

# Proiect component TADARAV

---

- Raport științific și tehnic în extenso -

Alexandru-Lucian Georgescu, Cristian Manolache, Gheorghe Pop, Dan Onea ță,  
Horia Cucu, Dragoș Burileanu, Corneliu Burileanu

**Program:** PNCDI III - Programul 1 - Dezvoltarea sistemului național de CD

**Proiect complex:** Resurse și tehnologii pentru dezvoltarea interfețelor om-mașină în limba română (ReTeRom)

**Proiect component:** Tehnologii pentru adnotarea automată a datelor audio și pentru realizarea interfețelor de recunoaștere automată a vorbirii (TADARAV)

**Data:** 29.11.2019

**Etapa:** 2 / 2019

**Activitatea / activitățile:**

- Activitatea 2.11 - Proiectarea și implementarea unei soluții de bază de filtrare și aliniere a transcrierilor aproximative cu semnalul de vorbire
- Activitatea 2.12 - Proiectarea și implementarea unei soluții de bază pentru generarea de scoruri de încredere pentru RAV
- Activitatea 2.13 - Îmbunătățirea soluției de adnotare automată a semnalului de vorbire utilizând sisteme de RAV complementare
- Activitatea 2.14 - Diseminare

**Număr contract:** 73PCCDI / 2018

**Acord de colaborare:** 30/20.02.2018 ICIA, 4726/01.03.2018 UTCN, 3950/07.03.2018 UPB, 3805/06.03.2018 UAIC

**Autoritatea contractantă:** Unitatea Executivă pentru Finanțarea Învățământului Superior, a Cercetării, Dezvoltării și Inovării

**Conducător proiect component:** Universitatea POLITEHNICA din București

**Conducător proiect complex:** ICIA

**Responsabil proiect component:** Conf. Horia Cucu

**Responsabil proiect complex:** Prof. Corneliu Burileanu

## Cuprins

<b>Rezumatul etapei</b>	<b>4</b>
<b>Descrierea științifică și tehnică a activităților</b>	<b>5</b>
Introducere	5
Seturi de date	5
Activitatea 2.11 - Proiectarea și implementarea unei soluții de bază de filtrare și aliniere a transcrierilor aproximative cu semnalul de vorbire	7
Descrierea metodei	7
Evaluarea cantitativă și evaluarea cantitativă metodei	9
Optimizarea procesului de aliniere și filtrare	9
Utilizarea transcrierilor aproximative în generarea de date	13
RAV utilizând corpusul nou creat	14
Concluzii	14
Activitatea 2.12 - Proiectarea și implementarea unei soluții de bază pentru generarea de scoruri de încredere pentru RAV	15
Evaluarea metodelor de decodare	15
Evaluarea calitativă a metodei: corelarea scorurilor de încredere cu corectitudinea transcrierii RAV	16
Utilizarea scorurilor de încredere în generarea de date	18
RAV utilizând corpusul nou creat	18
Activitatea 2.13 - Îmbunătățirea soluției de adnotare automată a semnalului de vorbire utilizând sisteme de RAV complementare	19
Introducere	19
Sisteme de RAV inițiale îmbunătățite	20
Evaluarea calitativă a metodei: complementaritatea sistemelor de RAV inițiale	22
Utilizarea sistemelor RAV complementare în generarea de date	23
RAV utilizând corpusul nou creat	23
Activitatea 2.14 - Diseminare	24
<b>Structura ofertei de servicii de cercetare și tehnologice</b>	<b>25</b>
<b>Locuri de muncă susținute prin program</b>	<b>25</b>
<b>Valorificarea și îmbunătățirea competențelor și resurselor existente la nivelul consorțiului</b>	<b>26</b>
<b>Anexe</b>	<b>27</b>

# 1 Rezumatul etapei

A doua etapă a proiectului TADARAV a avut trei obiective principale:

1. evaluarea posibilității utilizării transcrierilor aproximative ale materialelor ce conțin vorbire, împreună cu un sistem de recunoaștere automată a vorbirii (RAV) inițial, pentru a produce în mod automat transcrieri precise pentru o parte a semnalului de vorbire;
2. evaluarea posibilității utilizării scorurilor de încredere generate de un sistem de RAV inițial pentru a produce în mod automat transcrieri precise pentru o parte a semnalului de vorbire;
3. îmbunătățirea soluției de adnotare automată a semnalului de vorbire utilizând sisteme de RAV complementare.

Toate metodele de adnotare automată au fost evaluate și în contextul reantrenării sistemului de RAV inițial cu datele nou generate. Cele trei obiective au fost realizate în proporție de 100%, în urma activităților întreprinse rezultând toate livrabilele asumate de consorțiu la începutul acestei etape.

Concret, în urma activităților A2.11, A2.12 și A2.13 din etapa 2/2019 a proiectului TADARAV, au rezultat următoarele livrabile:

- Soluție de bază de filtrare și aliniere a transcrierilor aproximative cu semnalul de vorbire (TRL3), funcțională
- Soluție de bază pentru generarea de scoruri de încredere pentru RAV (TRL3), funcțională
- Soluție îmbunătățită de adnotare automată a semnalului de vorbire utilizând sisteme de RAV complementare (TRL4), funcțională

Diseminarea rezultatelor proiectului a fost realizată: în cadrul consorțiului în workshopul organizat la Cluj-Napoca pe 18 noiembrie 2019 și în comunitatea științifică la trei conferințe internaționale de prestigiu. Suplimentar, unele dintre rezultate au fost publicate într-un articol în Buletinul Științific al Universității Politehnica din București. De asemenea, progresul înregistrat în această etapă a fost diseminat prin intermediul website-ului proiectului: <https://tadarav.speed.pub.ro>. Articolele publicate în conferințe și jurnale științifice sunt următoarele:

1. Alexandru-Lucian Georgescu, Horia Cucu, Corneliu Burileanu, "[Progress on automatic annotation of speech corpora using complementary ASR systems](#)," in the Proceedings of the 42nd International Conference on Telecommunications and Signal Processing (TSP), 2019, Budapest, Hungary.
2. Gheorghe Pop, Serban Mihalache, Dragos Burileanu, "[Forensic Recognition of Narrowband AMR Signals](#)," in the Proceedings of the 10th Conference on Speech Technology and Human-Computer Dialogue (SpeD), Timișoara, Romania, 2019.
3. Alexandru-Lucian Georgescu, Horia Cucu, Corneliu Burileanu, "[Kaldi-based DNN architectures for speech recognition in Romanian](#)," in the Proceedings of the 10th Conference on Speech Technology and Human-Computer Dialogue (SpeD), Timișoara, Romania, 2019.
4. Gheorghe Pop and Dragos Burileanu, "[Speech Enhancement for Forensic Purposes](#)," in UPB Scientific Bulletin, Series C, Vol. 81, Issue 3, pp. 41-52, 2019.
5. Florin Iordache, Alexandru-Lucian Georgescu, Dan Oneață, Horia Cucu, "Romanian Automatic Diacritics Restoration Challenge", in the Proceedings of the 14th International Conference on Linguistics Resources and Tools for Natural Language Processing, Cluj-Napoca, Romania, 2019.

Dintre articolele listate mai sus, două sunt deja indexate în Web of Science (Thompson Reuters - ISI), altele două sunt deja indexate IEEE Xplore și în curs de indexare în Web of Science (Thompson Reuters - ISI), iar al cincilea a apărut în volumul conferinței și este în curs de indexare în Web of Science (Thompson Reuters - ISI).

## 2 Descrierea științifică și tehnică a activităților

### 2.1 Introducere

Modelele acustice bazate pe rețele neuronale profunde (*Deep Neural Network* – DNN) obțin performanțe direct proporționale cu cantitatea de date folosite la antrenarea rețelei. Prin urmare, dat fiind faptul că adnotarea manuală a resurselor audio presupune o investiție consistentă de efort și timp, interesul față de tehnicile de adnotare automată a vorbirii a crescut semnificativ. Adnotarea automată a vorbirii presupune colectarea de vorbire în format brut și folosirea unei metode automate pentru a produce transcrieri cât mai precise pentru cel puțin o parte din corpusul inițial.

#### 2.1.1 Seturi de date

Activitățile A2.11, A2.12 și A2.13 presupun (i) utilizarea unor seturi de date de vorbire deja existente pentru antrenarea și evaluarea unor sisteme de RAV necesare în aplicarea metodelor de adnotare automată și (ii) utilizarea unor seturi de date brute, neadnotate sau adnotate parțial ca date de intrare pentru cele trei metode de adnotare automată. Aceste seturi de date sunt sumarizate în tabelele 2.1.a și 2.1.b.

#### Seturi de date de vorbire adnotată

Pentru antrenarea și evaluarea sistemelor de RAV, au fost folosite două seturi de date de vorbire în limba română: *Read Speech Corpus* (RSC), ce conține vorbire citită, colectată în condiții de laborator, fără zgomot de fundal și *Spontaneous Speech Corpus* (SSC), ce conține vorbire continuă, spontană, preluată de la posturi de radio și TV, uneori afectată de zgomot. Ambele corpusuri cuprind fișiere audio și transcrieri corespunzătoare și sunt divizate în seturi de antrenare și seturi de evaluare. RSC-train este setul de antrenare din RSC, ce conține 100 ore de vorbire citită, cuvinte izolate sau fraze de la 157 de vorbitori diferiți. RSC-eval este setul de evaluare din RSC; acesta conține vorbire de la 22 de vorbitori diferiți, însumând 5.5 ore de vorbire. SSC-train este setul de antrenare din SSC și conține 130 ore de vorbire spontană, majoritatea din emisiuni de știri și *talkshow*-uri. SSC-eval este setul de evaluare din SSC și însumează 3.5 ore de vorbire.

În etapa anterioară a proiectului, ca parte a activității A1.13 au fost obținute seturile de date de vorbire adnotată SSC-train3-compl și SSC-train4-compl. Ele sunt prezentate, alături de seturile de vorbire adnotată RSC și SSC în Tabelul 2.1.a.

**Tabelul 2.1.a** Seturile de vorbire adnotată folosite pentru antrenarea și evaluarea sistemelor de RAV și seturile de vorbire adnotată obținute în etapa anterioară (1/2018)

Corpus	Set	Durată	
Antrenare	RSC-train	94h, 46m	225h, 30m
	SSC-train	130h, 44m	
Evaluare	RSC-eval	5h, 29m	8h, 58m
	SSC-eval	3h, 29m	
SSC-train3-compl-2018	radio #1	6h, 20m	49h, 13m
	TV #1	10h, 00m	
	TV #2	32h, 53m	
SSC-train4-compl-2018	radio #1	25h, 16m	280h, 00m
	TV #1	66h, 02m	
	TV #2	188h, 42m	

## Seturi de date brute

Seturile de date brute, neadnotate sau adnotate parțial, utilizate ca date de intrare pentru cele trei metode de adnotare automată sunt denumite SSC-train3-raw și SSC-train4-raw și sunt prezentate în Tabelul 2.1.b. Primul set de date neadnotat, SSC-train3-raw, a fost achiziționat din mass-media românească, mai exact de pe 2 website-uri de știri și un post de radio, de-a lungul unei perioade de o lună calendaristică. Al doilea set de date de vorbire neadnotată, SSC-train4-raw, a fost achiziționat de asemenea din cele 3 surse din mass-media românească, de-a lungul unei perioade de nouă luni calendaristice.

Seturile de date au fost achiziționate cu o aplicație creată în cadrul proiectului, aplicație ce parcurge *feed*-urile RSS al acestor *website*-uri, identifică știrile noi și descarcă fișierele audio (eșantionate la 16 kHz, 16 biți pe eșantion) și textele corespunzătoare știrilor respective.

**Tabelul 2.1.b** Seturi de date de vorbire neadnotată (+ transcrieri aproximative) utilizate ca date de intrare pentru cele trei metode de adnotare automată. Numărul de cuvinte se referă la textul brut descărcat de pe fiecare website în parte

	SSC-train3-raw		SSC-train4-raw	
	# cuvinte	# ore	Nr. de cuvinte	Nr. de ore
radio #1	30.049	19,3	120.121	78,2
TV #1	357.926	51,5	2.241.389	331,6
TV #2	825.722	65,9	4.111.690	367,3
<b>Total</b>	1.213.697	136,7	6.473.200	777,2

## Seturi de date de vorbire adnotată rezultate în această etapă a proiectului

După aplicarea celor trei metode de adnotare automată au fost obținute seturile de date din Tabelul 2.1.c.

**Tabelul 2.1.c** Seturile de vorbire adnotată rezultate în urma aplicării metodelor de adnotare automată

Corpus	Sursa	Durață [# ore]	Eficiență aliniere [% ore]		
SSC-train3-compl-2019	radio #1	9h, 51m	79h, 52m	51.0%	58.4%
	TV #1	27h, 52m		54.0%	
	TV #2	42h, 9m		63.9%	
SSC-train4-compl-2019	radio #1	36h, 00m	452h, 32m	46.0%	58.2%
	TV #1	181h, 42m		54.7%	
	TV #2	234h, 50m		63.9%	
SSC-train3-trans-v3	radio #1	1,0	37,5	5,0%	27,4%
	TV #1	12,8		25,0%	
	TV #2	23,6		35,9%	
SSC-train4-trans-v3	radio #1	2,7	228,8	3,5%	29,4%
	TV #1	87,9		26,5%	
	TV #2	138,1		37,6%	
SSC-train3-conf	radio #1	5h, 30m	55h, 51m	28,5%	44,1%
	TV #1	19h, 07m		37,1%	
	TV #2	31h, 13m		47,4%	
SSC-train4-conf	radio #1	22h 03m	315h, 34m	28,2%	40,6%
	TV #1	124h 29m		37,5%	
	TV #2	169h 01m		46,0%	

## 2.2 Activitatea 2.11 - Proiectarea și implementarea unei soluții de bază de filtrare și aliniere a transcrierilor aproximative cu semnalul de vorbire

Activitatea A2.11 a avut ca scop evaluarea unei metode de generare de seturi de date de vorbire adnotată folosind materiale audio disponibile pe diverse website-uri de mass-media împreună cu știrile text de pe paginile respective. Folosind un singur sistem RAV s-a generat un set de transcrieri aproximative, urmând ca apoi acestea să fie aliniată cu transcrierile de pe site. Părțile identice dintre cele 2 seturi de transcrieri au fost considerate ca fiind corecte. Motivul pentru care nu s-au folosit doar transcrierile de pe site este că acestea nu conțin întotdeauna textul vorbit din fișierul audio; există cazuri în care lipsesc părți din vorbire precum și cazuri în care apar informații adiționale în transcriere. Astfel, prin efectuarea alinierii dintre transcrierile de pe site și cele obținute cu sistemul RAV sperăm să obținem secvențe audio-text cât mai precise.

### 2.2.1 Descrierea metodei

Metoda utilizată în această etapă are ca scop obținerea într-un mod automat, nesupervizat, a unei adnotări cât mai precise pentru un corpus de vorbire. Corpusul nou obținut s-a dorit a fi utilizat pentru antrenarea sistemelor de RAV existente, crescând astfel variabilitatea acustică a modelelor, îmbunătățind implicit și acuratețea transcrierilor. Pașii corespunzători metodei vor fi descriși în continuare, aceștia fiind totodată ilustrați în Figura 2.2.a.

Ideea principală a acestei metode de adnotare constă în utilizarea unui singur sistem RAV pentru a produce transcrieri pentru un corpus neadnotat, dar pentru care există transcrieri aproximative. În urma alinierii transcrierilor RAV cu transcrierile aproximative, vor fi selectate ca fiind corecte părțile identice dintre cele 2 seturi de transcrieri. În final, transcrierile selectate și segmentele de vorbire corespunzătoare sunt folosite pentru a forma un nou corpus adnotat de vorbire.

**Resursele brute.** Resursele brute utilizate în această metodă au fost prezentate în tabelul 2.1.b. Materialele brute conțin (i) vorbire neadnotată (audio) și (ii) transcrieri aproximative (text).

**Transcrierea vorbirii neadnotate.** Sistemul RAV folosit pentru transcrierea materialelor audio este sistemul HMM-DNN creat în activitatea A1.13 din etapa 1/2018. Mai multe informații tehnice despre acest sistem pot fi obținute consultând raportul etapei 1/2018. Transcrierile obținute în urma folosirii acestui sistem RAV conțin doar litere mici, nu conțin semne de punctuație sau cifre, iar cuvintele sunt însoțite de ștampile de timp (timpul de început al rostirii cuvântului și timpul de încheiere al rostirii cuvântului). Iată, cu titlul de exemplu, o astfel de transcriere RAV: bărbatul(3.71,4.14) de(4.14,4.25) treizeci(4.25,4.55) și(4.55,4.65) șase(4.65,4.93) de(4.93,5.05) ani(5.05,5.19) povestește(5.19,5.66) că(5.66,5.75) muncise(5.75,6.22) toată(6.22,6.62) noaptea(6.62,6.99).

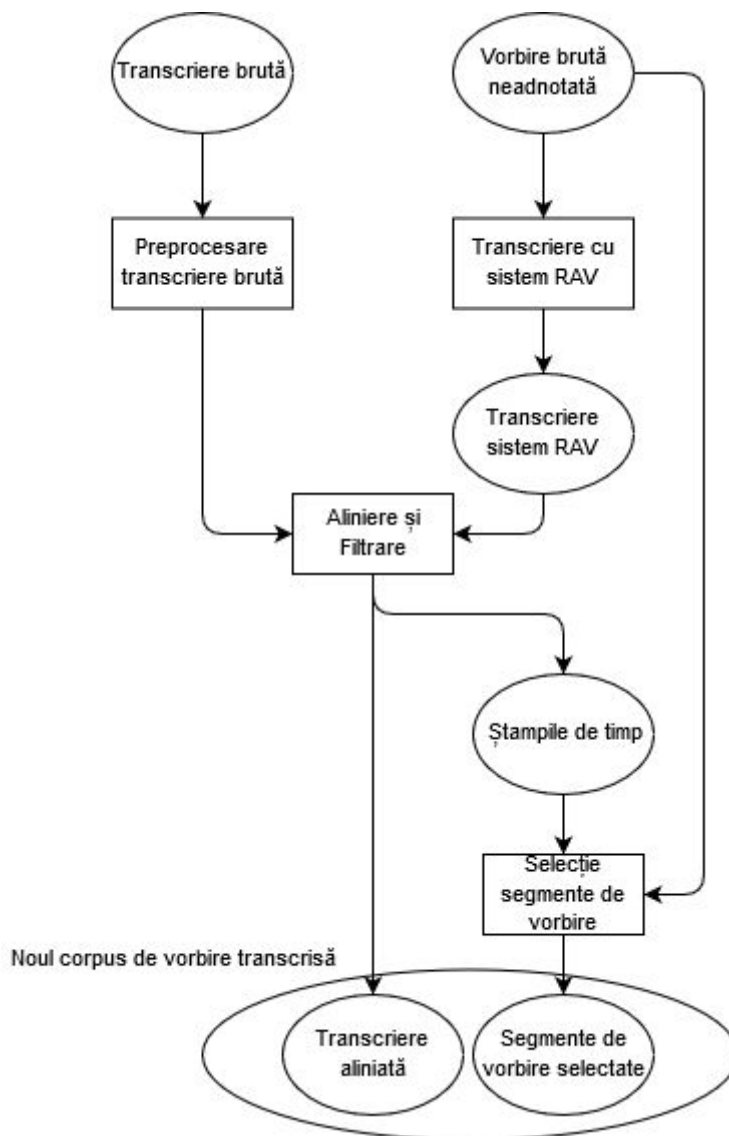
**Preprocesarea transcrierilor brute.** Materialele de pe site (vorbire și transcrieri aproximative) provin din *mass-media* (emisiuni, știri, interviuri, reportaje) și reprezintă o foarte bogată sursă de vorbire și text. Însă, transcrierile brute au o formă diferită față de transcrierile RAV, în sensul că acestea conțin litere mari, numere scrise cu cifre, abrevieri etc. Pentru a efectua procesul de aliniere, acestea trebuie aduse la o formă cât mai apropiată de transcrierile RAV. Astfel, s-au efectuat următoarele operații de preprocesare asupra transcrierilor brute: restaurarea de diacritice, înlocuirea URL-urilor cu forma lor vorbită, înlocuirea numerelor cu text, înlocuirea abrevierilor cu forma lor neabreviată, înlocuirea adreselor de email cu forma lor vorbită, mutarea textelor din paranteză pe linii separate și înlăturarea parantezelor, ștergerea liniilor din alte limbi, înlocuirea literelor mari cu litere mici.

**Alinierea și filtrarea transcrierilor.** Alinierea transcrierilor RAV cu transcrierile brute WEB s-a făcut folosind distanța Levenstein. Această metrică compară 2 secvențe de cuvinte ținând cont de numărul de substituții, inserții și ștergeri dintre cele 2 secvențe. După alinierea celor două transcrieri, selecția părților identice ce urmează să facă parte din corpusul nou de vorbire adnotată s-a făcut pe baza mai multor criterii, după cum urmează. Secvențe consecutive de cuvinte, ce conțin un număr de caractere mai mare decât un prag determinat experimental (8 caractere), sunt considerate a fi corect transcrise. Un alt criteriu utilizat la selecția transcrierilor este durata secvențelor audio, fiind necesar ca aceasta să depășească un anumit prag ales tot empiric (1 secundă). De asemenea, distanța în timp între două cuvinte consecutive este limitată superior la 2 secunde pentru a asigura faptul că nu există cuvinte intermediare netranscrise. În urma

efectuării alinierii și filtrării rezultă setul de transcrieri aliniate cu ștampile de timp, ștampile ce vor fi folosite pentru selecția segmentelor de vorbire corespunzătoare transcrierilor.

**Selecția segmentelor de vorbire.** La final, după ce secvențele de cuvinte corecte au fost selectate, ștampilele de timp asociate acestor cuvinte au fost folosite pentru tăierea secvențelor audio corespunzătoare din datele audio brute.

**Corpus nou de vorbire adnotată.** Corpusul nou de vorbire adnotată este format din transcrierile aliniate și segmentele de vorbire corespunzătoare. Corpusul poate fi folosit la reantrenarea sistemului RAV. Detalii privind corpusul obținut la finalul acestei activități au fost prezentate în tabelul 2.1.c.



**Figura 2.2.a.** Procesul de adnotare automată a vorbirii utilizând metoda filtrării și alinierii transcrierilor aproximative cu semnalul de vorbire

### 2.2.2 Evaluarea cantitativă și evaluarea cantitativă metodei

Evaluarea acestei metode de adnotare automată a fost realizată având în vedere cantitatea de vorbire adnotată obținută în urma alinierii, raportat la dimensiunea corpusului audio brut. Calitatea adnotării, măsurabilă în eroarea la nivel de cuvânt (WER) și/ sau caracter (ChER) nu poate fi măsurată din cauza lipsei unor transcrieri de referință. Totuși, presupunem că aceasta (calitatea adnotării) se traduce într-o performanță mai bună a unui sistem de RAV reantrenat pe baza corpusului de vorbire adnotată rezultat. În consecință, în cele ce urmează vom prezenta și rezultatele de transcriere de vorbire (exprimate ca erori la nivel de cuvânt -

WER) ale unor sisteme de RAV antrenate folosind corpusul de vorbire adnotată creat în cadrul acestei activități.

### 2.2.3 Optimizarea procesului de aliniere și filtrare

#### Procedura inițială

Procedura inițială de aliniere a fost următoarea:

1. Transcrierile brute au fost preprocesate prin efectuarea mai multor operații precum: restaurarea de diacritice, înlocuirea URL-urilor cu forma lor vorbită, înlocuirea numerelor cu text, înlocuirea abrevierilor cu forma lor neabreviată etc. Aceste operații au fost necesare pentru a aduce transcrierile brute la o formă cât mai apropiată transcrierilor RAV în vederea alinierii lor ulterioare.
2. Transcrierile RAV au fost reformatate (din forma JSON inițială) în vederea obținerii cuvintelor individuale din transcriere împreună cu timpii de început și de sfârșit din fișierul audio respectiv.
3. Transcrierile brute și preprocesate, obținute la pasul 1, sunt formate într-un mod similar cu cele RAV, introducându-se ștampile de timp (de început și sfârșit) preluate din transcrierile RAV. Inserarea ștampilelor de timp se face numai acolo unde cuvintele din cele două transcrieri se potrivesc; altfel se inserează zerouri.
4. Execuția programului de aliniere și filtrare, ce folosește CMU Sphinx (NISTAlign) pentru aliniere, realizează filtrarea și generează scripturi sox menite să taie fișierele audio acolo unde există o potrivire a unei secvențe de cuvinte între transcrierile brute formate și transcrierile RAV formate. Prin potrivire înțelegem că atât cuvintele cât și etichetele lor de timpi de început și sfârșit trebuie să fie identice. Secvența de cuvinte trebuie să aibă un număr minim de caractere, precum și o durată minimă. De asemenea, timpul între 2 cuvinte consecutive nu trebuie să depășească un anumit prag.

#### Rezultate inițiale

În urma aplicării procedurii de aliniere și filtrare prezentată în secțiunea anterioară pe seturile de date brute SSC-train3-raw și SSC-train4-raw, au fost obținute seturile de date denumite SSC-train3-trans-v1 și SSC-train4-trans-v1. Dimensiunile acestora, exprimate în număr de cuvinte, respectiv număr de ore de vorbire și eficiența procesului de adnotare automată, exprimată sub forma procentului de date brute ce au putut fi adnotate, raportat la dimensiunea datelor brute sunt prezentate în Tabelul 2.2.a pentru SSC-train3 și SSC-train4.

**Tabelul 2.2.a** Statistici pentru seturile de date SSC-train3-trans-v1 și SSC-train4-trans-v1

Corpus	Sursa	Durată [# ore]		Eficiență aliniere [% ore]		Dimensiune [# cuvinte]	Eficiență aliniere [% cuvinte]
SSC-train3-trans-v1	radio #1	0,4	19,3	2,1%	14,1%	2.942	9,8%
	TV #1	8,2		15,9%		76.106	21,3%
	TV #2	10,7		16,2%		90.830	11,0%
SSC-train4-trans-v1	radio #1	1,4	116,2	1,8%	15,5%	10.030	8,3%
	TV #1	56,3		17,0%		516.960	23,1%
	TV #2	58,6		15,9%		475.528	11,6%

#### Probleme identificate

După analiza rezultatelor de aliniere s-a identificat următoarea problemă: preluarea timpilor de început și de sfârșit de la pasul 3 se face doar pentru primele cuvinte întâlnite care se potrivesc. De exemplu, dacă în transcrierea RAV avem două apariții ale cuvântului “pe”, iar în cea aproximativă avem patru apariții ale



cuvântului “pe”, primele 2 apariții din transcrierea aproximativă vor prelua timpii de început și sfârșit, ceea ce nu este întotdeauna corect. Iată mai jos un exemplu concret:

**Transcriere brută formatată:**

bărbatul(3.71,4.14) **de(0.0,0.0)** treizeci(4.25,4.55) și(43.61,43.95) **șase(0.0,0.0)**  
**de(0.0,0.0) ani(0.0,0.0)** povestește(5.19,5.66) că(5.66,5.75) muncise(5.75,6.22) toată(6.22,6.62)  
noaptea(6.62,6.99) **și(45.57,45.69)** era(7.11,7.27) foarte(7.27,7.63) obosit(7.63,8.09)

**Transcriere RAV formatată:**

bărbatul(3.71,4.14) de(4.14,4.25) treizeci(4.25,4.55) și(4.55,4.65) șase(4.65,4.93) de(4.93,5.05)  
ani(5.05,5.19) povestește(5.19,5.66) că(5.66,5.75) muncise(5.75,6.22) toată(6.22,6.62)  
noaptea(6.62,6.99) și(6.99,7.11) era(7.11,7.27) foarte(7.27,7.63) obosit(7.63,8.09)

După cum se poate observa, cuvintele “de”, ”șase” și “ani” marcate cu roșu în transcrierea brută formatată nu conțin timp de început și sfârșit, deoarece aceștia au fost preluați de cuvinte întâlnite precedent, dar care nu au fost vorbite în fișierul audio (de exemplu în titlu). De asemenea, cuvântul “și”, marcat cu roșu a primit ștampile de timp de la o altă apariție a respectivului cuvânt în transcrierea RAV. Din cauza acestui fapt, în ultima etapă ce realizează alinierea, nu s-a efectuat alinierea întregii secvențe de cuvinte:

**Aliniere realizată:**

<s> povestește că muncise toată noaptea </s>

<s> era foarte obosit </s>

Textele complete pentru exemplul de mai sus sunt prezentate în ANEXA 1 tabelul A1.

**Soluții la probleme**

În primă fază, s-a modificat procedura de aliniere astfel încât la pasul 3 din procedura inițială timpii de început și sfârșit NU mai sunt preluați de la transcrierile RAV, rămânând doar text fără etichete de timp. Programul de aliniere de la pasul 4 a fost de asemenea modificat în sensul că alinierea între transcrierea aproximativă și transcrierea RAV se face doar pe baza textului din transcrieri și nu pe baza textului împreună cu etichetele de timp ca în procedura inițială. Reluarea exemplului de mai sus pentru alinierea v2:

**Transcriere brută formatată:**

bărbatul(3.71,4.14) de(4.14,4.25) treizeci(4.25,4.55) și(4.55,4.65) șase(4.65,4.93) de(4.93,5.05)  
ani(5.05,5.19) povestește(5.19,5.66) că(5.66,5.75) muncise(5.75,6.22) toată(6.22,6.62)  
noaptea(6.62,6.99) și(6.99,7.11) era(7.11,7.27) foarte(7.27,7.63) obosit(7.63,8.09)

**Transcriere RAV formatată:**

bărbatul de treizeci și șase de ani povestește că muncise toată noaptea și era foarte obosit

**Aliniere realizată:**

<s> bărbatul de treizeci și șase de ani povestește că muncise toată noaptea și era foarte obosit </s>

După cum se poate observa, se obține mai mult text aliniat față de versiunea inițială.

**Rezultate v2**

În urma aplicării procedurii de aliniere și filtrare prezentată în secțiunea anterioară pe seturile de date brute SSC-train3-raw și SSC-train4-raw, au fost obținute seturile de date denumite SSC-train3-trans-v2 și SSC-train4-trans-v2. Dimensiunile acestora, exprimate în număr de cuvinte, respectiv număr de ore de vorbire și eficiența procesului de adnotare automată, exprimată sub forma procentului de date brute ce au putut fi adnotate, raportat la dimensiunea datelor brute sunt prezentate în Tabelul 2.2.b pentru SSC-train3 și SSC-train4.

**Tabelul 2.2.b** Statistici pentru seturile de date SSC-train3-trans-v2 și SSC-train4-trans-v2

Corpus	Sursa	Durată [# ore]		Eficiență aliniere [% ore]		Dimensiune [# cuvinte]	Eficiență aliniere [% cuvinte]
SSC-train3-trans-v2	radio #1	1,0	39,0	5,1%	28,5%	9.036	30,1%
	TV #1	13,2		25,7%		138.762	38,8%
	TV #2	24,7		37,5%		256.270	31,0%
SSC-train4-trans-v2	radio #1	2,8	237,3	3,6%	30,5%	24.946	20,8%
	TV #1	90,5		27,3%		938.958	41,9%
	TV #2	144,0		39,2%		1.463.940	35,6%

### Alte probleme identificate

În urma unei analize a textului obținut la ieșirea procesului de aliniere a fost sesizată o problemă în procesul de aliniere: cuvintele ce apar în transcrierile RAV, dar nu și în transcrierile WEB sunt marcate cu ”\*” și apoi eliminate în alinierea finală. De exemplu, dacă avem:

**RAV:** iese un fum gros din camera alăturată în scurt timp vâlvătaia \* cuprins întreaga casă

**WEB:** iese un fum gros din camera alăturată în scurt timp vâlvătaia A cuprins întreaga casă

Se va obține un singur segment de text aliniat, astfel:

**Seg 1:** iese un fum gros din camera alăturată în scurt timp vâlvătaia cuprins întreaga casă.

Din punctul de vedere al alinierii ar trebui să se obțină 2 segmente separate, astfel:

**Seg 1:** iese un fum gros din camera alăturată în scurt timp vâlvătaia

**Seg 2:** cuprins întreaga casă

Alinierea a 2 secvențe de text ar trebui să genereze segmentele de text în care cele 2 secvențe coincid. În exemplul de mai sus, deoarece sistemul RAV nu a generat la ieșire litera “a”, deși aceasta se regăsește în transcrierea aproximativă de pe site, trebuie închis segmentul curent și căutat următorul pentru a evita apariția cuvintelor care nu sunt rostite în vorbire. Astfel, secvențele audio-text sunt mai precise.

Un exemplu similar cu cel anterior:

**RAV:** pentru a-și mai recupera \*\*\*\* ceva din bunuri

**WEB:** pentru a-și mai recupera CÂTE ceva din bunuri

### Soluții la probleme

Având în vedere lucrurile menționate mai sus, procedura de aliniere din programul de aliniere a fost modificată astfel: cuvintele marcate cu ”\*” în setul de ipoteză NU mai sunt eliminate în alinierea finală. Ele sunt păstrate și utilizate ca marcaje pentru întreruperi de segmente, evitându-se astfel crearea de segmente eronate precum în exemplele de mai sus. În urma acestei noi modificări, ne așteptăm să obținem mai puține cuvinte/ore alinate, dar mai precise.

### Rezultate v3

În Tabelul 2.2.c este prezentat exemplul de aliniere îmbunătățită (v3) comparat cu procedura anterioară (v2).

**Tabelul 2.2.c** Exemplu rezultate alinieri v2, respectiv v3.

<p>&lt;s&gt; bărbatul de treizeci și șase de ani povestește că muncise toată noaptea și era foarte obosit copilul se &lt;/s&gt; (61934940_00)</p> <p>&lt;s&gt; la televizor iar la un moment dat a văzut cum iese un fum gros din camera alăturată în scurt timp <b>vâlvătaia cuprins</b> întreaga casă &lt;/s&gt; (61934940_01)</p> <p>&lt;s&gt; abia trezit din somn omul a scos copilul din casă și a încercat să stingă flăcările cu zăpadă &lt;/s&gt; (61934940_02)</p> <p>&lt;s&gt; suferit arsuri pe față și pe mâini și a fost transportat cu ambulanța la spitalul din găești unde medicii &lt;/s&gt; (61934940_03)</p> <p>&lt;s&gt; refuzat însă să rămână internat și s-a &lt;/s&gt; (61934940_04)</p> <p>&lt;s&gt; acasă pentru a-și mai <b>recupera ceva</b> din bunuri &lt;/s&gt; (61934940_05)</p> <p>&lt;s&gt; pompierii au reușit să înlătore la timp pericolul și au stins focul oamenii nu mai au însă nimic locuiau cu chirie iar acum au rămas și fără haine aparatură și alte bunuri care au ars în incendiu &lt;/s&gt; (61934940_06)</p>
<p>&lt;s&gt; bărbatul de treizeci și șase de ani povestește că muncise toată noaptea și era foarte obosit copilul se &lt;/s&gt; (61934940_00)</p> <p>&lt;s&gt; la televizor iar la un moment dat a văzut cum iese un fum gros din camera alăturată în scurt timp <b>vâlvătaia</b> &lt;/s&gt; (61934940_01)</p> <p>&lt;s&gt; <b>cuprins</b> întreaga casă &lt;/s&gt; (61934940_02)</p> <p>&lt;s&gt; abia trezit din somn omul a scos copilul din casă și a încercat să stingă flăcările cu zăpadă &lt;/s&gt; (61934940_03)</p> <p>&lt;s&gt; suferit arsuri pe față și pe mâini și a fost transportat cu ambulanța la spitalul din găești unde medicii &lt;/s&gt; (61934940_04)</p> <p>&lt;s&gt; refuzat însă să rămână internat și s-a &lt;/s&gt; (61934940_05)</p> <p>&lt;s&gt; acasă pentru a-și mai <b>recupera</b> &lt;/s&gt; (61934940_06)</p> <p>&lt;s&gt; pompierii au reușit să înlătore la timp pericolul și au stins focul oamenii nu mai au însă nimic locuiau cu chirie iar acum au rămas și fără haine aparatură și alte bunuri care au ars în incendiu &lt;/s&gt; (61934940_07)</p>

În exemplul din Tabelul 2.2.c avem următoarele situații:

**RAV:** vâlvătaia \* cuprins întreaga casă

**WEB:** vâlvătaia A cuprins întreaga casă

**RAV:** pentru a-și mai recupera \*\*\*\* ceva din bunuri

**WEB:** pentru a-și mai recupera CÂTE ceva din bunuri

Deoarece cuvintele "a" și "câte" nu se regăsesc în transcrierea RAV, în locul lor ar trebui să existe o întrerupere de segment. În Tabelul 2.2.c, aceste cuvinte sunt marcate cu roșu în alinierea v2 și reprezintă punctul în care ar trebui să se încheie un segment și să înceapă alt segment nou. În cazul primului exemplu, o segmentare corectă ar presupune ca după cuvântul "vâlvătaia" să se încheie primul segment, iar cuvântul "cuprins" să reprezinte începutul unui segment separat. Cuvintele marcate cu verde din alinierea V3 indică o segmentare corectă.

Mai jos prezentăm un alt exemplu de aliniere eronată (în varianta v2), versus aliniere corectă (în varianta v3):

**Transcriere aproximativă de pe web:**

Astfel de situații sunt foarte dese în familiile din România, spun specialiștii. **Mii de victime suferă, deși nu sunt neapărat lovite sau rănite.** A dat cu piciorul unui trai prosper pentru o viață "îneacă" în

alcool. Povestea lui Gelu, bărbatul care a pierdut totul. După doi ani de relație, o femeie spune că a simțit cum lumea se prăbușește cu fiecare telefon, fiecare amenințare din partea partenerului.

**Aliniere V2 segmentul 1:** mii de victime suferă deși nu sunt neapărat lovite sau rănite după doi ani de relație

**Aliniere V3 segmentul 1:** mii de victime suferă deși nu sunt neapărat lovite sau rănite

**Aliniere V3 segmentul 2:** după doi ani de relație

În acest exemplu, segmentele marcate cu albastru reprezintă părțile comune dintre transcrierea RAV și transcrierea WEB; propoziția dintre cele 2 segmente nu a fost rostită în materialul vorbit și de aceea aceasta nu se regăsește în transcrierea RAV. Segmentul marcat cu roșu reprezintă segmentul alăturat în mod eronat de alinierea V2, dar separat corect de alinierea V3.

Mai jos prezentăm un alt exemplu de aliniere eronată (în varianta v2), versus aliniere corectă (în varianta v3):

**Transcriere aproximativă de pe web:**

Proiectul de lege propune ca toate aceste forme de violență psihologică să fie considerate infracțiuni. Pedepsa este închisoarea de la 3 luni la 1 an sau amendă. Specialiștii spun însă că fapta va fi greu de probat în fața instanței. Claudiu Dimitriu, președintele Alianței împotriva abuzurilor: "Cu atât mai periculoasă, pentru că nu apar urme fizice, dar urmele psihice sunt mai greu de vindecat. Este și o vorbă: rană de cuțit se vindecă, dar rana de vorbă e mult mai greu de contracarat."

**Aliniere V2:** an sau amendă specialiștii spun însă că fapta va fi greu de probat în fața instanței cu atât mai

**Aliniere V3:** an sau amendă specialiștii spun însă că fapta va fi greu de probat în fața instanței

Acest exemplu este similar cu cel precedent. În acest caz, segmentul aliniat eronat de alinierea v2 (marcat cu roșu) nu se mai regăsește în alinierea v3, deoarece durata segmentului nou care ar fi trebuit să rezulte nu depășește pragul impus.

#### 2.2.4 Utilizarea transcrierilor aproximative în generarea de date

Rezultatele complete folosind ultima variantă de aliniere sunt prezentate în Tabelul 2.2.d. Comparativ cu numărul de ore aliniat precedent în activitatea 1.13 (compl-2018), în această etapă (trans-v3) s-au aliniat mai puține ore per total (Tabelul 2.1c). Putem observa totuși că pentru setul de date TV #1 s-au aliniat mai multe ore în această etapă, față de seturile TV #2 și radio #1 unde s-a aliniat un număr de ore mai mare în activitatea precedentă (1.13). Pentru setul radio #1 de exemplu, textul de pe pagina web este foarte succint în comparație cu ceea ce se vorbește în materialul de voce; astfel, se poate înțelege de ce numărul de ore aliniat este mai mare în activitatea precedentă ce utiliza 2 sisteme RAV pentru transcrierea materialelor de voce și care efectua apoi alinierea celor 2 seturi de transcrieri obținute, pe când, în această etapă, s-a realizat alinierea transcrierilor obținute cu un singur sistem RAV cu transcrierile brute de pe paginile web care sunt scurte și nu cuprind tot ce se vorbește.

**Tabelul 2.2.d** Statistici pentru seturile de date SSC-train3-trans-v3 și SSC-train4-trans-v3, obținute în urma aplicării metodei alinierii transcrierilor aproximative cu transcrierile RAV

Corpus	Sursa	Durată [# ore]		Eficiență aliniere [% ore]		Dimensiune [# cuvinte]	Eficiență aliniere [% cuvinte]
SSC-train3-trans-v3	radio #1	1,0	37,5	5,0%	27,4%	8.833	29,4%
	TV #1	12,8		25,0%		135.874	38,0%
	TV #2	23,6		35,9%		249.271	30,2%
SSC-train4-trans-v3	radio #1	2,7	228,8	3,5%	29,4%	24.345	20,3%
	TV #1	87,9		26,5%		920.785	41,1%
	TV #2	138,1		37,6%		1.426.334	34,7%

### 2.2.5 RAV utilizând corpusul nou creat

Sistemul RAV bazat pe HMM-DNN folosit anterior în activitatea A1.13 a fost antrenat folosind seturile de date RSC-train și SSC-train, obținând un WER de 2.87% pe setul de evaluare RSC-eval, respectiv 15.87% pe setul de evaluare SSC-eval. Sistemul a fost reantrenat ulterior folosind corpusurile SSC-train3-compl-2018 și SSC-train4-compl-2018 rezultate din activitatea 1.13 împreună cu cele inițiale. Sistemul reantrenat a obținut rezultate puțin mai bune după cum se poate observa în Tabelul 2.2.e. Acesta a avut o îmbunătățire relativă a WER de 8.36% pe setul RSC-eval, respectiv 12.03% pe setul SSC-eval.

Tabelul 2.2.e Performanța sistemelor RAV după reantrenare

Corpus antrenare	Model acustic	WER [%]		Îmbunătățire relativă a WER [%]	
		RSC-eval	SSC-eval	RSC-eval	SSC-eval
RSC-train + SSC-train	HMM-DNN	2.87	15.87	n/a	n/a
+ SSC-train3-compl-2018 + SSC-train4-compl-2018	HMM-DNN	2.63	13.96	8.36	12.03
+ SSC-train3-trans + SSC-train4-trans	HMM-DNN	2.41	12.97	16.03	18.27

Același sistem RAV a fost reantrenat folosind corpusurile obținute în această activitate împreună cu cele deja existente. Noul sistem a obținut un WER de 2.41% pe setul RSC-eval, respectiv 12.97% pe setul SSC-eval. Comparativ, sistemul RAV rezultat din această activitate are o îmbunătățire relativă a WER față de sistemul inițial de 16.02%, pe când sistemul obținut anterior în activitatea 1.13 are o îmbunătățire relativă a WER de 8.36%. În cazul setului SSC-eval, noul sistem a obținut o îmbunătățire relativă a WER de 18.27% față de doar 12.03% a sistemului anterior din activitatea 1.13.

### 2.2.6 Concluzii

În concluzie, s-a prezentat o metodă de aliniere inițială unui set de transcrieri brute obținute de pe site-uri mass-media cu un set de transcrieri obținute prin transcrierea materialelor de voce corespunzătoare transcrierilor brute folosind un sistem RAV. S-au adresat de asemenea și problemele întâlnite pentru procedura inițială și s-au discutat versiunile îmbunătățite ale acesteia. Alinierea V2 (Tabelul 2.2.b) a adus creșteri semnificative în termeni de număr de ore, respectiv număr de cuvinte aliniat comparativ cu procedura de aliniere inițială (Tabelul 2.2.a). Alinierea V3 (Tabelul 2.2.d) în schimb, a obținut rezultate

puțin mai mici față de alinierea V2 (Tabelul 2.2.b), dar, față de transcrierile obținute folosind alinierea V2, transcrierile obținute folosind alinierea V3 nu conțin segmente alăturate în mod eronat.

Din punctul de vedere al utilității corpului de vorbire rezultat din această activitate, tabelul 2.2.e indică faptul că s-au făcut progrese semnificative față de situația de la finalul activității A1.13 din etapa 1/2018. Mai concret, sistemul RAV antrenat cu corpusul obținut în urma acestei activități a obținut rezultate mai bune atât pentru vorbire citită (RSC-eval), cât și pentru vorbire spontană (SSC-eval).

## 2.3 Activitatea 2.12 - Proiectarea și implementarea unei soluții de bază pentru generarea de scoruri de încredere pentru RAV

Majoritatea sistemelor de recunoaștere automată a vorbirii (RAV) oferă pe lângă transcrierea fișierului audio și o secvență de scoruri de încredere. Fiecare scor corespunde unui cuvânt și reprezintă gradul de încredere al sistemului de RAV în transcrierea cuvântului respectiv. Aceste scoruri sunt de obicei între 0 și 1 și pot fi interpretate ca probabilități – cu cât scorul este mai mare cu atât este mai probabil ca transcrierea furnizată să fie corectă.

Pentru a construi baze de date de vorbire într-un mod automat folosim scorurile de încredere astfel: (i) pornim de la un set de vorbire neadnotat pe care îl trecem prin sistemul de RAV pentru a produce o transcriere și secvența aferentă de scoruri de încredere; (ii) transcrierea este filtrată pe baza unui prag  $\tau$  aplicat scorurilor de încredere: dacă un cuvânt are scorul asociat mai mare sau egal cu pragul atunci este păstrat, altfel este ignorat. Repetând acest procedeu pentru fiecare fișier audio din setul de date, construim o nouă bază de date adnotată într-un mod total automat. Pragul  $\tau$  controlează compromisul dintre cantitatea și corectitudinea datelor generate: un prag mic rezultă în multe date, dar incerte din punct de vedere al transcrierilor; invers, un prag mare rezultă în puține date, dar corecte.

În această secțiune prezentăm rezultate experimentale pentru această metodă de generare de date. Începem prin a evalua o variantă a sistemului de RAV care ne permite să obținem scoruri de încredere (și anume, sistemul RAV ce folosește decodarea de tip MBR). Apoi evaluăm scorurile de încredere comparându-le cu transcrieri a căror corectitudine o cunoaștem în prealabil. În final, arătăm rezultate pentru utilizarea metodei pentru sarcina de interes, și anume, generarea de baze de date în mod automat.

### 2.3.1 Evaluarea metodelor de decodare

Sistemul de RAV de bază este construit peste utilitarul Kaldi. Pentru a obține scoruri de încredere în Kaldi este necesar să folosim decodarea de tip Minimum Bayes Risk (MBR); mai precis, script-ul `lattice-to-ctm-conf` cu opțiunea `--decode-mbr true`. Spre deosebire, de metoda de decodare standard (bazată pe script-ul `lattice-best-path`), această metodă obține transcrierea  $w^*$  prin optimizarea unui funcții de risc de tipul următor:

$$w^* = \underset{w'}{\operatorname{argmin}} \sum p(w'|x) L(w, w'),$$

unde  $p(w|x)$  indică probabilitatea unei secvențe de cuvinte  $w$  dat semnalul audio  $x$ , iar  $L(w, w')$  indică distanța Levenshtein între două secvențe de cuvinte.

Primul experiment a constat în a ne asigura că rezultatele pe care le obținem folosind tipul de decodare MBR sunt comparabile cu rezultatele obținute anterior cu decodarea de tip best path. Pentru acest experiment am utilizat seturile de date de testare RSC-eval și SSC-eval. Rezultatele sunt prezentate în tabelul 2.3.a și indică faptul că metoda utilizată pentru decodare nu este critică – obținem performanțe similare cu cele două tipuri de decodare.

**Tabelul 2.3.a** Eroarea la nivel de cuvânt (word error rate; WER) pentru seturile de date RSC-eval și SSC-eval folosind modelul de bază cu două tipuri de decodare: best path și MBR.

Decodare	WER[%] RSC-eval	WER[%] SSC-eval
best path	4.27	19.71
MBR	4.26	19.60

### 2.3.2 Evaluarea calitativă a metodei: corelarea scorurilor de încredere cu corectitudinea transcrierii RAV

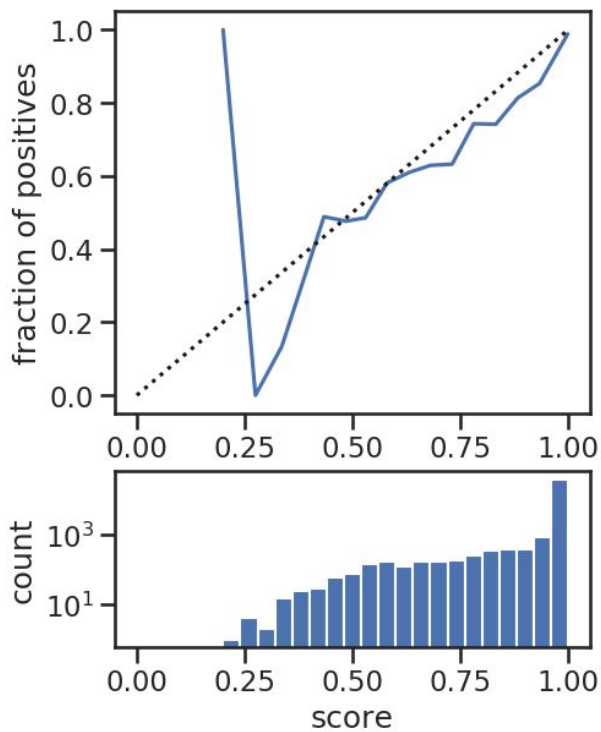
Ideal ne dorim ca scorurile de încredere să reflecte realitatea: valoarea 0 să corespundă cuvintelor transcrise greșit, iar valoarea 1 cuvintelor transcrise corect. Altfel, în cazul în care scorurile sunt în intervalul deschis (0, 1) am vrea ca acestea să reprezinte probabilități *calibrate*: de exemplu, un scor de 0.25 să însemne că un sfert din cuvintele marcate cu scorul de 0.25 sunt corecte, iar restul (de trei sferturi) greșite. În această secțiune oferim o astfel de evaluare a metodei utilizate a scorurilor de încredere. Folosim trei metrici care sunt des întâlnite și în literatura pe acest subiect, vezi de exemplu [Li, 2019]: *normalized cross-entropy* (NCE), *area under the curve precision-recall* (AUC PR), *area under the curve receiver operator characteristics* (AUC ROC); aceste metrici au fost prezentate și descrise în raportul anterior. Rezultatele sunt prezentate în tabelul 2.3.b.

**Tabelul 2.3.b** Evaluarea scorurilor de încredere în funcție de trei metrici pe seturile de date RSC și SSC.  
Pentru toate cele trei metrici valorile mai mari reprezintă rezultate mai bune.

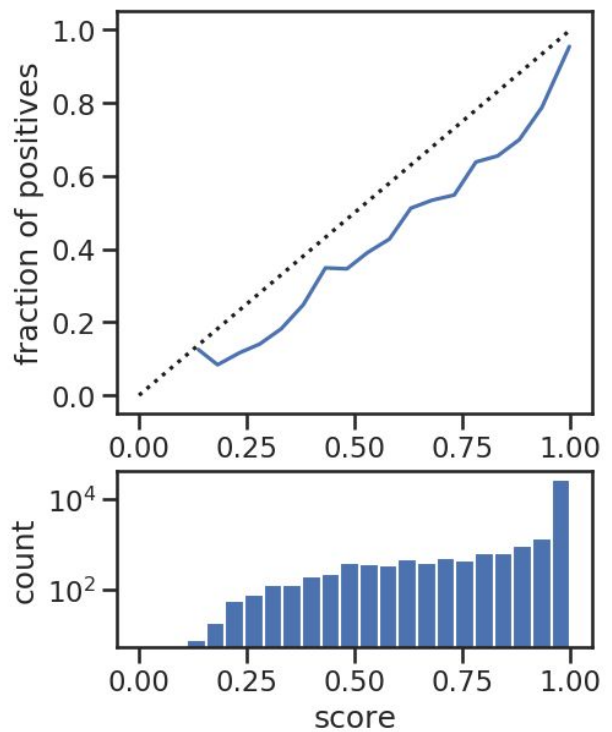
Set de date	NCE ↑	AUC PR ↑	AUC ROC ↑
RSC-eval	-0.1961	0.9953	0.8749
SSC-eval	-0.5604	0.9744	0.8456

Rezultatele pentru calibrarea scorurilor este ilustrată în figura 2.3.a. Sunt de observat trei aspecte:

1. scorurile sunt destul de bine calibrate, valorile situându-se foarte aproape de diagonala de 45 de grade, care indică calibrarea perfectă;
2. majoritatea scorurilor iau valoarea maximă (1.0), cum este indicat de histogramele de sub graficele de calibrare – acest lucru este de așteptat dată performanța bună a sistemelor de RAV;
3. graficul de calibrare pentru SSC atinge un punct de maxim pentru scoruri de aproximativ 0.2 – acesta este doar un caz de zgomot pentru că, după cum se poate observa pe histogramă, este un singur cuvânt care deși a avut un scor mic, de 0.20, este totuși corect; pentru cazurile cu mai multe cuvinte curba de calibrare arată mai neted.

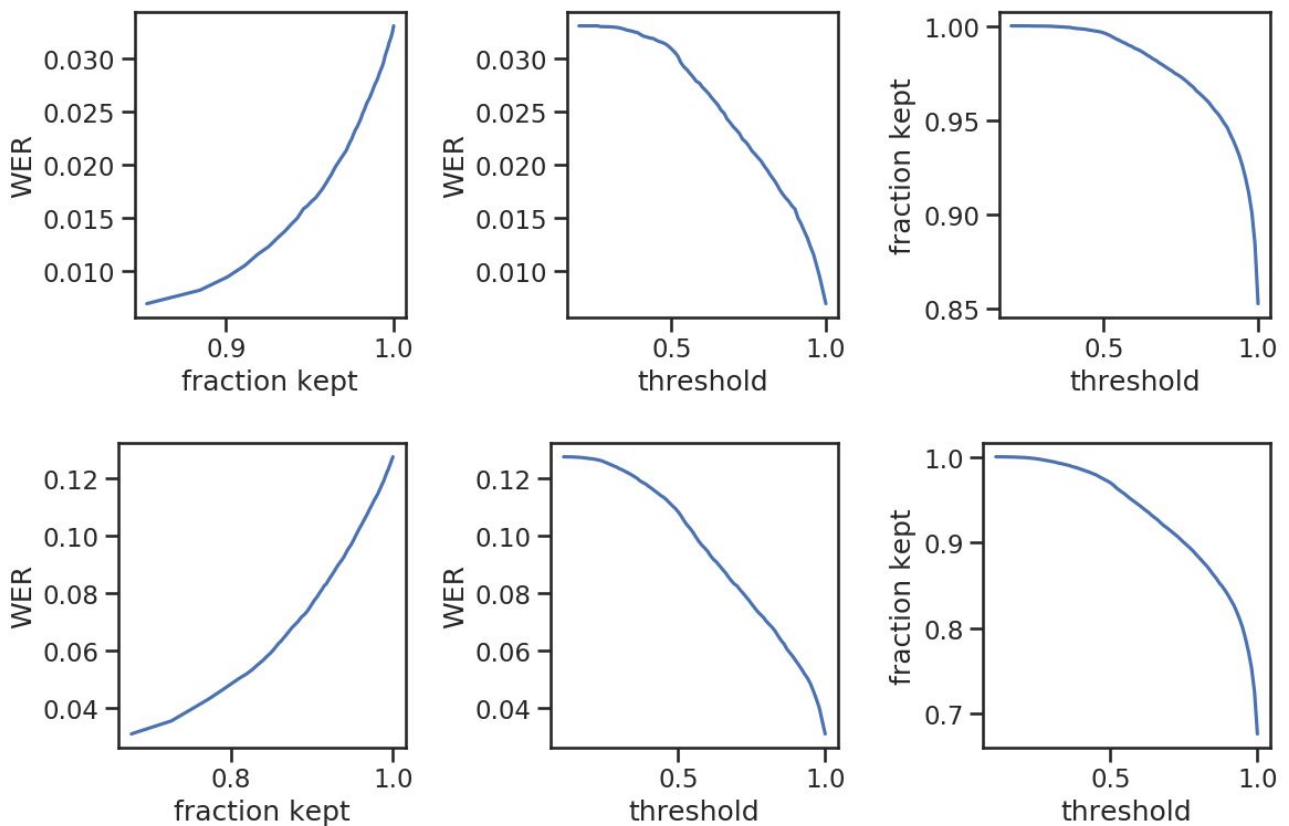


SSC-eval



RSC-eval

**Figura 2.3.a** Grafice de calibrare a scorurilor de încredere pentru cele două seturi de date SSC și RSC. Linia diagonală punctată indică un scor de încredere perfect calibrat, linia albastră indică rezultatele cu scorurile de încredere utilizate. Figurile de dedesubt indică o histogramă a scorurilor de încredere; de remarcat că axa y este logaritmică – majoritatea cuvintelor au un scor de încredere de 1.



**Figura 2.3.b Stânga:** Cantitatea versus calitatea datelor reținute folosind metoda scorurilor de încredere pe setul de date RSC-eval (*sus*) și SSC-eval (*jos*); cantitatea este măsurată în fracția de cuvinte reținute (*fraction*



*kept*), iar calitatea în eroarea la nivel de cuvânt (*WER*). Curba se obține baleind pragul de filtrare  $\tau$  în intervalul  $[0, 1]$ . Ideal ne dorim să ne situăm în colțul din dreapta jos. *Centru*: Calitatea datelor versus valoarea pragului. *Dreapta*: Cantitatea datelor versus valoarea pragului.

De asemenea, oferim rezultate care să arate compromisul între cele două cantități de interes: cantitatea de date păstrate și acuratețea sistemului (word error rate; *WER*). Aceste rezultate sunt prezentate în figura 2.3.b. Observăm că putem transcrie automat peste 90% din setul de date cu erori mai mici de 1% și 8% pentru cele două seturi de date, RSC-eval, respectiv SSC-eval. Aceste rezultate se observă și în tabelul 2.3.c în care prezentăm cantitativ rezultate pentru trei praguri de filtrare care sunt folosite în secțiunea următoare.

**Tabelul 2.3.c** Rezultate numerice pentru calitatea și cantitatea datelor reținute folosind trei praguri de filtrare  $\tau \in \{0.9, 0.95, 1.0\}$ .

Set evaluare	RSC-eval			SSC-eval		
Valoare prag	$\tau = 0.9$	$\tau = 0.95$	$\tau = 1$	$\tau = 0.9$	$\tau = 0.95$	$\tau = 1$
WER [%]	1.58	1.23	0.69	5.66	4.86	3.11
Durată	4h 9m (95.1 %)	4h 4m (93.2 %)	3h 47m (86.7 %)	2h 43m (85.2 %)	2h 36m (81.5 %)	2h 14m (70.1 %)
Număr cuvinte	39 608 (94.6 %)	38 742 (92.5 %)	35 698 (85.3 %)	28 749 (83.9 %)	27 424 (80.0 %)	23 189 (67.6 %)

### 2.3.3 Utilizarea scorurilor de încredere în generarea de date

Aplicând procedura descrisă anterior pe seturi de date pentru care nu avem transcrieri manuale, SSC-train3 și SSC-train4, și utilizând diferite praguri de filtrare  $\tau \in \{0.9, 0.95, 1.0\}$  obținem noi seturi de date; de asemenea, am exclus cuvintele mai scurte de 200 ms. Cantitatea de date rezultată pentru fiecare dintre aceste configurații este descrisă în tabelul 2.3.d. Aceste date sunt apoi utilizate pentru a augmenta setul de date standard și pentru a reantrena sistemul de RAV.

**Tabelul 2.3.d** Cantitatea de date obținute după filtrarea folosind scorurile de încredere. Prezentăm atât valori absolute (în ore *h* și minute *m*), cât și valori relative (în procente %) raportate la cantitatea totală de date.

Filtrarea s-a realizat pe baza a diferite praguri  $\tau \in \{0.9, 0.95, 1.0\}$  – cu cât pragul este mai mare cu atât se obțin mai puține date, dar mai corecte din punctul de vedere al transcrierilor.

Sursa	SSC-train3-conf			SSC-train4-conf		
	$\tau = 0.9$	$\tau = 0.95$	$\tau = 1$	$\tau = 0.9$	$\tau = 0.95$	$\tau = 1$
radio #1	8h 12m (42.5%)	7h 21m (38.1%)	5h 30m (28.5%)	33h 28m (42.8%)	29h 39m (37.9%)	22h 3m (28.2%)
TV #1	27h 58m (54.3%)	25h 15m (49.0%)	19h 7m (37.1%)	181h 40m (54.8%)	164h 24m (49.6%)	124h 29m (37.5%)
TV #2	42h 7m (63.9%)	39h 1m (59.2%)	31h 13m (47.4%)	229h 13m (62.4%)	212h 0m (57.7%)	169h 1m (46.0%)
Total	78h 17m (61.8%)	71h 38m (56.5%)	55h 51m (44.1%)	444h 22m (57.2%)	406h 4m (52.2%)	315h 34m (40.6%)

### 2.3.4 RAV utilizând corpusul nou creat

Rezultatele de bază de la care pornim în această evaluare sunt aceleași care au mai fost prezentate și în tabelul 2.2e:

- rezultatele sistemului RAV inițial (antrenat numai pe seturile de date RSC-train și SSC-train);

- rezultatele sistemului RAV obținut în activitatea A1.13 din etapa anterioară (antrenat pe seturile de date de mai sus și, suplimentar, pe seturile de date SSC-train{3,4}-compl-2018 rezultate în activitatea A1.13).

Același sistem de RAV, bazat pe arhitectura HMM-DNN din toolkit-ul Kaldi, a fost reantrenat folosind ca date de intrare seturile de date RSC-train și SSC-train împreună cu seturile de date SSC-train{3,4}-conf-{090,095,100}, prezentate succint în tabelul 2.3.d. Codurile 090, 095, respectiv 100 reprezintă pragul  $\tau$  ales pentru selecția datelor considerate a fi corecte. Rezultatele celor trei sisteme de RAV sunt prezentate în tabelul 2.3.e.

**Tabelul 2.3.e** Performanța sistemelor RAV după reantrenare

Corpus antrenare	Model acustic	WER [%]		Îmbunătățire relativă a WER [%]	
		RSC-eval	SSC-eval	RSC-eval	SSC-eval
RSC-train + SSC-train	HMM-DNN	2.87	15.87	n/a	n/a
+ SSC-train3-compl-2018 + SSC-train4-compl-2018	HMM-DNN	2.63	13.96	8.36	12.03
+ SSC-train3-conf-090 + SSC-train4-conf-090	HMM-DNN	2.67	14.88	6.97	6.24
+ SSC-train3-conf-095 + SSC-train4-conf-095	HMM-DNN	2.59	15.01	9.76	5.42
+ SSC-train3-conf-100 + SSC-train4-conf-100	HMM-DNN	2.76	14.93	3.83	5.92

Rezultatele experimentale indică mai multe aspecte:

1. metoda prezentată și evaluată mai sus poate fi utilizată pentru generare de date pentru antrenarea RAV, sistemele rezultate obținând rezultate mai bune decât sistemul RAV inițial;
2. metoda sistemelor RAV complementare, evaluată în activitatea A1.13 de anul trecut produce sisteme RAV mai performante decât metoda prezentată și evaluată în această secțiune;
3. este nevoie de metode mai precise de estimare a scorurilor de încredere pentru a produce seturi de date mai corecte, dar și pentru a putea selecta date cu o incertitudine mai mică în vederea reantrenării RAV.

## Referințe

Li, Qiuji, et al. "Bi-Directional Lattice Recurrent Neural Networks for Confidence Estimation." *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019.

## 2.4 Activitatea 2.13 - Îmbunătățirea soluției de adnotare automată a semnalului de vorbire utilizând sisteme de RAV complementare

### 2.4.1 Introducere

Proiectarea și implementarea inițială a metodei de adnotare automată a semnalului de vorbire utilizând sisteme de RAV complementare a făcut obiectul activității A1.13 din etapa 1/2018. Activitatea curentă a vizat dezvoltarea suplimentară și îmbunătățirea metodei folosind ca punct de pornire concluziile activității de anul trecut.

Ideea principală a acestei metode de adnotare automată constă în utilizarea a două sisteme RAV pentru a produce transcrieri pentru un corpus neadnotat, urmând ca apoi transcrierile să fie aliniat, iar părțile identice să fie selectate ca fiind corecte. În final, transcrierile selectate și segmentele de vorbire corespunzătoare sunt folosite pentru a forma un nou corpus adnotat de vorbire.

Pentru ca această metodă să funcționeze este esențial ca cele două sisteme RAV să fie complementare. Mai exact, erorile celor două sisteme RAV trebuie să fie necorelate. Există câteva opțiuni care fac ca acest lucru

să fie posibil: tipurile de modele acustice sau lingvistice să fie diferite, modelele să fie antrenate pe date diferite, algoritmi de decodare să fie diferiți etc.

În cadrul activității A1.13 din etapa 1/2018 au fost utilizate două sisteme de RAV inițiale care difereau prin următoarele caracteristici:

- Tipul modelului acustic (HMM-GMM vs. HMM-DNN);
- Dimensiunea vocabularului (64k cuvinte vs. 200k cuvinte);
- Modelul de limbă folosit la decodare (3-gram vs. 2-gram);
- Utilizarea tehnicii de reevaluare lingvistică (fără reevaluare vs. reevaluare folosind model de limbă 4-gram).

Am arătat atunci că cele două sisteme fac erori diferite, necorelate: practic numai 1.0% - 1.3% din datele adnotate în mod automat cu această metodă sunt adnotate greșit. Restul transcrierilor sunt realizate corect, iar datele nou create pot fi utilizate pentru reantrenarea sistemului de RAV.

Cu toate acestea, experimentele au arătat că datele nou generate ajută foarte puțin la creșterea performanțelor celui mai bun sistem de RAV inițial: eroarea la nivel de cuvânt (WER) a scăzut:

- de la 4.50% la 4.33% pentru vorbire citită și
- de la 20.20% la 18.41% pentru vorbire spontană

Rezultatele sumarizate ale sistemelor de RAV inițiale și ale sistemului de RAV îmbunătățit obținut în A1.13 din etapa anterioară sunt prezentate în tabelul 2.4.a.

Tabelul 2.4.a Performanța sistemelor RAV inițiale și a sistemului RAV îmbunătățit din A1.13, etapa 1.

Model acustic		Model lingvistic	WER [%]		Îmbunătățire relativă a WER [%]	
Corpus antrenare	Tip model		RSC-eval	SSC-eval	RSC-eval	SSC-eval
RSC-train + SSC-train	HMM-GMM	Decodare RAV: 64k cuvinte, 3-gram	12.60	32.30	-	-
RSC-train + SSC-train	HMM-DNN (TDNN2)	Decodare RAV: 200k cuvinte, 2-gram Reev. lingv.: 200k cuvinte, 4-gram	4.50	20.20	-	-
+ SSC-train3-compl-2018 + SSC-train4-compl-2018	HMM-DNN (TDNN2)	Decodare RAV: 200k cuvinte, 2-gram Reev. lingv.: 200k cuvinte, 4-gram	4.33	18.41	3.78	8.86

Pornind de la rezultatele de mai sus, în activitatea A2.13 din această etapă am decis să abordăm următoarele sarcini:

- crearea unor noi sisteme complementare de RAV, cu performanțe inițiale mai bune, similare cu performanțele celui mai bun sistem inițial de anul trecut și
- combinarea mai multor transcrieri de RAV în procesul de aliniere și selecție a transcrierilor cu scopul obținerii mai multor date adnotate.

## 2.4.2 Sisteme de RAV inițiale îmbunătățite

Din punctul de vedere al caracteristicilor cheie și al componentelor constitutive, sistemele RAV inițiale utilizate în A1.13 din etapa anterioară pot fi caracterizate astfel:

- Sistemul RAV #1: creat cu toolkitul CMU Sphinx, model acustic HMM-GMM, model de limbă pentru decodare RAV de tip 3-gram cu vocabular de 64k cuvinte, performanțe scăzute;
- Sistemul RAV #2: creat cu toolkitul Kaldi, model acustic HMM-DNN (TDNN2), model de limbă pentru decodare RAV de tip 2-gram cu vocabular de 200k cuvinte, model de limbă pentru reevaluare lingvistică de tip 4-gram cu vocabular de 200k cuvinte.

În cadrul A2.13 din etapa curentă au mai fost dezvoltate alte două sisteme de RAV cu următoarele caracteristici:

- Sistemul RAV #3: creat cu toolkitul Kaldi, model acustic HMM-DNN (TDNN3), model de limbă pentru decodare RAV de tip 2-gram cu vocabular de 200k cuvinte, model de limbă pentru reevaluare lingvistică de tip RNN cu istorie de 5 cuvinte și vocabular de 200k cuvinte;
- Sistemul RAV #4: creat cu toolkitul NVIDIA OpenSeq2Seq, model acustic și model de limbă pentru decodare integrate într-o singură rețea neurală de tip DeepSpeech, model de limbă pentru reevaluare lingvistică de tip 4-gram cu vocabular de 200k cuvinte.

### **Detalii privind arhitectura TDNN3 și modelul de limbă de tip RNN.**

Rețeaua TDNN3 [Kaldi, 2019] este folosită la antrenarea modelului acustic. Intrarea acesteia este compusă din două tipuri de trăsături: 40 coeficienți mel-cepstrali (MFCC) [Davis and Mermelstein, 1980], extrași din ferestre cu lungimea de 25 ms și 10 ms deplasare, împreună cu i-vectori 100-dimensionali [Dehak, 2010], calculați din segmente de 150 de ferestre consecutive (echivalent cu 1500 ms de vorbire). Trei astfel de vectori MFCC împreună cu i-vectorii corespunzători unui segment sunt concatenați, formând un vector final de trăsături 220-dimensional. Asupra acestuia se aplică analiza discriminatorie liniară (LDA) pentru a decorela componentele, în timp ce dimensionalitatea este păstrată. Rezultatul acestei operații, un vector 220-dimensional, reprezintă intrarea rețelei.

Trunchiul rețelei cuprinde 6 blocuri de tip TDNN [Peddinti, 2015]. În afară de blocul 1 TDNN, care primește la intrare trăsăturile descrise în paragraful anterior, celelalte blocuri TDNN operează asupra ieșirii 540-dimensională a blocului precedent. Fiecare bloc cuprinde o succesiune de operații specifice rețelelor neuronale, cum ar fi transformări afine, activări ReLU și normări la nivel de lot.

Așa cum numele lor sugerează, fiecare bloc cu întârziere în timp (cu excepția primului bloc) efectuează o convoluție temporală 1-D, lucrând cu vectorul de intrare curent, dar și cu unii vectori dinainte sau după acesta. Așadar, blocul 2 TDNN procesează vectorii de intrare de la momentele de timp  $t-1$ ,  $t$  și  $t+1$ . Blocul 3 TDNN procesează vectorii de intrare de la momentele de timp  $t-1$ ,  $t$ ,  $t+1$  și  $t+2$ . Blocurile 4 și 5 procesează vectorii de intrare de la momentele de timp  $t-3$ ,  $t$  și  $t+3$ . Blocul 6 TDNN procesează vectorii de intrare de la momentele de timp  $t-6$ ,  $t-3$  și  $t$ . Blocurile 4, 5 și 6 folosesc tehnica sub-eșantionării: unele cadre temporale sunt excluse în timpul convoluțiilor, permițând rețelei să acceseze un context mai larg. Acest context este format din cadrul de la momentul curent, împreună cu 14 cadre din trecut și 12 cadre din viitor.

Rețeaua dispune de două blocuri de ieșire: primul bazat pe funcția cost LF-MMI [Povey, 2016], în timp ce al doilea utilizează entropia încrucișată. Fiecare bloc de ieșire este compus din straturi ce efectuează transformări afine, activări ReLU, normare la nivel de lot, și din nou transformări afine. Diferența principală dintre aceste două blocuri de ieșire este dată de prezența sau absența stratului final de tip log-softmax. Ambele blocuri sunt folosite la antrenare, în timp ce inferența utilizează numai blocul fără softmax. Ieșirea rețelei are dimensiunea 3760 și reprezintă numărul probabilităților stărilor acustice.

Pe lângă reevaluarea lingvistică efectuată cu model de limbă probabilistic, de tip n-gram, a fost utilizat și un model de tip RNN-LM [Mikolov, 2010], bazat pe rețele neuronale, ce folosește un context de 5 cuvinte. În timp ce un model n-gram învață dintr-un context restrâns și fixat, RNN-LM are capacitatea teoretică de a învăța dintr-un context infinit. Totuși, din motive computaționale, contextul este limitat la câteva cuvinte învecinate. Rețeaua este compusă din 3 blocuri TDNN și conține straturi ce efectuează transformări afine, activări ReLU și normarea loturilor. Între aceste blocuri au fost inserate două straturi de tip LSTM.

### **Detalii privind arhitectura DeepSpeech.**

Intrarea rețelei de tip DeepSpeech [OpenSeq2Seq, 2018] constă în trăsături 160-dimensionale de tip spectrogramă, extrase dintr-un cadru de semnal cu lungimea de 20 ms și 10 ms deplasare. Segmentul de semnal procesat de rețea la o iterație este egal cu 160 de cadre. Arhitectura este considerată a fi recurentă, dar primele straturi din rețea sunt de tip convolutional și au mai mult un rol de preprocesare a semnalului. Prin urmare, primele două straturi efectuează operația de convoluție atât în domeniul timp, cât și în spațiul trăsăturilor. Primul strat aplică 32 de filtre de dimensiune  $41 \times 11$ , cu un pas egal cu 3 și 2, respectiv o bordare ale cărei dimensiuni sunt de 20 și 5, asupra unei intrări de dimensiune  $160 \times 160 \times 1$ . Al doilea strat primește ieșirea de la stratul precedent, un volum de dimensiune  $54 \times 81 \times 32$  și realizează o operație de

convoluție, folosind 32 de filtre de dimensiune 21 x 11. Pasul este egal cu 1 și 2, în timp ce dimensiunea bordării ia valorile 10 și 5. Acest strat scoate la ieșire un volum cu dimensiunea 54 x 41 x 32.

Următoarele 3 straturi sunt straturi recurente bidirecționale. Lungimea secvenței este 41 și corespunde dimensionalității trăsăturilor în urma operațiilor de convoluție. Intrarea primului strat recurent are dimensiunea 3776, fiind egală cu lungimea în timp a segmentului în urma operațiilor de convoluție, 54, înmulțit cu numărul de filtre, 32, la care se adaugă și dimensiunea stratului RNN, 2048. Intrarea celorlalte două straturi recurente este 4096, fiind suma dintre dimensiunea stratului RNN curent și dimensiunea stratului RNN precedent. După fiecare strat se aplică o operație de normare a loturilor. Mecanismul de tip CTC este utilizat în aceasta rețea, acționând cu rol de funcție cost la antrenare, dar fiind util și în procesul de decodare, făcând posibilă ieșirea caracterelor și formarea cuvintelor, fără a fi cunoscute limitele temporale ale fonemelor, mai precis alinierea lor cu ferestrele de semnal.

Performanțele acestor două noi sisteme de RAV sunt prezentate în tabelul 2.4.b. Așa cum se observă sistemul RAV #3 are performanțe net superioare sistemului RAV #2 (cel mai performant sistem RAV inițial din etapa anterioară). Concret, eroarea la nivel de cuvânt (WER) a acestui sistem este de 2.87% pentru vorbire citită (față de 4.50% pentru SRAV #2), respectiv de 15.87% pentru vorbire spontană (față de 20.20% pentru SRAV #2). În plus, în cadrul acestei activități am reantrenat SRAV #3 și cu setul de date generat anul trecut (SSC-train{3,4}-compl-2018), iar sistemul astfel rezultat a fost evaluat și mai bine: eroarea la nivel de cuvânt de 2.63% pentru vorbire citită, respectiv de 13.96% pentru vorbire spontană.

Sistemul de RAV #4 s-a dovedit a avea performanțe foarte slabe. Acesta este chiar mai puțin performant decât sistemul RAV #1, utilizat în A1.13 din etapa anterioară, SRAV bazat pe o tehnologie veche (CMU Sphinx cu modele acustice de timp HMM-GMM). Concluzia pe care o putem trage din acest experiment este că tehnologia de RAV de tip end-to-end (model acustic și model de limbă integrate într-o singură rețea neurală profundă de tip sequence-to-sequence) nu este încă suficient de matură pentru a putea fi utilizată în practică. Dat fiind această concluzie, SRAV #4 nu a mai fost utilizat în continuare în această activitate. Nu a fost evaluată nici complementaritatea lui față de celelalte SRAV inițiale și nici nu a fost folosit pentru a genera noi seturi de date adnotate automat.

## Referințe

[Kaldi, 2019]:  
[https://github.com/kaldi-asr/kaldi/blob/master/egs/tedlium/s5\\_r2/local/chain/tuning/run\\_tdnns\\_1d.sh](https://github.com/kaldi-asr/kaldi/blob/master/egs/tedlium/s5_r2/local/chain/tuning/run_tdnns_1d.sh)

[Davis and Mermelstein 1980] Davis, Steven, and Paul Mermelstein. "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences." *IEEE transactions on acoustics, speech, and signal processing* 28.4 (1980): 357-366.

[Dehak, 2010] Dehak, Najim, et al. "Front-end factor analysis for speaker verification." *IEEE Transactions on Audio, Speech, and Language Processing* 19.4 (2010): 788-798.

[Mikolov, 2010] Mikolov, Tomáš, et al. "Recurrent neural network based language model." *Eleventh annual conference of the international speech communication association*. 2010.

[OpenSeq2Seq, 2018]:  
[https://github.com/NVIDIA/OpenSeq2Seq/blob/master/example\\_configs/speech2text/ds2\\_small\\_1gpu.py](https://github.com/NVIDIA/OpenSeq2Seq/blob/master/example_configs/speech2text/ds2_small_1gpu.py)

[Peddinti, 2015] Peddinti, Vijayaditya, Daniel Povey, and Sanjeev Khudanpur. "A time delay neural network architecture for efficient modeling of long temporal contexts." *Sixteenth Annual Conference of the International Speech Communication Association*. 2015.

[Povey, 2016] Povey, Daniel, et al. "Purely sequence-trained neural networks for ASR based on lattice-free MMI." *Interspeech*. 2016.

**Tabelul 2.4.b** Performanța sistemelor RAV inițiale din A2.13, etapa 2/2019. Performanța sistemului RAV inițial reantrenat folosind și setul de date generat în cadrul A1.13 din etapa 1/2018.

Model acustic		Model lingvistic	WER [%]		Îmbunătățire relativă a WER [%]	
Corpus antrenare	Tip model		RSC-eval	SSC-eval	RSC-eval	SSC-eval
RSC-train + SSC-train	HMM-DNN (TDNN3)	Decodare RAV: 200k cuvinte, 2-gram Reev. lingv.: RNN 5-gram	2.87	15.87	-	-
RSC-train + SSC-train	DeepSpeech	Decodare RAV: integrat DeepSpeech Reev. lingv.: 200k cuvinte, 4-gram	15.12	43.61	-	-
+ SSC-train3-compl-2018 + SSC-train4-compl-2018	HMM-DNN (TDNN3)	Decodare RAV: 200k cuvinte, 2-gram Reev. lingv.: RNN 5-gram	2.63	13.96	9.67	21.65

### 2.4.3 Evaluarea calitativă a metodei: complementaritatea sistemelor de RAV inițiale

În contextul situației descrise mai sus (renunțarea la SRAV #4 din motive de performanță scăzută), sistemele de RAV inițiale au fost considerate ca fiind SRAV #1, SRAV #2 și SRAV #3. Complementaritatea perechii de sisteme (SRAV #1 - SRAV #2) a fost evaluată în etapa anterioară. S-a demonstrat atunci că aplicarea metodei folosind cele două sisteme RAV ca sisteme inițiale conduce la adnotarea automată a 48%, respectiv 20% din seturile de date RSC-eval, respectiv SSC-eval și că eroarea de adnotare se plasează în gama 1.0% - 1.3%. Aceste rezultate au fost reluate în tabelul 2.4.c.

În continuare, în această etapă a fost evaluată complementaritatea perechii de sisteme (SRAV #2 - SRAV #3). Rezultatele evaluării complementarității și implicit a eficienței și calității metodei de adnotare automată sunt sumarizate, pentru comparație, tot în tabelul 2.4.c. Se poate observa că sistemele analizate sunt mai asemănătoare: ele generează transcrieri similare și, implicit, fac și mai multe greșeli identice. Acest lucru este indicat de eroarea la nivel de cuvânt mai mare (2.6%, respectiv 2.7%) comparativ cu eroarea la nivel de cuvânt obținută pentru perechea de sisteme SRAV #1 - SRAV #2. Pe de altă parte, cantitatea de date selectată prin aplicarea metodei folosind sistemele SRAV #2 și #3 ca sisteme inițiale este semnificativ mai mare (79%, respectiv 73%) comparativ cu situația de anul trecut (48%, respectiv 20%).

Putem concluziona că utilizând perechea de SRAV #2 + #3 reușim să adnotăm automat o cantitate de date 2 ori mai mare, cu o eroare de adnotare de aproximativ 2 ori mai mare. În ce măsură acest lucru este benefic se va vedea în experimentele ulterioare.

**Tabelul 2.4.c** Calitatea și cantitatea datelor obținute prin aplicarea metodei folosind ca SRAV inițiale perechile (SRAV #1 - SRAV #2), respectiv (SRAV #2 - SRAV #3).

Set evaluare	RSC-eval		SSC-eval	
	SRAV #1 - SRAV #2	SRAV #2 - SRAV #3	SRAV #1 - SRAV #2	SRAV #2 - SRAV #3
Pereche SRAV				
WER [%]	1.0	2.6	1.30	2.7
ChER [%]	0.3	0.7	0.4	1.0
Durată	2h, 37 m (48 %)	4 h, 14 m (79 %)	0h, 41 m (20 %)	2 h, 33 m (73 %)

### 2.4.4 Utilizarea sistemelor RAV complementare în generarea de date

Aplicând metoda descrisă anterior pe seturile de date pentru care nu avem transcrieri manuale (SSC-train3-raw și SSC-train4-raw) obținem noi seturi de date. Cantitatea de date rezultată pentru fiecare set de date în parte și fiecare sursă de date din fiecare set este descrisă în tabelul 2.4.d. Aceste date sunt apoi utilizate pentru a augmenta setul de date inițial și pentru a reantrena sistemul de RAV.

**Tabelul 2.4.d** Statistici pentru seturile de date SSC-train3-compl-2019 și SSC-train4-compl-2019, obținute în urma aplicării metodei adnotării automate folosind sisteme de RAV complementare.

Corpus	Sursa	Durată [# ore]		Eficiență aliniere [% ore]	
SSC-train3-compl-2019	radio #1	9h, 51m	79h, 52m	51.0%	58.4%
	TV #1	27h, 52m		54.0%	
	TV #2	42h, 09m		63.9%	
SSC-train4-compl-2019	radio #1	36h, 00m	452h, 32m	46.0%	58.2%
	TV #1	181h, 42m		54.7%	
	TV #2	234h, 50m		63.9%	

### 2.4.5 RAV utilizând corpusul nou creat

După obținerea seturilor de date SSC-train{3,4}-compl-2019, sistemul de RAV inițiale mai performant (SRAV #3) a fost reantrenat folosind seturile de date inițiale (RSC-train și SSC-train) împreună cu aceste seturi de date nou create. Rezultatele evaluării sistemelor nou create sunt prezentate în tabelul 2.4.e.

**Tabelul 2.4.e** Performanța sistemelor RAV inițiale vs. performanța sistemelor de RAV după reantrenare

Model acustic		Model linvistic	WER [%]		Îmbunătățire relativă a WER [%]	
Corpus antrenare	Tip model		RSC-eval	SSC-eval	RSC-eval	SSC-eval
RSC-train + SSC-train	HMM-DNN (TDNN2)	Decodare RAV: 200k cuvinte, 2-gram Reev. lingv.: 200k cuvinte, 4-gram	4.50	20.20	-	-
RSC-train + SSC-train	HMM-DNN (TDNN3)	Decodare RAV: 200k cuvinte, 2-gram Reev. lingv.: RNN 5-gram	2.87	15.87	-	-
+ SSC-train3-compl-2018 + SSC-train4-compl-2018	HMM-DNN (TDNN3)	Decodare RAV: 200k cuvinte, 2-gram Reev. lingv.: RNN 5-gram	2.63	13.96	8.4	12.0
+ SSC-train3-compl-2019 + SSC-train4-compl-2019	HMM-DNN (TDNN3)	Decodare RAV: 200k cuvinte, 2-gram Reev. lingv.: RNN 5-gram	2.52	13.22	12.2	16.7

Rezultatele experimentale indică mai multe aspecte:

1. metoda prezentată și evaluată mai sus poate fi utilizată pentru generare de date pentru antrenarea RAV, sistemele rezultate obținând rezultate mai bune decât sistemul RAV inițial;
2. sistemul SRAV #3 reantrenat (cu SSC-train{3,4}-compl-2019) este mai performant decât sistemul reantrenat în anul 2018, folosind seturile de date SSC-train{3,4}-compl-2018.

## 2.5 Activitatea 2.14 - Diseminare

Diseminarea rezultatelor proiectului a fost realizată: în cadrul consorțiului în workshopul organizat la Cluj-Napoca pe 18 noiembrie 2019 și în comunitatea științifică la trei conferințe internaționale de prestigiu: 42nd International Conference on Telecommunications and Signal Processing, 10th Conference on Speech Technology and Human-Computer Dialogue și 14th International Conference on Linguistics Resources and Tools for Natural Language Processing. Suplimentar, unele dintre rezultate au fost publicate într-un articol în Buletinul Științific al Universității Politehnica din București. De asemenea, progresul înregistrat în această etapă a fost diseminat prin intermediul website-ului proiectului: <https://tadarav.speed.pub.ro>.

Dintre publicațiile menționate mai jos, articolele 1 și 4 sunt deja indexate în Web of Science (Thompson Reuters - ISI), articolele 2 și 3 sunt deja indexate IEEE Xplore și în curs de indexare în Web of Science

(Thompson Reuters - ISI), iar articolul 5 a apărut în volumul conferinței și este în curs de indexare în Web of Science (Thompson Reuters - ISI). În toate aceste articole numele finanțatorului este menționat în secțiunea Acknowledgement, conform indicațiilor din contractul de finanțare.

Lista completă a publicațiilor din etapa 2/2019 este următoarea:

1. Alexandru-Lucian Georgescu, Horia Cucu, Corneliu Burileanu, "[Progress on automatic annotation of speech corpora using complementary ASR systems](#)," in the Proceedings of the 42nd International Conference on Telecommunications and Signal Processing (TSP), 2019, Budapest, Hungary.
2. Gheorghe Pop, Serban Mihalache, Dragos Burileanu, "[Forensic Recognition of Narrowband AMR Signals](#)," in the Proceedings of the 10th Conference on Speech Technology and Human-Computer Dialogue (SpeD), Timișoara, Romania, 2019.
3. Alexandru-Lucian Georgescu, Horia Cucu, Corneliu Burileanu, "[Kaldi-based DNN architectures for speech recognition in Romanian](#)," in the Proceedings of the 10th Conference on Speech Technology and Human-Computer Dialogue (SpeD), Timișoara, Romania, 2019.
4. Gheorghe Pop and Dragos Burileanu, "[Speech Enhancement for Forensic Purposes](#)," in UPB Scientific Bulletin, Series C, Vol. 81, Issue 3, pp. 41-52, 2019.
5. Florin Iordache, Alexandru-Lucian Georgescu, Dan Oneață, Horia Cucu, "Romanian Automatic Diacritics Restoration Challenge", in the Proceedings of the 14th International Conference on Linguistics Resources and Tools for Natural Language Processing, Cluj-Napoca, Romania, 2019.

### 3 Structura ofertei de servicii de cercetare și tehnologice

Laboratorul de cercetare *Speech and Dialogue* (SpeD) din cadrul Universității Politehnica din București (UPB), reprezentantul UPB în proiectul TADARAV, oferă pe platforma ERRIS serviciile de cercetare și tehnologice enumerate în Tabelul 3.

**Tabelul 3.** Servicii de cercetare și tehnologice oferite de Laboratorul de cercetare *Speech and Dialogue*

Serviciu	Detalii
Serviciu și aplicație web de transcriere de documente ce conțin vorbire în limba română	<a href="https://transcriptions.speed.pub.ro">https://transcriptions.speed.pub.ro</a>
Serviciu și aplicație web de identificare de cuvinte cheie în documente ce conțin vorbire în limba română	<a href="https://keywords.speed.pub.ro">https://keywords.speed.pub.ro</a>
Serviciu și aplicație web de restaurare de diacritice în limba română	<a href="https://diacritics.speed.pub.ro">https://diacritics.speed.pub.ro</a>
Proiectarea și implementarea de aplicații personalizate de transcriere a vorbirii continue	La cerere
Proiectarea și implementarea de aplicații personalizate de identificare de cuvinte și termeni de interes	La cerere
Proiectarea și implementarea de aplicații personalizate de sinteză de vorbire pornind de la text	La cerere
Proiectarea și implementarea de sisteme de recunoaștere de pattern-uri folosind inteligență artificială	La cerere

Laboratorul de cercetare *Speech and Dialogue* (SpeD) este prezent pe platforma ERRIS la adresa <https://erris.gov.ro/Speed---UPB>.

### 4 Locuri de muncă susținute prin program

Echipele de cercetare a Universității Politehnica din București pentru proiectul component TADARAV este prezentată în Tabelul 4.



**Tabelul 4. Echipa de cercetare UPB**

<b>Nr.</b>	<b>Nume</b>	<b>Calitatea</b>	<b>Poziția</b>	<b>Normă</b>
1	Horia CUCU	Conf. Univ.	Responsabil proiect component	Parțială
2	Corneliu BURILEANU	Prof. Univ.	Membru cercetător	Parțială
3	Dragoș BURILEANU	Prof. Univ.	Membru cercetător	Parțială
4	Alexandru-Lucian GEORGESCU	ACS	Membru cercetător	Parțială
5	Dan Theodor ONEAȚĂ	CS	<b>Membru cercetător nou</b>	<b>Întreagă</b>
6	Gheorghe POP	ACS	<b>Membru cercetător nou</b>	<b>Întreagă</b>
7	Cristian MANOLACHE	ACS	<b>Membru cercetător nou</b>	<b>Întreagă</b>

## **5 Valorificarea și îmbunătățirea competențelor și resurselor existente la nivelul consorțiului**

La nivelul proiectului component TADARAV CEC-urile nu au fost valorificate.

## 6 Anexe

Tabelul A1. Exemplu de aliniere a transcrierii RAV (prima celula) cu transcrierea aproximativă (a doua celula) și rezultatul alinierii (a patra celula). Celula a treia prezintă transcrierea aproximativă formatată în vederea alinierii.

fure(0.71,1.24) și(1.24,1.38) un(1.38,1.55) master(1.55,1.96) cu(1.96,2.24) bărbatul(3.71,4.14)  
de(4.14,4.25) treizeci(4.25,4.55) și(4.55,4.65) șase(4.65,4.93) de(4.93,5.05) ani(5.05,5.19)  
povestește(5.19,5.66) că(5.66,5.75) muncise(5.75,6.22) toată(6.22,6.62) noaptea(6.62,6.99)  
și(6.99,7.11) era(7.11,7.27) foarte(7.27,7.63) obosit(7.63,8.09) copilul(8.23,8.64) se(8.64,8.75)  
uita(8.75,8.95) la(8.95,9.04) televizor(9.04,9.59) iar(9.59,9.76) la(9.76,9.86) un(9.86,9.96)  
moment(9.96,10.23) dat(10.23,10.4) a(10.4,10.47) văzut(10.47,10.84) cum(10.84,11.03)  
iese(11.03,11.34) un(11.34,11.53) fum(11.53,11.8) gros(11.8,12.15) din(12.15,12.32)  
camera(12.32,12.65) alăturată(12.65,13.23) în(13.46,13.66) scurt(13.66,14.11)  
timp(14.11,14.35) vâlvătaia(14.35,14.95) cuprins(14.95,15.32) întreaga(15.32,15.72)  
casă(15.72,16.21) per(16.35,16.57) elev(16.57,16.93) și(17.76,18.02) volumic(18.02,18.6)  
fine(18.6,18.97) o(18.97,19.34) materii(19.63,20.23) și(20.26,20.39) cu(20.39,20.51)  
toată(20.51,20.81) tăria(20.81,21.18) lor(21.18,21.36) de(21.36,21.53) ce(22.81,23.11)  
vedem(23.14,23.41) ori(23.41,23.99) iese(24.23,24.57) fum(24.57,24.89) fum(25.24,25.7)  
nu(25.7,25.83) e(25.83,25.92) rețezi(25.92,26.56) și(26.56,26.74) rama(26.74,26.97)  
unei(26.97,27.25) morți(27.25,27.72) fără(28.33,29.23) fund(29.63,30.09) în(30.09,30.17)  
sus(30.17,30.56) abia(31.11,31.46) trezit(31.46,31.8) din(31.8,31.98) somn(31.98,32.31)  
omul(32.34,32.6) a(32.6,32.66) scos(32.66,32.99) copilul(32.99,33.38) din(33.38,33.54)  
casă(33.54,33.86) și(33.86,34.0) a(34.0,34.04) încercat(34.04,34.48) să(34.48,34.6)  
stingă(34.6,35.06) flăcările(35.06,35.68) cu(35.71,35.85) zăpadă(35.85,36.57) foto(36.7,37.33)  
suferit(37.33,37.7) arsuri(37.7,38.06) pe(38.06,38.18) față(38.18,38.57) și(38.57,38.71)  
pe(38.71,38.85) mâini(38.85,39.16) și(39.16,39.25) a(39.25,39.3) fost(39.3,39.5)  
transportat(39.5,39.99) cu(39.99,40.1) ambulanța(40.1,40.65) la(40.65,40.75)  
spitalul(40.75,41.23) din(41.23,41.4) găești(41.4,41.75) unde(41.75,41.93) medicii(41.93,42.37)  
au(42.37,42.49) cordat(42.49,42.93) îngrijiri(42.93,43.58) și(43.61,43.95) l-a(43.95,44.08)  
refuzat(44.08,44.47) însă(44.47,44.66) să(44.66,44.79) rămână(44.79,45.05)  
internat(45.05,45.57) și(45.57,45.69) s-a(45.69,45.89) reîntors(45.89,46.37) acasă(46.37,46.79)  
pentru(46.79,47.13) a-și(47.13,47.27) mai(47.27,47.42) recupera(47.42,47.86)  
ceva(47.86,48.13) din(48.13,48.3) bunuri(48.3,48.63) și(48.63,48.81) flamanzi(48.84,50.01)  
nu(50.01,51.23) fi(52.54,52.7) pur(52.81,53.05) și(53.05,53.16) simplu(53.16,53.5)  
fiica(53.5,53.89) nu(53.97,56.69) fac(56.72,57.14) dar(58.33,58.55) a(59.1,59.19)  
venit(59.19,59.42) foarte(60.06,60.41) repede(60.41,60.71) pompierii(61.17,61.64)  
au(61.64,61.74) reușit(61.74,62.12) să(62.12,62.25) înlătore(62.25,62.63) la(62.63,62.75)  
timp(62.75,63.04) pericolul(63.04,63.55) și(63.55,63.67) au(63.67,63.77) stins(63.77,64.1)  
focul(64.1,64.45) oamenii(64.71,65.23) nu(65.23,65.4) mai(65.4,65.6) au(65.6,65.7)  
însă(65.7,65.96) nimic(65.96,66.4) locuiau(66.44,66.89) cu(66.89,67.02) chirie(67.02,67.38)  
iar(67.38,67.53) acum(67.53,67.73) au(67.73,67.84) rămas(67.84,68.19) și(68.19,68.36)  
fără(68.36,68.56) haine(68.56,68.92) aparatură(68.95,69.47) și(69.47,69.63) alte(69.63,69.89)  
bunuri(69.89,70.19) care(70.19,70.46) au(70.49,70.74) ars(70.77,71.1) în(71.1,71.16)  
incendiu(71.16,71.71)

Un copil de 6 ani și-a salvat tatăl de la moarte, după ce locuința lor a fost cuprinsă de flăcări

Ultimul update: Miercuri 17 Ianuarie 2018 17:54

Data publicării: Miercuri 17 Ianuarie 2018 16:56

+ 0

Caz impresionant, miercuri dimineață, în Găești, județul Dâmbovița. Un copil de 6 ani și-a salvat tatăl de la moarte, după ce locuința lor a fost cuprinsă de flăcări.

Omul, sudor de meserie, venise din schimbul trei și dormea adânc. Băiatul se uita liniștit la televizor și când a simțit miros de fum și a văzut flăcările nu a fugit din casă, ci și-a strigat tatăl până l-a trezit.

Vecinii au auzit și ei țipetele băiatului, și-au dat seama de pericol și au chemat imediat pompierii. Bărbatul, de 36 de ani, povestește că muncise toată noaptea și era foarte obosit. Copilul se uita la televizor, iar la un moment dat a văzut cum iese un fum gros din camera alăturată. În scurt timp vâlvătaia a cuprins întreaga casă.

Abia trezit din somn, omul a scos copilul din casă și a încercat să stingă flăcările cu zăpadă.

Citește și

Familie din Bistrița, pe drumuri după ce un incendiu s-a extins și la casa lor

Bărbatul a suferit arsuri pe față și pe mâini și a fost transportat cu ambulanța la spitalul din Găești, unde medicii i-au acordat îngrijiri. A refuzat însă să rămână internat și s-a întors acasă pentru a-și mai recupera câte ceva din bunuri.

Pompierii au reușit să înlăture la timp pericolul și au stins focul. Oamenii nu mai au însă nimic. Locuiau cu chirie, iar acum au rămas și fără haine, aparatură și alte bunuri, care au ars în incendiu.

Șocant

O bătrână a fost găsită fără viață în casa cuprinsă de foc: "Avea obiceiul să își aprindă tămâie"

Top citite acum

un(1.38,1.55) copil(0.0,0.0) de(4.14,4.25) șase(4.65,4.93) ani(5.05,5.19) și-a(0.0,0.0) salvat(0.0,0.0) tatăl(0.0,0.0) de(4.93,5.05) la(8.95,9.04) moarte(0.0,0.0) după(0.0,0.0) ce(22.81,23.11) locuința(0.0,0.0) lor(21.18,21.36) a(10.4,10.47) fost(39.3,39.5) cuprinsă(0.0,0.0) de(21.36,21.53) flăcări(0.0,0.0) ultimul(0.0,0.0) update(0.0,0.0) miercuri(0.0,0.0) șaptesprezece(0.0,0.0) ianuarie(0.0,0.0) două(0.0,0.0) mii(0.0,0.0) optsprezece(0.0,0.0) șaptesprezece(0.0,0.0) și(1.24,1.38) cincizeci(0.0,0.0) și(4.55,4.65) patru(0.0,0.0) de(0.0,0.0) minute(0.0,0.0) data(0.0,0.0) publicării(0.0,0.0) miercuri(0.0,0.0) șaptesprezece(0.0,0.0) ianuarie(0.0,0.0) două(0.0,0.0) mii(0.0,0.0) optsprezece(0.0,0.0) șaisprezece(0.0,0.0) și(6.99,7.11) cincizeci(0.0,0.0) și(17.76,18.02) șase(0.0,0.0) de(0.0,0.0) minute(0.0,0.0) caz(0.0,0.0) impresionant(0.0,0.0) miercuri(0.0,0.0) dimineață(0.0,0.0) în(13.46,13.66) găești(41.4,41.75) județul(0.0,0.0) dâmbovița(0.0,0.0) un(9.86,9.96) copil(0.0,0.0) de(0.0,0.0) șase(0.0,0.0) ani(0.0,0.0) și-a(0.0,0.0) salvat(0.0,0.0) tatăl(0.0,0.0) de(0.0,0.0) la(9.76,9.86) moarte(0.0,0.0) după(0.0,0.0) ce(0.0,0.0) locuința(0.0,0.0) lor(0.0,0.0) a(32.6,32.66) fost(0.0,0.0) cuprinsă(0.0,0.0) de(0.0,0.0) flăcări(0.0,0.0) omul(32.34,32.6) sudor(0.0,0.0) de(0.0,0.0) meserie(0.0,0.0) venise(0.0,0.0) din(12.15,12.32) schimbul(0.0,0.0) trei(0.0,0.0) și(20.26,20.39) dormea(0.0,0.0) adânc(0.0,0.0) băiatul(0.0,0.0) se(8.64,8.75) uită(0.0,0.0) liniștit(0.0,0.0) la(40.65,40.75) televizor(9.04,9.59) și(26.56,26.74) când(0.0,0.0) a(34.0,34.04) simțit(0.0,0.0) miros(0.0,0.0) de(0.0,0.0) fum(11.53,11.8) și(33.86,34.0) a(39.25,39.3) văzut(10.47,10.84) flăcările(35.06,35.68) nu(25.7,25.83) a(59.1,59.19) fugit(0.0,0.0) din(31.8,31.98) casă(15.72,16.21) ci(0.0,0.0) și-a(0.0,0.0) strigat(0.0,0.0) tatăl(0.0,0.0) până(0.0,0.0) l-a(43.95,44.08) trezit(31.46,31.8) vecinii(0.0,0.0) au(42.37,42.49) auzit(0.0,0.0) și(38.57,38.71) ei(0.0,0.0) țipetele(0.0,0.0) băiatului(0.0,0.0) și-au(0.0,0.0) dat(10.23,10.4) seama(0.0,0.0)

de(0.0,0.0) pericol(0.0,0.0) și(39.16,39.25) au(61.64,61.74) chemat(0.0,0.0) imediat(0.0,0.0) pompierii(61.17,61.64) bărbatul(3.71,4.14) de(0.0,0.0) treizeci(4.25,4.55) și(43.61,43.95) șase(0.0,0.0) de(0.0,0.0) ani(0.0,0.0) povestește(5.19,5.66) că(5.66,5.75) muncise(5.75,6.22) toată(6.22,6.62) noaptea(6.62,6.99) și(45.57,45.69) era(7.11,7.27) foarte(7.27,7.63) obosit(7.63,8.09) copilul(8.23,8.64) se(0.0,0.0) uită(0.0,0.0) la(62.63,62.75) televizor(0.0,0.0) iar(9.59,9.76) la(0.0,0.0) un(11.34,11.53) moment(9.96,10.23) dat(0.0,0.0) a(0.0,0.0) văzut(0.0,0.0) cum(10.84,11.03) iese(11.03,11.34) un(0.0,0.0) fum(24.57,24.89) gros(11.8,12.15) din(33.38,33.54) camera(12.32,12.65) alăturată(12.65,13.23) în(30.09,30.17) scurt(13.66,14.11) timp(14.11,14.35) vâlvătaia(14.35,14.95) a(0.0,0.0) cuprins(14.95,15.32) întreaga(15.32,15.72) casă(33.54,33.86) abia(31.11,31.46) trezit(0.0,0.0) din(41.23,41.4) somn(31.98,32.31) omul(0.0,0.0) a(0.0,0.0) scos(32.66,32.99) copilul(32.99,33.38) din(48.13,48.3) casă(0.0,0.0) și(48.63,48.81) a(0.0,0.0) încercat(34.04,34.48) să(34.48,34.6) stingă(34.6,35.06) flăcările(0.0,0.0) cu(1.96,2.24) zăpadă(35.85,36.57) familie(0.0,0.0) din(0.0,0.0) bistrița(0.0,0.0) pe(38.06,38.18) drumuri(0.0,0.0) după(0.0,0.0) ce(0.0,0.0) un(0.0,0.0) incendiu(71.16,71.71) s-a(45.69,45.89) extins(0.0,0.0) și(53.05,53.16) la(0.0,0.0) casa(0.0,0.0) lor(0.0,0.0) bărbatul(0.0,0.0) a(0.0,0.0) suferit(37.33,37.7) arsuri(37.7,38.06) pe(38.71,38.85) față(38.18,38.57) și(63.55,63.67) pe(0.0,0.0) mâini(38.85,39.16) și(68.19,68.36) a(0.0,0.0) fost(0.0,0.0) transportat(39.5,39.99) cu(20.39,20.51) ambulanța(40.1,40.65) la(0.0,0.0) spitalul(40.75,41.23) din(0.0,0.0) găești(0.0,0.0) unde(41.75,41.93) medicii(41.93,42.37) i-au(0.0,0.0) acordat(0.0,0.0) îngrijiri(42.93,43.58) a(0.0,0.0) refuzat(44.08,44.47) însă(44.47,44.66) să(44.66,44.79) rămână(44.79,45.05) internat(45.05,45.57) și(69.47,69.63) s-a(0.0,0.0) întors(0.0,0.0) acasă(46.37,46.79) pentru(46.79,47.13) a-și(47.13,47.27) mai(47.27,47.42) recupera(47.42,47.86) câte(0.0,0.0) ceva(47.86,48.13) din(0.0,0.0) bunuri(48.3,48.63) pompierii(0.0,0.0) au(63.67,63.77) reușit(61.74,62.12) să(62.12,62.25) înlătore(62.25,62.63) la(0.0,0.0) timp(62.75,63.04) pericolul(63.04,63.55) și(0.0,0.0) au(65.6,65.7) stins(63.77,64.1) focul(64.1,64.45) oamenii(64.71,65.23) nu(50.01,51.23) mai(65.4,65.6) au(67.73,67.84) însă(65.7,65.96) nimic(65.96,66.4) locuiau(66.44,66.89) cu(35.71,35.85) chirie(67.02,67.38) iar(67.38,67.53) acum(67.53,67.73) au(70.49,70.74) rămas(67.84,68.19) și(0.0,0.0) fără(28.33,29.23) haine(68.56,68.92) aparatură(68.95,69.47) și(0.0,0.0) alte(69.63,69.89) bunuri(69.89,70.19) care(70.19,70.46) au(0.0,0.0) ars(70.77,71.1) în(71.1,71.16) incendiu(0.0,0.0) o(18.97,19.34) bătrână(0.0,0.0) a(0.0,0.0) fost(0.0,0.0) găsită(0.0,0.0) fără(68.36,68.56) viață(0.0,0.0) în(0.0,0.0) casă(0.0,0.0) cuprinsă(0.0,0.0) de(0.0,0.0) foc(0.0,0.0) avea(0.0,0.0) obiceiul(0.0,0.0) să(0.0,0.0) își(0.0,0.0) aprindă(0.0,0.0) tămâie(0.0,0.0) top(0.0,0.0) citite(0.0,0.0) acum(0.0,0.0)

<s> povestește că muncise toată noaptea </s> (61934940\_00)

<s> era foarte obosit copilul </s> (61934940\_01)

Tabelul A2. Exemplu de aliniere îmbunătățită (V2) a transcrierii RAV (prima celula) cu transcrierea aproximativă (a doua celula) și rezultatul alinierii (a patra celula). Celula a treia prezintă transcrierea aproximativă formatată în vederea alinierii.

fure(0.71,1.24) și(1.24,1.38) un(1.38,1.55) master(1.55,1.96) cu(1.96,2.24) bărbatul(3.71,4.14) de(4.14,4.25) treizeci(4.25,4.55) și(4.55,4.65) șase(4.65,4.93) de(4.93,5.05) ani(5.05,5.19) povestește(5.19,5.66) că(5.66,5.75) muncise(5.75,6.22) toată(6.22,6.62) noaptea(6.62,6.99) și(6.99,7.11) era(7.11,7.27) foarte(7.27,7.63) obosit(7.63,8.09) copilul(8.23,8.64) se(8.64,8.75) uita(8.75,8.95) la(8.95,9.04) televizor(9.04,9.59) iar(9.59,9.76) la(9.76,9.86) un(9.86,9.96)

moment(9.96,10.23) dat(10.23,10.4) a(10.4,10.47) văzut(10.47,10.84) cum(10.84,11.03) iese(11.03,11.34) un(11.34,11.53) fum(11.53,11.8) gros(11.8,12.15) din(12.15,12.32) camera(12.32,12.65) alăturată(12.65,13.23) în(13.46,13.66) scurt(13.66,14.11) timp(14.11,14.35) vâlvătaia(14.35,14.95) cuprins(14.95,15.32) întreaga(15.32,15.72) casă(15.72,16.21) per(16.35,16.57) elev(16.57,16.93) și(17.76,18.02) volumic(18.02,18.6) fine(18.6,18.97) o(18.97,19.34) materii(19.63,20.23) și(20.26,20.39) cu(20.39,20.51) toată(20.51,20.81) tăria(20.81,21.18) lor(21.18,21.36) de(21.36,21.53) ce(22.81,23.11) vedem(23.14,23.41) ori(23.41,23.99) iese(24.23,24.57) fum(24.57,24.89) fum(25.24,25.7) nu(25.7,25.83) e(25.83,25.92) rezezi(25.92,26.56) și(26.56,26.74) rama(26.74,26.97) unei(26.97,27.25) morți(27.25,27.72) fără(28.33,29.23) fund(29.63,30.09) în(30.09,30.17) sus(30.17,30.56) abia(31.11,31.46) trezit(31.46,31.8) din(31.8,31.98) somn(31.98,32.31) omul(32.34,32.6) a(32.6,32.66) scos(32.66,32.99) copilul(32.99,33.38) din(33.38,33.54) casă(33.54,33.86) și(33.86,34.0) a(34.0,34.04) încercat(34.04,34.48) să(34.48,34.6) stingă(34.6,35.06) flăcările(35.06,35.68) cu(35.71,35.85) zăpadă(35.85,36.57) foto(36.7,37.33) suferit(37.33,37.7) arsuri(37.7,38.06) pe(38.06,38.18) față(38.18,38.57) și(38.57,38.71) pe(38.71,38.85) mâini(38.85,39.16) și(39.16,39.25) a(39.25,39.3) fost(39.3,39.5) transportat(39.5,39.99) cu(39.99,40.1) ambulanța(40.1,40.65) la(40.65,40.75) spitalul(40.75,41.23) din(41.23,41.4) găești(41.4,41.75) unde(41.75,41.93) medicii(41.93,42.37) au(42.37,42.49) cordat(42.49,42.93) îngrijiri(42.93,43.58) și(43.61,43.95) l-a(43.95,44.08) refuzat(44.08,44.47) însă(44.47,44.66) să(44.66,44.79) rămână(44.79,45.05) internat(45.05,45.57) și(45.57,45.69) s-a(45.69,45.89) reîntors(45.89,46.37) acasă(46.37,46.79) pentru(46.79,47.13) a-și(47.13,47.27) mai(47.27,47.42) recupera(47.42,47.86) ceva(47.86,48.13) din(48.13,48.3) bunuri(48.3,48.63) și(48.63,48.81) flamanzi(48.84,50.01) nu(50.01,51.23) fi(52.54,52.7) pur(52.81,53.05) și(53.05,53.16) simplu(53.16,53.5) fiica(53.5,53.89) nu(53.97,56.69) fac(56.72,57.14) dar(58.33,58.55) a(59.1,59.19) venit(59.19,59.42) foarte(60.06,60.41) repede(60.41,60.71) pompierii(61.17,61.64) au(61.64,61.74) reușit(61.74,62.12) să(62.12,62.25) înlătore(62.25,62.63) la(62.63,62.75) timp(62.75,63.04) pericolul(63.04,63.55) și(63.55,63.67) au(63.67,63.77) stins(63.77,64.1) focul(64.1,64.45) oamenii(64.71,65.23) nu(65.23,65.4) mai(65.4,65.6) au(65.6,65.7) însă(65.7,65.96) nimic(65.96,66.4) locuiau(66.44,66.89) cu(66.89,67.02) chirie(67.02,67.38) iar(67.38,67.53) acum(67.53,67.73) au(67.73,67.84) rămas(67.84,68.19) și(68.19,68.36) fără(68.36,68.56) haine(68.56,68.92) aparatură(68.95,69.47) și(69.47,69.63) alte(69.63,69.89) bunuri(69.89,70.19) care(70.19,70.46) au(70.49,70.74) ars(70.77,71.1) în(71.1,71.16) incendiu(71.16,71.71)

Un copil de 6 ani și-a salvat tatăl de la moarte, după ce locuința lor a fost cuprinsă de flăcări

Ultimul update: Miercuri 17 Ianuarie 2018 17:54

Data publicării: Miercuri 17 Ianuarie 2018 16:56

+ 0

Caz impresionant, miercuri dimineață, în Găești, județul Dâmbovița. Un copil de 6 ani și-a salvat tatăl de la moarte, după ce locuința lor a fost cuprinsă de flăcări.

Omul, sudor de meserie, venise din schimbul trei și dormea adânc. Băiatul se uita liniștit la televizor și când a simțit miros de fum și a văzut flăcările nu a fugit din casă, ci și-a strigat tatăl până l-a trezit.

Vecinii au auzit și ei țipetele băiatului, și-au dat seama de pericol și au chemat imediat pompierii.

Bărbatul, de 36 de ani, povestește că muncise toată noaptea și era foarte obosit. Copilul se uita la televizor, iar la un moment dat a văzut cum iese un fum gros din camera alăturată. În scurt timp vâlvătaia a cuprins întreaga casă.

Abia trezit din somn, omul a scos copilul din casă și a încercat să stingă flăcările cu zăpadă.

Citește și

Familie din Bistrița, pe drumuri după ce un incendiu s-a extins și la casa lor

Bărbatul a suferit arsuri pe față și pe mâini și a fost transportat cu ambulanța la spitalul din Găești, unde medicii i-au acordat îngrijiri. A refuzat însă să rămână internat și s-a întors acasă pentru a-și mai recupera câte ceva din bunuri.

Pompierii au reușit să înlătore la timp pericolul și au stins focul. Oamenii nu mai au însă nimic. Locuiau cu chirie, iar acum au rămas și fără haine, aparatură și alte bunuri, care au ars în incendiu.

Șocant

O bătrână a fost găsită fără viață în casa cuprinsă de foc: "Avea obiceiul să își aprindă tămâie"

Top citite acum

un copil de șase ani și-a salvat tatăl de la moarte după ce locuința lor a fost cuprinsă de flăcări  
ultimul update miercuri șaptesprezece ianuarie două mii optsprezece șaptesprezece și cincizeci  
și patru de minute data publicării miercuri șaptesprezece ianuarie două mii optsprezece  
șaisprezece și cincizeci și șase de minute caz impresionant miercuri dimineață în găești județul  
dâmbovița un copil de șase ani și-a salvat tatăl de la moarte după ce locuința lor a fost cuprinsă  
de flăcări omul sudor de meserie venise din schimbul trei și dormea adânc băiatul se uită liniștit  
la televizor și când a simțit miros de fum și a văzut flăcările nu a fugit din casă ci și-a strigat tatăl  
până l-a trezit vecinii au auzit și ei țipetele băiatului și-au dat seama de pericol și au chemat  
imediat pompierii bărbatul de treizeci și șase de ani povestește că muncise toată noaptea și era  
foarte obosit copilul se uită la televizor iar la un moment dat a văzut cum iese un fum gros din  
camera alăturată în scurt timp vâlvătaia a cuprins întreaga casă abia trezit din somn omul a scos  
copilul din casă și a încercat să stingă flăcările cu zăpadă familie din bistrița pe drumuri după ce  
un incendiu s-a extins și la casa lor bărbatul a suferit arsuri pe față și pe mâini și a fost  
transportat cu ambulanța la spitalul din găești unde medicii i-au acordat îngrijiri a refuzat însă să  
rămână internat și s-a întors acasă pentru a-și mai recupera câte ceva din bunuri pompierii au  
reușit să înlătore la timp pericolul și au stins focul oamenii nu mai au însă nimic locuiau cu chirie  
iar acum au rămas și fără haine aparatură și alte bunuri care au ars în incendiu o bătrână a fost  
găsită fără viață în casă cuprinsă de foc avea obiceiul să își aprindă tămâie top citite acum

<s> bărbatul de treizeci și șase de ani povestește că muncise toată noaptea și era foarte obosit  
copilul se </s> (61934940\_00)

<s> la televizor iar la un moment dat a văzut cum iese un fum gros din camera alăturată în scurt  
timp vâlvătaia cuprins întreaga casă </s> (61934940\_01)

<s> abia trezit din somn omul a scos copilul din casă și a încercat să stingă flăcările cu zăpadă  
</s> (61934940\_02)

<s> suferit arsuri pe față și pe mâini și a fost transportat cu ambulanța la spitalul din găești unde  
medicii </s> (61934940\_03)

<s> refuzat însă să rămână internat și s-a </s> (61934940\_04)

<s> acasă pentru a-și mai recupera ceva din bunuri </s> (61934940\_05)

<s> pompierii au reușit să înlătore la timp pericolul și au stins focul oamenii nu mai au însă nimic  
locuiau cu chirie iar acum au rămas și fără haine aparatură și alte bunuri care au ars în incendiu  
</s> (61934940\_06)