Article

Language and Speech

Phonetic Cues in Auditory Identification of Bulgarian, Czech, Polish, and Russian Language of Origin

Language *and* Speech
2023, Vol. 66(3) 606–624

DOI: 10.1177/00238309221119098
journals.sagepub.com/home/las
SAGE

Jacek Kudera, Irina Stenger, Bernd Möbius, Tania Avgustinova
Department of Language Science and Technology, Saarland University, Germany

Dietrich Klakow
Department of Language Science and Technology, Saarland University, Germany; Spoken Language Systems (LSV), Saarland University, Germany

Abstract
This work presents the results of an auditory language of origin identification experiment. Disyllabic and trisyllabic logatomes were recorded by speakers of Bulgarian, Czech, Polish, and Russian, and presented to L1 speakers of the abovementioned Slavic languages. The goals of the test were to verify the ability of lay listeners to recognize the linguistic origin of speakers, based on spoken samples with limited segmental and suprasegmental information, and to correlate the signal features with the subjects' performance. It was found that position of word stress is not an important predictor in language recognition. However, inherent vowel characteristics such as duration and vowel space computed by the means of Pillai scores correlate with subjects' performance. Both the linguistic profile and the familiarity with closely related languages also appear to be relevant predictors of listeners' performance. Finally, the information-theoretic notion of surprisal applied on regular cross-linguistic sound correspondences was correlated with recognition scores; though, the correlations did not reach the threshold of statistical significance. We conclude that auditory identification of linguistic origin by lay persons, native speakers of closely related languages, is possible even when exposed to limited segmental information, which can serve as a cue in the identification of linguistic origin.

Keywords
Slavic languages, information theory, phonetics, language identification, LADO

Corresponding author:
Jacek Kudera, Department of Language Science and Technology, Saarland University, Campus C7 2, 66123 Saarbrücken, Germany.
Email: kudera@coli.uni-saarland.de

1 Introduction

Spoken language identification (henceforth: LID) is a complex process of perceptual recognition or automatic identification of a language from a spoken sample. Recently, fine-grained LID, including areal and dialectological investigations, has become widely applicable in a procedure known as LADO: Language Analysis for the Determination of Origin (Patrick et al., 2012). Also in forensics, machine-based and perceptual LID is an important stage of spoken evidence validation and speaker profiling. The scientific debate on the relevance of non-linguist native speakers in LADO or LOID (Linguistic Origin Identification) is vibrant and polarized with arguments both in favor of such an approach when supervised by professionals (Cambier-Langeveld, 2016; Nolan, 2012; Wilson & Foulkes, 2014), and against the involvement of lay listeners altogether (Eades, 2005; Fraser, 2011; Patrick, 2010). The arguments for the involvement of lay persons are the lack of independent supervision of experts conducting LADO, as well as monopolist practices. Furthermore, the question of reliable counter-expertise contributes to the engagement of lay listeners in the task of L1 identification. The lack of academic literature on a particular vernacular can justify the engagement of non-trained native speakers in the identification of linguistic origin. Furthermore, a comparison of the accuracy of LADO professionals, academic phoneticians, phonetics students, and untrained native speakers has indicated that the last group performed best, whereas the LADO professionals' results were at chance level only (Foulkes & Wilson, 2011). The authors suggest that academic linguists may pay attention to different cues than native speakers. In contrast, the arguments for the exclusion of native speakers in LADO procedures touch upon non-experts' unfamiliarity with the proper terminology, or their possible bias in providing an opinion. In addition, there is high variability in listener accuracy and it is difficult to predict which listeners might perform better than others. Fraser (2009) has pointed out that untrained listeners often fail to identify dialects correctly. Since trained professionals and lay persons often pay attention to different cues and may have developed different skills due to training versus natural acquisition, a collaborative involvement of both parties seems to be an ideal solution.

It is also interesting to explore the relation of linguistic fluency and performance of lay listeners in LID tests, as well as the phonetic and phonological correspondences between listeners' L1 and perceived speech. The importance of phonological insights in auditory LID has previously been stressed by Peperkamp and Dupoux (1992), who suggested that listeners' sensitivity to stress cues depends on the stress function in their native language. Even when a particular feature associated with word stress is not phonemic in participants' L1 and does not play a role in lexical discrimination in one's own language, listeners might still be more sensitive to stress distinctions in a foreign language. The effect of stress predictability was also reported in the perception of non-words for speakers of languages exhibiting phonemic stress. Regarding the participants in this study, a weak "stress deafness" has previously been observed among Polish native speakers (Peperkamp et al., 2010).

Regardless of the uneven functional load of the units used in LID tasks, several studies have confirmed the possibility of identifying a language when exposed to limited information and distorted speech. Hence, apart from the inherent phonetic features, syllable structure can also serve as a cue in LID, especially in Slavic languages, which are known to exhibit complex onset clusters such as CCCCVC, in both Polish and Russian. Information conveyed in phonotactic rules of one language can intuitively lead to successful language identification by humans as well as by automatic LID systems (Navrátil, 2001; Zissman & Berkling, 2001). Prosodic cues are similarly meaningful in auditory language identification. Several techniques of limiting the spectral information in perceptual LID tasks have previously been applied, including spectral envelope removal and temporal envelope modulation. Studies involving modifications of speech signals have confirmed that prosodic cues play an important role in LID, even when separated from the segmental information (Ohala & Gilbert, 1981). Even though prosody is rarely, or never, decisive in contemporary LOID assessments (Hoskin, 2018), research on suprasegmental cues in LID has led to the proposal of a rhythmic model of language identification by Rouas et al. (2005). Regardless of signal distortion, the accuracy of language recognition still varies depending on the source data type: read or spoken, with higher accuracy on the former. For instance, a domain-dependency in a machine LID on the same set of Slavic languages was reported by Abdullah et al. (2020).

1.1 Aims and hypotheses

In the present work, a human-based auditory language identification task was conducted using spoken material from four Slavic languages: Bulgarian, Czech, Polish, and Russian. The main goals of this work are to define segmental and suprasegmental piece of information required for correct recognition of speakers' L1 in human-based LID; to correlate the phonetic and typological similarities of participants' L1 and the respective stimulus with their test performance; to investigate the relation of the three-dimensional (3D) vowel overlap measured by means of Pillai score (Hay et al., 2006; Nycz & Hall-Lew, 2013) with the subjects' performance in LID tests; and to examine the effect of sound variation on lay listeners' ability of language identification from an information-theoretic perspective (Shannon, 1948). The study combines the following methodological components. First, the acoustic-phonetic component provides an explanation of cross-linguistic influence of vowel space overlap on test performance. Second, the study explains the role of listeners' L1 in the recognition of linguistic origins of Slavic speakers. Third, an information-theoretic approach attempts to quantify cues in LID and compare the recognition scores with the information-theoretic notion of surprisal.

We assume that lay listeners whose L1 is closely related to the language of the recording can correctly identify the linguistic origin of speakers in the auditory modality even from logatomes with limited semantic information; the alternations in stress position and length of the logatomes influence the performance, with better performance on longer logatomes; spectral characteristics of the signal, such as cross-linguistic vowel space and duration overlap, are correlated with human performance in LID tests; and the mean logatome identification surprisal (LIS) values between the tested languages are correlated with the experimental results, that is, the lower the mean LIS values between two languages, the higher the LID scores. Moreover, it is worth investigating which phonetic features and their cross-lingual overlap improve listeners' performance. What is the optimal threshold of phonetic and acoustic information required for correct identification of speakers' origins? Do suprasegmental features and language-specific word stress distribution play a role in LID tasks? How can the information-theoretic notion of surprisal help us determine and predict human performance at identification of closely related languages? These issues will be addressed in this study with a focus on spectral and temporal properties of segments that are common in the phonological inventories of the investigated Slavic languages.

1.2 LADO and auditory language identification

Perceptual language recognition is a complex operation, involving several cognitive processes. It is a multidimensional action without discrete component stages, in which graded information flows in a cascade (McQueen et al., 2003). This process is based on various sources of acoustic, linguistic, and extralinguistic information. Perceptual as well as machine-based LID techniques can be improved with training. Studies in contemporary cognitive linguistics have shown that the ability of humans to identify a language can be significantly improved by training or exposure to a particular language or vernacular, similar to automatic LID, in which the size of training data often predicts overall system performance (Muthusamy et al., 1994). Both methods, perceptual and machine-based LID, are currently widely applied not only for commercial purposes, such as translation and localization, but also in the field of jurisdiction and forensics. Governments and intelligence bureaus often take advantage of associated or external agencies to perform LID for the needs of criminal investigations. Furthermore, LADO tests are widely used to verify the origin of asylum seekers, by closely investigating their speech using the methodological apparatus of acoustic phonetics, linguistics, and dialectology. In such a procedure, there is an underlying assumption of an existing connection between how people speak and their ethnic or national origin, or rather the place of their socialization, which is a natural consequence of language acquisition in a certain linguistic community. Nevertheless, lay listeners' expertise is often valuable in linguistic, or more precisely, dialectal background evaluation. "Guidelines for the use of language analysis in relation to national origin in refugee cases" suggest that lay persons should not be treated as experts when evaluating a speaker's origin on the basis of their speech (Language and National Origin Group, 2004). On the other hand, especially for cases involving languages with limited digital resources, agencies specialized in conducting LADO or LOID tests often make use of lay persons in cooperation with linguists to estimate the target's place of origin (Hoskin, 2018). Such a practice is partly pragmatic, having in some cases to do with lack of descriptions or linguists specializing in the languages concerned. This is also founded on the widely accepted principle that native speakers are in general the most knowledgeable informants.

The so-called "intelligent guessing" (Meissner, 2018) of a language depends on listeners' linguistic and extralinguistic knowledge, exposure to other languages, associative competence, as well as short-term and long-term memory storage and, above all, the quality and quantity of information perceived in the process of recognition. It has been shown that even when a signal for auditory identification is highly degraded, subjects are aware of the cues they perceive in LID tasks (Muthusamy et al., 1994). Furthermore, information provided in LID itself can be limited to a particular subsystem of a language (Schultz & Waibel, 1998). For instance, it is possible to distinguish two languages based solely on the presence of a phone in one language and its absence in another (Harper & Maxwell, 2007). Strictly phonetic knowledge in LID based on the characteristics of vowel systems was proposed by Pellegrino and André-Obrecht (2000). Articulatory classes have also been investigated as delimitation features in language recognition (Kirchhoff & Parandekar, 2001). In this application, several distinguishing labels have been defined, for example, manner of articulation, consonantal place of articulation, vocalic place of articulation, lip rounding, front-back tongue position, voicing, and nasality. Phonotactic rules in combination with labeling of broad phonetic classes can also constitute a kernel of the language identification process. This approach was proposed by House and Neuburg (1977). More fine-grained analyses involving strictly acoustic signal characteristics seem to be of relevance to automatic LID. Perceptual behavioral studies concerning acoustic data and LID are canonically gravitating toward a correlation of characteristics of formant dynamics, voice onset time (VOT), and center of gravity (CoG) along with their fluctuations in the signal with language-specific data. Hence, we wonder whether there is a cross-linguistic correlation of spectral and temporal features of vowel systems with the performance in auditory recognition of speakers' origins (Cambier-Langeveld, 2016; Nolan, 2012).

2 Methods

The task was presented as a game in which subjects played the role of investigators in a bank robbery case. Their task was to identify the origin of a speaker by listening to an artificial language made up by the speakers-suspects to mislead the investigators. This setup provides the rationale for the application of pseudowords in the LID sessions and draws the participants' attention to non-lexical cues.

The auditory language identification task was given to 228 lay persons—native speakers of four Slavic languages. To control for segmental and suprasegmental information in real words of the test languages, the test was based on pseudowords. These logatomes consisted of CVCV and CVCVCV (25 each) sequences with equal distribution of stress placements to reflect the lexical stress distribution patterns of the Slavic languages. The participants were asked to select one of the four languages which they believed to be the native language of the speaker in the recording, similar to common LADO procedure. This setup avoids the effect of associating lexical knowledge and language-specific consonant structures with a particular Slavic language. The experimental setup was based on a quasi-LADO paradigm, in which the performance of lay listeners is an important element taken into consideration during the investigation process.

2.1 Speakers

In total, 40 native speakers (five males and five females per each tested language) of Bulgarian, Czech, Polish, and Russian were asked to read a list of nonsense disyllabic CVCV and trisyllabic CVCVCV items according to the accentuation patterns of their native languages. Based on a questionnaire distributed after the recording sessions, the speakers whose voices were selected for use in the experiment were profiled as speakers of the standard language variety, of middle socioeconomic status, having completed secondary education or currently enrolled at a university program, aged 21 to 36, and having experienced no surgical operations in the ear, nose, and throat region nor having required speech or hearing therapy.

2.2 Participants

In total, 228 speakers of four Slavic languages (50 Bulgarian, 53 Czech, 66 Polish, and 59 Russian speakers) participated in the task. Before the experiment, participants filled out a questionnaire with basic demographic information and questions about their linguistic background. The participants were asked about their linguistic proficiency (CEFR – Common European Framework of Reference for Languages scale), multilingualism within their family, their language of everyday communication, education background, and years spent abroad. None of the participants reported any hearing difficulties. Since previous studies revealed that multilingualism significantly influences performance in identifying an unfamiliar language (Muthusamy et al., 1994), the results were post hoc correlated with the linguistic profiles from the questionnaires. The participants had no formal training in other Slavic languages. Data from subjects with a background in Slavistics, linguistics, forensics, or phonetics ($<1\%$) were excluded from analysis.

2.3 Design of materials

To avoid the possibility of overlapping with existing lexemes of Slavic languages and associating with meaningful tokens, the NUP (nonword uniqueness point) had to be achieved for every item presented in the identification task to ensure that logatomes do not resemble meaningful lexemes in tested languages (Cutler, 2012). Furthermore, the stimuli contained only vowels and consonants that are present in the phonological inventories of all four investigated languages. Stimuli with a fixed stress position were recorded in line with the natural stress distribution rules in these languages, that is, initial syllable stressed in Czech, penultimate in Polish, and flexible stress in

Bulgarian and Russian. To avoid the effect of unnatural articulation and audibly perceptible reading difficulty, the set of stimuli consisted only of sequences that yielded effortless and natural pronunciation. The pseudowords were marked as easy to read by the speakers, which ensured effortless articulation and natural rendering of the spoken samples (Bonatti et al., 2005).

The set of stimuli was created according to the following rules: (1) all logatomes were in line with the open-syllable principle, the common law in Slavic before the vocalizations of the semi-vowels; (2) the items consisted of stops /k/, /g/, /p/, /b/, /t/, /d/ and a combination of five common vowels /a/, /ɛ/, /i/, /ɔ/, /u/ which was justified by the degree of interference of plosives with the adjacent vocalic segment (Ladefoged & Maddieson, 1996; Stevens & House, 1963), as well as by the results from previous studies using logatomes in perceptual discrimination, which showed that the diphones "pa," "si," and "ki" are among the most discriminable elements (Pascal et al., 1989); (3) to control for a fixed stress position and length of the pseudowords, both bisyllabic CVCV and trisyllabic CVCVCV sequences were used in the test; (4) no zero-onset was present in the tokens, even though this structure is possible in all investigated languages; (5) primarily non-palatalized segments were used due to the unequal distribution and frequency of palatalized CV sequences in the investigated languages; (6) to avoid a priming effect, only non-nasalized units were taken into consideration due to frequent synchronous and asynchronous nasals in Polish (Kudera, 2018) as opposed to other Slavic languages.

2.4 Recording procedure

A sex-balanced group of 40 speakers of the Slavic languages was given a self-paced reading task in an acoustically controlled setting. The readers were native speakers of one of the four Slavic languages, who evaluated themselves as users of the standard variety of the language. To avoid possible uncontrolled effects and paralinguistic distractors, such as recording-related anxiety causing different speech rates, coughing, each list of pseudowords was read and recorded twice, which resulted in 4,000 tokens (40 readers × 50 tokens × 2 sessions). The recordings were randomized and intervals between the tokens were standardized. These bisyllabic and trisyllabic samples were used as the audio stimuli in the LID task.

2.5 Speech analysis procedure

Before the experiment, the stimuli were automatically segmented and annotated using the BAS annotation tool (Kisler et al., 2017), visually inspected, and, if necessary, manually corrected in Praat (Boersma & Weenink, 2020). The F1 and F2 values in critical band rate (Bark) were extracted from the midpoint of the vowels using LPC (Linear Predictive Coding) Burg's method. The results of these estimations were visually inspected and compared with the corresponding wideband FFT (Fast Fourier Transform) spectra. The manual verification and alignment with the nearest zero-crossing correction allowed for precise duration measures (in ms). Then, the vowel space overlap of the five investigated vowels was calculated by means of the Pillai method (Hay et al., 2006; Nycz & Hall-Lew, 2013) as a multivariate analog of the F ratio from the analysis of variance (ANOVA; F1 + F2 + duration ~ vowel) based on ANOVA (Bray & Maxwell, 1985; Pillai, 1954). As suggested by recent results from the Monte Carlo simulation, this method outperforms the other vowel overlap measures such as SOAM (spectral overlap assessment metric), VOACH (vowel overlap analysis with convex hulls), or Euclidean distance between centroids of respective vowel spaces (Kelley & Tucker, 2020). Furthermore, computed Pillai scores also take into consideration a 3D plane including the duration of segments, a factor highly relevant in the LID process and important when languages with significant differences in vowel lengths are considered.

2.6 Experimental procedure

In the first step, we examined the acoustic properties of segments in the spectral and temporal domains. Then, we introduced the fixed versus free stress-position factor, by implementing more complex stimuli with variations in stress positioning: namely, on the initial and penultimate syllables versus free stress distribution. The randomized disyllabic and trisyllabic logatomes were binaurally presented to the subjects. To exclude any orthographic influence, no visual input was given. The participants were provided with a controlled amount of acoustic information, excluding semantics. They were allowed to listen to the samples three times. Before the identification task began, subjects were given the opportunity to practice on trial samples, at which time they could adjust the volume to a comfortable level. During the entire session, participants were exposed to 64 randomized samples containing long and short logatomes. After each sample, they were asked to identify the language of the speaker by choosing from a closed set of Slavic languages: Bulgarian, Czech, Polish, and Russian. This setup corresponds to the speaker verification task, a common practice in LADO procedures. A 5-point Likert-type scale was used to rate the subjects' confidence in their choice. After having completed the entire session, the subjects were presented with their overall accuracy scores.

2.7 Analysis

A language confusion matrix was created from the results of the auditory LID task. Afterward, a similarity index (SI) of the investigated Slavic languages was calculated (Johnson, 2003). The perceptual similarity index (PSI) of Bulgarian, Czech, Polish, and Russian was calculated from the similarity scores (Thomas, 2011). In the next step, the performance of lay listeners was correlated with acoustic segmental similarity measures by means of the 3D Pillai scores across the four languages. The results of the Slavic language of origin recognition task were compared afterward with the amount of acoustic and phonetic information given in the samples. For the purposes of the LIS calculation, the logatomes were narrowly transcribed in the IPA (International Phonetic Alphabet) 2020 standard and were then paired with their equivalents in the other three languages. As a result, 19,164 logatome pairs were transcribed (3,283 Bulgarian-Czech logatome pairs, 3,305 Bulgarian-Polish, 3,304 Bulgarian-Russian, 3,084 Czech-Polish, 3,083 Czech-Russian, and 3,105 Polish-Russian). The numbers are different for each language pair due to some mispronunciations by the readers, yielding new logatomes with no equivalents. A logistic regression analysis was conducted to address the research question regarding differences in recognition based on logatome length. Then, a Pearson's χ^2 test was conducted to test the relation between Slavic L1 and participants' performance, and to correlate the LIS with the similarity scores. The performance scores were correlated with the subjects' linguistic profiles in terms of their mother tongue and fluency in non-Slavic languages. Language similarity sequences for disyllabic and trisyllabic logatomes were compared with respect to stress distribution patterns in Bulgarian, Czech, Polish, and Russian. Finally, the results in the language confusion matrix (the LID scores) were correlated with the mean LIS values for each language pair, to investigate the effect of regularities in cross-linguistic sound correspondences on listeners' performance in auditory identification of language of origin.

3 Results

The following sections introduce the results of the experiment according to each methodological component.

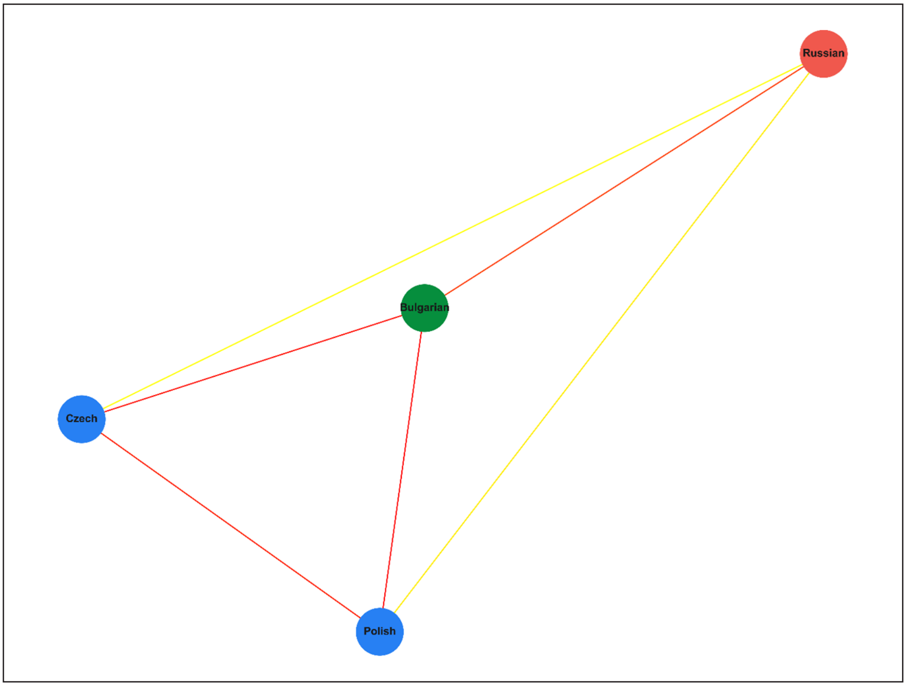

Figure 1. Distance based on perceptual similarity scores. The colors of the language nodes represent the genealogical groupings within the Slavic family (green—South Slavic, blue—West Slavic, red—East Slavic). The edges represent the similarity of each pair of languages.

3.1 Perceptual similarity score

The confusion matrix generated from the results of the language identification task served as the basis for the calculation of the PSIs. The pairwise PSI from the overall performance was calculated as Equation 1:

$$PSI = -\ln\left(\frac{P_{xy} + P_{yx}}{P_{xx} + P_{yy}}\right)$$

in which P_{xy} stands for proportion of Language X identified as Language Y, P_{yx} equals to proportion of Language Y identified as Language X, P_{xx}—proportion of Language X identified as X, and P_{yy}—proportion of Language Y identified as Y. The similarity scores obtained from the overall performance were as follows: for Polish and Russian, 0.25; for Czech and Polish, 0.48; for Bulgarian and Czech, 0.52; for Polish and Bulgarian, 0.55; for Czech and Russian, 0.21; and for Bulgarian and Russian, 0.47. The results are illustrated in the network plot (Figure 1). The distance between language nodes in the network plot reflects the perceptual similarity scores (PSI) between the languages. The network plot was created by projecting language nodes into a two-dimensional space using a force-directed graph function and multidimensional scaling, based upon the perceptual similarity distance scores yielded from the experiment. The colored edges connecting the language nodes are scaled according to the degree of similarity of each pair of languages, with

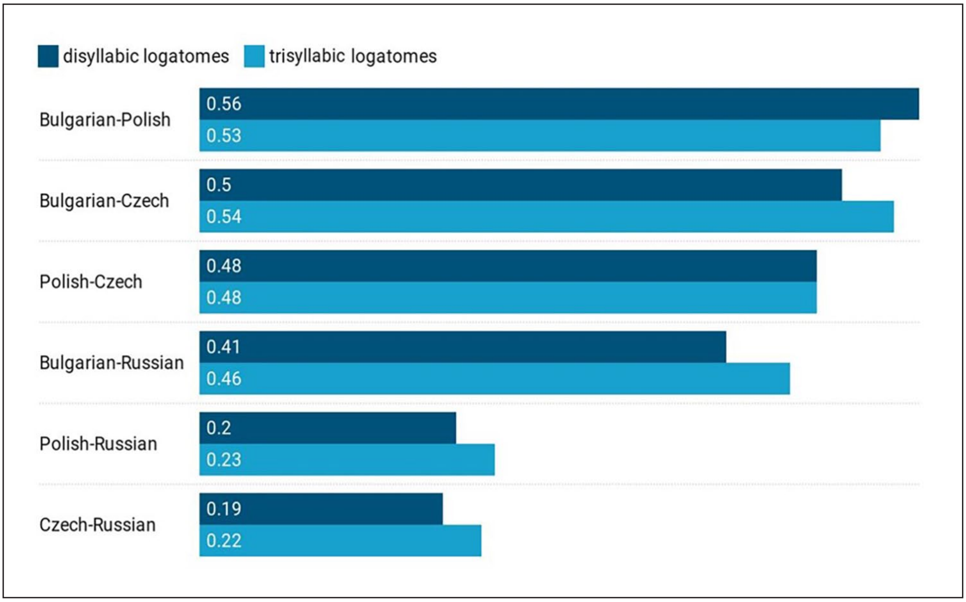

Figure 2. Pairwise similarity scores for disyllabic and trisyllabic logatomes.

darker colors representing greater similarity. The alternations in similarity scores between the disyllabic and trisyllabic sequences are presented in Figure 2.

3.2 Logatome length and stress position

A logistic regression analysis was conducted to address the research question regarding the differences in identification based on logatome length. A threshold of statistical significance $\alpha = .05$ was applied. The hypothesis of more accurate recognition of longer trisyllabic sequences with respect to disyllabic logatomes was rejected, $\chi^2 = 2.28$; $p = .131$. The length of the logatomes was not a relevant predictor of correct identification, Cox and Snell's $R^2 < .01$. The correct identifications of disyllabic stimuli reached 46.8% ($n = 3,397$), whereas the accuracy of recognition after CVCVCV sequences was 45.6% ($n = 3,304$). Detailed results are presented in Table 1. The overall perceptual similarity for all tested languages is shown in Figure 2. Nuances in stimuli length resulted in different clustering based on subjects' performance. The divergence of Russian compared with the other three Slavic languages is constant among the sequences. The similarity among Bulgarian, Czech, and Polish, however, differs with respect to the length of the logatomes. The results from the overall performance indicate that Polish is perceptually closer to Bulgarian than to Czech. A closer examination of the scores with respect to the length of the pseudowords reveals alternations between Polish and Bulgarian and shows a greater similarity between Czech and Bulgarian instead (Figure 3).

3.3 Native Slavic language and non-Slavic L2

Test participants were lay persons—native speakers of four Slavic languages: Bulgarian, Czech, Polish, and Russian. Their fluency in non-Slavic languages was documented in the questionnaire. Interestingly, fluency in non-Slavic languages also appeared to be a predictor of their LID

Table 1. Results of Logistic Regression Analysis for Correct Responses Depending on Logatome Length.

	B	*SE*	*Z*(1)	*p*	*OR*	95% *CI* for *OR*	
						LL	UL
Logatome length	0.05	0.03	2.28	.131	1.05	0.99	1.12
Const.	−0.18	0.02	56.37	<.001	0.84		

Note. B: unstandardized coefficient; *SE*: standard error; *Z*: Wald test; *p*: *p* value; *OR*: odds ratio; *CI*: confidence interval, LL: lower limit, UL: upper limit.

performance. The results showed that the model fits the research question well, $\chi^2 = 126.18$; $p < .001$; Cox and Snell's $R^2 = .009$.

Apart from Slavic languages, fluency in Romanian and Azeri appeared to accurately predict the LID scores, even though their systems of monophthongs differ from those of Slavic languages. L1 speakers of Polish had 39% higher correct identification, odds ratio (*OR*) = 1.39. Fluency in Czech increased performance by 41%, *OR* = 1.41, whereas knowledge of Russian improved the accuracy by 17%, *OR* = 1.17. The effect of fluency in Bulgarian did not reach the threshold of statistical significance. Regarding languages that are typologically unrelated to Slavic languages, fluency in Romanian increased the performance in the LID task 1.6 times, *OR* = 2.60, whereas the fluency in Azeri lowered the performance by 77%, *OR* = 0.23. Results including all languages are presented in Table 2.

3.4 Native Slavic language and identification scores

In the next step, the relation between listeners' Slavic L1 and accuracy of identification was investigated by means of a Pearson's χ^2 test. The analysis revealed statistically significant results for each group of respondents. The scores are presented in Table 3 and illustrated in Figure 4. In the group of Bulgarian native speakers, the best recognition scores were observed for Czech, $\chi^2(3) = 42.29$; $p < .001$; $V = 0.12$. Polish was identified in 33.8% of cases; correct identifications of Russian reached 40.1%; whereas the group L1 identification scores, 46.3%, were similar to Czech and Russian. In the group of Czech native speakers, the highest discrimination scores, 66.3%, were measured for Czech, $\chi^2(3) = 224.55$; $p < .001$; $V = 0.25$. Less accurate identifications were observed for speakers whose L1 was Russian 52.5%; Polish, 38.6%; and Bulgarian, 34.2%. Polish native speakers accurately recognized speakers' dominant language as Polish in 60.7% of cases, $\chi^2(3) = 176.81$; $p < .001$; $V = 0.20$. The second most correctly identified language was Russian, 56.8%. The recognition scores of Bulgarian and Czech ranged from 36% to 42%. The native speakers of Russian identified Russian in 70.2% of cases, $\chi^2(3) = 317.75$; $p <,001$; $V = 0.31$. The least accurately recognized linguistic origins were Czech, 41.4%; Bulgarian, 34.3%; and Polish, 32.4%.

3.5 Acoustic and perceptual measures: vowel overlap

Pillai scores were computed to discover the relation between vowel overlap and performance in language identification. The results in Table 4 correspond to the cross-lingual overlap of five vowels calculated using the Pillai method, including durations of segments and their F1 and F2 values (F1 + F2 + duration ~ vowel). The Pillai scores reflect the 3D similarity of vocalic segments in Bulgarian, Czech, Polish, and Russian. Higher Pillai scores indicate more diverged (less overlapping) segments. The results show that the overlap of low-central /a/ tokens was not correlated with

Table 2. Results of Logistic Regression Analysis for All Subjects' Languages.

	B	*SE*	*Z*(1)	*p*	*OR*	95% *CI* for *OR*	
						LL	UL
Hebrew	0.18	0.17	1.23	.268	1.20	0.87	1.66
Swedish	0.68	0.44	2.38	.123	1.98	0.83	4.73
Belarusian	0.59	0.54	1.18	.276	1.80	0.62	5.21
Norwegian	0.60	0.31	3.79	.052	1.82	1.00	3.33
Spanish	0.01	0.16	0.01	.928	1.01	0.74	1.39
Portuguese	0.18	0.19	0.84	.358	1.19	0.82	1.74
Greek	0.10	0.27	0.14	.713	1.10	0.66	1.85
Serbian	0.36	0.37	0.94	.331	1.43	0.70	2.92
Croatian	−0.19	0.44	0.20	.656	0.82	0.35	1.93
Polish	0.33	0.08	16.26	<.001	1.39	1.19	1.64
Japanese	0.24	0.27	0.78	.377	1.27	0.75	2.15
Romanian	0.95	0.24	16.48	<.001	2.60	1.64	4.11
Ukrainian	−0.30	0.28	1.20	.273	0.74	0.43	1.27
English	0.04	0.11	0.14	.711	1.04	0.85	1.28
Italian	−0.09	0.17	0.25	.615	0.92	0.66	1.29
Finnish	0.33	0.27	1.52	.218	1.40	0.82	2.37
French	−0.17	0.19	0.84	.360	0.84	0.59	1.21
German	−0.01	0.10	0.01	.940	0.99	0.81	1.21
Czech	0.34	0.08	18.15	<.001	1.41	1.20	1.65
Russian	0.16	0.08	3.99	.046	1.17	1.00	1.37
Afrikaans	0.10	0.20	0.25	.617	1.10	0.75	1.62
Azeri	−1.46	0.54	7.23	.007	0.23	0.08	0.67
Bulgarian	0.10	0.09	1.48	.224	1.11	0.94	1.31
Const.	−0.41	0.08	28.48	<.001	0.66		

Note. B: unstandardized coefficient, *SE*: standard error, *Z*: Wald test, *p*: *p* value, *OR*: odds ratio, *CI*: confidence interval, LL: lower limit, UL: upper limit.

correct language recognition, whereas types of /ɛ/, $OR=0.43$, and /u/, $OR=52.97$, were highly correlated with Pillai scores. More divergent vowels correlated with greater accuracy in the recognition of the language it belongs to. The detailed results are given in Table 5. Although vowel space is rarely mentioned as discriminable on the word level in classical models of speech processing, it appears to have had an influence when sample discrimination is performed on highly distorted spoken samples.

3.6 Perceptual measures: LIS

As suggested by Skirgård et al. (2017), languages that sound differently appear easier to distinguish from one another than phonetically close languages. An information-theoretic notion of surprisal is one method for quantifying these differences. In this context, surprisal metrics quantify the informativity of cross-linguistic unit correspondences in bits. In the current LADO setting, sound identification surprisal (SIS) is computed according to Equation 2:

$$SIS\left(L1=s1\mid L2=s2\right)=-\log_2 P\left(L1=s1\mid L2=s2\right)$$

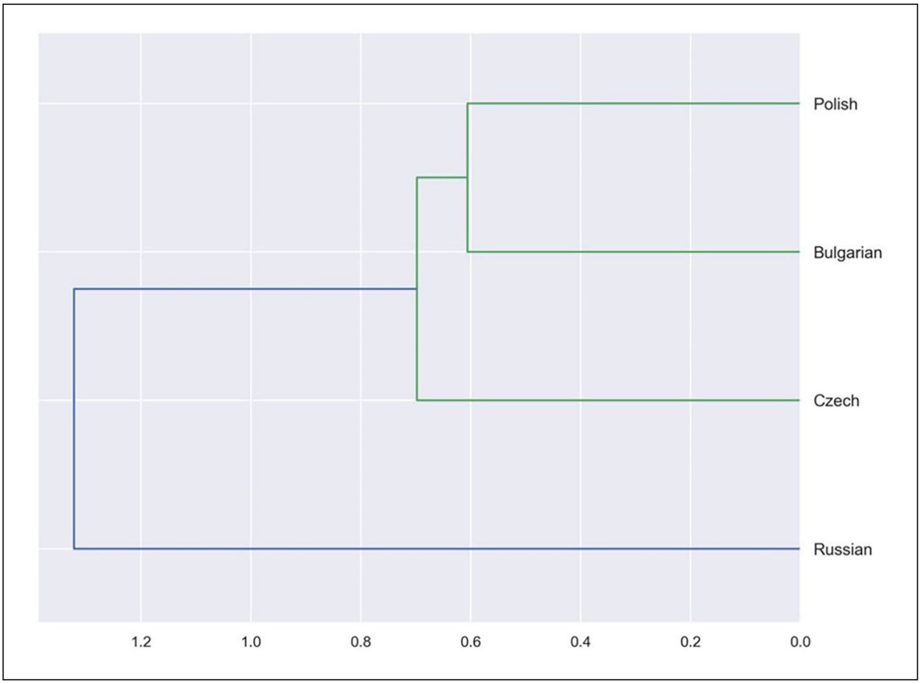

Figure 3. Dendrogram clustering language similarities based on subjects' performance. Language clustering based on the overall results.

Table 3. Overall Scores With Regard to Slavic L1.

Native language	Correctness	Response language							
		Bulgarian		Czech		Polish		Russian	
		n	%	*n*	%	*n*	%	*n*	%
Bulgarian	−	$395_{a,b}$	53.7	374_{b}	50.8	487_{c}	66.2	$441_{a,c}$	59.9
	+	$341_{a,b}$	46.3	362_{b}	49.2	249_{c}	33.9	$295_{a,c}$	40.1
Czech	−	580_{a}	65.9	296_{b}	33.7	540_{a}	61.4	418_{c}	47.5
	+	300_{a}	34.1	583_{b}	66.3	340_{a}	38.6	462_{c}	52.5
Polish	−	704_{a}	64.1	629_{b}	56.7	431_{c}	39.3	473_{c}	43.2
	+	394_{a}	35.9	481_{b}	43.3	667_{c}	60.7	621_{c}	56.8
Russian	−	559_{a}	65.7	502_{b}	58.6	577_{a}	67.6	254_{c}	29.8
	+	292_{a}	34.3	355_{b}	41.4	277_{a}	32.4	599_{c}	70.2

Note. a, b, c indexes—differences on the level $p < .05$ with Bonferroni correction.

in which $L1$ stands for response (decoder) language, $s1$ corresponds to sound of the response (decoder) language, $L2$—exposure (stimulus) language, and $s2$—sound of the exposure (stimulus) language.

The SIS values obtained for combinations of stimulus and decoder languages allowed the quantification of the (un)expectedness of cross-linguistic sound correspondences and of the

Table 4. Vowel Overlap in Pillai Scores.

	Vowels				
Languages	*a*	*e*	*i*	*o*	*u*
Cz-Pl	0.031	0.075	0.100	0.032	0.031
Cz-Ru	0.017	0.060	0.041	0.013	0.100
Cz-Bg	0.015	0.359	0.032	0.058	0.043
Pl-Ru	0.071	0.202	0.069	0.026	0.055
Pl-Bg	0.063	0.488	0.006	0.059	0.037
Ru-Bg	0.045	0.163	0.025	0.017	0.027

Note. Bg: Bulgarian, Cz: Czech, Pl: Polish, Ru: Russian.

Table 5. Vowel Overlap in Pillai Score and Subjects' Performance.

						95% *CI* for *OR*			
	B	*SE*	*Z*(1)	*p*	*OR*	LL	UL	Cox and Snell's R^2	χ^2
a	−0.50	0.46	1.19	.276	0.61	0.25	1.49	<0.01	1.19
const.	−0.35	0.05	48.34	<.001	0.71				
e	−0.85	0.18	21.13	<.001	0.43	0.30	0.62	<0.01	21.26***
const.	−0.16	0.05	8.42	.004	0.86				
i	1.00	0.56	3.13	.077	2.72	0.90	8.21	<0.01	3.14
const.	−0.51	0.06	83.87	<.001	0.60				
o	−0.86	0.65	1.74	.187	0.42	0.12	1.52	<0.01	1.74
const.	−0.25	0.06	15.95	<.001	0.78				
u	3.97	0.68	34.58	<.001	52.97	14.11	198.89	0.01	34.75***
const.	−0.67	0.07	89.47	<.001	0.51				

Note. B: unstandardized coefficient; *SE*: standard error; *Z*: Wald test; *p*: *p* value; *OR*: odds ratio; *CI*: confidence interval, LL: lower limit, UL: upper limit.
***$p < .001$.

pseudoword pairs used in the experiment. The overall LIS is calculated as the sum of the SIS. This provides a quantification of the overall (un)expectedness of each logatome's phonetic form given a corresponding logatome. Since LIS between two logatomes is computed by summing up the SIS values of the sounds from the aligned logatome pair, it depends on the number of available tokens. The normalized LIS values in Table 6 were obtained by dividing the LIS value by the length of the logatome. An important property of surprisal-based modeling is that it can reveal asymmetries in the overall identification difficulties depending on the direction of processing. This means that the Bulgarian–Polish exposure–response logatome pair may not have the same normalized LIS value as the Polish–Bulgarian pseudoword pair. To examine possible confusion asymmetries, the mean LIS values between the tested languages were calculated using the incom.py toolbox (Mosbach et al., 2019) and presented in Table 6.

First, the LID scores were correlated with the LIS values between pairs of exposure–response languages for all logatomes, regardless of length. A negative but not significant correlation was found: the lower the surprisal values, the larger the LID scores, Pearson's $r = -0.36$; $R^2 = .13$; $p = .25$. Furthermore, the mean LIS values for CVCV and CVCVCV logatomes, respectively, were correlated with the subject's performance for CVCV and CVCVCV sequences. The negative

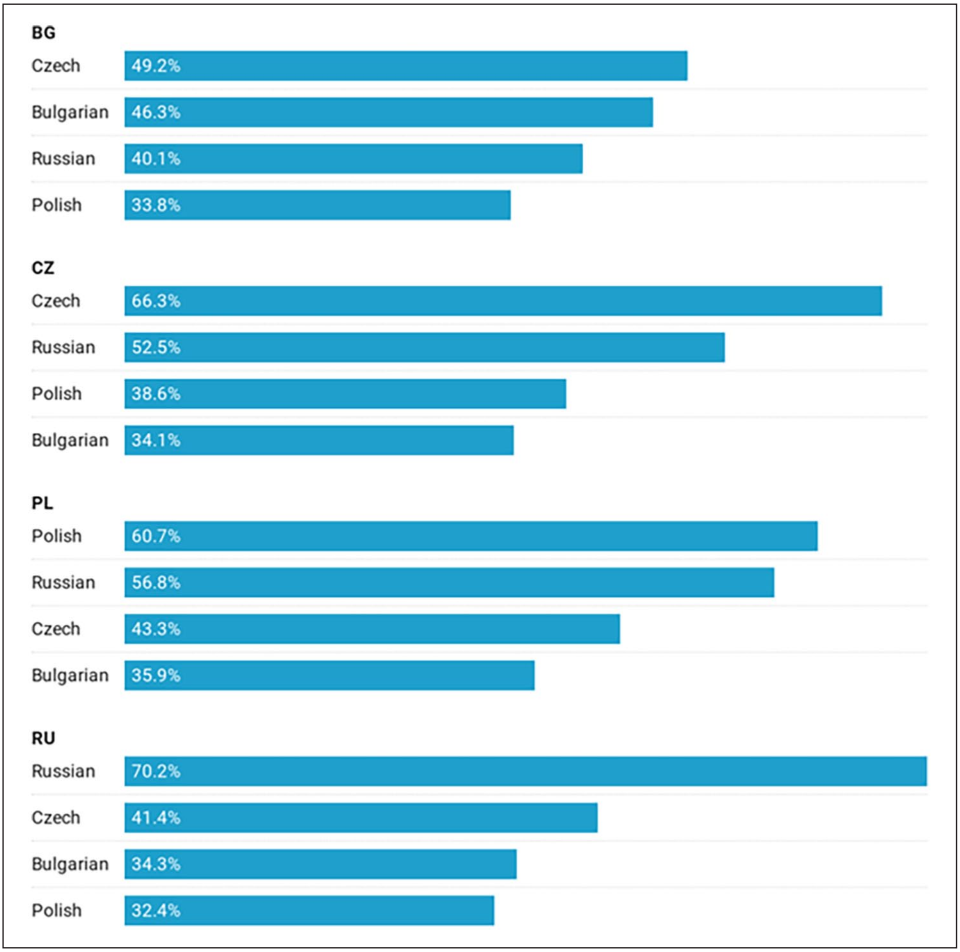

Figure 4. Overall identification scores in Slavic L1 groups.

Table 6. Mean Logatome Identification Surprisal Values in Bits.

Exposure	Response	All sequences	CVCV	CVCVCV
Bulgarian	Czech	0.43342744	0.42931416	0.40465307
Bulgarian	Polish	0.24398879	0.22550414	0.25974784
Bulgarian	Russian	0.62378925	0.58503154	0.63851192
Czech	Bulgarian	0.44885918	0.42195107	0.43798181
Czech	Polish	0.32994016	0.31232876	0.34074236
Czech	Russian	0.74938138	0.69254521	0.76796627
Polish	Bulgarian	0.41876436	0.40292569	0.41646455
Polish	Czech	0.48768054	0.49781193	0.46237080
Polish	Russian	0.61333058	0.55175746	0.64546872
Russian	Bulgarian	0.40262913	0.38082098	0.40771237
Russian	Czech	0.51112285	0.49624590	0.50177142
Russian	Polish	0.21727449	0.16997498	0.25811850

correlation for CVCV logatomes, Pearson's $r=-0.40$; $R^2=.16$; $p=.20$, was found to be stronger than for CVCVCV logatomes, Pearson's $r=-0.27$; $R^2=.07$; $p=.39$, but neither correlation reached the threshold of statistical significance.

4 Discussion

In this study, the ability to recognize a speakers' linguistic origin was investigated. The analysis focused on the spectral and temporal properties of segments shared by the phonological inventories of four Slavic languages: Bulgarian, Czech, Polish, and Russian. In this experimental setup, methodologies from acoustics, phonetics, and information theory were combined to discover which cues (language-specific stress distribution patterns, vowel space overlap, pseudoword length, or LIS) are relevant to lay listeners for determining a speaker's L1. In addition, the importance of fluency in non-Slavic languages on performance of the listeners was evaluated. The fixed versus flexible word stress did not appear to be informative enough for the speakers of Slavic languages to influence their performance in the perceptual task. The data suggest that stress distribution is not a discriminable factor. Languages with fixed and flexible word stress were not clustered according to accentuation patterns. Interestingly, Polish (with a fixed penultimate word stress) and Czech (with a fixed initial word stress) are placed in different groups depending on the stress distribution, which can suggest that stress distribution patterns are not informative enough in the L1 identification task. Polish and Czech also did not pattern in the same group according to perceptual similarity, despite their typological proximity—instead, Polish was clustered with Bulgarian. This finding can be explained by significant differences in vowel length among the languages. Czech exhibits vowel lengthening, which distinguishes it from Polish and Bulgarian and hence results in an alternative taxonomy. Furthermore, participants' fluency in a typologically related language significantly influenced their performance in the presented LID task.

Regarding the subjects' linguistic repertoire, the genetic relatedness of their L1 to the stimulus was, intuitively, a valid and strong predictor of performance in the L1 identification task; however, fluency in other non-Slavic languages also had an impact on the LID scores, though, in line with expectations, with a lower effect size. A closer look at the phonetic inventories of the non-Slavic languages correlating with test performance reveals that the interpretation of this finding should be limited to general linguistic knowledge and intuition of lay listeners, rather than viewed as a transfer or mapping between the phonological units of non-Slavic L2 to Slavic L1. The improvement of performance related to fluency in Romanian may be attributed to areal linguistics, as an effect of contact with Slavic languages and its membership in the Balkan Sprachbund. In contrast, fluency in Azeri, a Turkic language, which appeared to diminish recognition scores, may be attributed to the distinctiveness of its vowel inventory, which might result in decreased sensitivity to particular vowel characteristics. The analysis of correlations of Slavic L1 with LID performance demonstrates that listeners were often able to correctly identify the origin of speakers whose L1 was the same as their own, with the exception of the Bulgarian group. The Bulgarian native speakers correctly identified speakers whose L1 was Czech more often than fellow Bulgarian L1 speakers. This finding can be related to a dialectological landscape of Bulgarian marked by the West–East isogloss, which defines the quality of the reflex of the Proto-Slavic yat vowel. Such a difference in vowel quality can cause confusion and thus influence the performance in an identification task in which vowel quality serves as a primary cue.

Overall, it can be concluded that even without lexical, semantic, or syntactic information, the identification of a speaker's linguistic origin is nevertheless possible based exclusively on the phonetic and phonotactic subsystems. As shown in this study, the "optimized deduction" or "intelligent guessing" strategies were confirmed on the material of four Slavic languages. Furthermore, it appeared that it is possible to identify the language of a speaker's origin when exposed to delexicalized audio stimuli as long as the subject's native language and the stimulus language are phonetically close. At the same time, the data did not support the hypothesis of differences caused by disyllabic and trisyllabic sequences exhibiting various stress patterns. Interestingly, an overlap in a 3D vowel space appeared to be a valid predictor of lay listeners' performance. On the other hand, the correlations between the LIS values and LID scores were negative, but low and insignificant. Eventually, other acoustic dimensions such as formant dynamics should be taken into consideration in future studies concerning correlations of signal properties with subject's performance in the LID tasks. The surprisal-based results did not reach the threshold of statistical significance; hence, the formulated hypothesis could not be confirmed on the basis of the gathered material. Obviously, other linguistic and non-linguistic factors such as attitudes toward the test languages (Gooskens & van Heuven, 2020) can influence the LID performance and should be included in the evaluation of language identification results.

5 Conclusion

This study shed light on the question of involvement of native speakers without training in perceptual identification of linguistic origin when exposed to highly limited information. Work with so-called under-represented languages and vernaculars can be advanced by applying methods which combine human- and machine-based LID, especially when the available data are limited, or the training baseline is not sufficient. Since cues that are important for humans are significantly different from those for machines, it is obvious that the LID methodology should be not only source-specific but also recognizer-dependent, with respect to the domain of the training data. Even if the signal distortion reaches extreme levels, as it was in this study, in which semantic, lexical, and syntactic cues were not available, the accurate identification of the origin of a speaker was still possible by lay listeners. Therefore, the opinions of lay persons in the LADO framework should not be neglected. This study provides a clear argument for the involvement of lay persons in LADO/LOID procedures. Nevertheless, the improvement of origin verification tests using phonetic analysis remains uncontroversial. It appears that highly limited signals can cause an attention shift toward typically less relevant features in spoken language perception such as vowel quality in the spectral and temporal domains. These findings should be considered in the procedure of LADO/LOID tests, as well as in forensic procedures. It appears that speakers of closely related languages can successfully identify the linguistic origin of a user of another Slavic language, which contributes to the debate on involvement of non-experts in the LADO procedures. It also demonstrates the impressive human capability to identify the origin of a speaker with exposure to even highly limited acoustic information.

Acknowledgements

We wish to thank Hasan Md Tusfiqur Alam and Philip Georgis for their support in implementing the LADO experiment and associated data analysis.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 232722074—SFB 1102.

ORCID iD

Jacek Kudera https://orcid.org/0000-0003-3678-1067

Data availability

The experimental results and scripts are publicly available in the following Open Science Framework repository: https://osf.io/b4jfg/.

References

Abdullah, B., Avgustinova, T., Möbius, B., & Klakow, D. (2020, October 25–29). *Cross-domain adaptation of spoken language identification for related languages: The curious case of Slavic languages* [Conference paper]. Interspeech 2020, Shanghai, China. https://www.isca-speech.org/archive/pdfs/interspeech_2020/abdullah20_interspeech.pdf

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.16). Phonetic Sciences, University of Amsterdam. https://www.fon.hum.uva.nl/praat/

Bonatti, L. L., Pena, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*(6), 451–459. https://doi.org/10.1111/j.0956-7976.2005.01556.x

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance. Quantitative applications in the social sciences* (*Vol. 54*). SAGE. https://doi.org/10.4135/9781412985222

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/ijsll.v23i1.17539

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT.

Eades, D. (2005). Applied linguistics and language analysis in asylum seeker cases. *Applied Linguistics*, *26*(4), 503–526. https://doi.org/10.1093/applin/ami021

Foulkes, P., & Wilson, K. (2011, August 17–21). *Language analysis for the determination of origin: An empirical study* [Conference paper]. 17th International Congress of Phonetic Sciences, Hong Kong, China. http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Foulkes/Foulkes.pdf

Fraser, H. (2009). The role of "educated native speakers" in providing language analysis for the determination of the origin of asylum seekers. *International Journal of Speech, Language & the Law*, *16*(1), 113–138. https://doi.org/10.1558/ijsll.v16i1.113

Fraser, H. (2011). The role of linguistics and native speakers in language analysis for the determination of origin: A response to Tina Cambier-Langeveld. *International Journal of Speech, Language and the Law*, *18*(1), 121–130. https://doi.org/10.1558/ijsll.v18i1.121

Gooskens, C., & van Heuven, V. J. (2020). How well can intelligibility of closely related languages in Europe be predicted by linguistic and non-linguistic variables?. *Linguistic Approaches to Bilingualism*, *10*(3), 351–379. https://doi.org/10.1075/lab.17084.goo

Harper, M., & Maxwell, M. (2007). Spoken language characterization. In J. Benesty, M. Sondhi, & Y. Huang (Eds.), *Springer handbook of speech processing* (pp. 797–809). Springer.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001

Hoskin, J. (2018). Native speaker non-linguists in LADO: An insider perspective. In I. Nick (Ed.), *Forensic linguistics: Asylum-seekers, refugees and immigrants* (pp. 23–40). Vernon.

House, A., & Neuburg, E. (1977). Toward automatic identification of the language of an utterance: Preliminary methodological considerations. *The Journal of the Acoustical Society of America*, *62*(3), 708–713. https://doi.org/10.1121/1.381582

Johnson, E. K. (2003). *Acoustic and auditory phonetics* (2nd ed.). Blackwell.

Kelley, M. C., & Tucker, B. V. (2020). A comparison of four vowel overlap measures. *The Journal of the Acoustical Society of America*, *147*(1), 137–145. https://doi.org/10.1121/10.0000494

Kirchhoff, K., & Parandekar, S. (2001, September 3–7). *Multi-stream statistical N-gram modeling with application to automatic language identification* [Conference paper]. Eurospeech 2001, Aalborg, Denmark. https://www.isca-speech.org/archive/pdfs/eurospeech_2001/kirchhoff01_eurospeech.pdf

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347. https://doi.org/10.1016/j.csl.2017.01.005

Kudera, J. (2018). Synchronised and non-synchronised articulation of Polish nasals ą and ę in child and adult speech. *Govor*, *35*(2), 161–188. https://hrcak.srce.hr/214239

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell.

Language and National Origin Group. (2004). Guidelines for the use of language analysis in relation to questions of national origin in refugee cases. *International Journal of Speech, Language and the Law*, *11*(2), 261–266. https://doi.org/10.1558/ijsll.v11i2.261

McQueen, J., Dahan, D., & Cutler, A. (2003). Continuity and gradedness in speech processing. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 39–78). Mouton de Gruyter.

Meissner, F. J. (2018). Modelling plurilingual processing in intercomprehension. In S. Garbarino & Ch. Degache (Eds.), *Intercompréhension en réseau. Scénarios, médiations, évaluations* (pp. 279–283). CRTT.

Mosbach, M., Stenger, I., Avgustinova, T., & Klakow, D. (2019, September 2–4). *incom.py—A toolbox for calculating linguistic distances and asymmetries between related languages* [Conference paper]. International Conference on Recent Advances in Natural Language Processing, Varna, Bulgaria. https://doi.org/10.26615/978-954-452-056-4_094

Muthusamy, Y. K., Jain, N., & Cole, R. A. (1994, April 19–22). *Perceptual benchmarks for automatic language identification* [Conference paper]. International Conference on Acoustics, Speech and Signal Processing, Adelaide, SA, Australia. https://doi.org/10.1109/ICASSP.1994.389288

Navrátil, J. (2001). Spoken language recognition—A step toward multilinguality in speech processing. *IEEE Transactions on Speech and Audio Processing*, *9*(6), 678–685. https://doi.org/10.1109/89.943345

Nolan, F. (2012). Degrees of freedom in speech production: An argument for native speakers in LADO. *International Journal of Speech, Language & the Law*, *19*(2), 263–289. https://doi.org/10.1558/ijsll.v19i2.263

Nycz, J., & Hall-Lew, L. (2013). Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics*, *20*, 2–19. https://doi.org/10.1121/1.4894063

Ohala, J., & Gilbert, J. (1981). Listeners' ability to identify languages by their prosody. In P. Léon & M. Rossi (Eds.), *Problèmes de prosodie: Vol. 2, expérimentations, modèles et fonctions* (pp. 123–131). Didier.

Pascal, D., Thill, C., & Boyer, M. (1989). *Perceptive similarity of male voices: correlation with acoustic measures* [Conference paper]. Eurospeech 1989, Paris, France. https://www.isca-speech.org/archive_v0/archive_papers/eurospeech_1989/e89_2123.pdf

Patrick, P. (2010). Language variation and LADO. In K. Zwaan, M. Verrips, & P. Muysken (Eds.), *The role of language in European asylum procedure: A linguistic and legal survey* (pp. 73–87). Wolf Legal.

Patrick, P., Schmid, M., & Zwaan, K. (2012). *Language analysis for the determination of origin: Current perspectives and new directions*. Springer.

Pellegrino, F., & André-Obrecht, R. (2000). Automatic language identification: An alternative approach to phonetic modelling. *Signal Processing*, *80*(7), 1231–1244. https://doi.org/10.1016/S0165-1684(00)00032-3

Peperkamp, S., & Dupoux, E. (1992). A typological study of stress "deafness." In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (*Vol. 7*, pp. 203–240). Mouton de Gruyter. https://doi.org/10.1515/9783110197105.1.203

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430. https://doi.org/10.1016/j.wocn.2010.04.001

Pillai, K. C. (1954). *On some distribution problems in multivariate analysis*. Department of Statistics, North Carolina State University. https://repository.lib.ncsu.edu/bitstream/handle/1840.4/2164/ISMS_1954_88.pdf?sequence=1

Rouas, J. L., Farinas, J., Pellegrino, F., & André-Obrecht, R. (2005). Rhythmic unit extraction and modelling for automatic language identification. *Speech Communication*, *47*(4), 436–456. https://doi.org/10.1016/j.specom.2005.04.012

Schultz, T., & Waibel, A. (1998, February 8–11). *Multilingual and crosslingual speech recognition* [Conference paper]. DARPA Broadcast News Workshop 1998, Lansdowne, VA, United States. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.2484&rep=rep1&type=pdf

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Skirgård, H., Roberts, S. G., & Yencken, L. (2017). Why are some languages confused for others? Investigating data from the Great Language Game. *PLOS ONE*, *12*(4), Article e0165934. https://doi.org/10.1371/journal.pone.0165934

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, *6*(2), 111–128. https://doi.org/10.1044/jshr.0602.111

Thomas, E. R. (2011). *Sociophonetics: An introduction*. Palgrave MacMillan.

Wilson, K., & Foulkes, P. (2014). Borders, variation, and identity: Language analysis for the determination of origin (LADO). In D. Watt & C. Llamas (Eds.), *Language, borders and identity* (pp. 218–229). Edinburgh University Press.

Zissman, M., & Berkling, K. (2001). Automatic language identification. *Speech Communication*, *35*(1–2), 115–124. https://doi.org/10.1016/S0167-6393(00)00099-6