



D3.17. Integrare tehnologie nouă și demonstrarea în realizarea interfețelor om-mașină pentru sinteza text – vorbire.

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:	D3.17. Integrare tehnologie nouă și demonstrarea în realizarea interfețelor om-mașină pentru sinteza text – vorbire.
Termen:	Noiembrie 2020
Editor:	Adriana STAN
Adresa de eMail editor:	adriana.stan@com.utcluj.ro
Autori, în ordine alfabetică:	Mircea GIURGIU, Beata LORINCZ, Maria NUȚU, Adriana STAN
Ofițer de proiect:	Cristian STROE

Rezumat:

Acest raport prezintă interfața online pentru demonstrarea sistemelor de sinteză text-vorbire pe bază de rețele neuronale profunde, dezvoltate în cadrul proiectului P4. Interfața este disponibilă online la adresa: <http://speech.utcluj.ro/ronna> și poate fi utilizată pe baza unei chei de API obținută de la administratorii platformei.

Structura hardware pe care este implementat demonstratorul este de ultimă generație și are la bază un server performant dedicat procesării paralele a datelor folosind GPU (Graphical Processing Unit).

1. Demonstrarea unei interfețe online pentru sinteza text – vorbire în limba română folosind rețele neuronale profunde și procesarea paralelă a datelor pe servere cu GPU (Graphical Processing Unit)

Acest raport prezintă interfața web denumită RoNNA - Romanian Neural Network API și care integrează rezultatele proiectului de până la acest moment. Aceste rezultate sunt preponderent regăsite în extinderea tehnologiilor de sinteză text-vorbire în limba română bazate pe modele Markov și disponibile în website-ul www.romaniantts.com, la tehnologiile bazate pe rețele neuronale multistrat și utilizarea informațiilor lingvistice suplimentare în transcrierea textului de intrare.

Sistemele de sinteză disponibile în platforma API:

- 1) Sistem bazat pe rețele convoluționale (DC-TTS) - voce Mara
- 2) Sistem bazat pe rețele convoluționale (DC-TTS) - voce BAS
- 3) Sistem bazat pe rețele convoluționale (DC-TTS) - voce EME
- 4) Sistem bazat pe rețele convoluționale (DC-TTS) - voce CAU
- 5) Sistem bazat pe rețele convoluționale (DC-TTS) - voce DDM
- 6) Sistem bazat pe rețele convoluționale (DC-TTS) - voce DCS
- 7) Sistem bazat pe rețele convoluționale (DC-TTS) - voce FDS
- 8) Sistem bazat pe rețele convoluționale (DC-TTS) - voce IPS
- 9) Sistem bazat pe rețele convoluționale (DC-TTS) - voce PCS
- 10) Sistem bazat pe rețele convoluționale (DC-TTS) - voce SAM
- 11) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce Mara
- 12) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce Mara + informații lingvistice
- 13) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce EME
- 14) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce NLL
- 15) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce DOL
- 16) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce DLL
- 17) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce ZPL
- 18) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce SRL
- 19) Sistem bazat pe rețele recurente (Tacotron2) și vocoder Waveglow - voce FEM
- 20) Sistem bazat pe fluxuri de normalizare (Flowtron) - voce Mara

O captură de ecran cu interfața RoNNA este prezentată în figura de mai jos. Deoarece serverele pe baza de GPU care deservește demonstrația online sunt folosite în mod curent de grupul de cercetare pentru diferite experimente, utilizatorii au nevoie de o cheie de API ce poate fi obținută de la coordonatorii P4. În interfața web sunt disponibile mostre audio cu cele 20 de sisteme. Suplimentar se pot regăsi mostre audio ale sistemelor de sinteză descrise în livrabilele D3.15 și D3.16, dar care datorită calității scăzute nu sunt incluse în sistemul interactiv de sinteză.



Audio samples

	Natural	Sample 1	Sample 2
DC-TTS - MARA			
DC-TTS - IPS			
DC-TTS - SAM			
Tacotron2 - MARA			
Tacotron2 - EME			
Tacotron2 - DLL			
Tacotron2 - ZPL			
Tacotron2 - SRL			
Tacotron2 - NLL			

Online demo

API key:

System Voice

Textul de sinteză

Fig. 1. Interfața web a API-ului RoNNA www.speech.utcluj.ro/ronna/

De asemenea, alte demonstratoare privind implementarea tehnologiei pe diverse paliere ale acestora (evaluarea augmentării datelor text cu informații lexicale precise din textul de intrare, diferite structuri de rețele neuronale care implementează sistemul de sinteză: Tacotron2, DC-TTS, Flowtron) sunt accesibile la adresa:

<https://gitlab.utcluj.ro/speech/tts-samples/-/wikis/home>.