



DIPLOMSKI RAD br. 1043

NEURALNE MREŽE U IZDVAJANJU GOVORNOG SIGNALA IZ ZVUČNOG ZAPISA

Stjepan Henc

Zagreb, srpanj 2015.

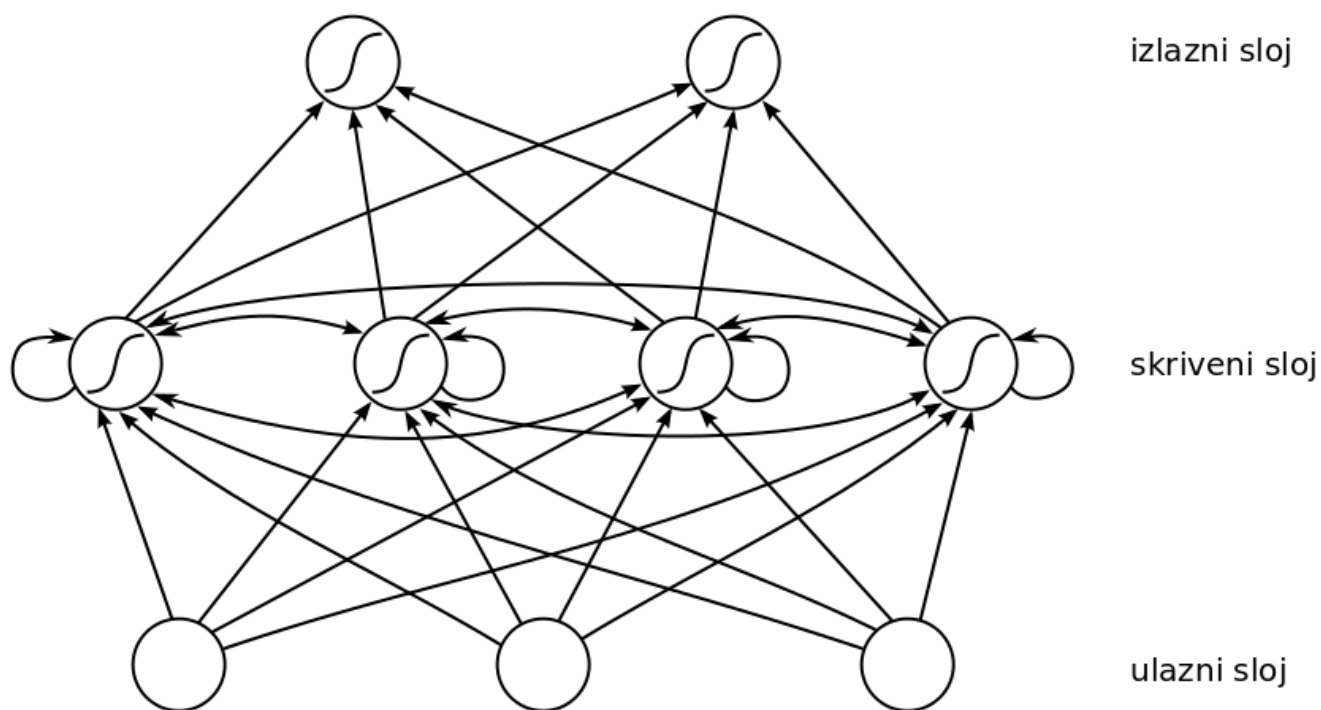
Izdvajanje govora se radi kako bi se poboljšala točnost računalnog prepoznavanja govora.

Zadatak:

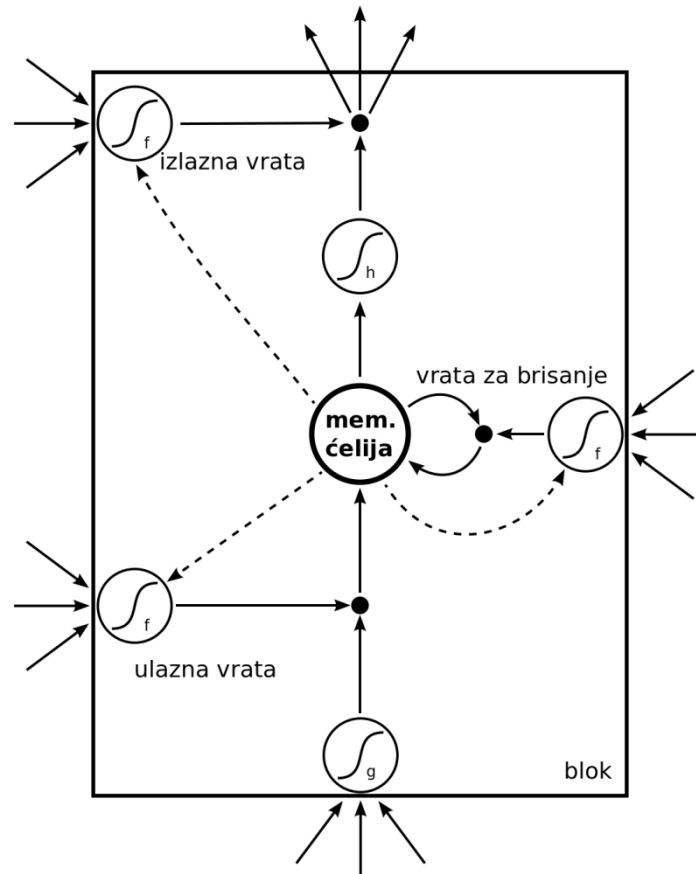
- odabrati programski paket
- ispitati koliko je poboljšanje prepoznavanja govora

- neuronske mreže prisutne u izdvajanju govora (*engl. Speech Extraction*) od 80-ih godina – PCA, ICA
- krajem 90-ih i početkom 2000-ih pojavljuju se RNN s BLSTM blokovima
- u zadnjih nekoliko godina sve popularnije duboke neuronske mreže

RNN

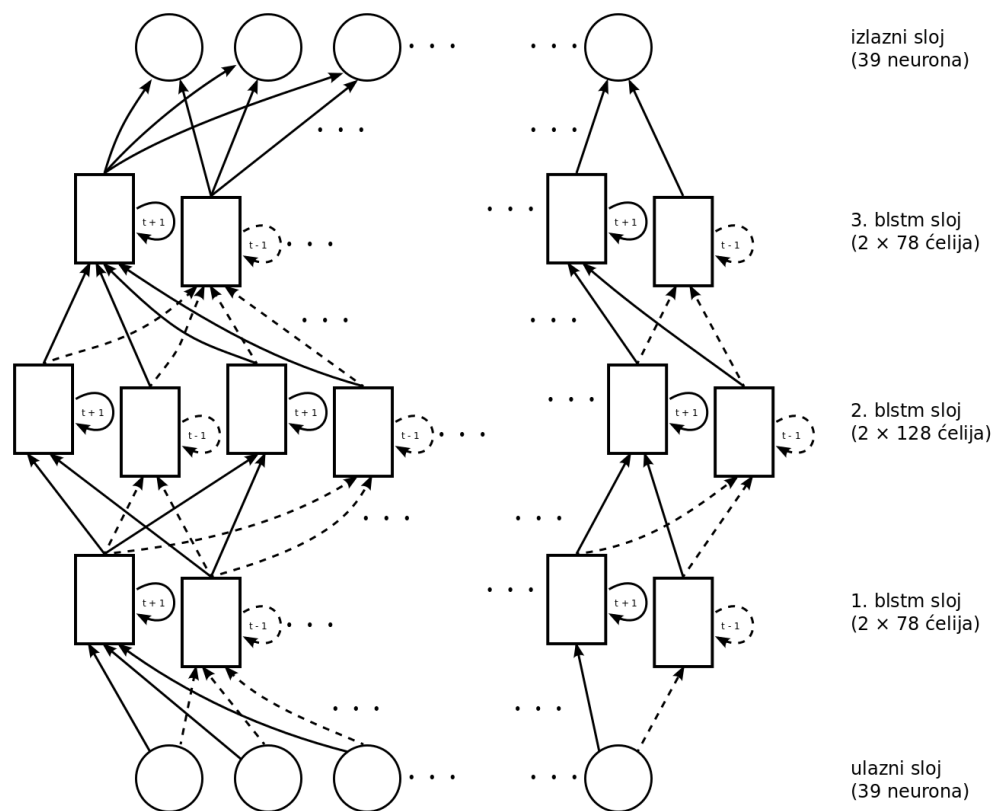


LSTM



Odabrana arhitektura mreže

Zavod za automatiku i računalno inženjstvo



Odabir programskog paketa

Zavod za automatiku i računalno inženjerstvo

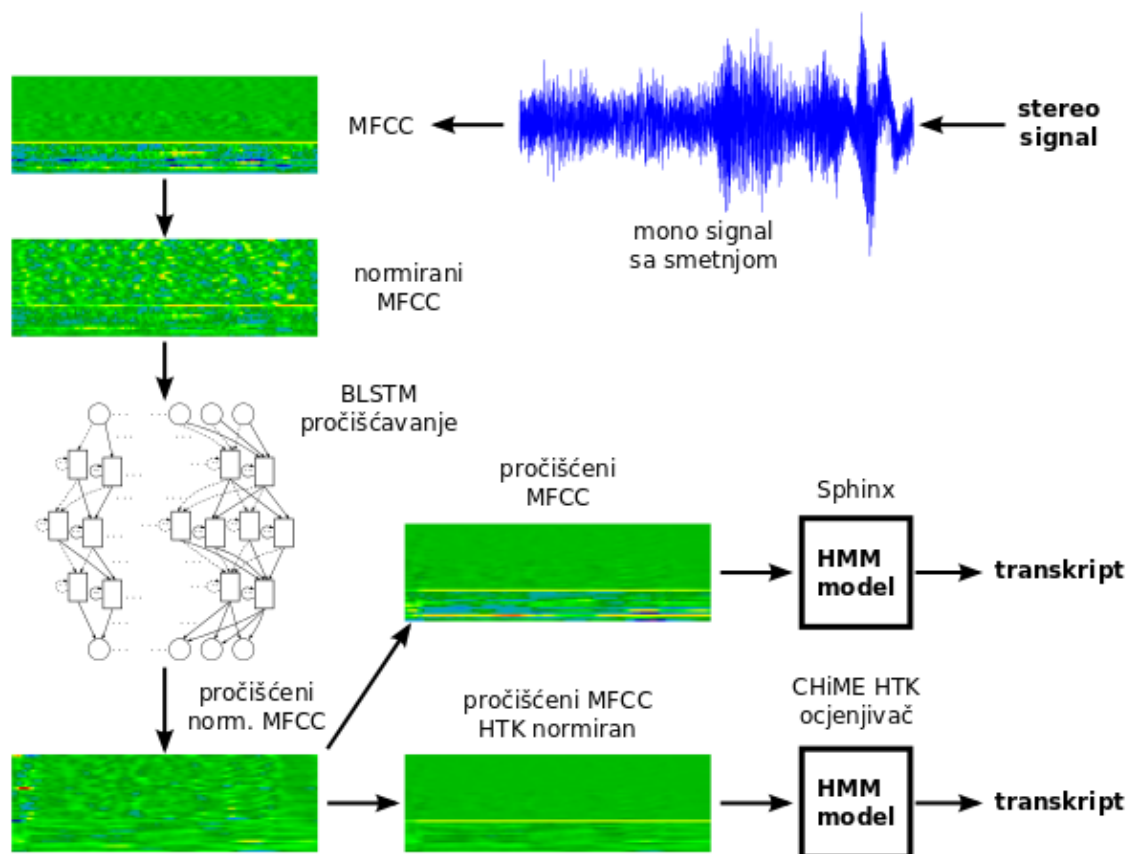


	pybrain	torch7	theano	rnnlib	CURRENNT
GPU	ne	da	da	ne	da
BLSTM	da	ne	ne	da	da
jezik	python	lua/c	python	c++	c++

- poboljšanje prepoznavanja je ispitano na podacima drugog CHiME natjecanja, (*engl. CHiME 2nd challenge Task 1*)
- korpus čitanog engleskog govora s malim rječnikom
- čisti govor izobličen odjekom sobe i realnim smetnjama

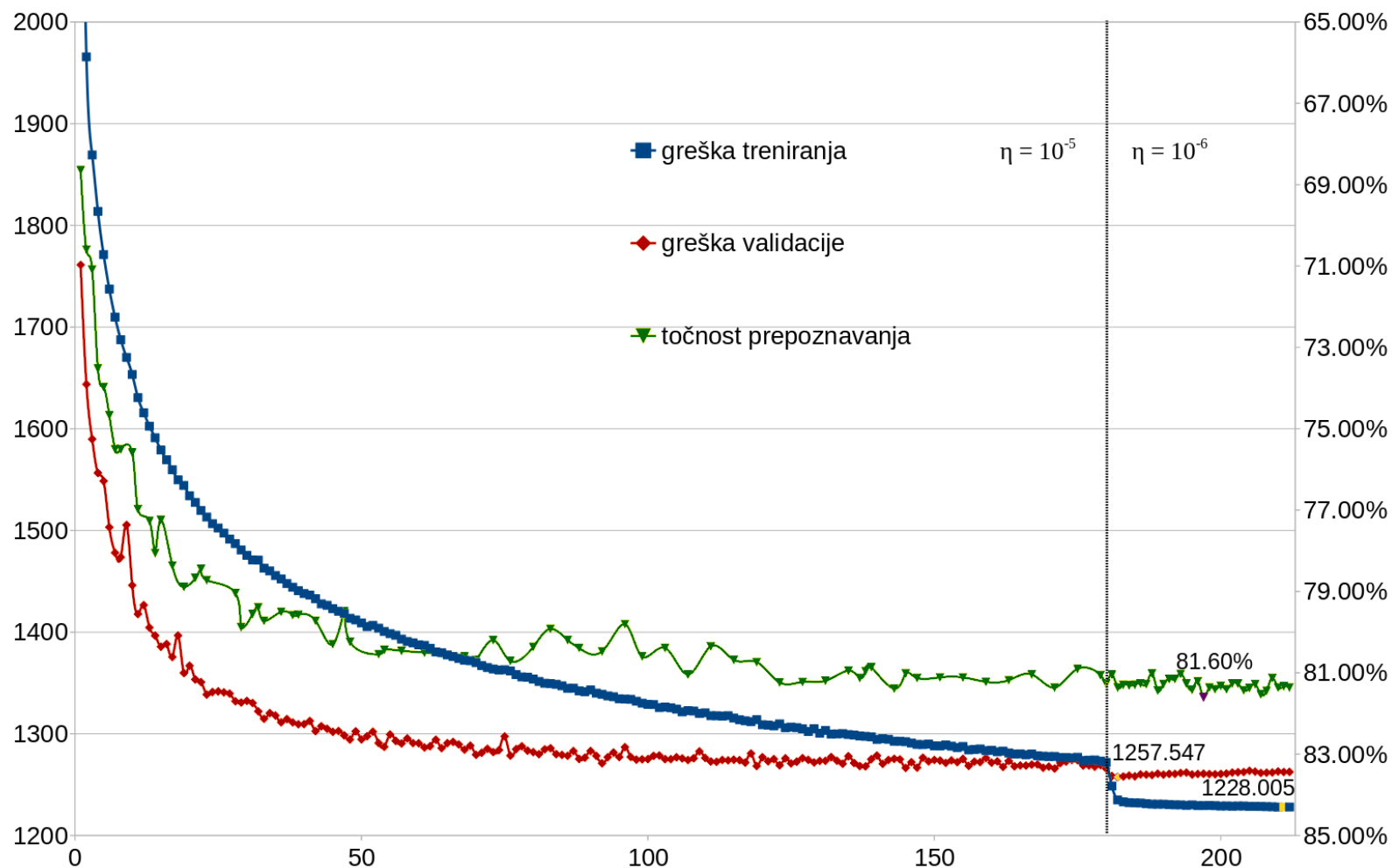
Pregled sustava

Zavod za automatiku i računalno inženjerstvo



Treniranje

Zavod za automatiku i računalno inženjstvo



Validacija		Test						
WA [%]	sr.vr.	-6dB	-3dB	0dB	3dB	6dB	9dB	sr.vr.
“Odjek” model, govor s bukom	56,9	32,2	38,3	52,1	62,7	76,1	83,8	57,5
“Buka” model, govor s bukom	68,6	49,3	58,7	67,5	75,1	78,8	82,9	68,7
“Odjek” model, izdvojeni govor	81,6	71,3	76,9	82,3	84,9	89,3	90,5	82,5
Prilagođeni model, izdvojeni govor	83,3	73,3	78,8	84,1	87,2	89,1	91,6	84,0
“Odjek” model, govor s jekom	93,8							

Sustav za prepoznavanje čistog govora korištenjem ovog pristupa za izdvajanje govora ima 25% veću točnost prepoznavanja riječi.

Za odnos govora i šuma od -6 dB, poboljšanje je čak 39.1%.

RNN-BLSTM je učinkoviti pristup za izdvajanje govora.

1. Graves, Alex: "Supervised Sequence Labelling with Recurrent Neural Networks", Springer, <http://www.cs.toronto.edu/~graves/preprint.pdf>, 2012., str. 20, 21, 26
2. Weninger, Felix; Geiger, Jurgen; Wollmer, Martin; Schuller, Bjorn; Rigoll, Gerhard: "The Munich feature enhancement approach to the 2nd CHiME challenge using BLSTM recurrent neural networks", Proceedings of the 2nd CHiME workshop on machine listening in multisource environments, Citeseer, 2013.
3. Weninger, Felix; Bergmann, Johannes; Schuller, Bjorn: "Introducing CURRENNT: The Munich Open-Source CUDA RecurREnt Neural Network Toolkit", Journal of Machine Learning Research, 2015.
4. "The 2nd 'CHiME' Speech Separation and Recognition Challenge: Small vocabulary track", http://spandh.dcs.shef.ac.uk/chime_challenge/chime2013/chime2_task1.html, 25.6.2015
5. Henc, Stjepan: "Scripts for creating netCDF databases for CURRENNT", https://github.com/sthenc/nc_packer, 25.6.2015