

Raport Științific și Tehnic Etapa a IV-a, an 2017: „Adaptare sistem”

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 Științific și Tehnic (Etapa a IV-a, 2017)
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, Mircea Giurgiu, Silviu Matu, Remus Pop, Adriana Stan, Alexandra Tomiță
Ofițer de proiect:	Andreea Matei

Rezumat:

Acest document prezintă o sinteză a realizărilor de natură științifică și tehnică obținute în a patra etapă de implementare a proiectului SWARA (perioada Ianuarie – Septembrie 2017). Realizările se referă la:

- dezvoltarea unui sistem preliminar 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.

Activitățile de cercetare desfășurate în etapa a patra de implementare a proiectului (2017) au condus la obținerea rezultatelor așteptate și ele sunt în concordanță cu obiectivele specifice ale etapei. Astfel, rezultatele raportate în acest document și descrise detaliat în cele 8 livrabile aferente perioadei de raportare, pregătesc cadrul pentru exploatarea și valorizarea comercială a sistemului de sinteză text vorbire disponibilizarea sistemului sub forma unui serviciu web de sinteză vocală accesibil în Cloud. De asemenea, acest raport prezintă detalii referitoare la activitățile de management și comunicare, precum și de diseminare a rezultatelor.

Cuprins

1. Activitățile etapei de raportare în contextul general al proiectului.....	4
2. Gradul de realizare a obiectivelor specifice pentru Etapa a 4-a	4
3. Rezultatele etapei și descrierea lor științifică și tehnică	6
3.1. Sistem de adaptare a vocilor sintetizate.....	6
3.1.1. Sistem preliminar de adaptare.....	6
3.1.2. Un nou model de adaptare pentru vocile sintetizate	7
3.2. Serviciu Cloud pentru sinteza text vorbire accesibil de pe mobil	9
3.2.1. Arhitectura sistemului	9
3.2.2. Componente functionale.....	10
3.3. Recunoașterea vizuală a vorbirii	11
3.3.1. Modelul propus	11
3.3.2. Rezultate experimentale privind recunoașterea vizuală a vorbirii	12
3.4. Evaluarea impactului psiho-social al vocilor sintetizate personalizate	13
3.4.1. Metodologia	13
3.4.2. Sinteza rezultatelor evaluării	13
3.5. Planul de exploatare și acord privind drepturile de proprietate intelectuală.....	15
3.5.1. Acordul privind proprietatea intelectuală	15
3.5.2. Planul de exploatare.....	15
4. Management si comunicare	17
5. Diseminarea rezultatelor.....	17
5.1. Pagina web a proiectului	17
5.2. Pagini web dedicate pentru demonstrarea online a unor funcționalități.....	18
5.3. Publicații științifice aferente etapei 2017	18
6. Concluzii	19
7. Referințe la livrabilele aferente etapei a patra, anul 2017 (Anexe la raport)	20

1. Activitățile etapei de raportare în contextul general al proiectului

În prima etapa a proiectului (2014) au fost indexate resursele (baze de date de semnal vocal, resurse de text și adnotări de natură lingvistică ale acestora, instrumente software utilizate în procesarea semnalului vocal și a textului aplicate în scopul sintezei din text a semnalului vocal). În etapa aferentă anului 2015 s-au desfășurat activități de dezvoltare experimentală pentru modulele sistemului de sinteză și pentru dezvoltarea unui sistem de sinteză experimental disponibil online. În etapa aferentă anului 2016 s-au desfășurat activități de integrare a modulelor sistemului de sinteză dezvoltate în etapa anterioară într-un sistem de sinteză de înaltă calitate a vocii sintetizate și realizarea a 2 demonstratoare disponibile online.

În etapa de raportare curentă (2017) s-au realizat activități pentru adaptarea sistemului de sinteză la noi vorbitori, elaborarea modelelor pentru crearea de noi voci sintetice, integrarea finală a serviciilor de sinteză cu pregătirea acestora pentru valorizare și exploatare de către partenerul industrial, respectiv evaluarea finală cu utilizatorii și elaborarea acordurilor de proprietate intelectuală conform cu contribuția partenerilor la realizarea sistemului final.

2. Gradul de realizare a obiectivelor specifice pentru Etapa a 4-a

Obiectivele specifice ale Etapei a 4-a, „Adaptare sistem”, împreună cu gradul lor de realizare, activitățile și principalele rezultate obținute în anul 2017 sunt prezentate mai jos.

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 SWARA¹, 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 online²)
- 1 demonstrator online cu cele 17 voci selectabile prin meniu³
- 1 model de combinare a parametrilor acustici ai vocilor sintetizate și implementarea acestui model folosind MLLR
- 1 demonstrator online pentru combinarea parametrilor acustici ai vocilor sintetizate și alegerea tipului de vocoder⁴
- 1 nou model de adaptare a vocilor folosind DNN (Deep Neural Networks)
- 2 articole științifice publicate la conferințe internaționale [1] [2]
- 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
- chestionar de evaluare
- rapoarte statistice privind impactul psiho-social al vocilor sintetizate
- 2 postere la conferințe internaționale în domeniul psiho-social [3][4][6]
- 1 comunicare orală la congresele internaționale ORL [8]
- 1 livrabil (D2.4) cu titlul „Raport științific asupra impactului psiho-social al vocilor sintetizate personalizate”.

¹ <http://speech.utcluj.ro/swarasc/>

² <http://speech.utcluj.ro/swarasc/samples/index.html>

³ <http://swara.fortech.ro/audio/speech/>

⁴ <http://romaniantts.com/swaramix/>

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”.

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

Grad realizare: Obiectiv realizat integral

Rezultate:

- Integrarea finală a arhitecturii software: 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 [5]
- 1 comunicare orală despre tehnologii asistive la congres ORL [7]
- un livrabil: (D6.2a) cu titlul „Versiune demonstrativă a sistemului de sinteză text vorbire în Cloud, accesibil de pe mobil”.

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 livrabililor 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”.

Obiectiv 4.f: *Diseminarea și exploatarea rezultatelor*

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 modulelor sistemului și a sistemului integral
- 2 articole prezentate și comunicate la conferințe internaționale [1][2]
- 4 postere prezentate la conferințe pe profil psiho-social [3][4][5][6]
- 2 comunicări orale la congrese internaționale ORL [7][8]
- 3 livrabile referitor la activitățile de diseminare, D7.4a „Publicații și tutoriale”, D7.5 „Acord final privind drepturile de proprietate intelectuală”, D7.6 „Planul de exploatare”.

3. Rezultatele etapei și descrierea lor științifică și tehnică

3.1. Sistem de adaptare a vocilor sintetizate

Rezultatele raportate în această secțiune corespund Obiectivului 4.a. din lista de obiective specifice Etapei a 4-a, iar ele sunt descrise în extenso în livrabilele D5.1 „Sistem preliminar de adaptare a vocii” și D5.2. „Un nou model de adaptare”. Corpusul SWARA a fost publicat în [1].

3.1.1. Sistem preliminar de adaptare

Manipularea parametrilor acustici ai sistemelor de sinteză text-vorbire în vederea modificării identității vorbitorului este cunoscută în literatură sub denumirea de adaptarea vocilor sintetizate. Această adaptare poate să intervină la momentul sintezei semnalului vocal, prin utilizarea unor filtre proiectate cu acest scop sau poate să fie realizată în mod offline prin adaptarea/modificarea parametrilor sistemului de sinteză.

Pentru a putea utiliza metodele existente de adaptare a vocilor sintetizate este nevoie de un corpus de semnal vocal pregătit în mod special pentru acest scop. Printre cerințele unui astfel de corpus se numără existența mai multor vorbitori, existența unui set de date suficient pentru fiecare vorbitor în parte, precum și ca segmentele de text înregistrate să fie preponderent comune tuturor vorbitorilor. În cadrul proiectului SWARA, am dezvoltat deja un astfel de corpus, intitulat SWARA-Corpus⁵.

Pentru a crea o voce sintetizată folosind un corpus de antrenare, o primă etapă este cea de obținere a informației lingvistice aferentă înregistrărilor audio. Această informație se referă la: transcriere fonetică, silabație, poziționarea accentului lexical, precum și o informație minimală legată de prozodie. Apoi, am realizat o aliniere a datelor audio cu textul folosind modele acustice bazate pe modele Markov antrenate iterativ (Tabel 1). Pe baza acestei alinieri, s-a trecut la antrenarea modelelor acustice folosind setul de unelte software HTS și folosind parametrii acustici ai vocoderului STRAIGHT. Modelele Markov conțin 5 stări conectate stângadreapta.

Modelele sunt antrenate separat pentru cele 3 caracteristici generate de vocoder și includ și coeficienții delta și delta-delta ai acestora. Pe lângă acest set de modele, sunt antrenate și modele ce permit estimarea duratei fiecărei stări din cele 5. Pentru gruparea modelelor Markov s-au folosit arbori de decizie bazați pe principiul Minimum Description Length. Clusterizarea modelelor dependente de context este realizată folosind arbori de decizie.

Tabel 1. Numărul de modele Markov utilizate în fiecare dintre cele 17 voci (extras)

Nr.crt.	Voce	Coeficienți cepstrali	F0	Coeficienți de aperiodicitate	Durată
1	BAS	1226	5879	1465	1306
2	CAU	1063	5980	1013	998
3	DCS	1179	5472	1215	1318
13	SAM	1540	4034	990	1101
14	SDS	948	4283	715	1009
15	SGS	907	3682	502	1365
16	TIM	683	5407	491	870
17	TSS	813	3734	708	1209

⁵ <http://speech.utcluj.ro/swarasc/>

Tabel 2. Numărul de modele Markov utilizate în vocea de tip eigen, pentru fiecare parametru al vocoderului.

Nr.crt.	Voce	Coeficienți cepstrali	F0	Coeficienți de aperiodicitate	Durată
1	EIGEN	3198	15554	4393	7137

Tabel 3. Durata seturilor de date pentru antrenarea vocii eigen (selectie vorbitori 1, 2, 3, 13, ...,17)

Nr. crt.	ID vorbitor	Sex	Durată
1	BAS	F	58'
2	CAU	F	1h 1'
3	DCS	F	1h 8'
13	SAM	F	1h 1'
14	SDS	M	53'
15	SGS	M	47'
16	TIM	F	1h 1'
17	TSS	M	53'

În cadrul metodelor de adaptare a vocilor sintetizate, pentru a fi necesar un set de date cât mai redus, am folosit modele acustice de tip eigen cu date de la mai multi vorbitori, astfel ca în aceste modele se pierde identitatea vorbitorilor. Numărul de modele Markov obținute în urma antrenării pentru fiecare parametru în parte este prezentat în Tabelul 2. Durata totală a datelor de antrenare este de 16 ore și 23 de minute, distribuită în funcție de vorbitori conform cu Tabelul 3. Ca și metode de adaptare la vorbitor au fost folosite: Maximum Likelihood Linear Regression (MLLR-Mean), Constrained Maximum Likelihood Linear Regression (CMLLR), Structural Maximum A posteriori Linear Regression (SMAPLR) și Constrained Structural Maximum A posteriori Linear Regression (CSMAPLR).

Deși numărul de modele Markov pentru vocea eigen este în medie de 3 ori mai mare decât numărul de modele ale vocilor dependente de vorbitori, datorită cantității mari de date de adaptare disponibile, rezultatele acestui proces sunt comparabile cu cele obținute în procesul de antrenare a vocilor individuale. Un demo este disponibil în acest sens.⁶ Cu toate că rezultatele obținute prin adaptare sunt similare cu cele obținute în cadrul vocilor dependente de vorbitor, ele însă nu sunt suficient de apropiate de identitatea reală a vocii persoanei care a realizat înregistrările audio. Astfel că este nevoie ca aceste metode să fie îmbunătățite prin propunerea unui nou model de adaptare.

3.1.2. Un nou model de adaptare pentru vocile sintetizate

Dezvoltările recente din domeniul sintezei de voce au adus în prim plan utilizarea sistemelor de sinteză bazate pe DNN (Deep Neural Networks), rețele neuronale profunde, folosind fie parametrizări ale semnalului vocal, fie chiar forma de undă. Aceste sisteme au o natură superioară celor bazate pe modele Markov, iar tendința generală este de a adopta această tehnologie.

Într-o primă abordare, pornind de la vocile sintetizate obținute pe baza înregistrărilor audio a fiecărui vorbitor în parte, am realizat un demonstrator online⁷ al sistemului de sinteză ce permite și combinarea parametrilor acustici ai vocilor sintetizate, permițând astfel o mai mare flexibilitate a ieșirii sistemului de sinteză.

Pentru a realiza un modul de adaptare a vocilor cu DNN s-a antrenat inițial rețeaua cu setul de date eigen, apoi s-a realizat o adaptare pentru un singur vorbitor. Ca și date de intrare s-au folosit adnotările lingvistice obținute pentru vocile bazate pe modele Markov. Datele de ieșire sunt parametrii vocoderului WORLD: coeficienți cepstrali, frecvența fundamentală și coeficienți de aperiodicitate. Structura rețelei DNN este alcătuită din 6 straturi cu câte 1024 de neuroni pe strat. Funcția de activare a neuronilor este funcția tangentă hiperbolică.

⁶ <http://speech.utcluj.ro/swara/adapt/>

⁷ <http://romanianntts.com/swaramix/>

SWARA Adaptation Samples

speech.utcluj.ro/swara/adapt/

SWARA ADAPTATION SAMPLES

Audio Samples

The following samples are used to compare speaker dependent voices with the adaptation of the HMM-based acoustic models, for two of the speakers in the SWARA corpus.

	Sample 1	Sample 2	Sample 3
Eigen	  	  	  
BAS - Natural	  	  	  
BAS - Speaker dependent	  	  	  
BAS - CSMAPLR Adaptation	  	  	  
CAU - Natural	  	  	  
CAU - Speaker dependent	  	  	  
CAU - CSMAPLR Adaptation	  	  	  

Figura 1. Demonstratorul online pentru adaptarea vocilor sintetizate

Romanian TTS - Online TTS

romaniantts.com

Romanian TTS

DEMO ABOUT DEVELOPERS RSS DB SWARA CORPUS CONTACT

Online demo

Speaker: Vocoder:

Textul de sinteză

Figura 2. Demonstratorul online pentru mixarea vocilor și alegerea vocoderului de sinteză

3.2. Serviciu Cloud pentru sinteza text vorbire accesibil de pe mobil

Rezultatele raportate în această secțiune corespund Obiectivului 4.d din lista de obiective specifice Etapei a IV-a, iar ele sunt descrise în extenso în livrabilul D6.2a „Raport asupra satisfacției utilizatorilor”. Rezultate au fost publicate la conferințe internaționale [7][8].

3.2.1. Arhitectura sistemului

Implementarea sistemului de sinteză în Cloud s-a realizat folosind o abordare top-down. Pornind de la vederea de ansamblu, sunt prezentate succint componentele majore ale sistemului, și sunt descrise conexiunile dintre ele. Fiecare componentă este descrisă în detaliu împreună cu sub-componentele ei în livrabilul D6.2a, iar aici doar o sinteză.

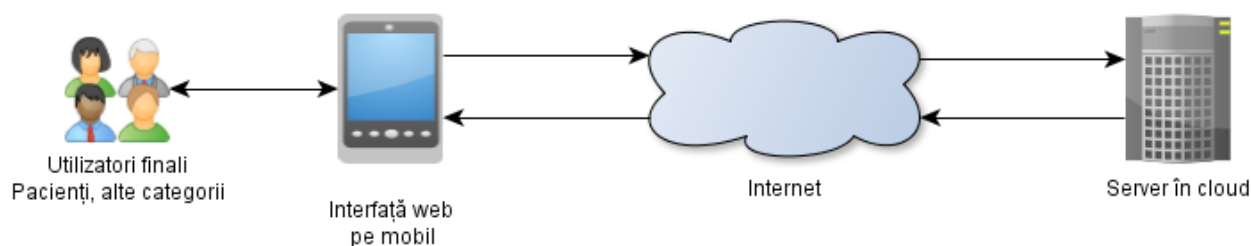


Figura 3. Vedere de ansamblu a arhitecturii sistemului

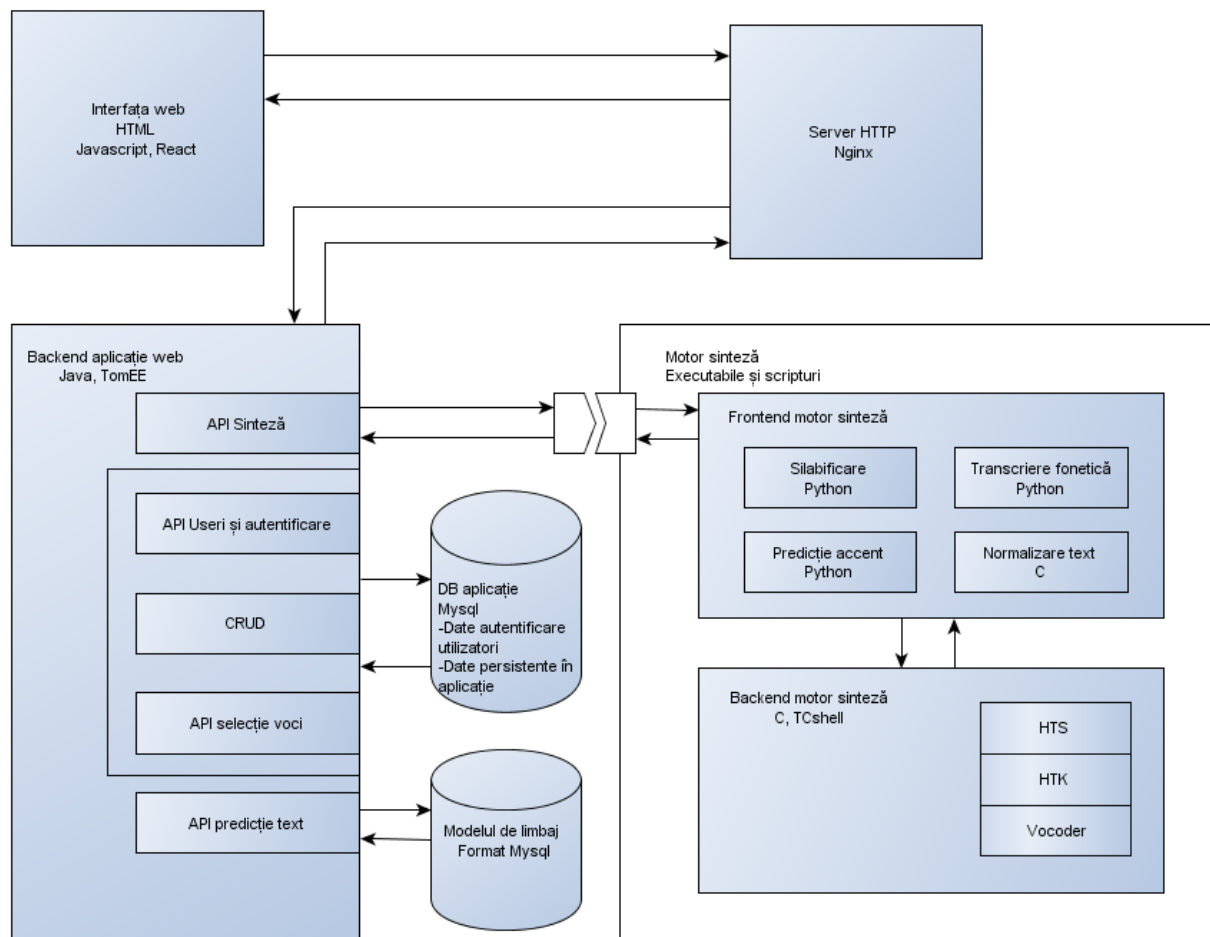


Figura 4. Componentele sistemului și fluxul de procesare

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

3.2.2. Componente functionale

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-grame pentru predicția textului.

Serviciul de securitate în Cloud implementează o serie de reconstrucții cum ar fi: reconstrucția 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. Un set de teste de evaluare și validare a acestui sistem au fost realizate de către colegii de la UMF și UBB în perioada Mai – Iulie 2017 și a căror rezultate sunt descrise în detaliu în livrabilul D2.4.

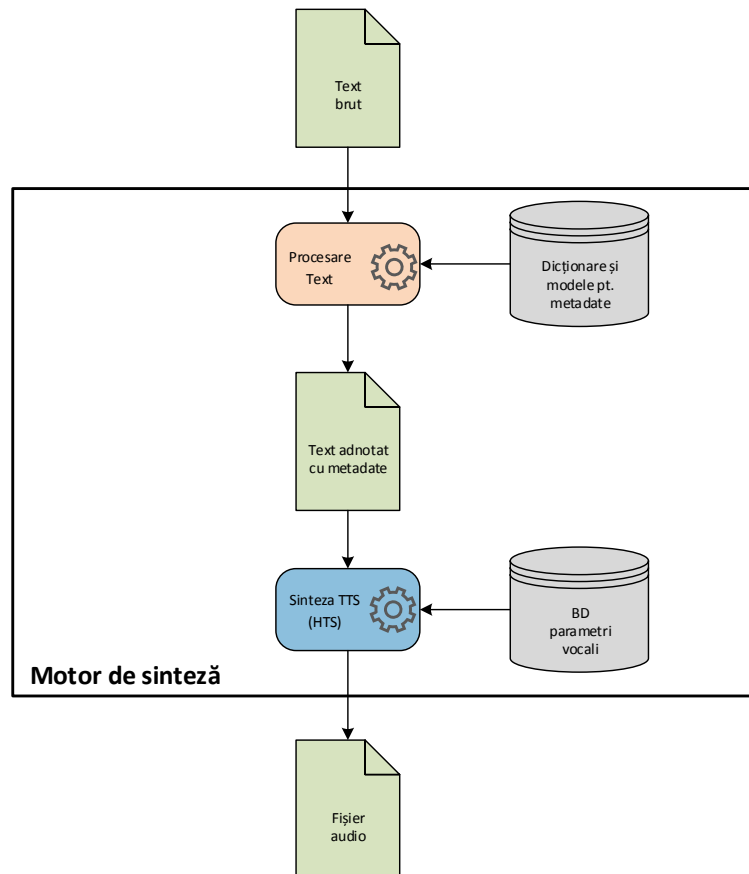


Figura 5. Schema bloc și fluxul de procesări pentru motorul de sinteză text vorbire în Cloud

3.3. Recunoașterea vizuală a vorbirii

Rezultatele raportate în această secțiune corespund Obiectivului 4.c din lista de obiective specifice Etapei a IV-a, iar ele sunt descrise în extenso în livrabilul D4.3b „Rezultate experimentale privind recunoașterea vizuală a vorbirii”.

3.3.1. Modelul propus

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 bazează pe interpretarea vizuală a mișcării buzelor, feței și limbii. Mișcări care sunt utilizate de obicei de oamenii care au probleme de auz, dar la nivel de subconștient sunt procesate și de persoanele cu auz normal.

În cadrul proiectului s-a proiectat și implementat software un sistem pentru recunoașterea vizuală a cuvintelor/vorbirii. Validarea sistemului s-a realizat pe un dicționar de cuvinte redus. Pentru realizarea dicționarului/ setului de date pentru experimentare, secvențele video au fost achiziționate folosind o cameră web, acestea fiind preprocesate prin - segmentarea secvenței pe cuvinte și salvarea lor. Fiecare cuvânt a fost înregistrat de câte două ori de fiecare vorbitor, iar înregistrările s-au realizat pe doi vorbitori. Sistemul prototip pentru recunoașterea vizuală a cuvintelor/vorbirii, se poate descrie în trei etape:

- Prima etapă constă în localizarea buzelor și realizarea unei secvențe video conținând zona buzelor. Detectia feței respectiv a regiunii buzelor se realizează cu ajutorul algoritmului Viola-Jones.
- Etapa a doua implică extragerea vectorului de trăsături/descrierea secvenței video optim, printr-un vector cât mai scurt. Vectorul de trăsături obținut pentru fiecare cuvânt este salvat sub forma unui fișier binare .dat într-un director. Fișierele binare sunt utilizate/necesare la antrenarea și validarea sistemului.
- Ultima etapă este cea de clasificare utilizând o mașină cu vectori suport (SVM). În urma clasificării vom avea ca rezultat forma scrisă a cuvântului, în cazul de față a numărului.

Sistemul a fost testat pe 20 de secvențe video (10 cuvinte – numerele de la unul la zece, doi vorbitori), iar rata de recunoaștere pe acest set este de peste 90%. Antrenarea sistemului a fost realizată tot pe 20 secvențe video din totalul de 40 secvențe înregistrate. Recunoașterea se poate face și pe baza unor fișiere binare care au fost create pentru validarea clasificatorului. Sistemul a fost implementat sub forma unei aplicații Windows de tip stand-alone și este ușor de utilizat și interactiv, permițând popularea dicționarului/ adăugarea de noi cuvinte.

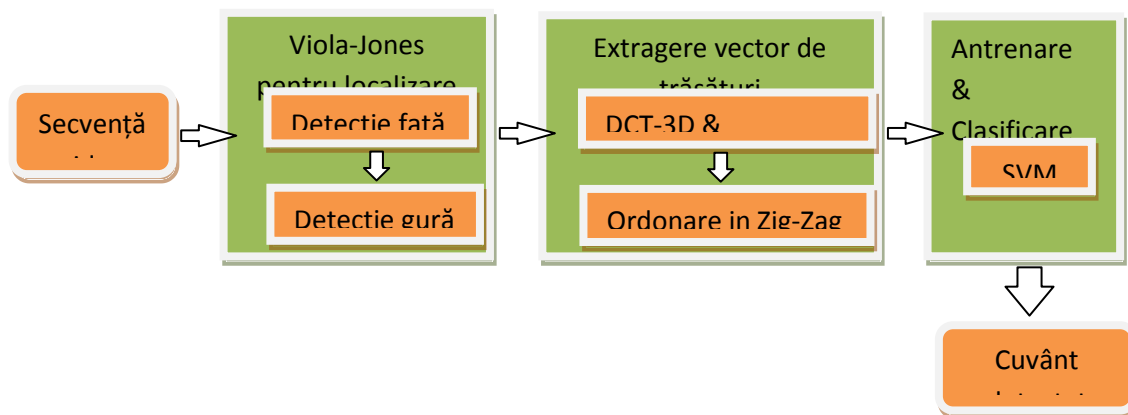


Figura 6. Arhitectura sistemului pentru recunoașterea vizuală a vorbirii

3.3.2. Rezultate experimentale privind recunoașterea vizuală a vorbirii

Aplicația implementată oferă posibilitatea de a localiza fața unei persoane într-o imagine făcută din profil, iar în zona de față se localizează buzele. Se permite o ajustare a regiunii buzelor detectate, pentru centrarea cât mai exactă a zonei de interes (rezultatul dat de algoritmul Viola&Jones nu conține obiectul centrat). Rezultatele clasificării depind foarte mult de ajustarea dreptunghiului în jurul gurii, astfel dacă nu este centrată bine regiunea de interes rezultatul va fi unul eronat. În condiția în care regiunea de interes a buzelor a fost aleasă corect rata de recunoaștere a programului este de 90%.



Figura 7. Secvența video care conține pronunția cifrei 7 de la vorbitorul masculin

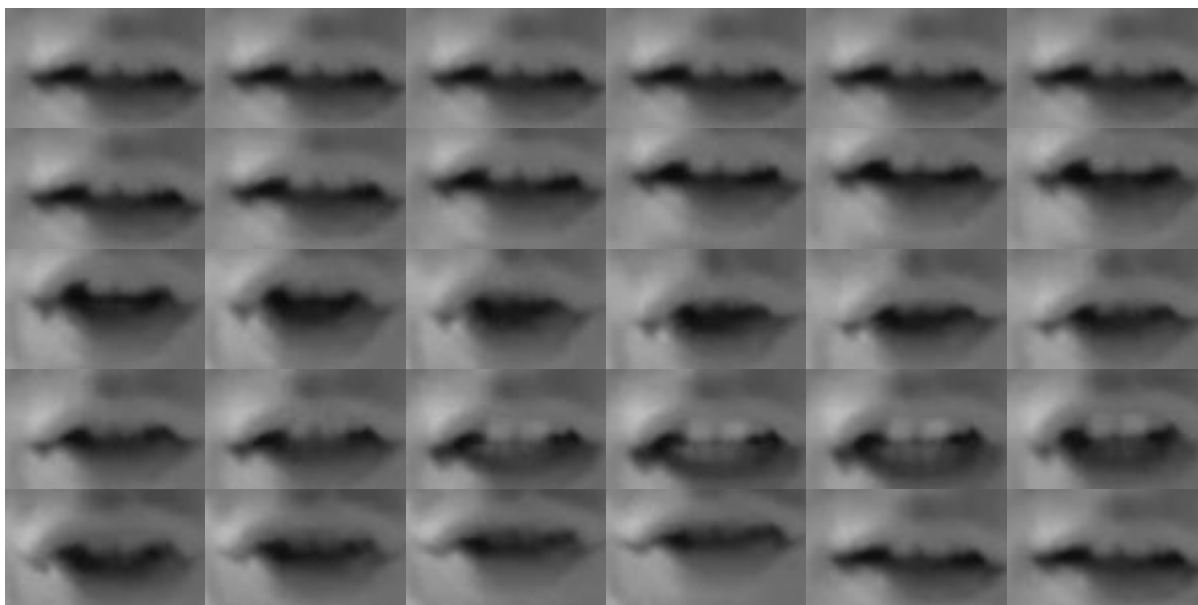


Figura 8. Secvența video care conține pronunția cifrei 5 de la vorbitorul feminin

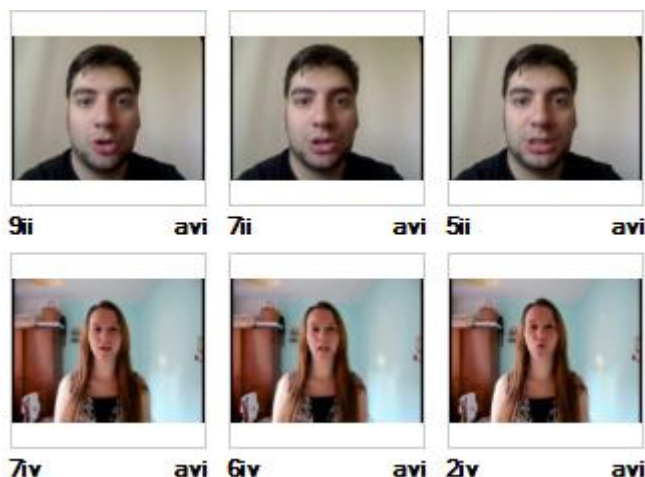


Figura 9. Secvențele video ce conțin imagini din directorul de test

3.4. Evaluarea impactului psiho-social al vocilor sintetizate personalizate

Rezultatele raportate în această secțiune corespund Obiectivului 4.b din lista de obiective specifice Etapei a IV-a, iar ele sunt descrise în extenso în livrabilul D2.4 cu titlul „Raport științific asupra impactului psiho-social al vocilor sintetizate personalizate”. Rezultatele au fost diseminate la 4 conferințe internaționale prin articolele [3][4][5][6][7][8].

3.4.1. Metodologia

Participanți. În lotul final au fost incluși 20 de participanți, cu vârste cuprinse între 25 și 74 de ani, (media fiind de 59 de ani) dintre aceștia 1 pacient a fost de sex feminin. Majoritatea sunt absolvenți de liceu (25 %), iar un procent mai mic (20 %) sunt absolvenți de studii de licență. Un procent de 70 % au declarat că provin din mediul urban. La momentul luării în studiu, 35 % sunt angajați, 50 % au fost pensionați ca limită legală de vârstă și 15% au fost pensionați pe motiv de invaliditate. Studiul clinic a inclus pacienți ce folosesc deja o metodă alternativă de asistare de voce (laringofon sau proteza ProVox) 60 % și pacienți fără nici o metodă de asistare de voce (40 %). Au fost incluși în studiu doar pacienții care dispun de telefoane mobile Smart și acces la rețeaua de internet. Pacienții au oferit un feedback referitor la sistemul de sinteză de voce în urma utilizării acestuia timp de 1 lună.

Instrumente. Pentru a evalua comparativ satisfacția percepută a vocilor sintetice cu cea a unui pacient cu laringectomie am folosit un chestionar pentru evaluarea sistemului mobil de sinteză de voce – SWARA. Pacienții au fost rugați să completeze chestionarul pentru a evalua beneficiile și problemele legate de sistemul de sinteză de voce. Întrebările au fost scurte, simple și ușor de răspuns. Întrebările au investigat capacitatea de comunicare în situații diferite: comunicare cu persoane cunoscute și necunoscute, comunicarea în mediul cu zgomot și satisfacția personală. Cele 4 voci sintetice folosite au fost generate de echipa UTCN și au cuprins două voci de gen feminin și două de gen masculin, fiecare dintre ele sintetizând nouă mesaje care reies din interacțiune a trei valențe emoționale (negativă, neutră, pozitivă) și trei contexte (familiar, persoană necunoscută, context medical). Ca termen de comparație s-a optat pentru utilizarea unei voci produse prin proteză vocală, ca și punct de referință pentru vocile sintetice. Mesajele sintetizate, respectiv transmise de vocea pacientului au conținut fiecare o structură propozițională enunțiativă, o structură propozițională exclamativă, respectiv una interogativă, pentru a controla pentru un potențial efect al tonalităților diferite implicate în cele trei tipuri de propoziții/fraze. Fiecare participant a ascultat vocea pacientului cu proteză vocală și în mod aleatoriu, una din 4 voci sintetice posibile.

3.4.2. Sinteza rezultatelor evaluării

Rezultatele evaluărilor 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, $t(81) = .76$, $p = .451$. Figurile 1 ilustrează aceste comparații între vocea pacientului și vocile sintetice pe cele trei dimensiuni. În continuare am repetat aceste comparații, luând însă în considerare tipul de voce sintetică pe care l-au ascultat participanții (voce de gen masculin vs. voce de gen feminin).

Rezultatele au urmat aceeași direcție indiferent de genul vocii sintetice care a fost folosită pentru comparație. Vocile feminine au fost percepute ca fiind mai clare decât vocea pacientului, $t(47) = -7.63$, $p < .001$, $d = 1.01$. Vocea pacientului cu proteză vocală fost percepută ca fiind mai naturală decât vocile sintetice feminine, $t(47) = 4.05$, $p < .001$, $d = .70$. Nu am identificat diferențe semnificative statistic între vocea pacientului și cele sintetice feminine în ceea ce privește atitudinea față de interacțiuni viitoare, $t(47) = -.14$, $p = .891$. În comparația cu vocile masculine, pentru claritate, vocile sintetice au fost evaluate cu scoruri mai bune decât vocea pacientului, $t(33) = -4.67$, $p < .001$, $d = .78$. Naturalitatea vocii generate cu proteza vocală a fost percepută mai bună decât cea a vocilor sintetice masculine, $t(33) = 3.22$, $p = .002$, $d = .56$. În fine, nu am găsit diferențe între vocea pacientului și vocile sintetice masculine în ceea ce privește atitudinea față de interacțiuni viitoare, $t(33) = -0.90$, $p = .377$.

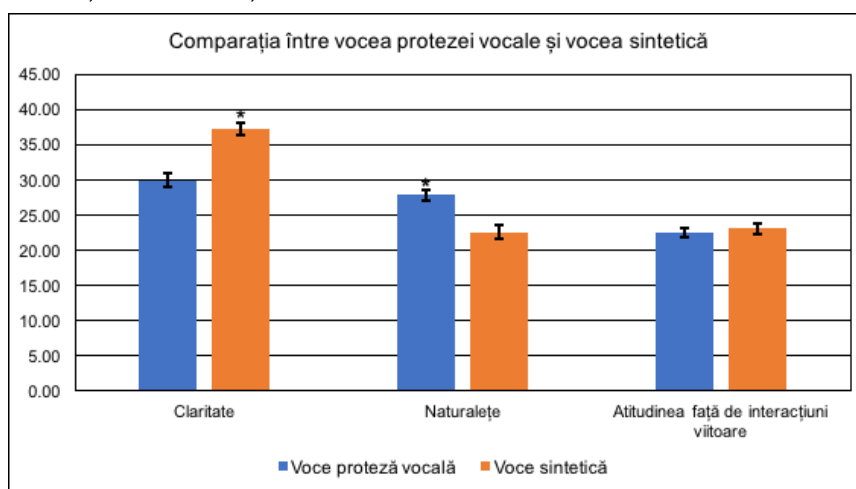


Figura 10. Comparația între vocea protezei vocale și cea sintetică pe trei dimensiuni: claritate, naturalitate și atitudinea față de interacțiuni viitoare. Scorurile mai mari indică o claritate și naturalitate mai bună, respectiv o atitudine mai pozitivă. Valorile marcate cu steluță indică scoruri semnificativ mai mari decât termenul de comparație

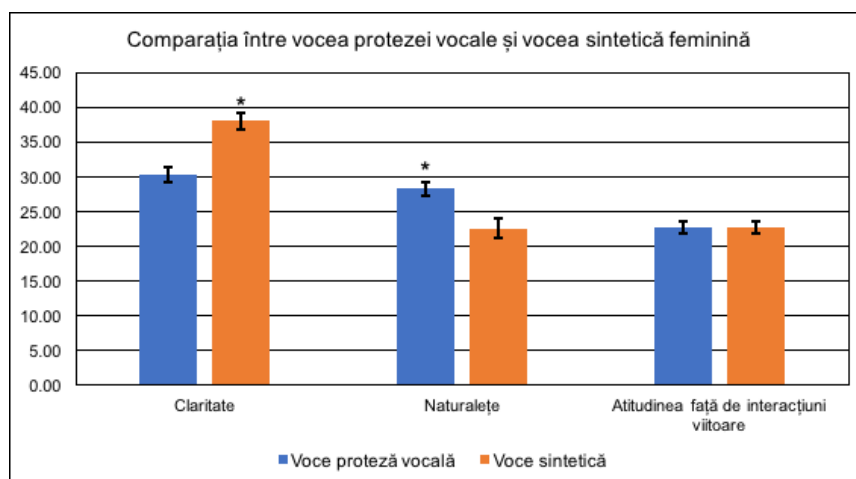


Figura 11. Comparația între vocea protezei vocale și cele sintetice feminine pe trei dimensiuni: claritate, naturalitate și atitudinea față de interacțiuni viitoare. Scorurile mai mari indică o claritate și naturalitate mai bună, respectiv o atitudine mai pozitivă. Valorile marcate cu steluță indică scoruri semnificativ mai mari decât termenul de comparație.

Rezultatele chestionarului privind calitatea vocilor sintetice, a arătat că: 16 (80 %) dintre pacienți au raportat că sunt satisfăcuți de calitatea acestora, în timp ce 4 (20 %) au fost oarecum mulțumiți. Cu privire la utilizarea aplicației, doar 2 (10 %) pacienți au utilizat sistemul de sinteză de voce în comunicarea cu alte persoane, în timp ce 18 (90 %) au renunțat la utilizarea acestei metode de asistare vocale din diverse motive

Principala dificultate a fost însă înregistrată la utilizarea de rutină a link-ului de accesare a vocilor produse de echipa tehnică, accesarea fiind greoaie, dependentă de funcționalitatea rețelelor de telefonie mobilă sau de telefonie fixă și internet și de faptul că site-ul nu este mobile-friendly. Viteza inconstantă de internet a făcut dificilă transformarea informației scrise în mesaj vorbit, cu latența sintezei de voce și dispariția cursivității în exprimare, rezultând nervozitate din partea utilizatorului și renunțare în utilizarea vocii sintetice ca modalitate de comunicare.

În ultima etapă a analizei am verificat atitudinile respondenților față de ideea ca un dispozitiv medical să preia o funcție naturală a corpului, respectiv atitudinea față de utilizarea vocilor sintetice pentru pacienții cu laringectomie. Rezultatele au arătat că majoritatea respondenților sunt de acord cu utilizarea dispozitivelor tehnologice care preiau funcțiile naturale ale corpului, condiția fiind ca acel dispozitiv să funcționeze constant, fără discontinuități care să inducă latență în cursivitatea comunicării verbale. În ceea ce privește atitudinea față de utilizarea vocilor sintetice, un procent de 55 % au considerat ideea utilizării vocilor sintetice pentru pacienții cu laringectomie ca fiind utilă dacă funcționează într-o aplicație ce poate oferi continuitate în exprimare. Un procent de 20 % din respondenți au raportat că această idee are o utilitate limitată; 25% din respondenți au considerat această modalitate de asistare ca fiind inutilă. Iar referitor la modul de introducere a textului, 80 % dintre corespondenți preferă labiolectura, în timp ce 20 % prin tastare.

3.5. Planul de exploatare și acord privind drepturile de proprietate intelectuală

Rezultatele raportate în aceasta secțiune corespund Obiectivului 4.f din lista de obiective specifice Etapei a IV-a, iar ele sunt descrise în extenso în livrabilele D7.6 cu titlul „Planul de exploatare” și D7.5 „Acord final privind drepturile de proprietate intelectuală”.

3.5.1. Acordul privind proprietatea intelectuală

În acest livrabilul D7.5. sunt prezentate principiile agreeate pentru respectarea proprietății intelectuale după finalizarea proiectului, în scopul exploatarea și valorizării rezultatelor acestuia. Se 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.

3.5.2. Planul de exploatare

În livrabilul D7.6 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. Acest plan a fost elaborat în limba engleză, deoarece aceasta este modalitatea de lucru în firma parteneră (Fortech SRL) în momentul în care se propun proiecte de comercializare.

Conform unui studiu recent⁸ 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:

- rapoarte ale companiile Hoovers, Bloomberg Business, Factiva, OneSource
- ofertele principalilor competitori
- adoptarea unei proceuri de analiză de tip bottom-up
- segmentarea piețelor
- interviuri oferite de CEP, VPs
- aplicarea unor metode de segmentare cum ar fi triangularea datelor

Principalii furnizori de sisteme TTS pentru care s-a făcut analiza sunt Nuance Communication (USA), Microsoft Corporation (USA), Sensory Inc. (USA), LumenVox LLC (USA), NeoSpeech (USA), Amazon.Com (USA), SESTEK (Turkey), iSpeech Inc. (USA), TextSpeak (USA), NextUp Technologies (USA), Acapela Group (France), CereProc (Scotland), ReadSpeaker (Netherlands), Nexmo (USA), and rSpeak (Netherlands).

Tehnologia dezvoltată în proiectul SWARA (Sistem Mobil de Asistare Vocală în Reintegrarea Persoanelor cu Afonii Chirurgicale) are principalul scop 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. Studiile arată că pacienții ce au suferit laringectomia totală refuză, evită sau nu reușesc să folosească protezele vocale oferite azi pe piață pentru a comunica verbal cu familia sau cu alți membrii ai societății. În plus, atât pacienții cât și familiile acestora, sau chiar prietenii apropiați au dificultăți în a accepta calitatea slabă a noii voci sau chiar a aparatului respectiv.

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. Aceste sisteme au totuși o serie de dezavantaje, printre care se află natura letăa vorbirii, numărul limitat de voci disponibile, precum și dificultatea de a personaliza aceste voci sintetice. De asemenea, echipamentele actuale se prezintă sub forma unui mini-computer costisitor, greu de utilizat pe durata întregii zile. În plus, editarea mesajului se face prin tastarea fiecărei litere, rezultând pauze în conversație și o interacțiune greoaie între utilizator și interlocutor. Astfel că, principalul scop al acestui proiect a fost crearea unei tehnologii de sinteză a vorbirii care să permită 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 a fost atins prin îndeplinirea următoarelor obiective:

- dezvoltarea unei noi tehnologii de sinteză text-vorbire cu voce personalizată, care să utilizeze vocea anterioară a pacientului sau o voce similară creată în mod artificial dintr-o bază de date cu mulți vorbitori, în cazul în care alterarea patologică a vocii pacientului nu permite înregistrarea ei prealabilă;
- implementarea a două modalități alternative inteligente de editare rapidă a textului, ce facilitează interacțiunea om-mașină prin predicția textului sau recunoașterea vizuală a vorbirii, astfel încât să nu existe pauze inutile și stânjenitoare pe durata conversației;
- implementarea tehnologiei de sinteză a vocii pe dispozitive mobile, corespunzător cu cerințele de funcționalitate și mobilitate ale unui astfel de sistem.

⁸http://www.marketsandmarkets.com/Market-Reports/text-to-speech-market-2434298.html?gclid=CjwKCAjwpfzOBRA5EiwAU0ccN6YHdUW4mQnSis-UzIRTMgQvrvBonaioi4Vsh9tJFTmQTdw31WQRwtBoCjEcQAvD_BwE

Tabel 4. Lista tehnologiilor cu potențial de comercializare (selecție a celor mai importante)

Tehnologie disponibilă pentru exploatare comercială	Cotă contribuție partener			
	UTCN	Fortech	UMF	UBB
D1.3. Sistem de sinteza 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 în Cloud, accesibil de pe mobil	25%	75%	0%	0%

4. Management și comunicare

Activitățile de management au fost orientate, în special către managementul grupurilor de cercetare constituite în jurul obiectivelor etapei și a interacțiunii dintre acestea în vederea integrării diferitelor componente software în sistemul final. Astfel, întâlniri ale întregului parteneriat au fost realizate în 20.01.2017 (pentru planificarea anuală) și în 06.09.2017 (pentru pregătirea raportării finale). Datorită bunei funcționări a comunicării prin Skype și eMail, dar și a faptului că deja activitățile și problematica de cercetare sunt acum bine cunoscute, grupurile de cercetare au avut doar întâlniri față în față trimestrial.

Este de notat faptul că s-a asigurat o bună comunicare și coordonare cu responsabilii financiari ai partenerilor, astfel ca documentele administrative legate de raportarea financiară și a auditului pe 2017 a fost foarte eficientă. Și în acest an 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 cu o regularitate adecvată și s-au derulat câteva activități privind achizițiile de obiecte de inventar necesare activităților de zi cu zi în proiect.

5. Diseminarea rezultatelor

O preocupare 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: a) actualizarea dinamică a paginii web cu rezultatele obținute incremental în proiect, inclusiv cu secțiuni demonstrative; b) elaborarea unui plan de diseminare pentru anul 2017; c) publicarea rezultatelor științifice la conferințe internaționale de prestigiu în domeniul proiectului (tehnic, 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 livrabilul D7.4a „Publicații și tutoriale”.

5.1. Pagina web a proiectului

Pagina web a proiectului⁹, 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 și legături către serviciile web care se vor dezvolta. Serviciul de monitorizare automată a site-ului, Google Analytics, a continuat să funcționeze și în 2017 (vezi mai jos raportul de monitorizare).

⁹ <http://speech.utcluj.ro/swara>

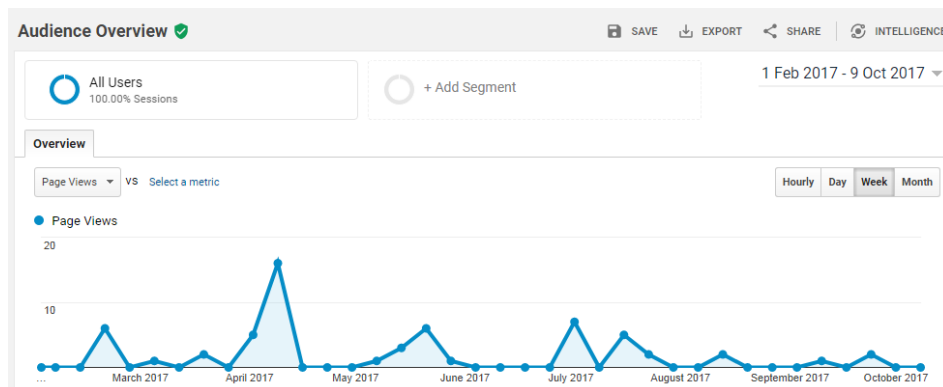


Figura 12. Rapoarte despre accesarea siste-ului folosind Google Analytics

5.2. Pagini web dedicate pentru demonstrarea online a unor funcționalități

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 o serie dintre modulele implementate, sistemul de sinteză online¹⁰, demonstratorul cu predicția textului¹¹, sa asculte mostre de semnal audio sintetizate cu sistem HTS¹² sau cu sistemul de sinteză text vorbire bazat pe DNN¹³ implementat în această etapă, respectiv corpusul SWARA¹⁴.

5.3. Publicații științifice aferente etapei 2017

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 și prezentate în livrabilul aferent diseminării D7.4b „Publicații și tutoriale”. Alte rezultate sunt în curs de publicare. Pentru vizibilitate directă, publicațiile științifice sunt listate mai jos și prezentate public pe pagina web a proiectului¹⁵.

[1] Adriana Stan; Florina Dinescu; Cristina Țiple; 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. <http://ieeexplore.ieee.org/document/7990428/>

[2] 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/>

[3] Matu S., Șoflău R., David D., Chirilă M., Țiple 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>

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

[4] Matu, S., Șoflău, R., David, D., Chirilă M., Țiple, 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, <http://www.iccp2017.org>

[5] Matu S., Șoflău R., David D., Chirilă M., Țiple 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 European Association for Behavioural and Cognitive Therapies. Ljubljana, 2017, <http://www.eabct2017.org/en/>

[6] Matu S., Șoflău R., David D., Chirilă M., Țiple 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, <http://www.eabct2017.org/en/>

[7] Magdalena Chirila, „Mobile system of vocal assistance for patients with surgical aphonia”, IFOS: INTERNATIONAL FEDERATION OF OTO-RHINOLARYNGOLOGICAL SOCIETIES, ENT World Congress, Paris, 24-28 June 2017, <http://ifosparis2017.org>

[8] Cristina Țiple, F Dinescu, S Matu, R Mureșan, R Șoflău, 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. Concluzii

Activitățile de cercetare desfășurate în etapa a patra de implementare a proiectului (2017) au condus la obținerea rezultatelor așteptate și ele sunt în concordanță cu obiectivele specifice ale etapei. Astfel, rezultatele raportate în acest document și descrise detaliat în cele 8 livrabile aferente perioadei de raportare (vezi Secțiunea 7 a acestui raport), pregătesc cadrul pentru exploatarea și valorizarea comercială a sistemului de sinteză text vorbire.

7. Referințe la livrabilele aferente etapei a patra, anul 2017 (Anexe la raport)

[1] Livrabil D5.1:	„Sistem preliminar de adaptare a vocii”, Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Septembrie 2017. <i>Nivel diseminare: Confidential</i>
[2] Livrabil D5.2:	„Un nou model de adaptare”, Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Septembrie 2017. <i>Nivel diseminare: Confidențial</i>
[3] Livrabil D6.2a.	„Versiune demonstrativă a sistemului de sinteză text vorbire în Cloud, accesibil de pe mobil”, Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Septembrie 2017. <i>Nivel diseminare: Public (inclus în acest raport)</i> <i>(http://swara.fortech.ro/audio) si</i> <i>(https://swara.fortech.ro/audio/predictionTests/medical.html)</i>
[4] Livrabil D2.4:	„Raport științific asupra impactului psiho-social al vocilor sintetizate personalizate”, Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Septembrie 2017. <i>Nivel diseminare: Public</i>
[5] Livrabil D4.3b:	„Rezultate experimentale privind recunoașterea vizuală a vorbirii”, Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Septembrie 2017. <i>Nivel diseminare: Confidențial</i>
[6] Livrabil D7.5	„Acord final privind drepturile de proprietate intelectuală”, Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Septembrie 2017. <i>Nivel diseminare: Confidențial</i>
[7] Livrabil D7.6.	„Planul de exploatare”, Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Septembrie 2017. <i>Nivel diseminare: Confidențial</i>
[8] Livrabil D7.4b:	„Publicații și tutoriale”, Proiect SWARA PN-II-PT-PCCA-2013-4-1660, Septembrie 2017. <i>Nivel diseminare: Public</i> <i>Public (http://speech.utcluj.ro/swara/results.html#publications)</i>
