



D2.19. 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”

© 2018-2020 – 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:	D2.19. Diseminare
Termen:	Noiembrie 2019
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 2019. Au fost identificate la începutul etapei conferințele de interes din domeniul aferent cercetărilor. Vom include lista articolelor și un rezumat cu principalele rezultate. Accesul la publicații (cf. 30.11.2019) este asigurat pe pagina web a proiectului în secțiunea dedicată.

De asemenea, este disponibilă și pagina web actualizată a sub-proiectului SINTERO (<http://speech.utcluj.ro/sintero/>), precum și o serie de pagini web cu demonstrații online de voci sintetice cu expresivitate (Corpusul Mara cu 11 ore de vorbire expresivă - <https://speech.utcluj.ro/marasc/>, demonstrator privind sinteza pe baza corpusului Mara și adaptarea la noi vorbitori - https://speech.utcluj.ro/lrec2020_mara/, evaluarea online a 7 sisteme de sinteză folosind testul Mushra - <http://romaniantts.com/lrec/>, demonstrator privind adaptarea sistemului de sinteză la noi vorbitori - <https://speech.utcluj.ro/sintero/dnn-samples/>).

Cuprins

1. Introducere	4
2. Identificarea posibilităților de publicare pe anul 2019 și realizări	4
3. Publicații științifice în anul 2019	4
4. Pagina web a proiectului SINTERO	6
5. Alte pagini web cu demonstrarea online a modului de control a prozodiei	6
6. Concluzii	8

1. Introducere

Acest livrabil prezintă o sinteză a articolelor publicate în anul 2019, pagina web a proiectului și o serie de demonstratoare online pentru controlul automat al expresivității vocilor sintetizate prin intermediul noilor modele proiectate și implementate în această etapă. Acestea prezintă un punct de plecare pentru comparații cu rezultate viitoare în proiect. De exemplu, în prezent sunt în cercetare și alte tehnici de adaptare a stilului de vorbire și a expresivității pe bază de structuri neuronale artificiale, respectiv prin folosirea unui set restrâns de date audio și de text.

2. Identificarea posibilităților de publicare pe anul 2019 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 2019 s-au identificat următoarele posibilități de publicare la conferințe internaționale:

- 2019 ISCA Interspeech 2019, 16-19 Sept 2019, Graz, Austria (<https://interspeech2019.org/>)
- 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Sept 2019, Cluj-Napoca, Romania (www.iccp.ro)
- The 10th Conference on Speech Technology and Human-Computer Dialogue, 10-12 Oct 2019, Timișoara, Romania (<https://sped.pub.ro/>)
- The 14th Edition of the International Conference on Linguistic Resources and Tools for Processing Romanian Language, 19-20 Nov 2019, Cluj-Napoca, Romania (<https://profs.info.uaic.ro/~consilr/>)
- Articol în pregătire pentru 12th Edition of Language Resources and Evaluation Conference, 11-16 Mai 2020, Marsilia, Franța (<https://lrec2020.lrec-conf.org/en/>).

3. Publicații științifice în anul 2019

Autori	Beata Lorincz, Maria Nuțu, Adriana Stan
Titlu	„Romanian Part of Speech Tagging using LSTM Networks”
Ref.	IEEE 15th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Sept 2019, Cluj-Napoca
Rezumat	<i>In this paper we present LSTM based neural network architectures for determining the part of speech (POS) tags for Romanian words. LSTM networks combined with fullyconnected output layers are used for predicting the root POS, and sequence-to-sequence models composed of LSTM encoders and decoders are evaluated for predicting the extended MSD and CTAG tags. The highest accuracy achieved for the root POS is 99.18% and for the extended tags is 98.25%. This method proves to be efficient for the proposed task and has the advantage of being language independent, as no expert linguistic knowledge is used in the input features.</i>

Autori	Maria Nuțu, Beata Lorincz, Adriana Stan
Titlu	„Deep Learning for Automatic Diacritics Restoration in Romanian”
Ref.	IEEE 15th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Sept 2019, Cluj-Napoca
Rezumat	<i>In this paper we address the issue of automatic diacritics restoration (ADR) for Romanian using deep learning strategies. We compare 6 separate architectures with various mixtures of recurrent and convolutional layers. The input consists in sequences of consecutive words stripped of their diacritic symbols. The network’s task is to learn to restore the diacritics for the entire sequence. No additional linguistic or semantic information is used as input to the networks. The best results were obtained with a CNN-based architecture and achieved an accuracy of 97% at word level. At diacritic-level the accuracy of the same architecture is 89%.</i>

Autori	Adriana Stan
Titlu	„Input Encoding for Sequence-to-Sequence Learning of Romanian Grapheme-to-Phoneme Conversion”
Ref.	10th Conference on Speech Technology and Human-Computer Dialogue, 10-12 Oct 2019, Timișoara, Romania
Rezumat	<i>This paper evaluates the use of sequence-to-sequence learning models for the Romanian grapheme-to-phoneme conversion. The strategies explore the use of different input feature encoding: one-hot letter encoding, additional embedding layer and grapheme embeddings learned from a large corpus of Romanian text. Additional lexical information, such as syllabification and lexical stress is also taken into consideration for augmenting the orthographic form of the word and providing more accurate phonetic transcriptions. The sequence-to-sequence models are also compared to a baseline decision tree algorithm in terms of both phone- and word-level accuracy. The best results are achieved by the model which uses grapheme embeddings and all additional linguistic information. Its accuracy is 97.90% at word-level, and 99.62% at phone-level. However, only minor differences exist between the tested systems.</i>

Autori	David A. Braude, Matthew P. Aylett, Caoimhin Laoide-Kemp, Simone Ashby, Kristen M. Scott, Brian O’Raghallaigh, Anna Braudo, Alex Brouwer, Adriana Stan
Titlu	„All Together Now: The Living Audio Dataset”
Ref.	Proceedings of Interspeech 2019, 16-19 Sept 2019, Graz, Austria
Rezumat	<i>The ongoing focus in speech technology research on machine learning based approaches leaves the community hungry for data. However, datasets tend to be recorded once and then released, sometimes behind registration requirements or paywalls. In this paper we describe our Living Audio Dataset. The aim is to provide audio data that is in the public domain, multilingual, and expandable by communities. We discuss the role of linguistic resources, given the success of systems such as Tacotron which use direct text-to-speech mappings, and consider how data provenance could be built into such resources. So far the data has been collected for TTS purposes, however, it is also suitable for ASR. At the time of publication audio resources already exist for Dutch, R.P. English, Irish, and Russian</i>

4. Pagina web a proiectului SINTERO

← → ↻ <https://speech.utcluj.ro/sintero/> 🔍 ☆

SINTERO EN DESCRIERE CONSORTIU ECHIPA REZULTATE CONTACT

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

Proiect finanțat de Ministerul Cercetării și Inovării, Program PN-III-P1-1.2.-PCCDI, nr. 73/2018, durata: 2018-2020

Proiect component al proiectului complex **RETEROM**

COBILIRO TEPROLIN TADARAV

Proiecte paralele

5. Alte pagini web cu demonstrarea online a modului de control a prozodiei

Pagina web (<https://speech.utcluj.ro/marasc/>) cu resurse audio și text din audiobook-ul Mara (nuvela de Ioan Slavici), conținând aproximativ 11 ore de vorbire și folosit pentru antrenarea unei voci cu naturalețe și expresivitate ridicate.

← → ↻ speech.utcluj.ro/marasc/ 🔍 ☆

Speech Processing GROUP ABOUT MARA DOWNLOAD AUDIO SAMPLES DEVELOPERS TEAM CONTACT

The MARA Speech Corpus

This corpus contains the audiobook **Mara** written by Ioan Slavici and published in 1906. The corpus consists of approximately 11 hours of speech recorded by a female speaker in a controlled environment.

The audiobook was kindly provided by Mihai Nae from Cartea Sonora.

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/). When using it in your work, please acknowledge Cartea Sonora.

An Expressive Romanian Speech Dataset

Pagină web (https://speech.utcluj.ro/lrec2020_mara/) cu demonstrator audio privind sinteza de voce folosind corpusul Mara.



Pagini web (<http://romaniantts.com/lrec/>) pentru evaluarea online, prin teste de ascultare cu utilizatorii, a naturaleții și expresivității sistemului de sinteză creat cu modulele implementate în etapa 2019.



Pagină web (<https://speech.utcluj.ro/sintero/dnn-samples/>) pentru demonstrarea online a sistemului de sinteză cu expresivitate bazat pe Tacotron GST (Global Style Tokens).

	Natural	Sample 1	Sample 2
Speaker 1	▶ 0:00 / 0:05	▶ 0:00 / 0:05	▶ 0:00 / 0:03
Speaker 2	▶ 0:00 / 0:06	▶ 0:00 / 0:05	▶ 0:00 / 0:03
Speaker 3	▶ 0:00 / 0:05	▶ 0:00 / 0:04	▶ 0:00 / 0:03

6. Concluzii

Accesul la publicațiile elaborate în anul 2019 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 control automat al expresivității, așa cum este prezentat în paginile web cu demonstratoarele.