a rezultatelor	14
3.2. Materiale promoționale	14
3.3. Implicarea studenților, masteranzilor și doctoranzilor	14
3.4. Implicarea utilizatorilor	15
3.5. Activitatea de publicare a rezultatelor	15
3.6. Demonstratoare online	16
3.7. Acord privind drepturile de proprietate intelectuală	16
3.8. Planul de exploatare a rezultatelor proiectului	16
3.9. Propuneri de colaborare din partea mediului economic.....	17
3.10. Rezultatul cel mai semnificativ	18
4. Concluzii	20
5. Lista articolelor științifice elaborate și publicate la conferințe de prestigiu	20
6. Lista livrabilelor elaborate pe durata proiectului	22

1. Obiectivele prevăzute ale proiectului

SWARA¹ (*Sistem Mobil de Asistare Vocală în Reintegrarea Persoanelor cu Afonii Chirurgicale*) are scopul de a ajuta persoanele care și-au pierdut abilitatea de a vorbi să fie capabile să ia parte la viața cotidiană, fără restricții privind comunicarea cu cei din jur. Laringectomia totală efectuată pentru un cancer laringian sau hipofaringian avansat este cea mai frecventă cauză de afonie. Sistemele de sinteză de voce sunt deja utilizate ca tehnologie de asistare a persoanelor care din diverse motive și-au pierdut capacitatea de a vorbi.

Astfel că, principalul scop al acestui proiect este crearea unei tehnologii de sinteză a vorbirii care le va permite persoanelor cu afonii, dar care nu au deficiențe motorii sau cognitive, să poată să comunice cu orice altă persoană, fără teama de a fi respins sau înțeles greșit. Acest deziderat este atins prin stabilirea unui set de obiective principale, ca mai jos.

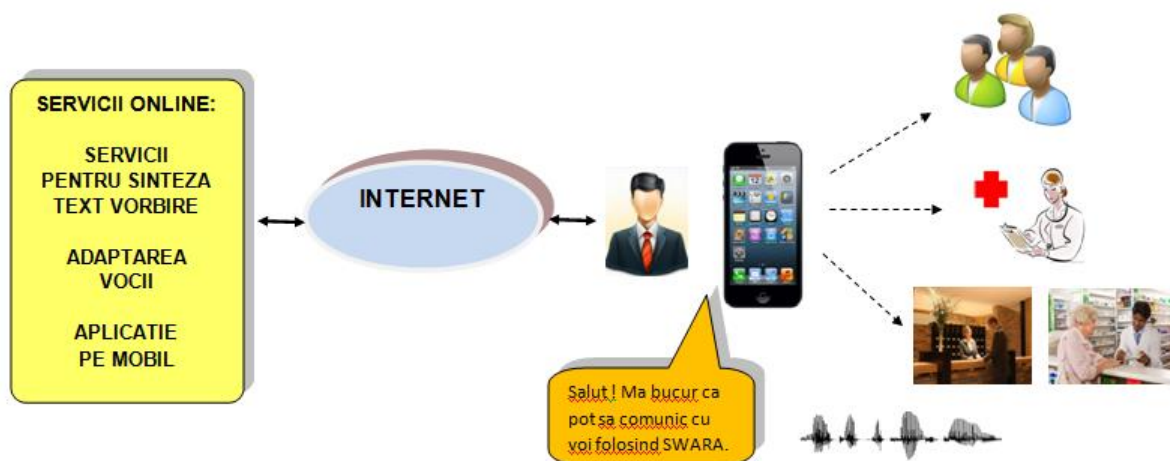


Figura 1. „SWARA” va ajuta pacienții laringectomizați să poată folosi sinteza din text a vorbirii ca o modalitate de comunicare cu familia, cu medicul, sau în activitățile de zi cu zi.

1.1. Obiectivul I: Sistem de sinteză text-vorbire personalizat

Sistemele de sinteză text-vorbire pot fi utilizate ca și tehnologie asistivă pentru persoanele cu deficiențe de vorbire. Totuși, aceste sisteme au dezavantajul că oferă un număr limitat de voci sintetizate, fapt ce face ca pacienții să nu perceapă sistemul ca fiind reprezentativ pentru ei. Vom proiecta și dezvolta un sistem ce va permite generarea vocilor personalizate în cadrul sistemului de sinteză. Acest obiectiv va fi atins prin utilizarea metodelor de adaptare la vorbitor de ultimă generație, pentru a crea voci ce se aseamănă cât mai bine cu vocea pacientului. În cazul în care nu există o voce de referință (pacientul nu a putut fi înregistrat înainte de a-și pierde vocea), pacientul va putea selecta sau crea o voce din cele disponibile în banca de voci. Pentru o mai bună înțelegere a modului în care a fost implementat acest obiectiv pe durata proiectului, prezentăm dezvoltarea obiectivelor specifice pe ani.

Tabel 1. Implementarea pe etape a Obiectivului I – Sistem de sinteză text-vorbire personalizat

Obiective specifice, conform cu numerotarea din rapoartele anuale	
Etapa 1, 2014	1.a. Indexarea resurselor și elaborarea specificațiilor pentru sistemul de sinteză 1.c. Crearea infrastructurii tehnice pentru baza de date audio-video
Etapa 2, 2015	2.a. Dezvoltarea unei versiuni preliminare a sistemului de sinteză text vorbire 2.c. Dezvoltarea primei versiuni a bazei de date audio-video

¹ „Swara” înseamnă „voce” în limba javaneză, limbă vorbită în Indonezia

	<i>2.g. Diseminarea rezultatelor intermediare</i>
Etapa 3, 2016	<i>3.a. Realizarea sistemului de sinteză text vorbire de înaltă calitate</i> <i>3.c. Dezvoltarea versiunii a 2-a a bazei de date audio-video</i>
Etapa 4, 2017	<i>4.a. Realizarea unui sistem de adaptare a vocilor sintetizate</i> <i>4.e. Dezvoltarea planului de exploatare și drepturile de proprietate intelectuală</i>

1.2. Obiectivul II: Tehnoredactare rapidă și eficientă

Daca numărul mediu de cuvinte rostite pe minut este de aproximativ 150 de cuvinte, în scriere acesta scade la 10-15 cuvinte pe minut. De aceea, este foarte important să se dezvolte metode alternative de editare a textului care să permită o medie de scriere apropiată de cea a vorbirii. Proiectul va dezvolta un sistem de predicție a textului în limba română, adaptabil la context, cu accent special asupra nevoilor și scenariilor de utilizare ale persoanelor cu deficiențe de vorbire.

Tabel 2. Implementarea pe etape a Obiectivului II – Tehnoredactare rapidă și eficientă

Obiective specifice, conform cu numerotarea din rapoartele anuale	
Etapa 1, 2014	<i>1.b. Identificarea funcționalităților sistemului pe baza unei analize psiho-cognitive</i>
Etapa 2, 2015	<i>2.b. Identificarea variabilelor psihic-sociale care trebuie personalizate</i> <i>2.d. Dezvoltarea unui model și a unui sistem preliminar pentru predicție text</i> <i>2.e. Identificarea metodelor de recunoaștere vizuală a vorbirii</i>
Etapa 3, 2016	<i>3.d. Realizarea demonstratorului pentru redactarea cu predicție a textului</i> <i>3.b. Evaluarea preliminară a satisfacției utilizatorilor</i>
Etapa 4, 2017	<i>4.c. Experimente privind recunoașterea vizuală a vorbirii</i>

1.3. Obiectivul III: Tehnologie portabilă

Echipamentele asistive de sinteză de voce trebuie să fie disponibile în orice situație pe parcursul activităților zilnice. Odată cu avansul tehnologie hardware, telefoanele inteligente și tabletele dețin resursele computaționale necesare dezvoltării de aplicații proprii, complexe. SWARA își propune să creeze o aplicație mobilă, inteligentă, pentru sistemul de sinteză. Aceasta va include optimizări atât ale algoritmilor, cât și a codului pentru aplicația software. Cu această alternativă la dispoziție, pacienții vor putea să utilizeze vocea sintetizată în orice situație conversațională folosind un serviciu de sinteză disponibil în Cloud.

Tabel 3. Implementarea pe etape a Obiectivului III – Tehnologie portabilă accesibilă în Cloud

Obiective specifice, conform cu numerotarea din rapoartele anuale	
Etapa 1, 2014	<i>1.d. Elaborarea specificațiilor pentru sistemul de sinteză în Cloud</i>
Etapa 2, 2015	<i>2.f. Dezvoltarea unui sistem experimental de sinteză accesibil în Cloud</i>
Etapa 3, 2016	<i>3.e. Diseminarea rezultatelor intermediare</i>
Etapa 4, 2017	<i>4.d. Dezvoltarea unui serviciu Cloud pentru sinteză text vorbire accesibil de pe mobil</i> <i>4.b. Evaluarea impactului psiho-social al vocilor sintetizate, 4.f. Diseminare.</i>

2. Realizarea obiectivelor și gradul de atingere a rezultatelor estimate

2.1. Realizarea obiectivelor aferente Etapei 1, „Resurse și specificații”

Obiectiv1.a: *Indexarea resurselor (audio, text si instrumente software) si elaborarea specificațiilor pentru sistemul de sinteza*

Grad realizare: Obiectiv realizat integral

Activități: Au fost identificate si pregătite corpusurile audio si de text necesare pentru

dezvoltarea sistemului de sinteza, s-a pregătit platforma hardware pentru stocarea resurselor (NAS – Network Attached Storage) și instalarea/testarea principalelor instrumente software la partenerii cu profil tehnic, s-au elaborat specificațiile pentru modulele sistemului de sinteza.

Rezultate:

- Resurse audio si de text adnotate, pregătite pentru dezvoltare
- Instrumente software de baza pentru construirea unui sistem de sinteza
- 1 livrabil (D1.1) sub forma de raport cu titlul „Resurse disponibile si specificațiile sistemului de sinteza”, cu specificarea cerințelor pentru sistemul de sinteza.

Obiectiv1.b: *Identificarea funcționalităților sistemului si a scenariilor de utilizare pe baza unei analize psiho-cognitive si medicale*

Grad realizare: Obiectiv realizat integral

Activități: S-a elaborat planul de chestionare a pacienților si aparținătorilor acestora, s-au elaborat chestionarele, s-a contactat grupul țintă, s-au organizat sesiuni de evaluare prin focus grupuri, s-au analizat si interpretat rezultatele.

Rezultate:

- Chestionare de evaluare
- Rezultatele sesiunilor de tip focus grup si interpretarea acestora
- 1 livrabil (D2.1) sub forma de raport cu titlul „Raport asupra funcționalităților sistemului si a scenariilor predefinite”

Obiectiv1.c: *Crearea infrastructurii tehnice pentru baza de date audio-video*

Grad realizare: Obiectiv realizat integral

Activități: S-a definit structura si s-a estimat conținutul înregistrărilor (conform cu standardul aplicat si in corpusul RSS), s-au făcut instalările hardware si software necesare pentru înregistrări audio-video, s-a elaborat un formular de acord a persoanelor înregistrate privind folosirea vocilor sintetice.

Rezultate:

- Infrastructura tehnica pregătită pentru achiziția de semnale audio-video.
- 1 livrabil (D3.1) sub forma de raport cu titlul „Cadrul tehnic al bazei de date”

Obiectiv1.d: *Elaborarea specificațiilor pentru sistemul de sinteza in Cloud*

Grad realizare: Obiectiv realizat integral

Activități: S-a analizat si testat de către firma partenera soluția propusa de către Coordonator si s-au identificat cerințele principale pentru dezvoltarea ca serviciu web, prin teste s-au identificat posibilitățile de optimizare a codului, s-au propus câteva soluții pentru componentele din interfața utilizator.

Rezultate:

- Rezultate ale testărilor si propuneri de optimizare
- 1 livrabil (D6.1) sub forma de raport cu titlul „Specificațiile sistemului de sinteza in Cloud”

Dezvoltarea sistemului de sinteza în Cloud a avut în vedere crearea unui ansamblu de servicii web pe o platforma de tip IAS (Internet as a Service) capabila sa integreze componentele software de analiza a textului si de sinteza de semnal vocal prin intermediul unei interfețe HTTP unificate, de tip REST (Representational State Transfer), pentru a fi accesata atât de aplicații web, cat si de pe dispozitive mobile conectate la Internet. S-a propus o astfel de soluție, deoarece expunerea serviciilor online și în Cloud permite accesul unei game largi de aplicații la sistemul de sinteză, de la aplicații bancare și medicale, la sisteme interactive de răspuns vocal, cititoare de cărți electronice, navigație și altele.

2.2. Realizarea obiectivelor aferente Etapei 2 „Dezvoltarea experimentală a componentelor”

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
- 5 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
- 1 modul pentru antrenarea modelelor acustice din datele audio și text
- 1 modul de sinteză de voce bazat pe sistemul HTS
- 1 demonstrator online de sinteză text vorbire în variantă preliminară
- 3 articole științifice publicate la conferințe internaționale [6] [7] [8]
- 1 livrabil (D1.2) cu titlul „Sistem preliminar de sinteză text vorbire”.

S-au dezvoltat versiuni îmbunătățite ale componentelor software necesare pentru procesarea textului: normalizare text, transcriere fonetică automată folosind o soluție hibridă ce combină lexiconul fonetic dezvoltat și arbori de decizie (acuratețe între 92% și 99%, în funcție de baza de date de antrenare și de fereastra de analiză), predicția accentului și silabificare folosind arhitecturi de tip Random Forest (acuratețe tipică de 98%), predicția automată a părții de vorbire folosind mai multe tehnologii pe baza de mașini cu vectori suport, regresie liniară și modele statistice de tip HMM - Hidden Markov Models (acuratețe între 93% și 97% în funcție de complexitatea dorită a etichetării). Pentru toate aceste module de procesare a textului, cercetările s-au axat pe studiul, dezvoltarea și experimentarea unor metode care au în vedere tehnici de învățare automată și de predicție, mai degrabă decât utilizarea de dicționare sau lexicoane, tocmai cu scopul de optimiza memoria necesară și timpul de procesare. Pentru toate componentele de procesare a textului s-au realizat ample testări și evaluări pentru a identifica modelul cu cea mai bună performanță.

Apoi, a fost optimizată și componenta de generare a semnalului vocal, prin identificarea parametrizării optime și a selectării unor combinații de parametri de configurare care să asigure o mai bună natură a sintezei. Pe baza corpusului de date cu semnal audio colectat de la doi vorbitori, respectiv dintr-un audiobook, au fost create suplimentar încă trei voci sintetice, față de sistemul de bază. Aceste voci sunt evaluate de către utilizatori prin disponibilitatea online a sistemului de sinteză.

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
- 5 articole științifice publicate la conferințe internaționale [1] [2] [3] [4] [5]
- 1 livrabil (D2.2) cu titlul „Raport asupra variabilelor psiho-sociale ce trebuie personalizate” și care descrie rezultatele menționate mai sus.

Rezultatele prezentate în acest livrabil oferă mai multe sugestii privind aspectele care merită luate în considerare cu privire la sistemul de sinteză. Se pare că vocile sintetice utilizate în acest moment par să sufere în ceea ce privește capacitatea lor de a simula valențe emoționale, dar depășesc vocile generate cu sisteme de asistare, în ceea ce privește claritatea.

Trei variabile psihosociale au reieșit ca având impact asupra percepției diferențiate ale celor două tipuri de voce, din care cel puțin două ar merita luate în considerare în vederea unei eventuale personalizări. Una dintre ele este valența emoțională. Vocea sintetică are cele mai scăzute performanțe în condițiile valenței pozitive și a celei neutre. Pe de altă parte, contextul în care este transmis mesajul afectează, de asemenea, percepția vocilor. De data acestea, vocea pacientului este percepută ca fiind mai naturală în orice condiție, dar vocea sintetică este

preferată pentru posibile interacțiuni viitoare cu persoane apropiate, fără a exista însă diferențe în ceea ce privește o astfel de atitudine pentru contextele mai formale. Practic, vocea sintetică ar putea aduce mai multă plus valoare mai ales în situațiile acestor contexte din urmă (de ex. în cazul interacțiunii cu medicul).

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

Într-o primă etapă am utilizat un corpus de text larg (câteva milioane de cuvinte extrase atât din Wikipedia, cât și din transcripțiile 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. Am constatat că deși avem un model de limbă 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.

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. Acest concepte au fost implementate în două experimente pilot: predicția cuvintelor folosind indexul inversat, respectiv o aplicație pe telefon mobil pentru predicția la nivel de propoziție.

Metoda propusă pentru predicția 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 cu câte 72.000, 102.000, respectiv 7,4 milioane de cuvinte. Performanța 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 îmbunătățește cu 35% capabilitățile sistemelor tradiționale pentru autocompletarea cuvintelor și conduce în final la o creștere a performanței acestora cu 18%. În plus, prin adoptarea tehnicii de căutare Bidirectional Group Boundary Identification timpul de căutare scade cu 80% față de tehnica de căutare liniară.

Obiectiv2.e: Identificarea metodelor de recunoaștere vizuală a vorbirii

Grad realizare: Obiectiv realizat integral

Rezultate:

- 3 tipuri de metode identificate și raportate
- 1 experiment de recunoaștere vizuală a vorbirii
- 1 livrabil: (D4.3a) cu titlul „Raport privind metodele de recunoaștere vizuală a vorbirii”.

Au fost identificate și prezentate succint o serie de metode de recunoaștere vizuală a vorbirii: metode bazate pe caracteristici geometrice, metode bazate pe aspect, metode bazate pe transformări de imagini, și metode hibride.

Pornind de la această analiză s-a proiectat și implementat un 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.

Metoda de recunoaștere experimentată constă din trei etape: 1) localizarea buzelor cu ajutorul algoritmului Viola-Jones și realizarea unei secvențe video conținând zona buzelor, 2) extragerea vectorului de trăsături folosind DCT și ordonarea coeficienților în zig-zag pentru optimizarea clasificării, 3) antrenarea sistemului și evaluarea acestuia folosind SVM.

Sistemul experimental a fost testat pe 20 de secvențe video, iar rata medie de recunoaștere pe acest set este de peste 90%, 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.

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

Grad realizare: Obiectiv realizat integral

Rezultate:

- 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
- 2 articole despre tehnologii asistive [4] [9]
- 1 livrabil: (D6.2) cu titlul „Versiune experimentală a sistemului de sinteză text vorbire în Cloud accesibil de pe mobil”.

Sistemul de sinteză a vorbirii a fost proiectat folosind modelul client-server și o arhitectură stratificată. Clientul este reprezentat de interfața web, serverul fiind alcătuit din următoarele componente: serverul HTTP, serverul de aplicație și motorul de sinteză. 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).

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 dezvoltate de parteneri și s-a găsit soluția tehnică de integrare a acestuia în infrastructura de tip Cloud. În acest sens, motorul de sinteză 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. Versiunea experimentală a sistemului de sinteză text vorbire este accesibilă online printr-o aplicație web pusă la dispoziția utilizatorilor în Cloud la adresa <http://swara.fortech.ro/audio/speech>.

Obiectiv2.g: *Diseminarea rezultatelor pe anul 2015*

Grad realizare: Obiectiv realizat integral

Rezultate:

- actualizarea dinamică și monitorizarea cu Google Analytics a site-ului
- materiale promoționale: pliant, prezentarea PPT a proiectului, 2 postere articole, pagina web pentru demonstratorul online, pagina web cu mostre de semnal vocal sintetizat (<http://speech.utcluj.ro/swara/>)
- 9 articole prezentate și comunicate la conferințe internaționale[1]...[9]
- 4 livrabile referitoare la diseminare: D7.1, D7.2, D7.3, D7.4.

2.3. Realizarea obiectivelor aferente Etapei 3 „Integrarea componentelor”

Obiectiv3.a: *Realizarea sistemului de sinteză text vorbire de înaltă calitate*

Grad realizare: Obiectiv realizat integral

Rezultate:

- noi resurse de date audio de înaltă calitate (12 noi vorbitori înregistrați, aproximativ 21 de ore de semnal audio) și de text
- îmbunătățirea și validarea finală a modelelor de procesare a textului
- 1 nou modul pentru antrenarea modelelor acustice din datele audio și text folosind tehnologii actuale performante de tip DNN
- 1 nou sistem integrat de sinteză text vorbire de înaltă calitate
- 1 demonstrator online de sinteză text vorbire
- 1 articol științific publicat la conferințe internaționale [11]
- un livrabil (D1.3) cu titlul „Sistem de sinteză text vorbire de înaltă calitate”, care descrie rezultatele menționate mai sus.

Noul sistem de sinteză text vorbire este bazat pe rețele neuronale de tip DNN (Deep Neural Networks). În primul rând au fost aduse îmbunătățiri la unele componente din modulul de procesare a textului. Aceste rezultate se referă la: restaurarea automată a diacriticelor și dezvoltarea unei metode inovative, performante, privind silabificarea. În al doilea rând, o activitate foarte importantă a fost implementarea sistemului de sinteză folosind o nouă tehnologie, DNN (Deep Neural Networks) pentru modelare acustică și aliniere cu textul. În final, noul sistem de sinteză pe baza de DNN a fost testat și a fost generată o nouă voce sintetică de înaltă calitate, folosind doar 1000 de propoziții (din total de 3500 din corpus) pentru antrenarea modelelor. Demonstrație cu vocile sintetizate http://speech.utcluj.ro/swara/listeningTest_DNN/, demonstratorul online <https://swara.fortech.ro/audio/speech/>.

Obiectiv3.b: *Evaluarea preliminară a satisfacției utilizatorilor*

Grad realizare: Obiectiv realizat integral

Rezultate:

- metodologie de evaluare a vocilor sintetice și a demonstratorului de predicție a textului pe telefon mobil și online
- rezultate ale evaluării preliminare a satisfacției utilizatorilor și recomandări
- 1 articol în abstract / poster la conferință de prestigiu [10]
- 1 livrabil (D2.3) cu titlul „Raport asupra satisfacției utilizatorilor” și care descrie rezultatele menționate mai sus.

Obiectiv3.c: *Dezvoltarea versiunii a 2-a a bazei de date audio-video*

Grad realizare: Obiectiv realizat integral

Rezultate:

- 12 noi vorbitori înregistrați în anul 2016 (aproape 21 ore de vorbire)
- 1 metodă îmbunătățită de aliniere automată a semnalului vocal cu textul, aplicată în activitatea de adnotare a corpusului audio, publicată [12]
- o nouă metodă nesupervizată și independentă de limbă pentru segmentarea acustică; metoda a fost testată cu succes pe corpusuri audio și de text în 3 limbi: Română, Engleză, Italiană [13]
- 1 articol la conferință internațională și care prezintă metoda de aliniere a semnalului vocal cu textul [12]
- 1 articol la conferință internațională și care prezintă metoda inovativă de segmentare acustică [13]
- 1 livrabil (D3.2b) cu titlul „Baza de date vers.2”,
- 1 livrabil (D3.3) cu titlul „Baza de date audio adnotată”

Obiectiv3.d: Realizarea demonstratorului pentru redactarea cu predicție a textului

Grad realizare: Obiectiv realizat integral

Rezultate:

- 1 corpus de text în limba română pentru modelare lingvistică dependentă de contextul utilizatorului (la medic): 11.644 cuvinte, 1.577 propoziții.
- 1 corpus de text în limba română (colectat de P1 – Fortech) pentru modelare lingvistică generală extras din Wikipedia (aprox 9GB de date)
- 1 corpus (sursa Google n-gram 1T) achiziționat cu licență și utilizat de către UTCN pentru validare modele lingvistice și posibile aplicații comerciale la P1
- module software pentru predicția rapidă a textului pe mobil și în Cloud
- 1 demonstrator pentru redactarea rapidă a textului accesibil online pe structura de servere în Cloud – disponibil public online
- <https://swara.fortech.ro/audio/predictionTests/medical.html>
- <https://swara.fortech.ro/audio/speech>
- 1 demonstrator pentru redactarea rapidă a textului pe telefon mobil – disponibil cu licență de utilizare ne-comercială de la UTCN
- 1 livrabil (D4.4.) cu titlul „Demonstrator pentru redactarea rapidă a textului”. Un scenariu de predicție folosind predicția pe 3 cuvinte:

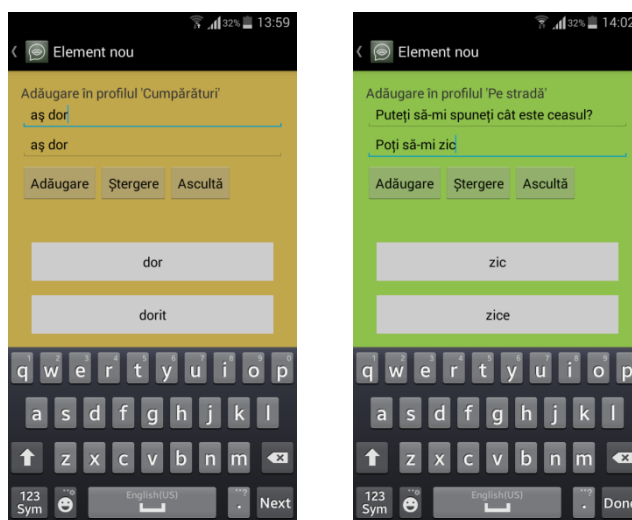
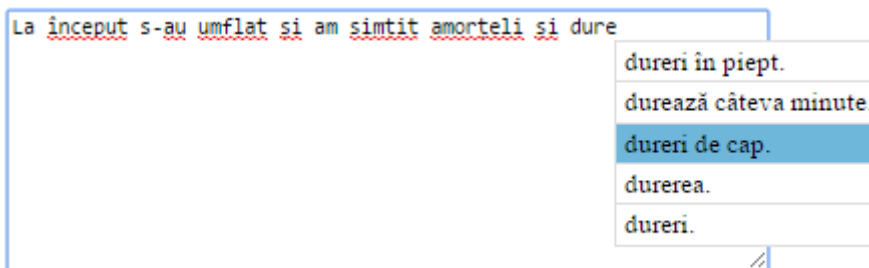


Figura 2. Interfața de predicție rapidă a textului pe mobil (jos) și online (sus)

Obiectiv3.e: Diseminarea rezultatelor pe anul 2016

Grad realizare: Obiectiv realizat integral

Rezultate:

- actualizarea dinamică și monitorizarea cu Google Analytics a site-ului
- planul de diseminare pentru anul 2016
- 1 pagină web pentru demonstrarea online a sistemului de sinteză care încorporează predicția de text, respectiv 1 pagină web cu mostre de semnal vocal sintetizate cu sistemul de sinteză bazat pe DNN
- 1 pagină web cu demonstratorul online de predicție rapidă a textului
- 4 articole prezentate și comunicate la conferințe internaționale [10]...[13]
- 1 livrabil (D7.4a) cu titlul „Publicații și tutoriale”

2.4. Realizarea obiectivelor aferente Etapei 4 „Adaptare sistem și integrarea lui în serviciul de sinteză în Cloud”

Obiectiv 4.a: *Realizarea unui sistem de adaptare a vocilor sintetizate*

Grad realizare: Obiectiv realizat integral

Rezultate:

- noi resurse de date audio pentru adaptare sistem - disponibile acum în corpusul SWARA², 17 vorbitori, 21 de ore de vorbire (acces gratuit pentru scopuri de cercetare)
- 17 noi voci sintetizate din noile date audio înregistrate (vocile pot fi ascultate online³)
- 1 demonstrator online cu cele 17 voci selectabile prin meniu⁴
- 1 model de combinare a parametrilor acustici ai vocilor sintetizate
- 1 demonstrator online pentru combinarea parametrilor acustici ai vocilor sintetizate și alegerea tipului de vocoder⁵
- 1 nou model de adaptare a vocilor folosind DNN (Deep Neural Networks)
- 2 articole științifice publicate la conferințe internaționale [14] [15]
- 2 livrabile, D5.1 „Sistem preliminar de adaptare a vocii” și D5.2. „Un nou model de adaptare”

Obiectiv 4.b: *Evaluarea impactului psiho-social al vocilor sintetizate*

Grad realizare: Obiectiv realizat integral

Rezultate:

- metodologie de evaluare a impactului psiho-social
- 1 chestionar de evaluare
- rapoarte statistice privind impactul psiho-social al vocilor sintetizate
- 2 postere la conferințe internaționale în domeniul psiho-social [16][17]
- 1 comunicare orală la congresele internaționale ORL [21]
- 1 livrabil (D2.4) cu titlul „Raport științific asupra impactului psiho-social al vocilor sintetizate personalizate”.

Evaluarea este orientată către preferința pacienților și a interlocutorilor acestora privind utilizarea vocilor sintetizate față de protezele vocale sau erigmofonația. Numărul de utilizatori evaluați în anul 2016 a fost de 82, iar în anul 2017 10, în prima fază a evaluării, iar ulterior 20. Rezultatele au arătat că vocile sintetizate sunt mai clare decât celelalte metode de vorbire, însă naturalitatea lor nu este cea așteptată de către pacienți. O posibilitate de îmbunătățire ar fi creșterea vitezei de procesare pe serverul în Cloud, respectiv creșterea vitezei de transfer.

Obiectiv 4.c: *Experimente privind recunoașterea vizuală a vorbirii*

Grad realizare: Obiectiv realizat integral

Rezultate:

- 1 metodologie de procesare a datelor video pentru recunoașterea vizuală a vorbirii bazată pe DCT (Discrete Cosine Transform) și clasificare cu SVM (Support Vector Machine)
- 1 experiment preliminar pentru recunoașterea vizuală a cifrelor
- 1 livrabil, D4.3b „Rezultate experimentale privind recunoașterea vizuală a vorbirii”.

² <http://speech.utcluj.ro/swarasc/>

³ <http://speech.utcluj.ro/swarasc/samples/index.html>

⁴ <http://swara.fortech.ro/audio/speech/>

⁵ <http://romaniantts.com/swaramix/>

Obiectiv 4.d: Dezvoltarea unui serviciu Cloud pentru sinteza text vorbire accesibil de pe mobil

Grad realizare: Obiectiv realizat integral

Rezultate:

- 1 tehnologie integrată de sinteză text vorbire accesibilă pe dispozitive mobile și conținând: interfața web, serverul HTTP, serverul de aplicație, integrarea motorului de sinteză vocală
- evaluarea online a sistemului experimental în Obiectivul 4.c
- 1 poster la conferințe internaționale în domeniul psiho-social [18]
- 1 comunicare orală despre tehnologii asistive la congres ORL [20]
- 1 livrabil: (D6.2a) cu titlul „Versiune demonstrativă a sistemului de sinteză text vorbire în Cloud, accesibil de pe mobil”.

Această tehnologie a fost testată, evaluată și validată cu pacienții larnigectomizați în perioada Mai – Iulie 2017, iar rezultatele evaluărilor sunt raportate în livrabilul D2.4. „Raport științific asupra impactului psiho-social al vocilor sintetice personalizate”.

Obiectiv 4.e: Dezvoltarea planului de exploatare și a acordului privind drepturile de proprietate intelectuală

Grad realizare: Obiectiv realizat integral

Rezultate:

- stabilirea cotelor de contribuție ale partenerilor în realizarea livrabilelor conform cu resursele financiare alocate
- 18 principii de aplicat privind proprietatea intelectuală
- stabilirea clauzelor de confidențialitate
- identificarea tehnologiilor cu potențial de valorizare comercială
- elaborarea planului de exploatare
- 3 colaborări cu companii private în afara consorțiului pentru utilizarea tehnologiei de sinteză text vorbire dezvoltate în proiectul SWARA
- 2 livrabile, D7.5 „Acord final privind drepturile de proprietate intelectuală” și D7.6 „Planul de exploatare”.

Acordul privind proprietatea intelectuală are în vedere: nominalizarea partenerilor interesați în valorificarea drepturilor de proprietate intelectuală, stabilirea cotelor de contribuție ale partenerilor în realizarea produselor / tehnologiilor generate în proiect, stabilirea principiilor privind proprietatea intelectuală, și stabilirea principiilor privind respectarea confidențialității. Pe baza lui s-a elaborat planul de exploatare în care sunt prezentate: obiectivele, lista aplicațiilor software din domeniul sintezei vocale cu potențial de comercializare, evaluarea impactului asupra utilizatorilor, iar în final în Anexă un plan de exploatare mai detaliat propus de partenerul industrial.

Obiectiv 4.f: Diseminarea și exploatarea rezultatelor pe anul 2017

Grad realizare: Obiectiv realizat integral

Rezultate:

- actualizarea dinamică și monitorizarea cu Google Analytics a site-ului
- planul de diseminare pentru anul 2017
- pagini web cu demonstrarea online a sistemului integral
- 2 articole prezentate și comunicate la conferințe internaționale [14][15]
- 4 postere prezentate la conferințe pe profil psiho-social [16][17][18][19]
- 2 comunicări orale la congrese internaționale ORL [20][21]
- 3 livrabile referitoare la activitățile de diseminare, D7.4a „Publicații și tutoriale”, D7.5 „Acord final privind drepturile de proprietate intelectuală”, D7.6 „Planul de exploatare”.

3. Impactul rezultatelor

3.1. Vizibilitatea online a rezultatelor

O preocupare a Consorțiului pe întreaga perioadă de derulare a proiectului a fost asigurarea vizibilității proiectului printr-o serie de măsuri specifice: a) actualizarea dinamică a paginii web cu rezultatele obținute incremental în proiect, inclusiv cu secțiuni demonstrative; b) elaborarea planurilor de diseminare anuale; c) publicarea rezultatelor științifice la conferințe internaționale de prestigiu în domeniul proiectului (tehnice, tehnologii medicale asistive), d) crearea unor pagini web dedicate pentru demonstrarea online a sintezei, precum și pentru predicția rapidă a textului. Detalii specifice sunt prezentate în livrabilele dedicate diseminării.

Pagina web⁶ a proiectului are un conținut dinamic, iar serviciul Google Analytics raportează un interes în creștere pentru informațiile disponibile pe acest site. Concluzia principală este ca disponibilitatea online a rezultatelor proiectului a făcut ca numărul de utilizatori să crească de la 244 (în 2015), la 842 (în 2016), respectiv 984 (în 2017), iar numărul de pagini vizitate de la 309 (în 2015), la 1605 (în 2016), respectiv 2062 (în 2017).

3.2. Materiale promoționale

Încă de la începutul proiectului au fost realizate următoarele materiale promoționale în scop de diseminare și de atragere a atenției persoanelor sau instituțiilor interesate de rezultatele proiectului

- Pliantul proiectului⁷ – prezintă obiectivele, activitățile, rezultatele estimate, partenerii implicați, modalități de utilizare a rezultatelor. Pliantul a fost distribuit la majoritatea conferințelor la care au participat membrii consorțiului.
- Posterul proiectului⁸ – prezintă obiectivele, rezultatele estimate ale proiectului și partenerii implicați. Posterul a fost afișat într-un spațiu special amenajat din cadrul Departamentului de Comunicații din Facultatea de Electronică, Telecomunicații și Tehnologia Informației.

3.3. Implicarea studenților, masteranzilor și doctoranzilor

Aparte de cercetătorii implicați direct în proiect, Consorțiul a mobilizat și tineri în formare, studenți, masteranzi sau doctoranzi care și-au realizat lucrări de diplomă sau de dizertație în legătură cu tematica SWARA, după cum urmează:

Proiecte de diplomă:

- (1) Ioana NEAGU, "Fast Predictive Text Input on Mobile Devices", Bachelor thesis, 2017
- (2) Alexandru Moldovan, "Improving Sentence-level Alignment of Speech with Imperfect Transcripts using Utterance Concatenation and VAD", Bachelor thesis, 2016
- (3) Laurentiu Rogean, "Predictive Text Input on Mobile Devices", Bachelor thesis, 2016
- (4) Ștefan Prișcă, "Language Independent, User Oriented Auto-completion", Bachelor Thesis, 2016
- (5) Cătălin Gabriel Stan, "Text autocompletion based on the user's input history", Bachelor Thesis, 2015
- (6) Anamaria Mihaela Beleiu, "Supervised learning for lexical stress discovery in Romanian", Bachelor Thesis, 2015

⁶ <http://speech.utcluj.ro/swara/>

⁷ <http://speech.utcluj.ro/swara/docs/Swara-brochure.pdf>

⁸ <http://speech.utcluj.ro/swara/docs/Swara-brochure.pdf>

- (7) Diana Maria Balci, "Supervised Learning for Romanian Language Syllabification", Bachelor Thesis, 2015
- (8) Alexandru Marchiș, "Romanian words syllabification based on pre-computed data search", Masters Thesis, 2015
- (9) Grigore Toth, "Software tools and comparative studies on POS tagging for Romanian and English languages. Application on text to speech synthesis", Bachelor Thesis, 2015
- (10) Răzvan Ghica, "Adaptive filters applied for speech enhancement", Bachelor Thesis, 2015
- (11) Daniela Domnici, "A web-based tool for the subjective evaluation of synthetic voices", Bachelor Thesis, 2015.
- (12) Dana Toșa, "Experiments on speech and text alignment of audiobooks data. Applications for text to speech synthesis", Bachelor Thesis, 2014.
- (13) Patricia Costin, "Automatic syllable segmentation of speech signals using the sonority function", Bachelor Thesis 2014.
- (14) Alin Rusu, "Prediction of words polarity and user feedback for sentiment analysis in text data", Bachelor Thesis, 2014.
- (15) Sebastian Rus, "Automatic voice selection of a target speaker from broadcasted news using a diarization toolkit", Bachelor Thesis, 2014
- (16) Marius Hantig, "Automatic recognition of speech emotions using prosodic features", Bachelor Thesis, 2014.

Stagii de practică în anul 2015:

- (1) Automatic Part of Speech Tagging for Romanian (Popp Lucian, Toth Nandor, 3rd year BSc Telecom.);
- (2) Text normalization for several categories of words (Rigan Levente, Bede Domokos, 3rd year BSc Telecom.)
- (3) Experiments using language models for text prediction (Stancovici Georgian, Rogean Laurentiu, Iovan Paul, 3rd year BSc Telecomm.)
- (4) Experiments using speech segmentation at syllable level (Ferenc Izsak, Sandy Endre, Comorasu Lucian, 3rd year BSc Electronics).

3.4. Implicarea utilizatorilor

Utilizatorii au fost implicați în testarea, evaluarea și validarea sistemului de către partenerii UBB, respectiv UMF, după cum urmează:

- în anul 2015 – 28 utilizatori pentru opinii privind elaborarea funcționalităților
- în anul 2016 – 82 de utilizatori pentru testarea și evaluarea intermediară
- în anul 2017 – 30 de utilizatori pentru evaluarea finală a sistemului.

3.5. Activitatea de publicare a rezultatelor

Pe durata proiectului au fost elaborate și publicate un total de 21 de articole științifice după cum urmează (ele sunt disponibile și online⁹, iar lista lor completă se găsește în Secțiunea „Lista articolelor științifice elaborate și publicate în proiect” a acestui raport):

- în anul 2015 – 9 articole
- în anul 2016 – 4 articole
- în anul 2017 – 8 articole

⁹ <http://speech.utcluj.ro/swara/results.html#publications>

3.6. Demonstratoare online

În scop de diseminare a rezultatelor către o audiență cât mai largă s-a luat decizia și s-au implementat o serie de pagini web prin intermediul cărora să se ofere utilizatorilor acces public gratuit pentru a testa și evalua modulele sau sistemele implementate:

- (1) demonstrator cu sistemul de sinteză accesibil online în Cloud (include și modulul de predicție a textului)¹⁰,
- (2) demonstrator online pentru predicția textelor din contextul la medic¹¹,
- (3) demonstrator online cu mostre de semnal audio sintetizate cu sistem HTS¹²
- (4) demonstrator online cu mostre de semnal audio sintetizate cu noile modele DNN¹³,
- (5) demonstrator online cu respectiv corpusul SWARA¹⁴,
- (6) demonstrator online pentru sistemul de adaptare a vocilor prin mixarea parametrilor acustici¹⁵.

3.7. Acord privind drepturile de proprietate intelectuală

În acordul privind drepturile de proprietate intelectuală au fost prezentate principiile agreeate pentru respectarea proprietății intelectuale după finalizarea proiectului, în scopul exploatării și valorizării rezultatelor acestuia. S-a avut în vedere: nominalizarea partenerilor interesați în valorificarea drepturilor de proprietate intelectuală, stabilirea cotelor de contribuție ale partenerilor în realizarea produselor / tehnologiilor generate în proiect, stabilirea principiilor privind proprietatea intelectuală, și stabilirea principiilor privind respectarea confidențialității.

3.8. Planul de exploatare a rezultatelor proiectului

În acest plan sunt prezentate în mod succint principalele aspecte ce țin de planul de exploatare a tehnologiei SWARA. Sunt prezentate obiective, lista aplicațiilor software din domeniul sintezei vocale cu potențial de comercializare, clauzele privind proprietatea intelectuală, evaluarea impactului asupra utilizatorilor, iar în final în Anexă un plan de exploatare propus de partenerul industrial. Conform unui studiu recent piața sistemelor de sinteză text vorbire este estimată la 1,4 miliarde USD și este așteptat să crească la 3,03 miliarde pentru anul 2022. Suplimentar s-au studiat alte surse de informații: (a) rapoarte ale companiile Hoovers, Bloomberg Business, Factiva, OneSource, (b) ofertele principalilor competitori, (c) adoptarea unei proceduri de analiză de tip bottom-up, (d) segmentarea piețelor, (e) interviuri, (f) aplicarea unor metode de segmentare cum ar fi triangularea datelor

Principalii competitori ai produsului realizat în cadrul acestui proiect sunt furnizorii de sisteme TTS pentru care s-a făcut analiza sunt Nuance Communication (SUA), Microsoft Corporation (SUA), Sensory Inc. (SUA), LumenVox LLC (SUA), NeoSpeech (SUA), Amazon.Com (SUA), SESTEK (Turcia), iSpeech Inc. (SUA), TextSpeak (SUA), NextUp Technologies (SUA), Acapela Group (Franța), CereProc (Scoția), ReadSpeaker (Olanda), Nexmo (SUA), and rSpeak (Olanda).

¹⁰ <https://swara.fortech.ro/audio/speech/>

¹¹ <https://swara.fortech.ro/audio/predictionTests/medical.html>

¹² <http://speech.utcluj.ro/swara/listeningTest/>

¹³ http://speech.utcluj.ro/swara/listeningTest_DNN/

¹⁴ <http://speech.utcluj.ro/swarasc/samples/index.html>

¹⁵ <http://romanianntts.com/swmix/>

Tabel 4. Lista tehnologiilor cu potențial de comercializare

Tehnologie disponibilă pentru exploatare comercială	Cotă contribuție partener			
	UTCN	Fortech	UMF	UBB
D1.3. Sistem de sinteză text vorbire de înaltă calitate	75%	25%	0%	0%
D3.3. Baza de date audio adnotată (SWARA)	90%	0%	5%	5%
D4.4. Demonstrator pentru redactarea cu predicție a textului	35%	65%	0%	0%
D5.2. Un nou sistem de adaptare a vocilor sintetizate	85%	15%	0%	0%
D6.2a. Versiune demonstrativă TTS in Cloud, accesibil de pe mobil	25%	75%	0%	0%

Rezultatele evaluărilor sistemului au arătat că vocile sintetice au fost în general percepute ca fiind mai clare decât vocea pacientului, $t(81) = -8.74$, $p < .001$, $d = .92$. În ceea ce privește naturalitatea însă, vocea pacientului a fost percepută ca fiind mai naturală decât cele sintetice, $t(81) = 5.19$, $p < .001$, $d = .55$. În ceea ce privește atitudinea față de posibile interacțiuni viitoare, nu am identificat diferențe semnificative statistic între vocea pacientului și cele sintetice, $t(81) = .76$, $p = .451$.

Rezultatele privind satisfacția interlocutorilor în ceea ce privește vocile sintetice, prin raportare la o voce asistată de o proteză vocală, arată următorul pattern: vocile sintetice sunt mai clare dar mai puțin naturale. Cu toate acestea, nu există diferențe în ceea ce privește atitudinea față de posibile interacțiuni viitoare cu orice tip de voce. Acest lucru sugerează că în ciuda lipsei de naturalitate a vocii, interlocutorii sunt dispuși să se angajeze în interacțiuni cu un pacient care folosește ca voce asistată o voce sintetică, în egală măsură cu o voce asistată de o metodă clasică.

Ca urmare a dezvoltărilor permanente din cadrul proiectului SWARA, sistemul de sinteză a atins un grad de maturitate ce permite exploatarea sa ca și serviciu web, atât de către pacienții implicați în proiect, cât și pentru eventuale dezvoltări ulterioare din cadrul companiei Fortech. Un prim set de teste ale acestui sistem au fost realizate de către colegii de la UMF și UBB și a căror rezultate sunt descrise în livrabilul D2.4 Raport științific asupra impactului psihosocial al vocilor sintetice personalizate.

3.9. Propuneri de colaborare din partea mediului economic

Până în prezent s-au primit 3 solicitări de colaborare cu mediul economic, după cum urmează:

Solicitarea #1

Nume companie:	<i>AutoVortex România (http://autovortex.ro)</i>
Descriere companie:	<i>AutoVortex este echipa națională de robotică a României. Sunt campioni mondiali la mai multe concursuri de robotică.</i>
Colaborare solicitată:	<i>Licență de utilizare a API-ului de sinteză text vorbire în Cloud pentru roboții pe care îi construiesc. Au primit o licență de utilizare gratuită pentru 1000 de apeluri de sinteză (August 2017)</i>

Solicitarea #2

Nume companie:	<i>Inspectoratul de Poliție Ialomița</i>
Descriere companie:	<i>Serviciul de înmatriculări auto</i>
Colaborare solicitată:	<i>Sinteza din text a mesajelor vocale pentru informarea cetățenilor la biroul de înmatriculări auto. Sistemul permite atât afișarea vizuală a informațiilor, precum și citirea acestora. Au primit gratuit drept de utilizare a sistemului (Septembrie 2017)</i>

Solicitarea #3

Nume companie:	AllView România (https://www.allview.ro/)
Descriere companie:	Allview este o companie românească ce produce, importa și distribuie telefoane mobile Dual SIM și produse electronice. Brandul Allview este recunoscut ca fiind unul cu o dezvoltare rapidă în România datorită serviciilor gratuite și facilităților atașate produselor sale. Telefoanele Allview oferă posibilitatea separării convorbirilor de business de cele personale sau efectuarea de apeluri în rețele diferite la cele mai mici costuri.
Colaborare solicitată:	Sinteza din text a vorbirii pe telefoanele mobile care le produce compania AllView și pentru Asistentul Vocal AllView În discuție (Octombrie 2017)

3.10. Rezultatul cel mai semnificativ

Rezultatul cel mai semnificativ obținut în proiectul SWARA este: „**Tehnologia de sinteză text vorbire în Cloud, accesibilă și pe terminale mobile**”

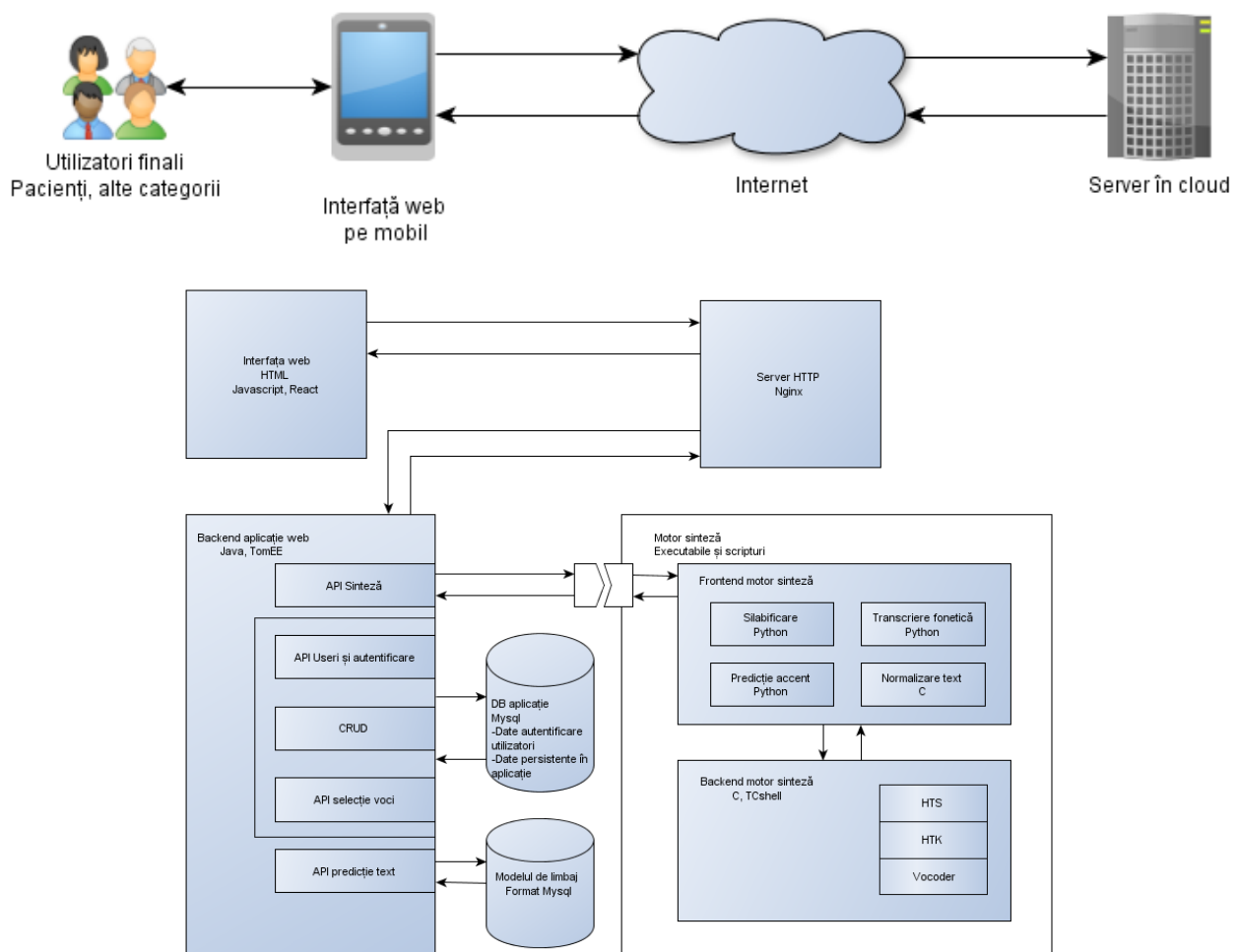


Figura 3. Arhitectura globală (sus) și componentele tehnologiei cu fluxul de procesări (jos)

Componentele majore ce intră în alcătuirea sistemului¹⁶ sunt: Interfața web accesibilă de pe mobil, Serverul HTTP, Serverul de aplicație, Modulul de sinteză, Modulul de predicție text.

Interfața web conține module din clasele: React, Bootstrap, React-Bootstrap, React-redux, jQuery, jQuery-textcomplete.

Serverul HTTP este de tip Nginx și reprezintă poarta de intrare către serverul de sinteză realizând comunicarea dintre clienți și aplicația web de sinteză în Cloud.

Serverul de aplicație are rolul de a genera conținutul dinamic către clienții web, iar pentru aceasta el execută o serie de servicii cum ar fi: serviciul de autentificare, serviciul pentru utilizatori, serviciul de predicție de text, serviciul de listare voci disponibile, serviciul pentru istoric audio, serviciul de sinteză.

Modulul de sinteză este alcătuit dintr-o colecție de scripturi Shell, Python, respectiv programe scrise în C cu scopul de a realiza sinteza din text pe baza cererilor primite de la serverul de aplicație.

Modulul de predicție a textului conține interfața jQuery-textcomplete și modelele statistice de tip n-gram pentru predicția textului.

Serviciul de securitate în Cloud implementează o serie de restricții cum ar fi: restricționarea utilizatorilor pe baza de autentificare, restricționarea serviciilor prin setări în Nginx / TomEE / Postfix, expunerea în rețea a serviciilor disponibile, metode specifice de filtrare a pachetelor, securizarea protocolului SSH, securizarea serviciului de reverse-proxy, serviciul Chkrootkit de raportare zilnică a problemelor de securitate, serviciul Rkhunter de atenționare prin eMail, serviciul ClamAV de protecție la viruși.

Ca urmare a dezvoltărilor permanente din cadrul proiectului SWARA, sistemul de sinteză a atins un grad de maturitate ce permite exploatarea sa ca și serviciu web, atât de către pacienții implicați în proiect, cât și pentru eventuale dezvoltări ulterioare din cadrul companiei Fortech.

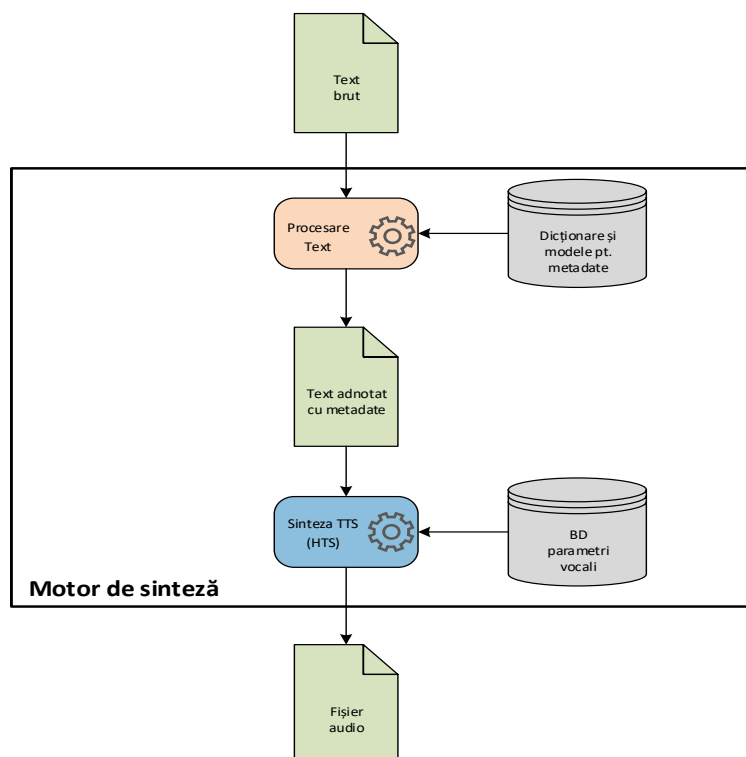


Figura 4. Schema bloc și fluxul de procesări pentru motorul de sinteză text vorbire în Cloud

¹⁶ <https://swara.fortech.ro/audio/speech/>

4. Concluzii

Activitățile de cercetare desfășurate în proiectul SWARA au condus la obținerea rezultatelor așteptate și ele sunt în concordanță cu obiectivele specifice ale fiecărei etape de implementare.

În Etapa 1, 2014, realizările se refera la indexarea bazelor de date de semnal vocal, a resurselor de text și adnotări de natura lingvistica, a instrumentelor software utilizate în procesarea semnalului vocal și a textului aplicate în scopul sintezei din text a vorbirii. Pe aceste fundamente au fost elaborate specificațiile funcționale pentru componentele software ale sistemului de sinteză, în conformitate cu obiectivele etapei.

În Etapa 2, 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 în 9 articole.

În Etapa 3, 2016, realizările se referă la: proiectarea unui sistem de sinteză text vorbire de înaltă calitate bazat pe rețele DNN (Deep Neural Networks), integrarea noului demonstrator de sinteză text vorbire în Cloud și evaluarea lui online, evaluarea satisfacției utilizatorilor raportat la vocile sintetice în interacțiune social, dezvoltarea versiunii finale a bazei de date audio-video: în total 21 de ore semnal vocal, dezvoltarea și aplicarea unor noi metode de adnotare automată a semnalului audio, realizarea unui demonstrator (pe telefon mobil și online) pentru predicția textului, diseminarea rezultatelor intermediare în 4 articole.

În Etapa 4, 2017, realizările se referă la: dezvoltarea unui sistem pentru adaptarea vocilor sintetizate, dezvoltarea unui nou sistem de adaptare a vocilor folosind DNN (Deep Neural Networks), finalizarea demonstratorului de sinteză text vorbire în Cloud și evaluarea lui online, evaluarea satisfacției utilizatorilor raportat la vocile sintetice personalizate, finalizarea adnotărilor și livrarea open source a corpusului SWARA, realizarea acordului privind drepturile de proprietate intelectuală, realizarea planului de exploatare a rezultatelor proiectului, diseminarea rezultatelor în 8 articole.

În prezent sistemul este funcțional online și există cereri de colaborare cu mediul economic. Sperăm să continuăm și să dezvoltăm pe mai departe această tehnologie în proiectele noastre viitoare.

5. Lista articolelor științifice elaborate și publicate la conferințe de prestigiu

[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", Published by : European Archives of ORL, Vol 272/2015m, ISSN : 0937-4477, Springer Berlin Heidelberg, pp.1-6, 2015.

[4] Matu S., Șoflău R., Cîmpean A., David D., Chirilă M., Tiple 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., Tiple 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 Balci, 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 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 Bioengineering, 21-25 Npvember, 2015, Iasi.
- [10] M.Chirila, C. Tiple, F. Dinescu, S. Matu, R. Soflau, M. Giurgiu, A. Stan, and D. David , „Voice-related quality of life in laryngectomies and the next generation of vocal assistive technologies”, In Proc. AHNS 9th International Conference on Head and Neck Cancer, July 16-20, Seattle, USA, 2016, Poster 167, Book of abstracts pag 205.
- [11] A.Bona, C. Lemnaru, R. Potolea, "Syllabification with frequent sequence patterns - A language independent approach", In Proc. of the 9th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2017), 1-3 November 2016, Funchal, Portugal, Poster Session 2.
- [12] Alexandru Moldovan, Adriana Stan, Mircea Giurgiu, "Improving Sentence-level Alignment of Speech with Imperfect Transcripts using Utterance Concatenation and VAD", Proc. of 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP 2016), 8-10 Sept. 2016, Cluj-Napoca, ISBN: 978-1-5090-3899-2, pp.171-174, DOI: 10.1109/ICCP.2016.7737141
- [13] Adriana Stan, Cassia Valentini-Botinhao, Bogdan Orza, Mircea Giurgiu, "Blind Speech Segmentation using Spectrogram Image-based Features and Mel Cepstral Coefficients", Proc. of 2016 IEEE Workshop on Spoken Language Technology, 13–16 December 2016, San Diego, California; Poster Sessions (not yet indexed).
- [14] Adriana Stan; Florina Dinescu; Cristina Tiple; Șerban Meza; Bogdan Orza; Magdalena Chirilă, Mircea Giurgiu, "The SWARA speech corpus: A large parallel Romanian read speech dataset", In Proc. 2017 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Bucuresti, 6-9 July 2017, ISBN: 978-1-5090-6497-7, DOI: 10.1109/SPED.2017.7990428 pag 1-6.
- [15] Ștefan-Adrian Toma ; Adriana Stan ; Mihai-Lică Pura ; Traian Bârsan, „MaRePhoR — An open access machine-readable phonetic dictionary for Romanian”, In Proc. 2017 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Bucuresti, 6-9 July 2017, ISBN: 978-1-5090-6497-7, DOI 10.1109/SPED.2017.7990435, pag 6-9, <http://ieeexplore.ieee.org/document/7990435/>
- [16] Matu S., Șoflău R., David D., Chirilă M., Tiple C., Dinescu V., Mureșan R., Giurgiu M., Stan A., „Using synthetic voices to facilitate the communication of patients with laryngectomy”, International Congress of Cognitive Psychotherapy, Cluj-Napoca, 2017, <http://www.iccp2017.org>
- [17] Matu, S, Șoflău, R. David, D. Chirilă M., Tiple, C., Dinescu, V., Mureșan, R., Giurgiu, M., Stan A. „A systematic review and meta-analysis on the psycho-social predictors of the quality of life of patients with laryngectomy”, International Congress of Cognitive Psychotherapy, Cluj-Napoca, 2017.
- [18] Matu S., Șoflău R., David D., Chirilă M., Tiple C., Dinescu V., Mureșan R., Giurgiu M., Stan A. „The potential of synthetic voices to facilitate communication and adaptation of patient with laryngectomy”, 47th Congress of the EABCT, Ljubljana, 2017
- [19] Matu S., Șoflău R., David D., Chirilă M., Tiple C., Dinescu V., Mureșan R., Giurgiu M., Stan A., „Psychological predictors of quality of life after laryngectomy: A systematic review and meta-analysis”, 47th Congress of the European Association for Behavioural and Cognitive Therapies. Ljubljana, 2017.
- [20] Magdalena Chirila, „Mobile system of vocal assistance for patients with surgical aphonia”, IFOS: INTERNATIONAL FEDERATION OF OTO-RHINOLARYNGOLOGICAL SOCIETIES, ENT World Congres, Paris, 24-28 June 2017, <http://ifosparis2017.org>
- [21] Cristina Tiple, F Dinescu, S Matu, R Mureșan, R Soflău, M Giurgiu, A Stan, D David, M Cosgarea, Magdalena Chirilă, „Mobile System for Rehabilitative Vocal Assistance of laryngectomees. A Pilot Study of Romanian patients with total laryngectomy”, CEORL Congress, Barcelona, 7-11 October, 2017, <http://www.ceorlhns2017.com/>

6. Lista livrabilelor elaborate pe durata proiectului

Lista livrabilelor și cotele de contribuție ale partenerilor conform cu efortul de timp alocat

Rezultat proiect	Cotă contribuție partener			
	UTCN	Fortech	UMF	UBB
D1.1. Resurse disponibile si specificațiile sistemului de sinteza	65%	35%	0%	0%
D1.2. Sistem preliminar de sinteza text vorbire	80%	20%	0%	0%
D1.3. Sistem de sinteza text vorbire de înalta calitate	75%	25%	0%	0%
D2.1. Raport asupra funcționalităților sistemului si a scenariilor	0%	0%	50%	50%
D2.2. Raport asupra variabilelor psiho-sociale personalizate.	20%	0%	30%	50%
D2.3. Raport asupra satisfacției utilizatorilor	25%	0%	5%	70%
D2.4. Raport științific asupra impactului psiho-social al vocilor sintetice	15%	0%	60%	25%
D3.1. Cadrul tehnic al bazei de date	50%	40%	10%	0%
D3.2a. Baza de date vers.1	50%	0%	40%	10%
D3.2b Baza de date vers.2	80%	0%	10%	10%
D3.3. Baza de date audio adnotata	90%	0%	5%	5%
D4.1. Model de context pentru predicție text	75%	25%	0%	0%
D4.2. Sistem preliminar de predicție text	80%	20%	0%	0%
D4.3a. Raport privind metodele de recunoaștere vizuala a vorbirii	90%	10%	0%	0%
D4.3b. Rezultate experimentale privind recunoașterea vizuala	80%	20%	0%	0%
D4.4. Demonstrator pentru redactarea cu predicție a textului	35%	65%	0%	0%
D5.1. Sistem preliminar pentru adaptarea vocii	75%	25%	0%	0%
D5.2. Un nou model pentru adaptare rapida si fiabila	85%	15%	0%	0%
D6.1. Specificațiile sistemului de sinteza in Cloud	30%	60%	0%	0%
D6.2. Versiune experimentală a sistemului in Cloud	10%	84%	3%	3%
D6.2a.Versiune demonstrativa in Cloud, accesibil de pe mobil	25%	75%	0%	0%
D7.1. Pagina web proiect	60%	20%	10%	10%
D7.2. Plan de diseminare	30%	30%	20%	20%
D7.3. Materiale promoționale	30%	30%	20%	20%
D7.4. Publicații si tutoriale – 2015	30%	10%	40%	20%
D7.4a. Publicații si tutoriale – 2016	30%	25%	25%	20%
D7.4b. Publicații si tutoriale - 2017	20%	30%	20%	30%
D7.5. Acord final privind drepturile de proprietate intelectuala	25%	25%	25%	25%
D7.6. Planul de exploatare	10%	75%	5%	10%