



D4.7. Diseminare

Aceste rezultate au fost obținute prin finanțare în cadrul Programului PN-III Proiecte complexe realizate în consorții CDI, derulat cu sprijinul MEN – UEFISCDI,
Cod: PN-III-P1-1.2-PCCDI-2017-0818, Contract Nr. 73 PCCDI/2018:

“SINTERO: Tehnologii de realizare a interfețelor om-mașină pentru sinteza text-vorbire cu expresivitate”

© 2018-2021 – SINTERO

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)
Institutul de Cercetări Pentru Inteligență Artificială “Mihai Drăgănescu”	ICIA	UNI	CO
Universitatea Tehnică din Cluj-Napoca	UTCN	UNI	P1
Universitatea Politehnică din București	UPB	UNI	P2
Universitatea "Alexandru Ioan Cuza" din Iași	UAIC	UNI	P3

**Date de identificare proiect**

Număr contract:	PN-III-P1-1.2-PCCDI-2017-0818, Nr. 73 PCCDI/2018
Acronim / titlu:	„SINTERO: Tehnologii de realizare a interfețelor om-mașină pentru sinteza text-vorbire cu expresivitate”
Titlu livrabil:	D4.7. Diseminare
Termen:	Aprilie 2021
Editor:	Mircea Giurgiu (Universitatea Tehnică din Cluj-Napoca)
Adresa de eMail editor:	Mircea.Giurgiu@com.utcluj.ro
Autori, în ordine alfabetică:	Mircea Giurgiu, Beata Lorincz, Maria Nuțu, Adriana Stan
Ofițer de proiect:	Cristian STROE

Rezumat:

În acest livrabil se prezintă articolele științifice susținute și publicate de către partenerul UTCN la conferințe internaționale în anul 2021, paginile web cu demonstratoare online privind sistemul final de sinteza, precum și modul în care au fost implicați studenții în tematica proiectului pe durata elaborării proiectelor de diplomă. Vom include lista articolelor și un rezumat cu principalele rezultate. Accesul la publicații (cf. 30.04.2021) este asigurat pe pagina web a proiectului în secțiunea dedicată. De asemenea, există o serie de pagini web cu demonstrații online de voci sintetice, respectiv din interfața interactivă online (RONNA).

Cuprins

1. Introducere	4
2. Identificarea posibilităților de publicare pe anul 2021 și realizări	4
3. Publicații elaborate și transmise spre publicare în anul 2021	4
4. Lucrări de licență în legătura cu tematica proiectului	5
5. Stagii de practică pentru studenți	5
6. Pagini web ale proiectului SINTERO	5
7. Pagini Wiki interne grupului de cercetare cu mostre audio generate de diferite versiuni ale sistemelor de sinteză text – vorbire implementare în etapa a IV-a.	7
8. Concluzii	10

1. Introducere

Acest livrabil prezintă o sinteză a articolelor elaborate și transmise spre publicare în anul 2021, pagina web a proiectului și o serie de demonstratoare online. Prin utilizarea arhitecturilor de rețele neuronale profunde de tip Tacotron și DC-TTS se obține o calitate înaltă a vocii sintetice pentru vorbitori în limba română, calitate comparabilă cu rezultate de dată recentă. De asemenea, se prezintă referințe către paginile web ale demonstratoarelor online și tematici de cercetare în care au fost implicați și studenții pe durata elaborării proiectelor de diplomă.

2. Identificarea posibilităților de publicare pe anul 2021 și realizări

Conform cu strategia de diseminare inclusă în formularul de aplicație, pentru fiecare an calendaristic s-au identificat posibilitățile de diseminare și de publicare de articole la conferințe științifice sau în jurnale.

Pentru anul 2021 s-au identificat următoarele posibilități de publicare la conferințe internaționale:

- 2021 ISCA Interspeech, 30 august – 3 septembrie 2021, Brno, Cehia, (<https://www.interspeech2021.org/>)
- The 11th Speech and Dialog Conference (SPED 2021), 13-15 octombrie 2021, Bucuresti (<https://sped.pub.ro/>).
- 25th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 8-10 septembrie 2021, Polonia (<http://kes2021.kesinternational.org/>) - online

3. Publicații elaborate și transmise spre publicare în anul 2021

Dan Oneață, Alexandru Caranica, Adriana Stan, Horia Cucu, "*An Evaluation of Word-level Confidence Estimation for end-to-end Automatic Speech Recognition*", In Proceedings of the 8th IEEE Spoken Language Technology Workshop (SLT 2021), Shenzhen, China, January 2021

Maria Nuțu, "*Automatic Romanian lemmatization through a deep learning approach.*", acceptat spre publicare la 25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES), 8-10 septembrie 2021, Polonia.

Beata Lorincz, Adriana Stan, Mircea Giurgiu, "*An objective evaluation of the effects of recording conditions and speaker characteristics in multi-speaker deep neural speech synthesis*", trimis spre recenzare la The 25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES), 8-10 septembrie 2021, Polonia.

Beata Lorincz, Adriana Stan, Mircea Giurgiu, "*Speaker verification-derived loss and data augmentation for DNN-based multispeaker speech synthesis*", trimis spre recenzare la The 29th European Signal Processing Conference, EUSIPCO 2021, Dublin, Ireland.

Beata Lorincz, Elena Irimia, Adriana Stan, Verginica Barbu-Mititelu, "*RoLEX: The development of an extended Romanian lexical dataset and its evaluation at predicting concurrent lexical information*", trimis spre recenzare la Computer Speech and Language.

Adriana Stan, Beata Lorincz, Maria Nuțu, Mircea Giurgiu, "The MARA corpus: Expressivity in end-to-end TTS systems using synthesised speech data", trimis spre recenzare la The 11th Conference on Speech Technology and Human-Computer Dialogue, in Bucharest, Romania, 13-15 octombrie 2021.

4. Lucrări de licență în legătura cu tematica proiectului

- Vlad Crehul - "Improving TTS for Romanian using Tacotron2 based on deep neural networks", lucrare de diplomă, iulie 2021.
- Vlad Hornai - "Adaptation of neural network-based automatic speech recognition system DeepSpeech for a specific application domain", lucrare de diplomă, iulie 2021.
- Bogdan Oros - "Cross-domain NLP adaptation using BERT and Transformers for Twitter text analysis", lucrare de diplomă, iulie 2021.
- Adrian Dobrescu - "Speaker recognition in noisy environments", lucrare de diplomă, iulie 2021.

5. Stagii de practică pentru studenți

- Oana Ranta (Master) - „Speech denoising with neural networks”.

6. Pagini web ale proiectului SINTERO

The screenshot shows the website for SINTERO. The header includes the SINTERO logo and a navigation menu with links for EN, DESCRIERE, CONSORTIU, ECHIPA, REZULTATE, and CONTACT. The main content area features a large title: "SINTERO: Tehnologii de realizare a interfețelor om-mașină pentru sinteza text-vorbire cu expresivitate". Below the title, there is a subtitle: "Proiect finanțat de Ministerul Cercetării și Inovării, Program PN-III-P1-1.2.-PCCDI, nr. 73/2018, durata: 2018-2020". A prominent button labeled "RETEROM" is visible. Below this, there is a section titled "Proiecte paralele" with three buttons: "COBILIRO", "TEPROLIN", and "TADARAV".



Rapoarte științifice

- **Raport științific etapa 1 - sinteză (2018) [pdf]**
 - **D1.15.** Identificare pattern-uri prozodice [pdf]
 - **D1.16.** Metode de clasificare a stilului de exprimare din text [pdf]
 - **D1.17.** Analiza metodelor de control și adaptare automată a expresivității [pdf]
 - **D1.18.** Implementarea modulului de control automat al prozodiei [pdf]
 - **D1.19.** Diseminare [pdf]
- **Raport științific etapa 2 - sinteză (2019) [pdf]**
 - **D2.15.** Implementarea modulului de identificare a stilului de vorbire și nivelului de expresivitate din analiza textului [pdf]
 - **D2.16.** Implementarea unui modul de adaptare la un nou vorbitor a sistemului de sinteză [pdf]
 - **D2.17.** Implementarea unui modul de transplantare a prozodiei unui vorbitor în sistemul de sinteză [pdf]
 - **D2.18.** Îmbunătățirea componentei de modelare și control al prozodiei; activități de testare, validare / demonstrare module software [pdf]
 - **D2.19.** Diseminare [pdf]
- **Raport științific etapa 3 - sinteză (2020) [pdf]**
 - **D3.15.** Dezvoltarea unei noi tehnologii pentru adaptarea vocii sintetice la stilul și expresivitatea unui nou vorbitor [pdf]
 - **D3.16.** Dezvoltarea unei noi metode de adaptare rapidă a vocii sintetice folosind date audio atipice [pdf]
 - **D3.17.** Integrare tehnologie nouă și demonstrarea în realizarea interfețelor om-mășină pentru sinteza text – vorbire. [pdf]
 - **D3.18.** Diseminare [pdf]



Publicații

- Beăta Lőrincz, Maria Nutu, Adriana Stan, Mircea Giurgiu "An Evaluation of Postfiltering for Deep Learning Based Speech Synthesis with Limited Data", IEEE 10th International Conference on Intelligent Systems (IS), Bulgaria, 2020 [pdf]
- Beăta Lőrincz, "Concurrent phonetic transcription, lexical stress assignment and syllabification with deep neural networks", Proceedings of the 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems KES2020, 2020 [pdf]
- Adriana Stan, "RECOApy: Data Recording, Pre-Processing and Phonetic Transcription for End-to-End Speech-Based Applications", In Proceedings of the Interspeech, Shanghai, China, 2020 [bib] | [pdf]
- Kristen M Scott, Simone Ashby, Adriana Stan "Designing a Synthesized Content Feed System for Community Radio", Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, Estonia, 2020 [pdf]
- Adriana Stan, "Input Encoding for Sequence-to-Sequence Learning of Romanian Grapheme-to-Phoneme Conversion", In Proceedings of the 10th IEEE International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Timisoara, Romania, 2019 [bib] | [pdf]
- Beăta Lőrincz, Maria Nutu, Adriana Stan, "Romanian Part of Speech Tagging using LSTM Networks", In Proceedings of the IEEE 15th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 2019 [bib] | [pdf]
- Maria Nutu, Beăta Lőrincz, Adriana Stan, "Deep Learning for Automatic Diacritics Restoration in Romanian", In Proceedings of the IEEE 15th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 2019. [bib] | [pdf]
- David A. Braude, Matthew P. Aylett, Caoimhin Laoide-Kemp, Simone Ashby, Kristen M. Scott, Brian O Raghallaigh, Anna Braudo, Alex Brouwer, Adriana Stan, "All Together Now: The Living Audio Dataset", In Proceedings of Interspeech, Graz, Austria, 2019. [bib] | [pdf]
- Adriana Stan, Mircea Giurgiu, A Comparison Between Traditional Machine Learning Approaches And Deep Neural Networks For Text Processing In Romanian, in Proc. of the 13th International Conference on Linguistic Resources and Tools for Processing Romanian Language, 22-23 November, Jassy, Romania [bib] | [pdf]



- **D3.18.** Diseminare [pdf]

Prezentări video ale articolelor diseminate la conferințe cu desfășurare virtuală:

- Beăta Lőrincz, Maria Nutu, Adriana Stan, Mircea Giurgiu "An Evaluation of Postfiltering for Deep Learning Based Speech Synthesis with Limited Data", IEEE 10th International Conference on Intelligent Systems (IS), Bulgaria, 2020
- Beăta Lőrincz, "Concurrent phonetic transcription, lexical stress assignment and syllabification with deep neural networks", Proceedings of the 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems KES2020
- Adriana Stan, "RECOApy: Data Recording, Pre-Processing and Phonetic Transcription for End-to-End Speech-Based Applications", In Proceedings of the Interspeech, Shanghai, China, 2020

The screenshot shows a web browser at the URL speech.utcluj.ro/ronna/. The page features a dark header with the 'ronna' logo and navigation links for 'DEMO', 'AUDIO SAMPLES', and 'CONTACT'. The main content area is titled 'Romanian Neural Networks API (RoNNA)' and 'Text-To-Speech Online demo'. Below this is a form for API key registration, including a text input field for the key, a 'Contact maintainers' link, and dropdown menus for 'System' and 'Voice'. A large text area is provided for the text to be synthesized in Romanian.

Audio samples

The screenshot displays a grid of audio player controls for various TTS systems. The grid is organized into three columns: 'Natural', 'Sample 1', and 'Sample 2'. Each row represents a different system, with five rows in total. Each cell in the grid contains a play button, a progress indicator (e.g., 0:00 / 0:03), a volume icon, and a settings menu icon.

	Natural	Sample 1	Sample 2
DC-TTS - MARA	▶ 0:00 / 0:03 - 🔊 ⋮	▶ 0:00 / 0:02 - 🔊 ⋮	▶ 0:00 / 0:02 - 🔊 ⋮
DC-TTS - BAS	▶ 0:00 / 0:04 - 🔊 ⋮	▶ 0:00 / 0:03 - 🔊 ⋮	▶ 0:00 / 0:04 - 🔊 ⋮
DC-TTS - CAU	▶ 0:00 / 0:04 - 🔊 ⋮	▶ 0:00 / 0:03 - 🔊 ⋮	▶ 0:00 / 0:04 - 🔊 ⋮
DC-TTS - DCS	▶ 0:00 / 0:04 - 🔊 ⋮	▶ 0:00 / 0:03 - 🔊 ⋮	▶ 0:00 / 0:04 - 🔊 ⋮
DC-TTS - DDM	▶ 0:00 / 0:04 - 🔊 ⋮	▶ 0:00 / 0:03 - 🔊 ⋮	▶ 0:00 / 0:04 - 🔊 ⋮

7. Pagini Wiki interne grupului de cercetare cu mostre audio generate de diferite versiuni ale sistemelor de sinteză text – vorbire implementare în etapa a IV-a.

Mostre audio aferente noilor voci din SWARA 2.0 sunt disponibile la adresa: <https://gitlab.utcluj.ro/speech/tts-samples/-/wikis/home>.

<https://gitlab.utcluj.ro/speech/tts-samples/-/wikis/Sisteme%20Tacotron2%20voci%20noi>

Sisteme Tacotron2 voci noi

Voice ID	Sample 1 - "Tezele ar putea fi eliminate în acest an școlar, iar Ministerul Educației analizează mai multe variante."	Sample 2 - "Astfel încât să fie aplicate aceleași reguli și pentru copiii care învață în scenariul verde,"
DLL		
DOL		
EME		

<https://gitlab.utcluj.ro/speech/tts-samples/-/wikis/Date%20atipice>

- Sistem Tacotron2 antrenat cu date preluate din segmentarea și transcrierea automată a unei resurse audio colectate de P1-CoBiLiRo, realizată în P3-TADARAV

Voice ID	Natural	Sample 1 - "Tezele ar putea fi eliminate în acest an școlar, iar Ministerul Educației analizează mai multe variante."	Sample 2 - "Astfel încât să fie aplicate aceleași reguli și pentru copiii care învață în scenariul verde,"
COB			

- Sistem Tacotron2 antrenat cu date din audiobook-ul Mara folosind doar transcrierea ortografică (Mara-letters) sau transcrierea fonetică și accentul (Mara-phoneAcc) rezultate din rețelele de predicție concurentă a informației lexicale

Voice ID	Natural	Sample 1 - "Tezele ar putea fi eliminate în acest an școlar, iar Ministerul Educației analizează mai multe variante."	Sample 2 - "Astfel încât să fie aplicate aceleași reguli și pentru copiii care învață în scenariul verde,"
MARA-letters			

<https://gitlab.utcluj.ro/speech/tts-samples/-/wikis/Sistem%20DC-TTS%20Date%20expressive%20neexpressive>

Model	Epoci	Sample 1	Sample 2
Forma ortografică & Neexpressiv	3000		
Forma ortografică & Neexpressiv & Expresiv	4500		
Forma ortografică & Neexpressiv & Expresiv	6000		

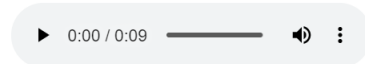
<https://gitlab.utcluj.ro/speech/tts-samples/-/wikis/Sistem%20DC-TTS%20Date%20lexicale>

Text sintetizat:

Sample 1: Ce se va întâmpla cu aceasta după ce vom ieși din starea de urgență.

Sample 2: Elevii vor putea intra în școală doar cu declarația pe propria răspundere completată de către părinți.

Referință (audio natural):



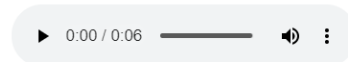
Date text	Sample 1	Sample 2
Forma ortografică		
Forma transcrisă fonetic		

<https://gitlab.utcluj.ro/speech/tts-samples/-/wikis/Sistem%20DC-TTS%20Englez%C4%83%20Singur%20Vorbitor%20Date%20lexicale>

Sistem DC TTS Engleză Singur Vorbitor Date lexicale

LJSpeech Voce

Referință (audio natural):



Text sintetizat:

Sample 1: The demands on the President in the execution of His responsibilities in today's world are so varied and complex

Sample 2: and the traditions of the office in a democracy such as ours are so deep-seated as to preclude absolute security.

Date text	Sample 1	Sample 2
Forma ortografică		
Forma transcrisă fonetic		
Forma transcrisă fonetic cu accent		

<https://gitlab.utcluj.ro/speech/tts-samples/-/wikis/Sistem%20DC-TTS%20Englez%C4%83%20Date%20lexicale>

Sistem DC TTS Engleză Date lexicale

LibriTTS voci

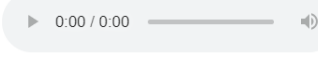
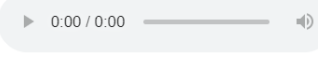
Text sintetizat:

Sample 1: Jimmie Dale slipped his mask into his pocket, and, with the parcel under his arm, stepped to the door and unlocked it.

Sample 2: He paused for an instant on the threshold for a single, quick, comprehensive glance around the room-then passed on out into the street.

Epoci	Sample 1	Sample 2
1000		
3000		

Voci masculine

Date text	Epoci	Sample 1	Sample 2
Forma ortografică	1000		

8. Concluzii

Accesul la publicațiile elaborate în anul 2021 este asigurat la adresa <http://speech.utcluj.ro/sintero/rezultate/>. Pagina web are un conținut dinamic, adaptat cu realizările din proiect, astfel că pentru această raportare se pot accesa și mostre cu semnal sintetic generat de modulul de adaptare la un noi vorbitori, așa cum este prezentat în paginile web cu demonstratoarele.