

Машинско и аудитивно препознавање словенских језика

Jacek Kudera

Trier University, Germany
E-mail: kudera@uni-trier.de
ORCID <https://orcid.org/0000-0003-3678-1067>
<https://ror.org/02778hg05>

Jovana Stevanović

University of Niš, Serbia
E-mail: jovanica.st@gmail.com
ORCID <https://orcid.org/0000-0001-6384-8568>
<https://ror.org/00965bg92>

Анотација. Предмет рада јесте машинско и перцепцијско препознавање језичког порекла. Спроведене су две студије за процену способности лаичких слушалаца и најсавременијег машинског приступа за идентификацију словенског Л1 из делексикализованих говорних узорака. У првој студији учествовало је 228 изворних говорника четири словенска језика (бугарски, чешки, пољски и руски). Треба рећи да говорници нису прошли никакву претходну обуку из словенске филологије, фонетике, лингвистике или форензичких наука. Њихов задатак је био да идентификују лингвистичко порекло говорника у ситуацији када су изложени ограниченим фонетским знацима. Стимулуси су се састојали од бесмислених логатома за контролу лексичке информације. Друга студија је користила машинску идентификацију говорног језика, засновану на два различита приступа: (1) формантној структури фонетског сигнала и (2) неуронској мрежи и векторској репрезентацији говорних узорака. Подаци су показали да су словенски изворни говорници, чак и када су изложени ограниченим слушним знацима, у стању да идентификују Л1 говорнике. Занимљиво је да је у контексту бугарског језика метод идентификације заснован на машинама био бољи од лаичких слушалаца. Резултати експеримента дају увид у предности хибридних приступа у истраживањима везаним за ЛАДО (Анализа језика за одређивање порекла). Штавише, резултати овог поређења могу допринети дебати о учешћу изворних говорника у поступцима идентификације Л1 за блиско сродне језике.

Кључне речи: машинска идентификација језичког порекла, аудитивно истраживање, поређење метода, словенски језици

Machine-Based and Auditory Identification of Slavic Languages

Summary. This paper presents a comparison of auditory and machine-based identification of linguistic origins. Two studies were conducted to assess the ability of lay listeners and a state-of-the-art machine approach to identify Slavic L1 from delexicalized speech samples. The first study involved 228 native

Received: 26.01.2024. **Accepted:** 15.04.2024.

Copyright © 2024 Jacek Kudera, Jovana Stevanović. Published by Vilnius University Press. This is an Open Access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

speakers of the four Slavic languages (Bulgarian, Czech, Polish and Russian) who had not received any prior training in Slavic philology, phonetics, linguistics, or forensic science. Their task was to identify the linguistic origins of speakers when exposed to limited phonetic cues. The stimuli consisted of meaningless logatomes to control for the lexical information. The second study employed machine-based identification of a spoken language, based on two distinct approaches: (1) formant structure of phonetic signal and (2) a neural network and vector representation of speech samples. The data showed that Slavic native speakers, even when exposed to limited auditory cues, are able to identify speakers' L1s. Interestingly, in the context of the Bulgarian language, the machine-based identification method performed better than the lay listeners. The results of the experiments provide insight into the advantages of hybrid approaches in investigations related to LADO (Language Analysis for the Determination of Origin). Furthermore, the outcomes of this comparison may contribute to the debate on the involvement of native speakers in L1 identification procedures for closely related languages.

Key words: machine identification of linguistic origin, auditory study, comparison, Slavic languages

Mašininis ir akustinis slavų kalbų identifikavimas

Santrauka. Šiame straipsnyje pateikiami garsinio ir mašininio slavų kalbų identifikavimo lyginmieji rezultatai. Iš viso buvo išbandyti 228 gimtakalbiai, atstovaujantys keturioms slavų kalboms (bulgarų, čekų, lenkų ir rusų). Du eksperimentai buvo skirti klausytojų gebėjimui atpažinti slavišką gimtąją kalbą iš deleksikalizuotos kalbos įvertinti ir apibūdinti moderniausio mašininio metodo veikimą. Pirmasis eksperimentas buvo susijęs su gimtosios kalbos identifikavimu, kurį atliko klausytojai, kurie buvo veikiami beprasmių logatomų. Antrajame eksperimente kalbos identifikavimo užduotis buvo atlikta naudojant (a) akustines-fonetines savybes, įskaitant formanto struktūrą (F1–F4) ir (b) neuroninį tinklą, apimančią kalbos pavyzdžių x-vektorių atvaizdavimą. Palyginus abiejų eksperimentų rezultatus, tyrime pateikiami pasiūlymai hibridinėms metodikoms LADO (angl. *Language Analysis for the Determination of Origin*) ir teismo medicinos taikymams.

Reikšminiai žodžiai: mašininis kalbų identifikavimas, akustiniai tyrimas, eksperimentinė fonetika, palyginimas, slavų kalbos

Машинско и аудитивно препознавање словенских језика – делови овога рада представљени су на Међународној конференцији IAFPA 2022: International Association for Forensic Phonetics, на Институту за фонетику Универзитета у Прагу.

1. Увод

Задатак идентификације језичког порекла јесте изузетно компликован када се ради о језицима који су типолошки и филогенетски сродни. Поступак препознавања матерњег језика (Л1) јесте од значаја не само за лингвистичка и фонетска истраживања, него и за практичну примену [Drygajlo et al. 2015; Jessen 2018; Hoskin 2018; van der Vloed & Cambier-Langeveld 2023]. Наиме, овај поступак се користи како у форензици, тако и у другим случајевима (захтев за азил нпр.) [Patrick 2010; Cambier-Langeveld 2016]. Савремена језичка технологија и примена неуралних мрежа омогућавају аугментативну реализацију задатка идентификације језичког порекла [Mateju et al. 2018]. Међутим, ако се ради о језицима са ограниченим

дигиталним ресурсима, људско ухо и лингвистичка интуиција постају кључни чиниоци у стручној анализи звучног материјала.

У овом раду се перцепцијски начин препознавања језичког порекла пореди са савременом машинском анализом говора изворних говорника словенских језика. Циљ истраживања јесте поређење резултата машинског и перцепцијског препознавања језичког порекла. У раду се примењује методологија слична форензичком поступку, који је спроведен уз помоћ хибридног методолошког апарата. Очекује се да машинско препознавање Л1 може знатно да побољша ефикасност анализе језичког порекла. Међутим, у перцепцијском делу поступка очекује се да изворни говорници словенских језика најтачније идентификују језичко порекло које је истоветно с њиховим Л1.

Први део перцепцијског поступка је заснован на идентификацији језичког порекла говорника. На основу тог теста претпоставља се да је идентификација језичког порекла могућа од стране изворних говорника словенских језика чак и у стимулусима који су безначајни. Накнадно се очекује да испитаници најуспешније препознају изворне говорнике свог матерњег језика. У машинском делу експеримента у коме је примењена неурална мрежа, претпоставља се да је систем спреман да идентификује језичко порекло говорника, узимајући у обзир анализу спектралне карактеристике сигнала. Сматра се да примењена архитектура неуралне мреже омогућава разликовање сигнала који чине фонетски слични и генетски сродни језици словенске групе и који садржи исте логатомске речи.

У првом делу рада описани су поступци у којима се често користи препознавање језичког порекла. Даље, разматра се природа задатка препознавања језичког порекла изворних говорника словенских језика. У методолошком делу представљен је задатак препознавања језика, као што је и архитектура неуралне мреже која се користи за машинско препознавање језичког порекла говорника. У том поглављу је, такође, представљен метод упоређивања сигнала на основу формантске структуре. Део који описује резултате истраживања подељен је на перцепцијски експеримент и машинско препознавање. Машинско препознавање се састоји од формантске структуре сигнала (F1–F4) и од вектора својстава (*x-vectors*). У закључном делу говори се о могућностима које проистичу из хибридне аутоматске и перцепцијске методологије.

2. Методологија

У овом истраживању се примењује хибридна методологија. Хибридна методологија утемељена је на: а) перцепцијском поступку и интуицији изворних говорника словенских језика; б) неуралној мрежи и машин-

ском препознавању језика које се заснива на векторима својстава и формантској структури сигнала.

2.1. Стимулуси

Стимулуси су били двосложни и тросложни логатоми, ткз. псеудоречи снимљене од стране 40 изворних говорника бугарског, чешког, пољског и руског језика. Списак стимулуса је обухватио шест плозива /k/, /g/, /p/, /b/, /t/, /d/ и пет вокала који су присутни у свим фонолошким саставима словенских језика /a/, /ɛ/, /i/, /ɔ/, /u/. Сви стимулуси били су састављени од консонантско-вокалних скупина (CV), те су били у складу са фонотактичним правилима тестираних словенских језика.

За говорнике руског језика речи су биле састављене графемом Э. Наиме, њихова структура у скупинама сугласник-самогласник (CVCV) није укључивала графему е, како би се узео у обзир утицај палатализације на претходни сугласнички сегмент. Такође, фонеме који су присутне у самогласничком саставу само једног испитиваног језика нису биле укључене у творбу псеудоречи, нпр. /i/ у пољском. Свака група обухвата по пет мушкараца и по пет жена, односно по 10 изворних говорника, старосне границе 21–36 година. Ниједан говорник нема говорну ману. Говорници су се служили стандардним језиком. Сваки говорник је имао задатак да прочита са списка 50 логатома. Сесија се састојала од два снимања, што је резултирало 4000 јединица: 40 говорника *50 логатомских јединица *2 снимка. Такви стимулуси су били основа за препознавање језичког порекла.

2.2. Перцепцијски поступак

У задатку идентификације језичког порекла учествовало је 228 испитаника: 50 изворних говорника бугарског, 53 чешког, 66 пољског и 59 руског језика. Испитаници су регрутирани путем сервиса *Prolific*. Њихови одговори су анонимни. Ономогућавање идентификације сваког испитаника на основу тих података спроводено је путем приписивања идентификатора свакоме испитанику. Да би се омогућило демографско профилисање, прикупљени су подаци о полу, годинама и језичким способностима испитаника. Дакле, сваки испитаник попунио је образац са основним демографским подацима и језичким способностима. У упитнику је требало означити следеће: владање страним језицима у Скали ЦЕФР А1-Ц2 [Council of Europe 2001], мултилингвизам у породици, језик свакодневне комуникације, године проведене у иностранству, ниво образовања и струка. Испитаницима је било плаћено учешће у експерименту.

Читав перцепцијски поступак био је осмишљен као компјутерска игра у којој су испитаници имали задатак да идентификују језичко порекло

особа које су осумњичене за пљачку банке. Коришћење логатомских изрза уместо правих речи се тумачило на следећи начин: осумњичени су претпостављали да су били прислушкивани, па су измислили једноставан језик за међусобну комуникацију који се заснива само на двосложним и тросложним речима. Инспектори су успели да дођу до њихових телефонских разговора. На основу телефонских разговора инспектори су дошли до закључка да су осумњичени изворни говорници неког словенског језика. Међутим, нису успели да утврде који словенски језик је у питању. Сумњали су на бугарски, чешки, пољски и руски. Ипак, на основу семантике измишљеног говора не може се утврдити порекло осумњичених. Задатак је био препознати матерњи језик на основу фонетских обележја говора [Kudera i dr. 2022]. Након стимулуса, учесници експеримента морали су да означе степен сигурности своје процене на Ликертовој скали (1–5).

2.3. Аутоматски поступак

Пре него што се примени неуронска мрежа ширења унапред (тзв. *feed-forward DNN*), на свим снимцима се аутоматски детектује глас уз помоћ алгорита аутоматске детекције гласовне активности (тзв. *Voice Activity Detection*). Аутоматски се искључују делови звука с релативно ниском енергијом. У претпроцесирању се проверавају квалитети снимака. Следећи корак јесте ексцерпција вектора својстава. На крају процеса моделирања снимака аутоматски се креира фајл који садржи нумеричке особине гласа у формату *.vmf* (*voice model file*).

2.3.1. Вектори својстава (*x-vectors*)

Процес аутоматског препознавања језика се састоји из следећих делова: ексцерпција акустичких мерила из аудио сигнала, моделирање, поређење и анализа резултата.

Архитектура дубоке неуралне мреже која је коришћена у истраживању типична је за поступак препознавања који се користи у програму *VOCALISE* [Kelly et al. 2019]. Архитектура мреже направљена је као компјутерски програм који омогућава приступ анализи и обради података кроз *GUI*.

У први део поступка спада ексцерпција мел-фреквентних цепстралних коефицијената кратког оквира, тзв. *MFCC* (*Mel-frequency cepstral coefficients*), који представљају особине људског говорног сигнала у распону од 20 до 30 милисекунди с 50% поклапања у Хаминговом прозору. Након завршетка Брзе Фуријерове Трансформације (тзв. *FFT*) која се врши с циљем да се сигнал из бременског домена анализира у спектралном до-

мену, сигнал је мултипликован у филтербанку, с циљем апроксимирања скале честотливости људског слушног апарата. Потом се примењују логаритмичка скала и дискретне косинусне трансформације (*DCT*), како би сигнал био заступљен у мањем броју коефицијената. Последњи процес у првом делу поступка јесте укључивање темпоралне димензије кроз низ тзв. делта и делта-делта коефицијената, која обухвата временске особине сигнала. Поступак је извршен према стандарду и упутима објављеним од стране девелопера компјутерског програма.

У другом делу аутоматског поступка који обухвата моделирање уз помоћ векторске репрезентације користи се дубока неурална мрежа (*Deep Neural Network*) типа ширења унапред са девет слојева и пет циклуса тренирања. У дубокој неуралној мрежи првих пет слојева моделирају статичка и временска својства гласа. Потом, резултат добијен од петог слоја мреже је статистички обрађиван на шестом слоју (тзв. *statistics pooling layer*), с циљем да се израчуна средина и стандардна девијација резултата у оквиру читавог снимка. На крају процеса, односно из излазних чворова, снимак је представљен као вектор својстава [Snyder 2018].

2.3.2. *Формантска структура*

Пре него што је векторски начин опонашања говорног сигнала постао распрострањена и примењена технологија, форензичка анализа се углавном ослањала на спектралним вредностима сигнала, као што су: основна фреквенција, формантска структура и мерила квалитета гласа [Becker et al. 2008]. И данас се у ручној анализи спектрограма користе наведена мерила, нарочито за снимке на којима се не добија пожељан резултат мел-цепстралне конверзије. Дакле, у том случају занимљиво је упоредити класичан начин анализе са савременим и тзв. *x-vectors*. Наиме, други аутоматски процес се заснивао на спектралним подацима, који обухватају прва четири форманта (F1–F4). Форманти су аутоматски ексцерпирани уз помоћ програма Праат (Praat) [Voersma et al. 2021].

Резултати

Поређење прикупљених података се заснива на тестирању хипотезе о најуспешнијем препознавању језичког порекла сопственог матерњег језика [Kudera 2022]. Нпр. изворни говорници пољског језика најуспешнији су у идентификовању пољског језика. Затим, говорници чији је матерњи језик чешки најуспешнији су у препознавању чешког језика итд. Са друге стране, пореде се и резултати добијени путем аудитивног тестирања са проценом машинске анализе звучног материјала.

2.4. Аудитивни поступак

За тестирање хипотезе о идентификацији свог језичког порекла спроведен је Пирсонов хи-квадратни тест. Анализа прикупљених података показује да су изворни говорници словенских језика најтачније идентификовали порекло осумњичених који су говорници истог матерњег језика.

У групи испитаника чији је матерњи језик пољски тачна идентификација језичког порекла достигла је 61% случајева. Чеси су, такође, најтачније идентификовали своје језичко порекло (66% случајева), као и Руси (70% случајева). Оно што је занимљиво тиче се бугарских изворних говорника. Наиме, бугарски изворни говорници су тачније идентификовали чешко порекло него бугарско. Занимљиво је, такође, упоредити асиметричне резултате између тестираних група. Дакле, код пољских испитаника други по степену идентификације је био руски језик (57%). Међутим, препознавање пољског порекла од стране руских изворних говорника било је најтеже и резултат је износио само 32%. За Пољаке је најтежа била идентификација бугарског језичког порекла (36%), док је за Бугаре најтеже било препознати пољско порекло (32%).

2.5. Аутоматски поступак – вектори својстава (x -vectors)

Аутоматска анализа је, за разлику од резултата перцепцијског поступка, тачно идентификовала Л1 за сваку језичку групу. За поређење вектора својстава користи се пробабилистичка линеарна дискриминантна анализа, тзв. ПЛДА – *Probabilistic Linear Discriminant Analysis* [Prince et al. 2007]. У тој анализи бодује се: процена максималне логаритмичке вероватноће, припадност звука са снимка једном од четири језика и процена да је језичко порекло говорника на снимку другачије. Дакле, у склопу ПЛДА анализе разматрају се две контрадикторне хипотезе унутар векторског простора. У делу претпроцесирања података 20-димензионални МФЦЦ су ексцерпирани из Хаминговог прозора дужине 32 милисекунде с 50% поклапања. Двадесет четири Мел-филтербанке (које служе за кепстралну репрезентацију обрађиваног сигнала) у спектралном оквиру од 1 до 4000 Hz служе за апроксимацију сигнала који би био сличан ономе сигналу, који је обрађиван од стране људског аудитивног састава.

На основу прикупљених података испоставило се да је примењена неурална мрежа била успешнија у задатку препознавања језика него изворни говорници словенских језика. Рангирани резултати ПЛДА анализе показују да је машински поступак делимично различит од резултата добијених у перцепцијском експерименту. Дакле, чешки стимулус био је тачно идентификован (ранг 1, ПЛДА 100), бугарски стимулус нашао се на другом месту поређења (ранг 2, ПЛДА 10,2), док се пољски стимулус

нашао на трећем (ранг 3, ПЛДА 6,6). На последњем месту, с проценом најмање вероватноће, јесте руски језик (ранг 4, ПЛДА -28,4). Бугарски стимулус је, такође, био тачно идентификован као бугарски (ранг 1, ПЛДА 107). Затим следи чешки са симетричним резултатима као код претходног поређења, па пољски (ПЛДА 5,8) и на крају руски (ПЛДА -2,9). Неурална мрежа је, такође, успешно идентификовала пољски снимак (ранг 1, ПЛДА 102). Са друге стране, неурална мрежа је успешно идентификовала и чешки снимак (ранг 2, ПЛДА 6,6). Најмање вероватна идентификација била је идентификација бугарског (ранг 3, ПЛДА 5,8) и руског језика (ранг 4, ПЛДА -13). Стимулус који је садржао руски језик био је с великом вероватноћом препознат као руски од стране неуралне мреже (ранг 1, ПЛДА 103). У другом рангу мрежа је препознала руски као бугарски (ПЛДА -2,9), а у трећем рангу је препознала пољски (ПЛДА -13). Најмања вероватноћа била је приписана чешком језику (ранг 4, ПЛДА 28,4).

2.6. Аутоматски поступак – формантска структура

За поређење резултата добијених путем анализе формантске структуре сигнала користи се косинусна удаљеност којом се упоређује сличност између састављених вектора својстава у векторском простору. Слично као у претходној методи, алгоритам процењује вероватноћу идентификације језика. Међутим, уместо ПЛДА анализе примењује косинусну удаљеност. Интерпретација дистанце захтева објашњење да већа косинусна сличност у распону $[0,1]$ означава већу вероватноћу исправне идентификације. На основу аутоматски ексцерпираних формантских вредности састав је успео да препозна да чешки снимак садржи чешки језик (косинусна удаљеност стимулуса и тестираног снимка износила је 0,99), па тек онда пољски (0,64) и бугарски (0,63). Формантска структура сигнала чешког снимка највише се разликовала од руског стимулуса (0,5). Формантска структура снимка који је садржао бугарски језик препозната је, такође, као бугарски језик са косинусном удаљеношћу израчунатом као 0,99. На другом месту по сличности формантске структуре нашао се чешки снимак (0,63), а на трећем месту пољски стимулус са косинусном удаљеношћу од 0,61. Особине формантске структуре сигнала који је представљао руски биле су најудаљеније изворном снимку и биле су типичне за руски, јер је косинусна удаљеност износила 0,56. Поређење аудио стимулуса с пољским језиком показало је успешну идентификацију Л1 (0,99). На другом месту је формантска структура чешког (0,64), на трећем формантска структура бугарског (0,61), а на последњем месту формантска структура руског језика (0,54). Изворни снимак с руским језиком био је, такође, препознат као руски Л1 с косинусном удаљеношћу

0,99. На другом месту се налази снимак са бугарским језиком (0,56), а на трећем месту снимак са пољским језиком (0,54). Најразличитија руском језику била је формантска структура чешког стимулуса. Косинусна удаљеност између ова два стимулуса износила је 0,5.

Након поређења два метода машинске анализе, испоставило се да су разлике између снимака мање у задатку машинског препознавања језика на темељу формантске структуре него онда када је говорни сигнал претворен у вектор својстава.

Закључак

На основу прикупљене грађе можемо закључити да је машинско препознавање језичког порекла тачније него перцепцијски поступак. Разлог за то треба тражити у компликованој дијалекатској слици, која утиче на квалитет самогласничких сегмената. Компликована дијалекатска слика, такође, смањује процену сигурности на Ликертовој скали. У следећем кораку било би пожељно направити детаљну прозодијску анализу која би обухватила све логатомске изразе, како би се утврдио утицај супрасегменталног нивоа на ефикасност препознавања језичког порекла. Примењена неурална мрежа је успела ефикасније да означи језичко порекло на основу анализе вектора својстава. Примењена савремена форензичка методологија показала је да је у случају машинске анализе разликовања фонетски и типолошки сродних језика неурална мрежа у стању да тачније препозна језичко порекло говорника. Акустички сигнал претворен у вектор својстава, као и формантска структура, омогућавају ефикасно препознавање језичког порекла. Наиме, вероватноћа с којом је машински систем препознао језик била је већа код репрезентације сигнала као вектора својстава него код поређења формантске структуре. Будући да у традиционалној форензици формантска структура и основна фреквенција играју важну улогу у поређењу извора говорног сигнала, парадигма која уводи векторе својстава може веома да допринесе квалитету истраживања. Даља истраживања у правцу препознавања језичког порекла би могла укључити и дијалектологију, а мрежу би требало тренирати да препознаје регионално порекло. Из овог рада можемо закључити да је хибридна методологија у поступцима који вежу лингвистику, акустичку фонетику и фоноскопију с машинским учењем успешна.

Аудитивни поступак		Машински поступак							
		Вектори својстава				Формантска структура – Косинусна удаљеност			
Л1	%	ПЛ	РУ	ЦС	БГ	ПЛ	РУ	ЦС	БГ
ПЛ	61	-	.55	.64	.61	-	.42	.23	.56
РУ	70	.55	-	.50	.57	.42	-	.12	.44
ЦС	66	.64	.50	-	.64	.23	.12	-	.27
БГ	-	.61	.57	.64	-	.56	.44	.27	-

Литература

BECKER, T., JESSEN, M., GRIGORAS, C., 2008. Forensic speaker verification using formant features and Gaussian mixture models. In *INTERSPEECH 2008: 9th Annual Conference of the International Speech Communication Association, Brisbane, Australia, September 22–26, 2008*. 1505–1508. DOI:10.21437/Interspeech.2008-432

BOERSMA, P., WEENINK, D., 2021. *Praat: doing phonetics by computer*. Version 6.1.42. <http://www.praat.org/>

CAMBIER-LANGEVELD, T., 2016. Language analysis in the asylum procedure: a specification of the task in practice, *International Journal of Speech, Language & the Law*, 23(1). 25–41. <https://doi.org/10.1558/ijssl.v23i1.17539>

Council of Europe, 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Press Syndicate of the University of Cambridge.

DRYGAJLO, A., JESSEN, M., GFROERER, S., WAGNER, I., VERMEULEN, J., NIEMI, T., 2015. *Methodological guidelines for best practice in forensic semiautomatic and automatic speaker recognition including guidance on the conduct of proficiency testing and collaborative exercises*. Frankfurt: Verlag für Polizeiwissenschaft. https://enfsi.eu/wp-content/uploads/2016/09/guidelines_fasr_and_fsasr_0.pdf

HOSKIN, J., 2018. Native speaker non-linguists in LADO: an insider perspective. In NICK, I. M. (ed). *Forensic Linguistics: Asylum-seekers, Refugees and Immigrants*. Malaga: Vernon Press, 23–40.

JESSEN, M., 2018. Forensic voice comparison. In VISCONTI, J. (ed.). *Handbook of Communication in the Legal Sphere*. Berlin, Boston: De Gruyter Mouton, 219–255.

KELLY, F., FORTH, O., KENT, S., GERLACH, L., ALEXANDER, A., 2019. Deep neural network based forensic automatic speaker recognition in VOCALISE using x-vectors. In *AES Conference on Audio Forensics, Porto, Portugal, June 18–20, 2019*. <https://secure.aes.org/forum/pubs/conferences/?elib=20477>

KUDERA, J., STENGER, I., MÖBIUS, B., AVGUSTINOVA, T., KLAJOW, D., 2022. Phonetic cues in auditory identification of Bulgarian, Czech, Polish, and Russian language of origin, *Language and Speech*, 66(3). 606–624. <https://doi.org/10.1177/00238309221119098>

KUDERA, J., 2022. *Slavic Receptive Multilingualism: Intercomprehension of Speech*. Saarbrücken: universaar, Saarland University Press. <https://dx.doi.org/10.22028/D291-36578>

MATEJU, L., CERVA, P., ZDANSKY, J., SAFARIK, R., 2018. Using deep neural networks for identification of Slavic languages from acoustic signal. In *Proceedings Interspeech 2018, Annual Conference of the International Speech Communication Association*. Hyderabad, India, 1803–1807. https://www.isca-archive.org/interspeech_2018/mateju18_interspeech.pdf

MORRISON, G. S., ENZINGER, E., HUGHES V., JESSEN, M., MEUWLY, D., NEUMANN, C., PLANTING, S., THOMPSON, W. C., van der VLOED, D., YPMA, R., ZHANG, C., ANONYMOUS, A., ANONYMOUS, B., 2021. Consensus on validation of forensic voice comparison, *Science & Justice*, 61(3). 299–309. <https://doi.org/10.1016/j.scijus.2021.02.002>

PATRICK, P., 2010. Language Variation and LADO. In ZWAAN, K., VERRIPS, M., MUYSKEN, P. (eds.). *The Role of Language in European Asylum Procedure: A Linguistic and Legal Survey*. Nijmegen: Wolf Legal Publishers, 73–87.

PRINCE, S., ELDER, J., 2007. Probabilistic Linear Discriminant Analysis for Inferences about identity. In *IEEE 11th International Conference on Computer Vision (ICCV)*, 1–8.

SNYDER, D., GARCIA-ROMERO, D., SELL, G., POVEY, D., KHUDANPUR, S., 2018. X-Vectors: Robust DNN Embeddings for Speaker Recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018*. 5329–5333. <https://doi.org/10.1109/ICASSP.2018.8461375>

van der VLOED, D., CAMBIER-LANGEVELD, T., 2023. How we use automatic speaker comparison in forensic practice, *International Journal of Speech, Language and the Law*, 29(2). 201–224. DOI: 10.1558/ijssl.23955