

DOI: 10.15514/ISPRAS-2019-31(5)-12



## Применение i-векторов для автоматизированного определения уровня близости языков

A.У. Берзинь, ORCID: 0000-0002-3313-5935 <ansis@latnet.lv>

Латвийский университет,  
LV-1003, Латвия, г. Рига, ул. Московская, д. 54

**Аннотация.** В статье рассказывается о результатах применения i-векторных методов распознавания речи для задания расстояния между языками. В качестве входных данных используются фонограммы спонтанной речи. Эксперименты проводятся на звукозаписях латышских и латгальских говоров, но методы применимы и к любым другим идиомам.

**Ключевые слова:** речь; идиома; язык; диалект; i-вектор; и-вектор; фонограмма; близость языков; расстояние между языками

**Для цитирования:** Берзинь А.У. Применение i-векторов для автоматизированного определения уровня близости языков. Труды ИСП РАН, том 31, вып. 5, 2019 г., стр. 153-164. DOI: 10.15514/ISPRAS-2019-31(5)-12

## Usage of i-Vectors for Automated Determination of a Similarity Level between Languages

A.A. Bērziņš, ORCID: 0000-0002-3313-5935 <ansis@latnet.lv>

University of Latvia,  
54, Moscow str., Riga, LV-1003, Latvia

**Abstract.** The article describes results of applying i-vectors-based (both LID and SID) speech identification methods to define a kind of a distance between languages (in a wide sense of the word – including dialects and any other forms of spoken language). Spontaneous speech recordings of many enough speakers of languages are used on the input of the method. The experiments were carried out at recordings of Latvian and Latgalian dialects, but the method is applicable to any other idioms. Cosine similarity, Euclidean metric, standardized Euclidean metric, Jordan (or Chebyshev) metric and city block (or  $L_1$ ) metric were tried out. Cosine similarity worked well for SID i-vectors, but for unknown reasons was senseless for LID i-vectors. Jordan metric worked well for LID, but was not good enough for SID i-vectors. Standardization of the Euclidean metric does not gave any improvement. Thus, the conclusions are: 1) both SID and LID vectors of full length recordings of spontaneous speech are characterizing and representing languages good enough to be used for detection of a distance between languages; 2) the best metrics for such tasks are Euclidean and  $L_1$  (for arithmetic mean vectors computed from i-vectors of all informants coordinate by coordinate).

**Keywords:** speech; idiom; language; dialect; i-vector; LID; SID; recording; proximity of languages; distance between languages

**For citation:** Bērziņš A.A. Usage of i-Vectors for Automated Determination of a Similarity Level between Languages. Trudy ISP RAN/Proc. ISP RAS, vol. 31, issue 5, 2019, pp. 153-164 (in Russian). DOI: 10.15514/ISPRAS-2019-31(5)-12

### 1. Введение

i-векторы – это относительно новый способ решения задач распознавания, который в настоящее время используется для распознавания объектов и других видов<sup>1</sup>, но первоначально их ввели при поиске новых методов распознавания речи.

Первая относительно широко известная публикация, в которой озвучили данную идею (в контексте идентификации говорящего), вышла в 2009 году<sup>2</sup>. Название i-векторов в то время еще не появилось, в статье их пространство именуется пространством признаков полной изменчивости. В начале 2010 года сочетание «i-вектор» появилось как дополнительное название<sup>3</sup>, но ко второй половине того же года им уже пользовались в полной мере, причём уже описывая именно задачи идентификации языка<sup>4</sup>.

<sup>1</sup> Например, при распознавании символов: *We propose a novel text classification approach based on iVector, a newly developed concept in speaker verification. To a given text line, the iVector is a fixed-length feature vector representation, transformed from a high-dimensional supervector based on means of Gaussian mixture model (GMM), where the text dependent component is separated from a universal background model (UBM) and can be represented by a lowdimensional set of factors. We classify the text lines with a discriminative classifier - support vector machine (SVM) in iVector space. A baseline approach of text classification using GMM in feature space is also presented for evaluation purpose. Experimental results on an Arabic document database show accuracy of 92.04% for text line classification using the proposed method. Furthermore, the relative word error rate (WER) of 9.6% is decreased in optical character recognition (OCR) when coupled with the proposed iVector-SVM classifier. The proposed iVector-SVM approach is language independent, thus, can be applied to other scripts as well.* [2]

<sup>2</sup> *This paper presents a new speaker verification system architecture based on Joint Factor Analysis (JFA) as feature extractor. In this modeling, the JFA is used to define a new low-dimensional space named the total variability factor space, instead of both channel and speaker variability spaces for the classical JFA.* [3]

<sup>3</sup> *Based on this, we proposed a new speaker verification system based on factor analysis as a feature extractor. The factor analysis is used to define a new low-dimensional space named total variability space. In this new space, a given speech utterance is represented by a new vector named total factors (we also refer to this vector as “i-vector” in this paper).* [4]

<sup>4</sup> *...a new language identification system is presented based on the total variability approach previously developed in the field of speaker identification. Various techniques are employed to extract the most salient features in the lower dimensional i-vector space..* [5]

*...we described the application of the i-vector or total variability space approach to the language identification task. The i-vector representation is a data-driven approach for feature extraction that provides an elegant and general framework for audio classification and identification. It consists of mapping a sequence of frames for a given utterance into a low-dimensional vector space, referred to as the total variability space, based on a factor analysis technique.* [5]

Метод i-векторов основан на представлении моделей выражений гауссовой смеси со скрытой маломерной переменной и использовании изображения этого выражения в качестве вектора признаков в языковом классификаторе<sup>5</sup>.

Бывают разные i-векторы в зависимости от того, какую лингвистическую информацию они содержат (например, акустическую, просодическую, фонотактическую), построены ли они на непрерывных или дискретных данных и предназначены ли для идентификации говорящего (SID), идентификации языка (LID) или других задач. Так что, на самом деле, мы могли бы даже говорить о целом ряде методов, но погружение в такие тонкости не является целью данной статьи.

## 2. Исходные данные

В нашем распоряжении были собранные (записанные) нами звукозаписи спонтанной речи пяти идиомов (латвийских говоров) – один из Курляндии: Дундажской волости, и четыре из Латгалии: Аулея, Бальтинова, Вилека и Рудзатов. Курляндия исторически была под немецким игмом, поэтому местные говоры подверглись влиянию (нижне)немецкого языка, а северокурляндские говоры, в том числе и дундажский, содержат большой субстрат ливонского языка (принадлежащего к прибалтийско-финской подгруппе финно-угорских языков). Латгалия, в свою очередь, была под поляками, поэтому в латгальских говорах присутствует влияние польского языка, также – в силу близкого соседства и наличия белорусских и старообрядческих деревень – белорусского и русского. Бальтиновский и вилекский являются говорами северолатгальскими, которые от западнолатгальского рудзатского и южнолатгальского аулейского отличаются существенно – и морфологически, и лексически.

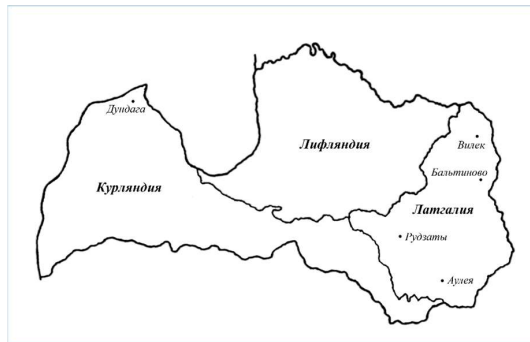


Рис. 1. Расположение записанных говоров на карте Латвии  
Fig. 1. The recorded dialects on the map of Latvia

<sup>5</sup> The most recent front-end subspace modeling technique known as iVector, which is a feature extraction model in the front-end of the language recognition system, has become the state-of-the-art technique in SID and was successfully adapted for the language recognition. The main idea of the iVector model in acoustic language recognition is to represent each utterance dependent GMM with a low-dimensional latent variable and use the low-dimensional representation of the utterance as a feature vector to the following language classifier. [6]

Все звукозаписи собирались согласно заданным нами принципам сбора информации для автоматизированного анализа фонограмм [1], т.е., все записи были однородными, записанными однотипной аппаратурой (использовался динамический микрофон одностороннего направления, фиксированный на голове информанта), в условиях уменьшенного влияния внешних шумов. Все записи были вручную вычищены, удалению подверглись все посторонние звуки и голоса, оставив только прямую речь информанта. Качество записи – 44,1 кГц / 16 битов. В зависимости от требований конкретных скриптов по вычислению i-векторов, для фонограмм выполнялось понижение частоты дискретизации.

Табл. 1. Характеристика набора фонограмм, используемого в эксперименте.  
Table 1. Characteristics of recordings used in the experiment

Говор	Минут	Информантов	Мужчин	Женщин
Аулея	95	14	8	6
Бальтиново	140	23	9	14
Дундага	161	17	4	13
Рудзаты	246	28	11	17
Вилек	238	30	11	19

Всех информантов просили рассказывать о быте, родителях, бабушках, дедушках, братьях, сёстрах, детях, других членах семьи, учёбе, работе, хозяйстве, службе в армии, свадьбах, праздниках, соседях и т. п. Т.е., в ходе сбора данных на традиционность и гомогенность лексики обращалось пристальное внимание.

Поэтому, учитывая однородность нашего многоговорного корпуса и в техническом, и в содержательном смысле, мы даже можем не постесняться его считать сопоставимым<sup>6</sup>. Доселе этот термин применялся только к текстовым корпусам, но мы считаем, что его можно применять и к речевым, и при таком применении наш корпус соответствует смыслу сопоставимости.

## 3. Эксперимент. Векторы SID

Поскольку в нашем распоряжении были скрипты вычисления i-векторов, разработанные Брненским Техническим университетом (БТУ), мы, конечно, ими воспользовались. В 2015 году Речевая группа БТУ выступила с предложением создать общий стандарт голосовой биометрии – *Voice Biometry Standart* или VBS, поскольку различные используемые в настоящее время технические стандарты не позволяют быстро предоставлять данные и обмениваться ими. Предложение подкреплялось скриптами,

<sup>6</sup> A comparable corpus is one which selects similar texts in more than one language or variety. There is as yet no agreement on the nature of the similarity, because there are very few examples of comparable corpora. ... The possibilities of a comparable corpus are to compare different languages or varieties in similar circumstances of communication, but avoiding the inevitable distortion introduced by the translations of a parallel corpus. [12]

A comparable corpus is a pair of corpora in two different languages, which come from the same domain. [13]

написанными на питоне, для вычисления i-векторов в предложенном стандарте<sup>7</sup> (на вход можно подавать необработанные фонограммы речи, но при подаче заранее определённых интервалов голосовой активности – VAD или *Voice Activity Detection* – результаты лучше), потому что встроенный определитель активности очень примитивен. В [7] достаточно подробно описано и теоретическое обоснование VBS, и его техническая реализация, поэтому мы не будем это переписывать. Понятно, что биометрический стандарт предназначен для задач идентификации говорящего, то есть он фокусируется на особенности речи (в том числе голоса) определённого лица, т.е. i-векторы, генерируемые этим пакетом, называются SID (от *Speaker IDentification*) i-векторами. Теоретически они менее подходят для нашей задачи, но мы решили их попробовать, так как открытый стандарт и общедоступность ПО являются ключевыми факторами при выборе технологии. i-векторы мы рассчитывали для полных фонограмм спонтанной речи информантов данного языка, таким образом ожидая, что они, в силу большой продолжительности звукозаписи, будут характеризовать язык в целом, т.е., и фонетические, и морфологические, и лексические, и даже синтаксические особенности. Косинусный коэффициент<sup>8</sup> (или мера Отиаи) был предложен в качестве наиболее перспективного метода оценки расстояния между i-векторами с самого начала – в изначальных публикациях об i-векторах<sup>9</sup>. Несмотря на постоянные поиски различных способов его улучшения<sup>10</sup>, он всё равно пока остаётся основным методом оценки близости i-векторов. Поэтому для всех пар i-векторов говорю мы сначала вычислили косинусный коэффициент.

<sup>7</sup> This standard is supposed to give formal description of the i-vector extraction algorithm. However, we provide a python demo package for i) better understanding of the properties and features of the extraction, and ii) for convenience, so that the user can immediately use the basic functions and do prompt customizations. [7]

<sup>8</sup> Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. [8]

<sup>9</sup> We have proposed two new systems based on this new speech representation. The first system is an SVM-based system which uses the cosine kernel to compute the similarity between the total factors. The second system directly uses the value of the cosine distance computed between the target speaker factors and test total factors as a decision score. In this scoring, we removed the SVM from the decision process. One important characteristic of this approach is that there is no speaker enrolment, unlike in other approaches like SVM and JFA, which makes the decision process faster and less complex. [4]

<sup>10</sup> This paper deals with the problem of processing of I-vectors in the text-independent speaker verification systems. A new generalized cosine similarity optimization technique is proposed. The optimization is performed over sets of orthogonal and diagonal matrices. [9]

It is known that the equal-error-rate (EER) performance of a speaker verification system is determined by the overlap region of the decision scores of true and imposter trials. Also, the cosine similarity scores of the true or imposter trials produced by the state-of-the-art i-vector front-end approximate to a Gaussian distribution, and the overlap region of the two classes of trials depends mainly on their between-class distance. Motivated by the above facts, this paper presents a cosine similarity learning (CML) framework for speaker verification, which combines classical compensation techniques and the cosine similarity scoring for improving the EER performance. CML minimizes the overlap region by enlarging the between-class distance while introducing a regularization term to control the within class variance, which is initialized by a traditional channel compensation technique such as linear discriminant analysis. [10]

Табл. 2. Косинусный коэффициент между SID i-векторами фонограмм (значения округлены)  
Table 2. Cosine similarity between SID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	1	0,07	-0,73	-0,16	-0,08
Baļtinova	0,07	1	-0,63	-0,33	0,68
Dundag	-0,73	-0,63	1	0,21	-0,54
Rudzātys	-0,16	-0,33	0,21	1	-0,45
Vileks	-0,08	0,68	-0,54	-0,45	1

Табл. 3. Углы в градусах от косинусного коэффициента между SID i-векторами наших фонограмм (значения округлены)

Table 3. Angles in degrees from cosine similarity on SID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	86	137	99	94
Baļtinova	86	0	129	109	47
Dundag	137	129	0	78	122
Rudzātys	99	109	78	0	116
Vileks	95	47	122	116	0

Косинусный коэффициент по сути – величина косинуса, а арккосинус от него возвращает углы, которые также характеризуют вычисляемое расстояние. Поэтому визуально вообразить, что к чему ближе, а что от чего дальше, лучше всего можно, перейдя на углы в градусах, т.е., вычислив арккосинус от косинусного коэффициента. Представьте себе нулевую линию, проведенную на плоскости и на ней точку или центр; тогда угол, образованный лучами от этого центра и нулевого луча (правый луч нулевой линии), представляет соответствующее расстояние между двумя кривыми – чем меньше угол, тем языки ближе.

Если таким образом проанализировать 3-ю таблицу, то видно, что результаты являются осмысленными, т.е., между более близкими языками углы меньше, а более далёкими – больше.

Так бальтиновский и вилекский говоры оказываются ближайшей парой (47°). Расстояние между южно- и западнолатгальскими говорами – рудзатским и аулейским – меньше, чем между ними и севернолатгальскими. Дундага наиболее удалена от Аулеи, Бальтинова и Вилека. Единственная оценка, которая кажется в корне неверной, это расстояние между рудзатским и дундажским – оно, безусловно, не должно было быть меньше, чем расстояние между рудзатским и тремя остальными латгальскими говорами.

Интереса ради, дабы было с чем сравнить, мы для определения расстояния между i-векторами решили попробовать также евклидову и другие векторные метрики (например, жорданову<sup>11</sup>). Под влиянием [11] мы решили применить тоже метрику городского квартала. Так как все вышеупомянутые метрики заданы на векторном пространстве, то потребовалось привести наши многовекторные характеристики к одному вектору: мы это

<sup>11</sup> Она же – метрика Чебышёва, расстояние Чебышёва, равномерная метрика, sup-метрика, бокс-метрика.

сделали для каждого языка, по координатно вычислив средний арифметический i-вектор из i-векторов информантов данного языка.

Табл. 4. Евклидова метрика между SID i-векторами наших фонограмм (значения округлены)

Table 4. Euclidean metric between SID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	8,04	11,42	8,71	9,00
Baļtinova	8,04	0	9,78	7,56	6,56
Dundag	11,42	9,78	0	9,27	9,13
Rudzātys	8,71	7,56	9,27	0	7,24
Vileks	9,00	6,56	9,13	7,24	0

Табл. 5. Нормализованная евклидова метрика между SID i-векторами наших фонограмм (значения округлены)

Table 5. Standardized Euclidean metric between SID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	0,480	0,663	0,515	0,542
Baļtinova	0,480	0	0,582	0,458	0,405
Dundag	0,663	0,582	0	0,547	0,549
Rudzātys	0,515	0,458	0,547	0	0,443
Vileks	0,542	0,505	0,549	0,443	0

Табл. 6. Жорданова метрика между SID i-векторами наших фонограмм (значения округлены)

Table 6. Jordan metric between SID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	1,20	1,39	1,24	1,36
Baļtinova	1,20	0	1,35	0,95	0,87
Dundag	1,39	1,35	0	1,10	1,04
Rudzātys	1,24	0,95	1,10	0	0,94
Vileks	1,36	0,87	1,04	0,94	0

Табл. 7. Метрика городского квартала или L1 между SID i-векторами наших фонограмм (значения округлены)

Table 7. City block or L1 metric between SID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	155	223	170	173
Baļtinova	155	0	188	149	127
Dundag	223	188	0	182	178
Rudzātys	170	149	182	0	142
Vileks	173	127	178	142	0

Из табл. 4–7 видно, что в нашем случае наихудшие результаты (хоть и не совсем плохие) показала жорданова метрика: Аулея и для Вилека, и для Рудзат оказалась гораздо дальше Дундаги.

Метрики L1 и евклидова (как нормализованная, так и обыкновенная, поскольку нормализация на результаты существенно не повлияла) обе выглядят одинаково хорошо и – главное – даже лучше, чем косинусный коэффициент: Вилек и Баљтиново – самые близкие, Дундага – по отношению ко всем латгальским говорам – самая дальняя.

Единственный вопрос, который возникает: почему Аулея к Баљтинову оказывается ближе чем к Рудзатам? Это может быть ошибкой метрики, неадекватностью данных, но также и объективной оценкой, которая учитывает некоторые диалектальные нюансы, которые в теоретических сравнениях обычно игнорируются. Чтобы ответить на этот вопрос, необходимы дополнительные эксперименты с большим количеством данных и большим количеством говоров.

#### 4. Эксперимент. Векторы LID

Другой вид i-векторов, предназначенный для идентификации языка, называется LID (от Language IDentification).

Сперва мы провели предварительный эксперимент, дабы убедиться, что i-векторный метод распознавания языка эффективен для наших диалектальных звукозаписей: во время стажировки в Брненском Техническом университете мы дали Олдриху Плоту, научному сотруднику Исследовательской группы по обработке речи наши фонограммы (он их попросил для своих экспериментов) и попросили заодно провести и опыты, интересующие нас.

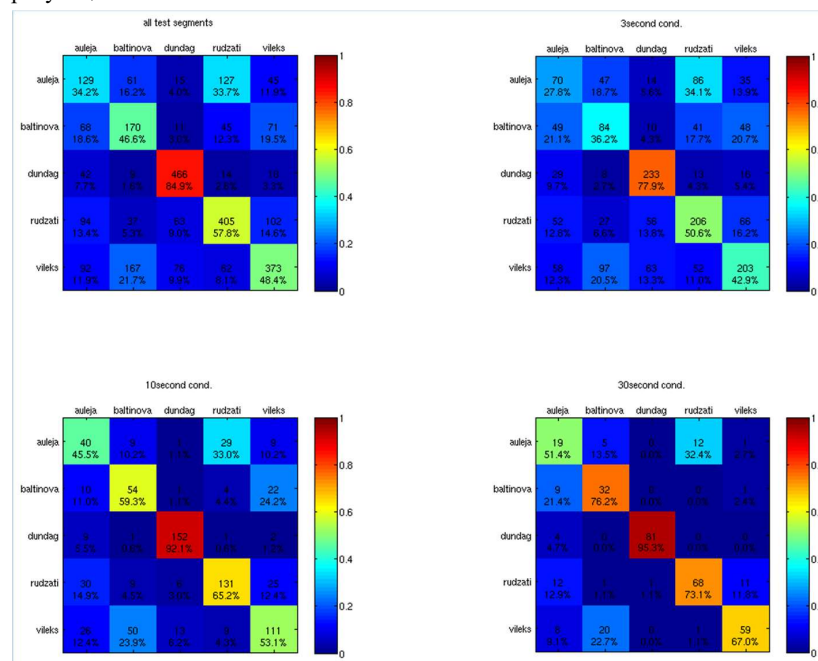


Рис. 2. Результаты классификации наших звукоданных, предоставленные Исследовательской группой по обработке речи Брненского Технического университета  
 Fig. 2. Classification results of our audio data provided by the Speech Processing Research Group of the Technical University of Brno.

Перед проведением эксперимента данные каждого говора были путём случайной выборки разделены на две части: большую, обучающую часть и меньшую, проверочную часть. Затем i-векторы были рассчитаны для каждой части отдельно. После этого гауссовский линейный классификатор<sup>12</sup> обучался на i-векторах обучающей части, а на i-векторах проверочной части он в свою очередь применялся.

На рис. 2 для речевых сегментов разной длительности показано процентное распределение того, сколько проверочных данных было распределено правильно (т.е., правильно определён говор фонограммы) и сколько – неправильно. Как видим, результаты весьма близки к реальности: Дундага как наиболее отличающаяся определяется лучше всего; Рудзаты тоже отражены достоверно, то, что они «отдают» часть другим латгальским говорам, вполне объяснимо; Балтиново и Вилек, учитывая их близость, также показывают относительно хорошие результаты, причём большая часть разницы «отдаётся» друг другу – между собой; единственное, что удивляет, это сравнительно плохие результаты Аулеи «в пользу» Рудзатов.

Учитывая хорошие результаты брэнцев, мы решили на их LID i-векторах, которые они нам любезно предоставили, провести свои эксперименты, которые мы до того провели на SID i-векторах. Мы ожидали, что результаты будут подобны SID i-векторным, но всё-таки чуть лучше, потому что LID i-векторы предназначены для решения задачи, более схожей с нашей.

Табл. 8. Косинусный коэффициент между LID i-векторами фонограмм (значения округлены)

Table 8. Cosine similarity between LID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	1	0,49	-0,06	0,57	-0,85
Baļtinova	0,49	1	-0,24	-0,05	-0,49
Dundag	-0,06	-0,24	1	0,06	-0,04
Rudzātys	0,57	-0,05	0,06	1	-0,82
Vileks	-0,85	-0,49	-0,04	-0,82	1

Табл. 9. Углы в градусах от косинусного коэффициента между LID i-векторами наших фонограмм (значения округлены)

Table 9. Angles in degrees from cosine similarity on LID i-vectors of our recordings (values rounded).

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	61	93	55	148
Baļtinova	61	0	104	93	120
Dundag	93	104	0	87	92

<sup>12</sup> The next step in an LID system is training of language models using the generated iVectors. ...

...it is enough to have class (language) likelihoods so that we can make an optimal Bayesian decision on the language of a trial. The optimal Bayesian decision could be made if our LID system delivers optimal likelihoods for the languages of interest. E.g. if the task is to minimize the probability of language misclassification, we can select the most likely language, where the language posteriors can be obtained using Bayes rule from the priors and likelihoods. Having iVectors, our back-end would be a single multi-class probabilistic classifier (e.g. multi-class logistic regression, Gaussian linear classifier and etc.) that takes iVectors as inputs and, by definition, delivers class likelihoods. ... [6] Далее в [6] выводятся формулы гауссовского линейного классификатора.

	Rudzātys	55	93	87	0	145
Vileks	148	120	92	145	0	

К нашему великому удивлению, результаты косинусного коэффициента оказались совершенно бессмысленными. Установить причины этого нам пока не удалось.

Несмотря на неудачу, мы решили на LID i-векторах испробовать и остальные расстояния, применённые для SID i-векторов. Для евклидовой и L<sub>1</sub> метрик результаты были похожими на SID i-векторные. Интересно, однако, что жорданова метрика, которая для SID i-векторов была не слишком адекватной, вела себя намного лучше на LID i-векторах – без каких-либо наглядных проблем, как в случае с SID i-векторами, и, можно сказать, почти так же хорошо, как евклидова и L<sub>1</sub>.

Табл. 10. Евклидова метрика между LID i-векторами наших фонограмм (значения округлены)

Table 10. Euclidean metric between LID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	6,07	9,59	6,22	6,76
Baļtinova	6,07	0	8,75	6,91	4,71
Dundag	9,59	8,75	0	8,39	8,45
Rudzātys	6,22	5,91	8,39	0	6,14
Vileks	6,76	4,71	8,45	6,14	0

Табл. 11. Нормализованная евклидова метрика между LID i-векторами наших фонограмм (значения округлены)

Table 11. Standardized Euclidean metric between LID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	0,381	0,586	0,390	0,428
Baļtinova	0,381	0	0,543	0,377	0,303
Dundag	0,586	0,543	0	0,519	0,528
Rudzātys	0,390	0,377	0,519	0	0,394
Vileks	0,428	0,303	0,528	0,394	0

Табл. 12. Жорданова метрика между LID i-векторами наших фонограмм (значения округлены)

Table 12. Jordan metric between LID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	0,75	1,13	0,79	0,87
Baļtinova	0,75	0	1,31	0,75	0,59
Dundag	1,13	1,31	0	1,11	1,06
Rudzātys	0,79	0,75	1,11	0	0,81
Vileks	0,87	0,59	1,06	0,81	0

Табл. 13. Метрика городского квартала или L1 между LID i-векторами наших фонограмм (значения округлены)

Table 13. City block or L1 metric between LID i-vectors of our recordings (values rounded)

	Auleja	Baļtinova	Dundag	Rudzātys	Vileks
Auleja	0	120	188	123	131
Baļtinova	120	0	171	115	91
Dundag	188	171	0	164	167
Rudzātys	123	115	164	0	119

Vileks	131	91	167	119	0
--------	-----	----	-----	-----	---

## 5. Выводы

Вследствие проведённых экспериментов мы убедились, что i-векторы достаточно хорошо характеризуют языки и поэтому могут использоваться для количественной оценки языковых различий, причём пользоваться можно как SID, так и LID i-векторами. Кроме того, несмотря на традицию применения косинусного коэффициента, более надёжно воспользоваться евклидовой или  $L_1$  метриками.

Метод i-векторов не единственный метод, разрабатываемый нами для решения задачи численной оценки близости языков, в том числе и по звукозаписям. Поэтому для нас актуальна задача сравнения этих методов, отбора лучших из них и даже создания «надметода», объединяющего наши разработки. Конечно, никаких «золотых стандартов»<sup>13</sup> в этой области не существует, особенно в контексте латышских говоров. Мы уже начали работу по разработке и применению метода экспертных оценок, который подходил бы под наши данные и методы, в том числе и описанный в этой статье. Но это уже тема отдельной публикации...

## Список литературы / References

- [1]. A.A. Bērziņš. The Principles of Collection of Information for Automated Analyse of Audio Recordings. Tbilisi, Meridiani, 2011, pp. 39–46 (in Georgian and Russian) / А.У. Берзинь. Принципы сбора информации для автоматизированного анализа фонограмм. Тбилиси, Меридиани, 2011 / ზერზინი ა. ინფორმაციის მოპოვების პრინციპები ფონოგრამების ავტომატიზირებული ანალიზისთვის. ქართული ენა და თანამედროვე ტექნოლოგიები, თბილისი, მერიდიანი, 2011
- [2]. Zha Sh., Peng X., Cao H., Zhuang X., Natarajan P., Natarajan P. Text Classification via iVector Based Feature Representation. In Proc. of the 11th IAPR International Workshop on Document Analysis Systems, 2014, pp. 151-155.
- [3]. Dehak N., Dehak R., Kenny P., Brummer N., Ouellet P., Dumouchel P. Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification. In Proc. of the Interspeech Conference, 2009, pp. 1559-1562.
- [4]. Dehak N., Kenny P.J., Dehak R., Dumouchel P., Ouellet P. Front-End Factor Analysis for Speaker Verification. IEEE Transactions on Audio, Speech, And Language Processing, vol. 19, no. 4, 2011, pp. 788-798.
- [5]. Dehak N., Torres-Carrasquillo P.A., Reynolds D., Dehak R. Language Recognition via I-vectors and Dimensionality Reduction. In Proc. of the Interspeech Conference, 2011, pp. 857-860.
- [6]. Soufifar M. Subspace Modeling of Discrete Features for Language Recognition. Doctoral theses, Trondheim, NTNU, 2014.
- [7]. Glembek O., Burget L., Matejka P. Voice Biometry Standard, Draft. Brno: Speech@FIT, 2015.
- [8]. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. 3rd Edition. Morgan Kaufmann, 2012, 800 p.
- [9]. Drgas Sz., Dąbrowski A. Generalized cosine similarity in I-vector based automatic speaker recognition systems. In Proc. of the International Conference on Signal Processing: Algorithms, Architectures, Arrangements, and Applications, 2013, pp. 73-77.

<sup>13</sup> Например, наподобие описанных в [14].

- [10]. Bai Zh., Zhang X.-L., Chen J. Cosine Metric Learning for Speaker Verification in the i-Vector Space. In Proc. of the Interspeech Conference, 2018, pp. 1126-1130.
- [11]. Ghosh S., Vijay Girish K.V., Sreenivas T.V. Relationship between Indian Languages Using Long Distance Bigram Language Models. In Proc of the 9th International Conference on Natural Language Processing, 2011, pp. 104-113.
- [12]. Preliminary recommendations on Corpus Typology. EAGLES – Expert Advisory Group on Language Engineering Standards Guidelines, 1996. Available at: <http://www.ilc.cnr.it/EAGLES96/corpusstyp/corpusstyp.html>, 05.11.2019.
- [13]. Comparable Corpora. MT Research Survey Wiki. University of Edinburgh. Available at: <http://www.statmt.org/survey/Topic/ComparableCorpora>, 05.11.2019.
- [14]. Similarity (State of the art). ACL Wiki for Computational Linguistics. The Association for Computational Linguistics. Available at: [https://aclweb.org/aclwiki/Similarity\\_\(State\\_of\\_the\\_art\)](https://aclweb.org/aclwiki/Similarity_(State_of_the_art)), 06.11.2019.

## Информация об авторе / Information about the author

Анс-Атаол Улдович БЕРЗИНЬ – магистр математических наук, завершающий свой труд над диссертацией по компьютерной лингвистике. Сферы научных интересов: лингвометрия, распознавание речи, машинный перевод, малые языки, лексикография, терминология, фольклор, сравнительное языкознание, этномузыкология, права человека, конституционное право, функциональный анализ.

Ansis Ataols BĒRZIŅŠ – Master of Mathematics, completing his work on thesis on computational linguistics. Research interests: linguometry, speech recognition, machine translation, endangered and low-resourced languages, lexicography, terminology, folklore, comparative linguistics, ethnomusicology, human rights, constitutional law, functional analysis.