УНИВЕРСИТЕТ ИТМО
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2016 Том 16 № 4 ISSN 2226-1494 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
July–August 2016 Vol. 16 No 4 ISSN 2226-1494 http://ntv.ifmo.ru/en

УДК 004.93

РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ ДЛЯ ЕГИПЕТСКОГО ДИАЛЕКТА АРАБСКОГО ЯЗЫКА В ТЕЛЕФОННОМ КАНАЛЕ

А.Н. Романенко[a, b]
[a] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[b] Центр Речевых Технологий, Санкт-Петербург, 196084, Российская Федерация
Адрес для переписки: romanenko@speechpro.com

Информация о статье
Поступила в редакцию 31.05.16, принята к печати 20.06.16
doi: 10.17586/2226-1494-2016-16-4-703-709
Язык статьи – русский

Ссылка для цитирования: Романенко А.Н. Разработка системы автоматического распознавания речи для египетского диалекта арабского языка в телефонном канале // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 4. С. 703–709. doi: 10.17586/2226-1494-2016-16-4-703-709

Аннотация
Приводится описание ряда систем автоматического распознавания речи для египетского диалекта арабского языка, построенных на основе набора данных CALLHOME Egyptian. Присутствует описание как классических систем, основанных на скрытых марковских моделях и смеси гауссовых распределений, так и акустических моделей на основе глубоких нейронных сетей. Продемонстрирован вклад от использования дикторозависимых акустических признаков (bottleneck), для извлечения которых были обучены три экстрактора на основе нейронных сетей. Для обучения экстракторов были использованы три набора данных на различных языках: русский, английский и различных диалектах арабского. Исследована возможность использования набора данных современного стандартного арабского языка малого объема для получения фонетических транскрипций. Эксперименты показали, что использование экстрактора, полученного на основе русскоязычного набора данных, позволяет значительно повысить качество распознавания арабской речи. Также установлено, что, хотя использование фонетических транскрипций, основанных на современном стандартном арабском, снижает качество распознавания, все же результаты работы системы остаются применимыми на практике. Дополнительно проведено исследование применения полученных моделей для решения задачи поиска ключевых слов. Полученные системы демонстрируют качество распознавания, сравнимое с современными опубликованными результатами. Предложены дальнейшие пути увеличения качества распознавания.

Ключевые слова
распознавание речи, арабский язык, египетский диалект, дикторозависимые признаки, ограниченные ресурсы

DEVELOPMENT OF AUTOMATED SPEECH RECOGNITION SYSTEM FOR EGYPTIAN ARABIC PHONE CONVERSATIONS

A.N. Romanenko[a,b]
[a] ITMO University, Saint Petersburg, 197101, Russian Federation
[b] Speech Technology Center, Saint Petersburg, 196084, Russian Federation
Corresponding author: romanenko@speechpro.com

Article info
Received 31.05.16, accepted 20.06.16
doi: 10.17586/2226-1494-2016-16-4-703-709
Article in Russian
For citation: Romanenko A.N. Development of automated speech recognition system for Egyptian Arabic phone conversations. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 4, pp. 703–709. doi: 10.17586/2226-1494-2016-16-4-703-709

Abstract
The paper deals with description of several speech recognition systems for the Egyptian Colloquial Arabic. The research is based on the CALLHOME Egyptian corpus. The description of both systems, classic: based on Hidden Markov and Gaussian Mixture Models, and state-of-the-art: deep neural network acoustic models is given. We have demonstrated the contribution from the usage of speaker-dependent bottleneck features; for their extraction three extractors based on neural networks were trained. For their training three datasets in several languages were used: Russian, English and different Arabic dialects. We

have studied the possibility of application of a small Modern Standard Arabic (MSA) corpus to derive phonetic transcriptions. The experiments have shown that application of the extractor obtained on the basis of the Russian dataset enables to increase significantly the quality of the Arabic speech recognition. We have also stated that the usage of phonetic transcriptions based on modern standard Arabic decreases recognition quality. Nevertheless, system operation results remain applicable in practice. In addition, we have carried out the study of obtained models application for the keywords searching problem solution. The systems obtained demonstrate good results as compared to those published before. Some ways to improve speech recognition are offered.

Keywords
speech recognition, Arabic language, Egyptian dialect, speaker-dependent features, limited resources

Введение

На сегодняшний день арабский язык является одним из наиболее распространенных в мире [1]. Общее число носителей по разным оценкам варьируется от 260 до 323 миллионов человек [2]. Говоря о распространенности, стоит отметить, что в повседневной жизни носители используют не современный стандартный арабский (Modern Standard Arabic – MSA), а его различные диалекты, например, египетский, иорданский, ливанский и др. Несмотря на это, число трудов, посвященных построению систем автоматического распознавания речи различных диалектов, довольно мало́ относительно работ, связанных с современным стандартным арабским языком. Этот факт связан со значительными трудностями, возникающими в процессе построения систем обработки натуральных языков (Natural Language Processing) для различных диалектов:

- как правило, диалекты в значительной степени отличаются от MSA морфологически, фонетически и лексически [3]:
  - высокая степень флективности MSA и диалектов обусловливает наличие огромного количества слов и словоформ, образованных от единого корня. Это значительно увеличивает количество слов, для которых не известна фонетическая транскрипция (out of vocabulary – OOV), и затрудняет оценку вероятностей языковых моделей;
  - количество фонем в MSA и диалектах зачастую различно. Кроме того, некоторые фонемы из MSA могут отсутствовать в определенных диалектах или заменяться на другие. Из-за различий в произношении одних и тех же слов в MSA и диалектах построение надежного лексикона невозможно;
  - в MSA и диалектах одно и то же слово может являться разной частью речи [4] – это является непреодолимым препятствием на пути лексического анализа, который мог бы помочь в установлении корректной фонемной транскрипции;
  - частое отсутствие огласовок [5] в арабских текстах, как диалектных, так и на MSA. Огласовки отвечают за однозначное определение таких особенностей произношения слов, как удвоение согласной, удлинение гласного звука и т.д.;
- очень ограниченный объем обучающих данных для диалектов. Если проблему нехватки материала для построения языковых моделей можно решить при помощи Интернет-источников [6], то вопрос получения речевых корпусов для различных диалектов остается открытым. Распознавание речи для диалектов является задачей с ограниченными ресурсами (low resource);
- подавляющее большинство существующих на сегодняшний день инструментов анализа арабского языка разработано специально для MSA [7], что делает их использование для диалектов затруднительным или невозможным, ввиду вышеперечисленных особенностей.

Кроме того, что перечисленные выше особенности в значительной степени осложняют построение систем автоматического распознавания речи, они также в значительной степени влияют на задачу поиска ключевых слов.

В последние годы акустические и языковые модели, основанные на различных типах нейронных сетей, продемонстрировали свою эффективность в разнообразных областях обработки натуральных языков. Так, для решения задачи распознавания речи различных диалектов арабского языка в работе [8] была использована техника последовательного обучения (sequen cetraining) глубоких нейронных сетей (deep neural network – DNN), что позволило значительно снизить уровень словной ошибки (word error rate – WER) относительно классических систем, основанных на скрытых марковских моделях и смеси гауссовых распределений (hidden Markov model – Gaussian mixture model (HMM-GMM)). В [9] использование комбинации ансамбля DNN и сверточной нейронной сети (convolution nalneural network – CNN), в совокупности с внушительным объемом обучающих данных (около 100 часов) и нейросетевой языковой моделью (neural network language model – NNLM), позволило авторам достичь впечатляющих результатов в распознавании разговорного египетского (conversational Egyptian Arabic). Некоторые ученые предпринимают попытки повысить качество распознавания за счет использования диалектных особенностей, перечисленных выше. Так, в [7] языковая модель, основанная на морфемном разделении и нейронных сетях (neural network morpheme-based feature-rich language model) позволила достичь снижения уровня словной ошибки на абсолютных 0,6–0,7%. Межъязыковое акустическое моделирование

(Phonemic Cross-Lingual Acoustic Modeling) [4] является попыткой получения корректных фонемных транскрипций для разговорного египетского при помощи данных MSA. Такой подход позволяет достичь относительного снижения уровня словной ошибки на 41,8%.

На сегодняшний день актуальной проблемой для языков с ограниченными ресурсами, кроме распознавания, является также и задача поиска ключевых слов. Такие проекты, как Open KWS, нацелены на быструю разработку систем распознавания (automatic speech recognition) и поиска (key word search) для новых языков, располагающих скромными ресурсами [10].

В настоящей работе описаны несколько вариантов системы распознавания речи и поиска ключевых слов для египетского диалекта арабского языка, построенных на основе корпуса CALLHOME Egyptian Arabic. Показан вклад от использования дикторозависимых акустических признаков, полученных при помощи данных русского, английского, а также различных диалектов арабского языков. Также была исследована возможность использования небольшого набора данных современного стандартного арабского для получения фонетических транскрипций. Для поиска ключевых слов был использован набор инструментов Kaldi [11], а именно утилиты индексированного поиска.

Описание использованных наборов данных

Целью работы была разработка системы автоматического распознавания речи и поиска ключевых слов для египетского диалекта арабского языка. Ввиду этого для обучения и тестирования моделей был использован набор данных CALLHOME Arabic (LDC97S45 – аудио, LDC97T19 – текстовки). Этот набор данных представляет собой записи спонтанных телефонных переговоров между носителями языка, а именно, разговорного египетского арабского, классифицируемого как каирский арабский диалект. Общий объем данных равен 120 записям, каждая из которых имела продолжительность до 30 мин. В каждой из записей произвольно был выбран фрагмент длительностью 5 или 10 мин, для которого были составлены текстовые транскрипции. Весь набор разделен на 3 части: обучение (training), настройка (development) и тестирование (evaluation). Обучающее подмножество состоит из 80 записей и представляет собой около 14 ч речи, содержащей примерно 130000 слов. 20 записей были использованы для формирования development-набора длительностью 3,5 ч, содержащего 32000 слов. Оставшиеся 20 записей составляли тестовый набор, содержащий около 14000 слов. Кроме того, был использован дополнительный материал, распространяемый Linguistic Data Consortium – это наборы LDC2002S37 (аудио) объемом около 2 ч и LDC2002T38 (текстовки), содержащие почти 16000 слов. Этот материал был добавлен к обучающим данным как для акустических, так и для языковых моделей.

Также были использованы данные Egyptian Colloquial Arabic Lexicon (LDC99L22) в качестве словаря произношений (фонемных транскрипций). Стоит упомянуть, что оригинальные наборы транскрипций представлены в двух вариантах: это романизированные транскрипции и арабская вязь. В данной работе были использованы текстовки на арабской вязи. Исходный словарь содержит почти 52000 романизированных представлений слов, в том числе множественные транскрипции. В работе был использован набор из 34 фонем, отличный от используемого в оригинальном словаре. Для этого были установлены однозначные соответствия между исходными фонемами и их заменами, после чего исходный словарь был переразмечен. В итоге объем словаря составил почти 57000 уникальных вхождений (романизированное представление, арабская вязь, транскрипция).

С целью обучения экстрактора bottleneck-признаков был использован набор данных Levantine Arabic QT Training Data Set 5 (LDC2006S29 – аудио, LDC2006T07 – текстовки). Этот корпус содержит 1660 записей телефонных переговоров на сиро-палестинском диалекте арабского языка общим объемом около 250 часов. Текстовки представляют собой арабскую вязь с указанием временных границ каждого произнесения. Стоит отметить, что, в отличие от корпуса CALLHOME Arabic, для данного набора не существует соответствующего словаря произношений (фонемных транскрипций). Для использования этого набора на материале NEMLAR Speech Synthesis Corpus, содержащем около 10 ч речи подготовленных дикторов, была обучена модель преобразования графем в фонемы (graphemetophoneme – G2P). Набор данных NEMLAR состоит из 2032 предложений, покрывающих примерно 42000 слов. Так как каждая запись сопровождается фонемной транскрипцией в SAMPA формате, была сформирована обучающая выборка для построения G2P модели. Выборка имела следующую структуру:

1. предложение (последовательность слов, разделенная пробелами);
2. фонемная транскрипция (последовательность фонем для каждого слова, разделенная пробелами, каждая последовательность отделена специальным символом).

Получившаяся выборка содержит около 12000 предложений (как с огласовками, так и без).

Описание системы

Для построения систем автоматического распознавания речи были обучены три экстрактора bottleneck-признаков:

1. экстрактор, обученный на данных русскоязычных телефонных переговоров;

2. экстрактор, обученный на данных англоязычных телефонных переговоров. Набор данных – The Switchboard-1 Telephone Speech Corpus;
3. экстрактор, обученный на данных арабских телефонных переговоров. Набор данных – Levantine Arabic QT Training Data Set 5.

В качестве акустических признаков для обучения экстракторов были использованы банки фильтров с их производными (первого и второго порядков), для которых была проведена нормализация среднего значения (cepstral mean normalization) [12] и добавлен 50-размерный i-vector [13] (вектор, содержащий информацию о дикторской канальной изменчивости). На вход экстрактору подавались 11 последовательных фреймов (1 центральный и ± 5 справа/слева). Выходными признаками являлись 80-мерные bottleneck [14]. Для выравнивания были построены HMM-GMM-модели с использованием максимизации правдоподобия линейной регрессии в пространстве признаков (feature-space Maximum Likelihood Linear Regression – fMLLR) [15] и дикtoроадаптированное обучение (speaker adapted training – SAT). Исходными акустическими признаками для HMM-GMM-систем были мел-кепстральные коэффициенты (Mel Frequency Cepstral Coefficient – MFCC) с их производными и нормализацией кепстрального среднего (Cepstral Mean Normalization – CMN).

На признаках, полученных от каждого экстрактора, были обучены DNN-модели, принимавшие на вход 11 последовательных фреймов (1 центральный и ±5 справа/слева), которые были объединены в вектор признаков размерностью 80×11=880. В данной работе были использованы DNN-модели из 4 слоев, по 1536 нейрона в каждом. В качестве функции активации нейронов, была использована сигмоида. Все DNN-модели были обучены по критерию кросс-энтропии. После этого для получения финальных моделей было проведено 2 итерации sequencetraining с критерием минимизации байесова риска в пространстве состояний (state-levelminimum bayes risk – sMBR) [16]. Все полученные модели были обучены с использованием Egyptian Colloquial Arabic Lexicon (кроме арабского экстрактора, который использовал фонемные транскрипции от G2P-модели, упомянутой ранее).

Для исследования применимости словаря произношений, полученного при помощи G2P-модели, обученной на MSA-данных, мы провели полный цикл обучения акустических моделей, от HMM-GMM до финальной DNN. Для языковой модели мы использовали данные обучающей выборки (LDC97T19 – train) и текстовки дополнительного набора данных (LDC2002T38). На этих текстах мы обучили трехграммную языковую модель. Для проведения поиска ключевых слов мы пользовались средствами Kaldi.

Эксперименты

Для определения качества полученных моделей были использованы два набора данных:
1. Evaluationset: продолжительность – 1,5 ч, количество слов около 15500, из них почти 13800 словарных и 1700 внесловарных;
2. Developmentset: продолжительность 3,5 ч, количество слов почти 33000, из которых 29100 словарных и 3900 внесловарных.

Качество распознавания для различных обученных акустических моделей приведено в табл. 1.

Модель	Признаки	Словарь	WER %	
			dev	eval
HMM-GMM fMLLR SAT	39MFCC+deltas+CMN	ECA	59,24	57,64
HMM-GMM fMLLR SAT	39MFCC+deltas+CMN	MSA	63,03	61,83
DNN 4×1536	RUS BN	ECA	52,27	52,91
	ENG BN		52,52	52,86
	ARA BN		54,02	54,48
DNN 4×1536 sMBR	RUS BN	ECA	49,20	49,84
	ENG BN		51,49	51,33
	ARA BN		51,77	52,26
DNN 4×1536	RUS BN	MSA	56,86	56,55
DNN 4×1536 sMBR			54,57	54,39

Таблица 1. Качество распознавания для различных конфигураций системы

Как видно из результатов, представленных в табл. 1, применение акустических моделей, основанных на глубоких нейронных сетях, значительно повышает качество распознавания относительно классических HMM-GMM. Кроме того, sequence-training последовательно снижает WER. Использование словаря фонемных транскрипций, построенного по данным MSA, значительно ухудшает качество распознавания, однако результаты остаются приемлемыми.

После получения результатов распознавания были проведены эксперименты по поиску ключевых слов. Поиск был выполнен на каждом из наборов данных development и evaluation. Для каждого из этих наборов были сформированы следующие поисковые запросы:

– 100uni включает в себя 100 наиболее частотных словарных униграм;
– 50uni_35bi_15tri состоит из 50 наиболее частотных словарных униграм, 35 биграм и 15 триграм.

Результаты поиска ключевых слов приведены в табл. 2.

Поисковые запросы	Признаки	Словарь	dev		eval	
			ATWV	MTWV	ATWV	MTWV
100uni	RUS BN	MSA	0,4010	0,6691	0,3350	0,3528
		ECA	0,5741	0,8204	0,5413	0,5728
	ENG BN		0,5687	0,8065	0,5211	0,5476
	ARA BN		0,5605	0,7979	0,5177	0,5431
50uni_35bi_15tri	RUS BN	MSA	0,4424	0,5951	0,4556	0,4920
		ECA	0,6558	0,7802	0,5871	0,6850
	ENG BN		0,5824	0,7214	0,6310	0,7092
	ARA BN		0,6419	0,7634	0,6073	0,6571

Таблица 2. Качество поиска ключевых слов для различных конфигураций системы

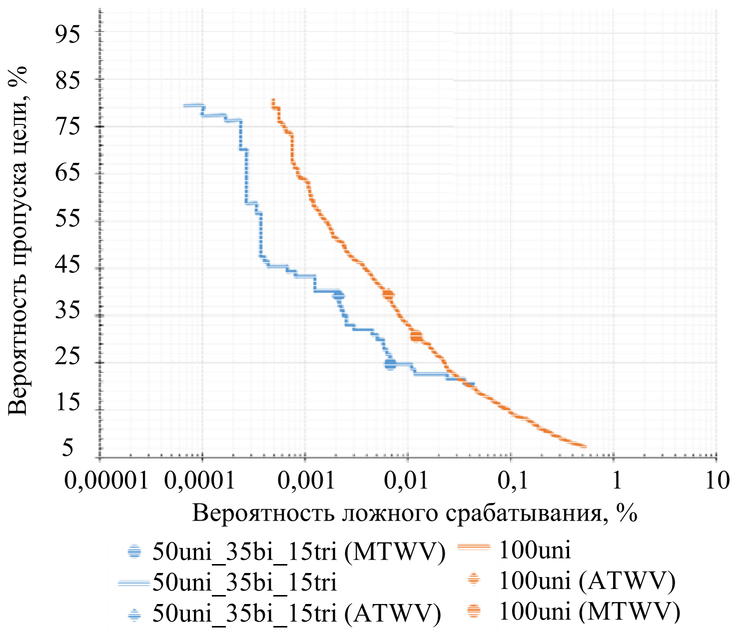

Рис. 1. Качество поиска для evaluationset

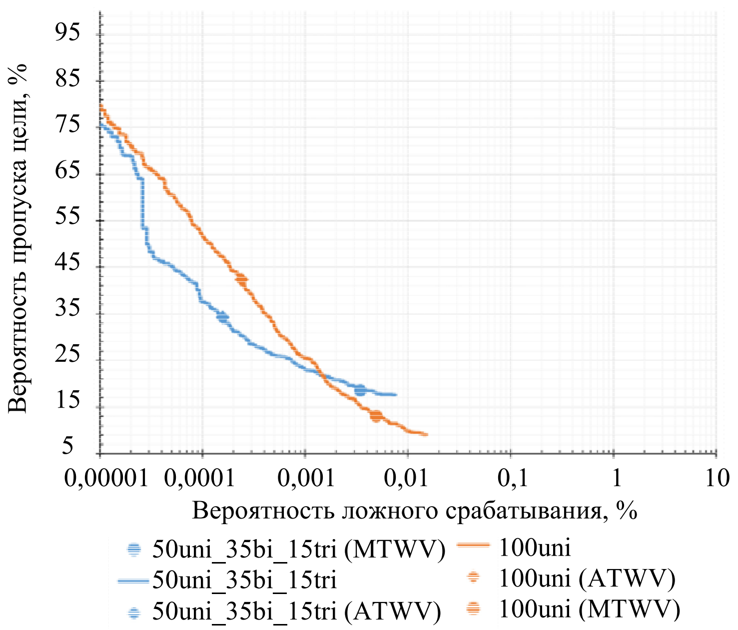

Рис. 2. Качество поиска для developmentset

Лучшие результаты, полученные с использованием русских bottleneck-признаков и Egyptian Colloquial Arabic Lexicon представлены на рис. 1, 2.

Заключение

В работе была продемонстрирована возможность разработки системы автоматического распознавания речи для языка с ограниченными ресурсами. Разработанная система обладает показателями качества, сравнимыми с ранее опубликованными. Использование bottleneck-признаков в совокупности с DNN акустическими моделями позволило получить современные результаты распознавания. Система, обученная на русских bottleneck-признаках продемонстрировала наилучший результат, который на абсолютных 6,16% превосходит результаты опубликованные в [5]. Стоит отметить, что в наборе инструментов Kaldi присутствуют результаты для рецепта CALLHOME Egyptian, где также был использован набор данных CALLHOME Egyptian. В данном рецепте акустическая модель, основанная на Time Delayed Neural Network и i-vector показывает результаты, уступающие продемонстрированному, абсолютных 2,26%.

Было проведено сравнение использования транскрипций Egyptian Colloquial Arabic Lexicon с транскрипциями, полученными G2P-моделью, обученной на малом объеме данных MSA. Полученное ухудшение качества в 4,5% является значительным, однако позволяет обходиться без специализированного египетского лексикона, и работать с более доступными данными MSA.

Из проведенных экспериментов следует, что арабские bottleneck показывают такое же качество, как и английские, в то время, как русские превосходят их на абсолютных 2%.

Показатели качества поиска ключевых слов, представленные в табл. 2, демонстрируют возможность использования полученной системы в реальных условиях.

В качестве дальнейших работ планируется исследовать использование long-shorttermmemoryи двунаправленные (bidirectional) нейронные сети в качестве акустических моделей и нейросетевые языковые модели с целью повышения качества распознавания речи. Также использование дополнительных объемов данных позволило бы провести более детальные исследования и значительно снизить WER.

References

1. Kirchhoff K., Bilmes J., Das S., Duta N., Egan M., Ji G., He F., Henderson J., Liu D., Noamany M., Schone P., Schwartz R., Vergyri D. Novel approaches to Arabic speech recognition: report from the 2002 Johns-Hopkins Summer Workshop. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP*. Hong Kong, 2003, vol. 1, pp. 344–347.
2. *Human Development Report 2006. Beyond Scarcity: Power, Poverty and Global Water Crisis*. Palgrave Macmillan, UK, 2006, pp. 297–300.
3. Habash N., Eskander R., Hawwari A. A morphological analyzer for egyptian Arabic. *NAACL-HLT 2012 Workshop on Computational Morphology and Phonology, SIGMOR-PHON2012*. 2012, pp. 1–9.
4. Elmahdy M., Hasegawa-Johnson M., Mustafawi E., Duwairi R., Minker W. Challenges and techniques for dialectal arabic speech recognition and machine translation. *Proc. Qatar Foundation Annual Research Forum*. Doha, 2011.
5. Elmahdy M., Hasegawa-Johnson M., Mustafawi E. Hybrid phonemic and graphemic modeling for arabic speech recognition. *International Journal of Computational Linguistics*, 2012, vol. 3, no. 1, pp. 88–96.
6. Ali A., Mubarak H., Vogel S. Advances in dialectal arabic speech recognition: a study using twitter to improve Egyptian ASR. *Proc. Int. Workshop on Spoken Language Translation, IWSLT 2014*. South Lake Tahoe, USA, 2014, pp. 156–162.
7. El-Desoky Mousa A., Kuo H.-K.J., Mangu L., Soltau H. Morpheme-based feature-rich language models using Deep Neural Networks for LVCSR of Egyptian Arabic. *Proc. 38th IEEE Int. Conf. on Acoustics Speech and Signal Processing, ICASSP*. Vancouver, Canada, 2013, pp. 8435–8439. doi: 10.1109/ICASSP.2013.6639311
8. Ali A., Zhang Y., Cardinal P., Dahak N., Vogel S., Glass J. A complete KALDI recipe for building Arabic speech recognition systems. *Proc. IEEE Workshop on Spoken Language Technology*. South Lake Tahoe, USA, 2014, pp. 525–529. doi: 10.1109/SLT.2014.7078629
9. Thomas S.W., Saon G., Kuo H.-K., Mangu L. The IBM BOLT speech transcription system. *Proc. 6th Annual Conference of the International Speech Communication Association*. Dresden, Germany, 2015, pp. 3150–3153.
10. Trmal J., Chen G., Povey D., Khudanpur S. et. al. A keyword search system using open source software. *Proc. IEEE Workshop on Spoken Language Technology*. South Lake Tahoe, USA, 2014, pp. 530–535.
11. Povey D., Ghoshal A. et al. The Kaldi speech recognition toolkit. *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU*. Waikoloa, Hawaii, USA, 2011.
12. Liu F., Stern R., Huang X., Acero A. Efficient cepstral normalization for robust speech recognition. *Proc. ARPA Workshop on Human Language Technology*. Princeton, 1993, pp. 69–74. doi: 10.3115/1075671.1075688

13. Senoussaoui M., Kenny P., Dehak N., Dumouchel P. An i-vector extractor suitable for speaker recognition with both microphone and telephone speech. *Odyssey 2010. The Speaker and Language Recognition Workshop*. Brno, Czech Republic, 2010, pp. 28–33.
14. Gehring J., Miao Y., Metze F., Waibel A. Extracting deep bottleneck features using stacked auto-encoders. *Proc. 38th IEEE Int. Conf. on Acoustics Speech and Signal Processing, ICASSP*. Vancouver, Canada, 2013, pp. 3377–3381. doi: 10.1109/ICASSP.2013.6638284
15. Xin L., Hamaker J., He X. Robust feature space adaptation for telephony speech recognition. *Proc. 9th Int. Conf. on Spoken Language Processing*. Pittsburgh, USA, 2006, pp. 773–776.
16. Vesely K., Ghoshal A., Burget L., Povey D. Sequence-discriminative training of deep neural networks. *Proc. 14th Annual Conf. of the International Speech Communication*. Lyon, France, 2013, pp. 2345–2349.

Романенко Алексей Николаевич – аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; младший научный сотрудник, Центр Речевых Технологий, Санкт-Петербург, 196084, Российская Федерация, romanenko@speechpro.com

Alexei N. Romanenko – postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation; junior researcher, Speech Technology Center, Saint Petersburg, 196084, Russian Federation, romanenko@speechpro.com