

Electronics and electrical engineering Elektronika ir elektros inžinerija

DIKTORIAUS IDENTIFIKAVIMO TIKSLUMO TYRIMAS NAUDOJANT NEURONINIUS TINKLUS

Laurynas DOVYDAITIS*, Vytautas RUDŽIONIS

Vilniaus universitetas, Vilnius, Lietuva

Gauta 2018 m. birželio 28 d.; priimta 2018 m. birželio 28 d.

Santrauka. Šiame straipsnyje nagrinėjami kai kurie diktoriaus identifikavimo problemos aspektai. Problemos aktualumas nulėmtas praktinių galimybių suteikti adaptuotas paslaugas konkrečiam asmeniui, žinant jo tapatybę. Straipsnyje aprašoma diktoriaus identifikavimo veiksmų seka ir išskiriami identifikavimo etapai. Apžvelgiami moksliniai akustinių modelių kūrimo darbai pasitelkiant neuroninius tinklus. Šiame straipsnyje siūlomos kelios neuroninių tinklų konfigūracijos, kurios gali būti naudojamos diktoriaus akustiniam modeliavimui. Teikiami pasiūlymai eksperimentiniu būdu patikrinami fiksuojant gaunamą diktoriaus identifikavimo tikslumą su LIEPA projekto metu surinktu garsynu.

Reikšminiai žodžiai: diktoriaus identifikavimas, neuroniniai tinklai, GRU, BGRU, LSTM, BLSTM, MFCC.

Įvadas

Diktoriaus identifikavimas – vienas patogiausių asmens tapatybės biometrinių identifikavimo būdų. Tačiau tai tiksliai atlikti yra sudėtinga, nes susiduriame su ypač dideliu signalo variatyvumu. Tai, kad kiekvieno diktoriaus kiekvienas žodžių ištartimas yra vis kitoks, kad balsas gali būti paveiktas ligos ar bandymų pamėgdžioti identifikuojamą asmenį, yra tik keli variatyvumo šaltinių pavyzdžiai. Variatyvumas ir yra pagrindinė priežastis, kodėl, nepaisant nemenkų pastangų šioje srityje, diktoriaus identifikavimo negalima laikyti tobulai išspręstu uždaviniu ir tema išlieka aktuali moksliniu požiūriu.

Diktoriaus identifikavimas ypač aktualus, kai komunikacija vykdoma balsu per atstumą. Klasikinis tokios sistemos pavyzdys yra telefoninis pokalbis. Turint galimybę tiksliai identifikuoti pašnekovą, galima pateikti daugiau personalizuotos, asmeninės informacijos, suteikti papildomų paslaugų.

Diktoriaus identifikavimas pagal balsą, naudojant giliuosius neuroninius tinklus, padeda spręsti šį uždavinį, t. y. leidžia pasiekti aukštesnį identifikavimo tikslumą. Problemos sprendimas reikalauja panaudoti daug skirtingų algoritimų, kurių kiekvienas iškelia nemažus iššūkius: nuo balso pavyzdžių surinkimo, unikalų požymių išskyrimo, diktoriaus akustinio modeliavimo iki gauto rezultato klasifikavimo, priskiriant jį konkrečiam diktoriui.

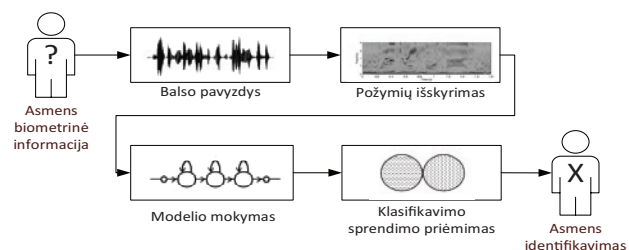
Diktoriaus identifikavimas

Norint atsakyti į klausimą, kaip vykdomas diktoriaus identifikavimas, galima sudaryti veiksmų seką, kuriuos reikia įvykdyti sprendžiant uždavinį. Šių veiksmų diagrama parodyta 1 paveiksle.

Atpažinimo kokybė priklauso nuo visų (1 paveikslas) atvaizduotų komponentų, tiek balso kokybės, išskirtų požymių unikalumo, nuo akustinio modelio sukūrimo ir nuo priimamo klasifikatoriaus sprendimo.

Diktoriaus identifikavimo etapai

Šiuolaikinėse diktoriaus atpažinimo sistemose taikomi įvairūs atpažinimo metodai. Šie metodai ir algoritmai skiriasi nuo atpažinimo etapo (Saon ir Chien, 2012).



1 paveikslas. Diktoriaus atpažinimo diagrama
Figure 1. Speaker recognition sequence

*Autorius susirašinėti. El. paštas laurynas.dovydaitis@gmail.com

Atlikus kiekvieno etapo mokslinės literatūros analizę, galima sugrupuoti skirtingas kalbos ir diktoriaus atpažinimo sistemas, jų pagrindinę klasifikaciją galima sudaryti taip (Molau et al., 2001; Šalna ir Kamarauskas, 2010; Hanun ir Case, 2014; Amodei et al., 2015):

1. Požymių išskyrimo etapas:
 - a. MFCC – melo dažnių kepstro koeficientai;
 - b. neuroninio tinklo tarpinių svorių požymiai (angl. *bottleneck*);
 - c. požymių apjungimo modeliai;
2. Diktoriaus modeliavimo etapas ir sprendimo priėmimo etapas:
 - a. Klasikiniai modeliavimo būdai:
 - i. HMM – paslėptieji Markovo modeliai;
 - ii. HMM–GMM – paslėptieji Markovo modeliai su Gauso mišiniais;
 - iii. UBM – universaliojo fono modeliai;
 - iv. SVM–GMM – atraminių vektorių mašinų ir Gauso mišinių modeliai;
 - v. i – vektorių sistemos;
 - b. Modeliavimas naudojant dirbtinius neuroninius tinklus:
 - i. ANN – dirbtiniai neuroniniai tinklai;
 - ii. DNN – gilieji neuroniniai tinklai;
 - iii. RNN – rekurentiniai neuroniniai tinklai;

Pateiktoje klasifikacijoje yra atskirti klasikiniai diktoriaus atpažinimo metodai, taip pat ir metodai, naudojantys gilųjų mokymą ir giliuosius neuroninius tinklus. Iš atliktos mokslinės analizės galima išskirti populiariausią požymių išskyrimo metodą – tai MFCC. Taip pat pažymėtina, kad yra atlikta mažai diktoriaus identifikavimo tyrimų, naudojant giliuosius neuroninius tinklus. Todėl tai suteikia papildomą motyvaciją sukurti sistemą, taikančią būtent tokį diktoriaus modeliavimo metodą.

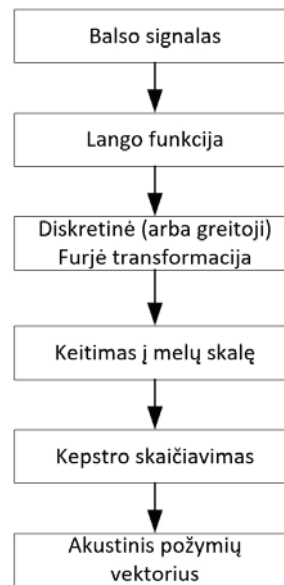
Požymių parinkimas

Kaip teigia įvairūs autoriai (Ravindran, Anderson ir Slaney, 2006; Tiwari, 2010), melo dažnių skalės kepstro koeficientai yra vienas atspariausių būdų aprašyti laike kintančias savybes. Šių koeficientų skaičiavimas apibendrintas 2 paveiksle (Molau et al., 2001; Aroon ir Dhonde, 2015; Rubal ir Mehan, 2017).

Pateiktame paveiksle matyti, kad gaunamas balso signalas yra apdorojamas: suskaidomas į įvairios trukmės langus, tada apskaičiuojamos Furjė transformacijos. Antrąjį ciklo dalyje viskas perskaičiuojama į melų skalę.

Dažniausiai naudojamas požymių rinkinys diktoriaus identifikuoti (Šalna ir Kamarauskas, 2010) sudarytas iš tokių komponentų:

- 12 MFCC koeficientų;
- 12 delta – MFCC koeficientų;
- 12 delta-delta – MFCC koeficientų;
- 1 (log) kadro energijos elemento;
- 1 (log) delta kadro energijos elemento;
- 1 (log) delta-delta kadro energijos elemento.



2 paveikslas. Melo dažnių kepstro koeficientų skaičiavimo seka
Figure 2. Mel frequency cepstral coefficients calculation workflow

Klasifikavimo tyrimai taikant grįžtamojo ryšio neuroninius tinklus

Grįžtamojo ryšio tinklas (GRT) (angl. *Recurrent Neural Network*) – vienas iš neuroninių tinklų tipų, kuriuos naudojant pastaruoju metu gaunami geri klasifikavimo rezultatai vaizdų, šnekos ir kitose atpažinimo srityse.

GRT, priešingai nuo tiesinio sklaidimo tipo neuroninių tinklų, įėjimo sluoksniui gali naudoti kintančių dimensijų įėjimo vektorius (Schuster ir Paliwal, 1997). Ši savybė suteikia geresnes sąlygas GRT panaudoti natūraliai kalbai suprasti bei apdoroti, taip pat kalbai generuoti, vaizdui apdoroti ir įvairioms kitoms užduotims atlikti (Gal ir Ghahramani, 2016).

LSTM tipo neuroniniai tinklai. Ilgas laikinosios atminties tinklas (LSTM) (angl. *Long-Short term memory*) – tai rekurentinis tinklas (RNN), naudojantis grįžtamąjį ryšį, siekiant išsaugoti („atsiminti“) paskutinius periodo įėjimo signalus. Keletas iš šio tinklo naudojimo pavyzdžių yra kalbos apdorojimas, muzikos kompozicija ir kt. (Hochreiter ir Schmidhuber, 1997).

Tyrimas naudojant LSTM ir dvikrypčio LSTM tinklą lietuviškai kalbančiam diktoriaus identifikuoti, naudojant skirtingas tyrimo duomenų apimtis, buvo atliktas Dovydaičio ir Rudžionio (2018a, 2018b), kuriame, naudojant šio tipo tinklus buvo gauti tikslesni identifikavimo rezultatai nei su klasikiniiais modeliavimo metodais, naudojant HMM.

GRU tipo neuroniniai tinklai. Sulaikomo pasikartojančio vieneto (GRU) (angl. *Gated Recurrent Unit*) tinklas yra panašus į LSTM tipo tinklą, nes taip pat naudoja loginius elementus, kuriais kontroliuojamas informacijos siuntimas, tačiau priešingai nuo LSTM tipo tinklo neturi atminties celių (Chung, Gulcehre, K. Cho ir Y. Cho, 2014).

Kaip ir kitų rekurentinių neuroninių tinklų konfigūracijoms, GRU yra galimybė naudoti dvikryptę architektūrą.

GRU tipo tinklai buvo palyginti su LSTM tipo tinklais (Chung et al., 2014) naudojant muzikos duomenų bases. Autoriai visais atvejais gavo geresnius rezultatus, nei naudodami LSTM tinklą. Šiame darbe gautos išvados skatina pabandyti patobulinti diktoriaus identifikavimo tikslumą patikrinant šio tinklo galimybes.

Tinklo hiperparametrų paieška

Pagal pateikiamą įvairių neuroninių tinklų konfigūracijų įvairovę nėra lengva parinkti vieno neuroninio tinklo konfigūracijos, kuri tiksliausiai vykdytų diktoriaus identifikavimo uždavinį. Dauguma autorių sutaria, kad neįmanoma automatizuoti neuroninio tinklo konfigūracijos sukūrimo ir automatinio nustatymo procedūros. Nors yra pavyzdžių, kai automatinis metamodeliavimas pritaikomas neuroninio tinklo topologijai parinkti, galinčiai atpažinti skaitmeninius vaizdus (Baker et al., 2017), ekspertinė analizė leidžia pasiekti geresnių rezultatų.

Tinklo hiperparametrai leidžia optimizuoti tinklo mokymo greitį ir veikia gaunamus mokymosi ir atpažinimo rezultatus. Todėl galima teigti, kad tinklo mokymo optimizavimas yra geriausių hiperparametrų parinkimas. Pagrindiniai giliųjų neuroninių tinklų atpažinimo parametrai yra mokymo dažnis, optimizavimo algoritmas, neuronų kiekis sluoksnyje, sluoksnių kiekis ir kiti parametrai.

Geriausios neuroninio tinklo konfigūracijos paieškai sudaryta 1 parametų lentelė.

Toliau apžvelgsime diktorių garsyno rinkinį ir reikalingą programinę įrangą eksperimentams atlikti.

Eksperimentui naudoto garsyno aprašas

Lietuvių šnekos garsynas LIEPA – fonetiškai reprezentatyvi lietuvių šnekos duomenų bazė, pritaikyta elektroninėms paslaugoms teikti ir šnekos technologijų moksliniams tyrimams ir konstravimo darbams (Laurinčiukaitė, Telksnys, Kasparaitis, Kliukienė ir Paukštytė, 2017). Garsyną sudaro šnekos atpažinimo paslaugų kūrimo tikslams parengta 100 val. (garsyno dalis Nr. 1) ir šnekos sintezavimo tikslams parengta 13 val. (garsyno dalis Nr. 2) trukmės gar-

syno dalys. Diktorių skaičius – 380. Garso įrašai: žodžiai, frazės, sakiniai ir jų dalys.

LIEPA garsyne naudojami 78 skirtingų tekstų įrašai, iš jų 45 apima nepertraukiamos kalbos pavyzdžius. Įrašai buvo renkami daugiau nei šešiose lokacijose su skirtinga įranga. Pirmojoje garsyno dalyje, iš kurios paimtos šiame tyrime naudotos mokymo ir testavimo pavyzdžių imtys, iš viso dalyvavo 376 diktoriai, iš kurių 116 diktorių buvo mokiniai ir 260 universiteto studentų, bei kviestiniai diktoriai, iš kurių iš viso 248 moterys ir 128 vyrai (Laurinčiukaitė et al., 2017).

Kaip teigia autoriai (Laurinčiukaitė et al., 2017), LIEPA garsynas gali būti sėkmingai naudojamas tiek įvairių atpažinimo sistemoms tirti, tiek praktiškai pritaikyti. Minėtame straipsnyje taip pat pateikiamos detalios garsyno charakteristikos.

Įrašo turinio iliustracijai pateikiamas nepertraukiamos kalbos turinio pavyzdys diktoriaus D02 ištartam tekstui: „Aprašymuose atsiskleidžia mauzoliejus ir kapa viečių grandioziškumas.“ Šis pavyzdys buvo suskaidytas į 586 langus, kurių kiekvienam bus skaičiuojami MFCC požymiai. Diktorius D02 viso ištarė 57 nepertraukiamos kalbos pavyzdžius.

Garsyno paruošimas ir programinė įranga

Prieš pradėdant tyrimą ir siekiant optimaliai išnaudoti turimus programinius ir aparatinius išteklius, nuspręsta tyrimo eigą suskirstyti į du etapus – testavimo etapą bei visos apimties, t. y. naudojant visus pavyzdžius.

Eksperimentui naudojama programinė įranga:

- požymiams išskirti – HTK ir „Python“ su SIDEKit paketas;
- neuroniniams tinklams kurti ir apmokyti – „Python“ su „Keras“ ir „Theano“ paketais.

Tyrimo imčių parametrai naudojami tokie pat kaip (Dovydaitis ir Rudžionis, 2018a) atliktame tyrime, todėl individualūs diktoriaus pavyzdžiai yra dalijami santykiu 70 % / 30 %, čia 70 % pavyzdžių naudojama klasifikatoriui mokytis, 30 % – identifikavimui testuoti. Geriausių modelių eksperimentams iš LIEPA garsyno yra sudaryti 5 diktorių pavyzdžių rinkiniai pagal (Dovydaitis ir Rudžionis, 2018b) atliktą tyrimą, kur kiekviename rinkinyje yra tik unikalūs diktoriai.

Diktoriaus identifikavimo rezultatai

Testavimo imties tyrimai

Pirmajame etape visos tinklo konfigūracijos be išmetimo sluoksnio gavo vidutiniškai 5 % blogesnius identifikavimo rezultatus. Modelių su neuronų išmetimo sluoksnių tikslumo rezultatai pateikiami 2 lentelėje.

Lyginant gautus rezultatus su darbe (Dovydaitis ir Rudžionis, 2018a) tirtais tinklais, visos konfigūracijos atskirai rodo prastesnius tikslumo rezultatus, nei darbe tirti LSTM tipo neuroniniai tinklai.

1 lentelė. Neuroninio tinklo hiperparametrų paieškos lentelė
Table 1. Neural network hyper parameter search grid

Eksperimentas	Parametrų paieška	Baigimo sąlyga
Tinklo architektūra	GRU, dvikryptis GRU	Bandymas
Išmetimo sluoksnis	Pridėti išmetimo sluoksnį prie geriausių variantų	Bandymas
Tinklo gylis	80 neuronų žingsnio didinimas kiekvienam tyrimui	Gaunamas tikslumo rezultatas yra prastesnis prieš tai buvusį

2 lentelė. Geriausi vieno sluoksnio neuroninio tinklo atpažinimo rezultatai su GRU ir BGRU architektūra

Table 2. DNN recognition results on GRU and BGRU architecture

Tinklo architektūra	Identifikavimo tikslumas testinei imčiai
GRU 80 drop	84,57 %
GRU 160 drop	88,20 %
GRU 240 drop	91,41 %
GRU 320 drop	90,24 %
GRU 480 drop	89,95 %
BGRU 80 drop	89,30 %
BGRU 160 drop	93,08 %
BGRU 240 drop	93,30 %
BGRU 320 drop	93,44 %

Atliekant kitą bandymą siekiama patikrinti, ar pakeičiant geriausiai veikiančių kelių sluoksnių tinklą yra galimybė gauti geresnius rezultatus. Tuo tikslu paimamas didžiausių našumą tyrime (Dovydaitis ir Rudžionis, 2018a) parodęs BLSTM 2*160 ir jo pirmasis sluoksnis pakeičiamas GRU 160 tinklu. Gauti rezultatai pateikiami 3 lentelėje.

Iš pateiktų lentelių matyti, kad tikslesnį rezultatą hibridinis šios konfigūracijos tinklas GRU 160 * LSTM 160 drop.

3 lentelė. Neuroninio tinklo atpažinimo tikslumas sujungus kelis sluoksnius

Table 3. Neural network accuracy dependence on network depth

Tinklo architektūra	Identifikavimo tikslumas testinei imčiai
BGRU 160 * BLSTM 160 drop	96,22 %
BLSTM 160 * BLSTM 160 drop	94,68 %

4 lentelė. LIEPA eksperimentų imties tikslumas su GRU tipo tinklais

Table 4. LIEPA identification accuracy results with GRU type network

Identifikavimo tikslumas	1 rinkinys	2 rinkinys	3 rinkinys	4 rinkinys	5 rinkinys
GRU 240 drop	78,04 %	60,91 %	82,31 %	92,26 %	91,89 %
BGRU 240 drop	85,80 %	76,41 %	88,41 %	96,38 %	94,26 %
2x GRU160 drop	92,00 %	78,55 %	91,74 %	97,94 %	96,47 %

5 lentelė. LIEPA eksperimentų imties tikslumas pagal Dovydaičio ir Rudžionio (2018b) tyrimus

Table 5. LIEPA identification accuracy results in Dovydaitis and Rudžionis (2018b) experiment

Identifikavimo tikslumas	1 rinkinys	2 rinkinys	3 rinkinys	4 rinkinys	5 rinkinys
BLSTM 240 drop	90,30 %	83,91 %	92,48 %	97,66 %	95,82 %
2xBLSTM 160	92,08 %	86,45 %	92,35 %	97,10 %	96,39 %

6 lentelė. LIEPA eksperimentų imties tikslumas naudojant hibridinės topologijos neuroninį tinklą

Table 6. LIEPA identification accuracy results with hybrid recurrent topology network

Identifikavimo tikslumas	1 rinkinys	2 rinkinys	3 rinkinys	4 rinkinys	5 rinkinys
GRU 160 LSTM 160 drop	94,02 %	87,23 %	93,09 %	98,05 %	97,54 %

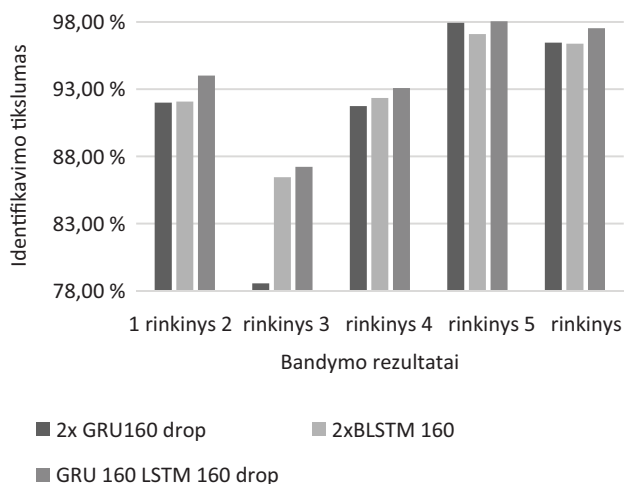
Visos apimties identifikavimo tyrimai

Tyrimai, naudojant visą duomenų bazę, atliekami su tiksliausiais ir hibridiniais variantais, kurie buvo nustatyti testuojant bandomąją imtį. Nustatyta, kad tiksliausias identifikavimo rezultatas gaunamas su šių tinklų konfigūracijomis, pateikiamomis 4 lentelėje.

5 lentelėje pateikiami tikslumo įverčiai iš Dovydaičio ir Rudžionio (2018b) tyrime atliktų duomenų.

Siūlomas hibridinis sujungimo variantas pakeičiant LSTM pirmą sluoksnį GRU tinklo sluoksniu. Gaunami rezultatai pateikiami 6 lentelėje.

Geriausių trijų rezultatų palyginimas pateikiamas 3 paveiksle.



3 paveikslas. Trijų geriausių variantų lyginamasis grafikas
Figure 3. Three most accurate NN models

Išvados

Atliekamų diktorius identifikavimo tyrimų skaičius rodo, kad minima problema yra aktuali ir suteikia papildomą motyvaciją atlikti tyrimus naujais metodais.

Apžvelgus kitus tyrimus ir atlikus eksperimentus nustatyta, kad tiksliausiai diktorių identifikuojanti gilaus neuroninio tinklo konfigūracija yra hibridinės konfigūracijos GRU 160 * LSTM 160 drop architektūros neuroninis tinklas.

Įvertintas neuroninio tinklo tikslumas ir tyrimo metu nustatyta, kad, tiriant diktorius identifikavimo uždavinį LIEPA duomenų bazės nevisos apimties atžvilgiu, geriausiai atpažinimą įvykdo gilus neuroninis tinklas, kurio architektūra yra GRU 160 * LSTM 160 drop. Ši neuroninio tinklo architektūra lietuviškai kalbantį diktorių identifikuoja 1 % tiksliau nei BLSTM 160 * BLSTM 160 drop neuroninio tinklo konfigūracija.

Literatūra

- Amodei, D., Ananthanarayanan, S., & et. al. (2015). *Deep speech 2: end-to-end speech recognition in English and Mandarin*. Retrieved from arXiv:1512.02595 [cs.CL]
- Aroon, A., & Dhonde, S. B. (2015). Speaker recognition system using Gaussian Mixture model. *International Journal of Computer Applications* (0975 – 8887), 130(14, November), 38-40. <https://doi.org/10.5120/ijca2015907193>
- Baker, B., Gupta, O., Naik, N., & et. al. (2017). Designing neural network architectures using reinforcement learning. *International Conference on Learning Representations, ICLR*. Vancouver, BC, Canada.
- Chung, J., Gulcehre, C., Cho, K., & Cho., Y. (2014). *Empirical evaluation of gated recurrent neural networks on sequence modeling*. Retrieved from arXiv:1412.3555v1 [cs.NE]
- Dovydaitis, L. ir Rudžionis, V. (2018a). Building Bi-directional LSTM neural network based speaker identification system. *Computational Science and Techniques* (pp. 574-580). Klaipėda university.
- Dovydaitis, L. ir Rudžionis, V. (2018b). Speaker identification accuracy improvement using blstm neural network. *Indian Journal of Computer Science and Engineering*, 9(2), 31-37.
- Gal, Y., & Ghahramani, Z. (2016). *A theoretically grounded application of dropout in recurrent neural networks*. University of Cambridge.
- Hannun, A., & Case, C. (2014). *Deep speech: scaling up end-to-end speech recognition*. Retrieved from arXiv:1412.5567v2 [cs.CL]
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Laurinčiukaitė, S., Telksnys, L., Kasparaitis, P., Kliukienė, R. ir Paukštytė, V. (2017). *Lithuanian speech corpus Liepa for*

the development of Lithuanian speech controlled equipment. DRAFT.

- Molau, S., Pitz, M., & et al. (2001). Computing mel-frequency cepstral coefficients on the power spectrum. *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01)*. IEEE. <https://doi.org/10.1109/ICASSP.2001.940770>
- Programinė įranga HTK. (n.d.). Prieiga per internetą: <http://htk.eng.cam.ac.uk/>
- Programinė įranga Python. (n.d.). Prieiga per internetą: <https://www.python.org/>
- Programinė įranga SIDEKit. (n.d.). Prieiga per internetą: <http://www-lium.univ-lemans.fr/sidekit/index.html>
- Programinė įranga Keras. (n.d.). Prieiga per internetą: <https://keras.io/>
- Programinė įranga Theano. (n.d.). Prieiga per internetą: <http://deeplearning.net/software/theano/>
- Ravindran, S., Anderson, V. D., & Slaney, M. (2006). *Improving the noise-robustness of mel-frequency cepstral coefficients for speech processing*. ISCA Tutorial and Research Workshop on Statistical And Perceptual Audition SAPA2006 16 September 2006, Pittsburgh PA.
- Rubal, & Mehan, V. (2017). Isolated word recognition and sign language detection using LPC and MFCC. *International Journal of Advance Research in Computer Science and Management Studies*, 5(7), 137-145.
- Saon, G., & Chien, J. (2012). Large-vocabulary continuous speech recognition systems. *IEEE signal processing magazine*, 29(6). <https://doi.org/10.1109/MSP.2012.2197156>
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673-2681. <https://doi.org/10.1109/78.650093>
- Šalna, B. ir Kamarauskas, J. (2010). Evaluation of effectiveness of different methods in speaker recognition. *Elektronika ir elektrotechnika*, 2(98), 67-70.
- Tiwari, V. (2010). MFCC and its applications in speaker recognition. *International Journal on Emerging Technologies*, 1(1), 19-22.

ACCURACY RESEARCH USING NEURAL NETWORKS FOR SPEAKER IDENTIFICATION

L. Dovydaitis, V. Rudžionis

Abstract

In this paper we present results on speaker identification task by using neural networks for acoustic modelling. Article is structured by describing speaker identification workflow, later identifying specific steps needed for speaker identification. Afterwards we identify number of different neural network configurations which can be used for speaker identification.

Keywords: speaker identification, neural nets, GRU, BGRU, LSTM, BLSTM, MFCC.