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 004.02

АКТУАЛЬНЫЕ ЗАДАЧИ И ДОСТИЖЕНИЯ СИСТЕМ ПАРАЛИНГВИСТИЧЕСКОГО АНАЛИЗА РЕЧИ

А.А. Карпов[a,b], Х. Кайа[c], А.А. Салах[d]

[a] Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация
[b] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[c] Университет Намык Кемаль, Чорлу / Текирдаг, 59860, Турция
[d] Босфорский Университет, Стамбул, 34342, Турция
Адрес для переписки: karpov@iias.spb.su

Информация о статье
Поступила в редакцию 04.05.16, принята к печати 14.06.16
doi: 10.17586/2226-1494-2016-16-4-581-592
Язык статьи – русский

Ссылка для цитирования: Карпов А.А., Кайа Х., Салах А.А. Актуальные задачи и достижения систем паралингвистического анализа речи // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 4. С. 581–592. doi: 10.17586/2226-1494-2016-16-4-581-592

Аннотация
Представлен аналитической обзор современных и актуальных задач, стоящих в области компьютерной паралингвистики, а также последних достижений автоматических систем паралингвистического анализа разговорной речи. Паралингвистика изучает невербальные аспекты человеческой коммуникации и речи: естественные эмоции, акценты, психофизиологические состояния, особенности произношения, параметры голоса диктора и т.д. Представлена архитектура базовой компьютерной системы акустического паралингвистического анализа, ее основные компоненты и используемые методы обработки речи. Приведена информация о международных соревнованиях по компьютерной паралингвистике Computational Paralinguistics Challenge (ComParE), которые с 2009 года проходят ежегодно в рамках международной конференции INTERSPEECH, организуемой международной ассоциацией по речевой коммуникации ISCA. Представлены задачи (конкурсы), которые решались в рамках данного соревнования в период с 2009 по 2016 г.г., а также компьютерные системы, победившие в каждом из проведенных конкурсов, и полученные результаты. Последние завершенные соревнования ComParE-2015 проходили в сентябре 2015 года в Германии и содержали следующие 3 конкурса: 1) распознавание дикторов, которые говорят на родном для них языке (DN); 2) предсказание наличия болезни Паркинсона по речи (PC); 3) автоматическое определение, ест ли человек (диктор) во время говорения или диалога, и классификация вида пищи (определение одного из 7 типов), которую он принимает в это время. В последнем конкурсе («Eating Condition Sub-Challenge», EC) победу одержала совместная турецко-российская команда авторов данной статьи, которая разработала наиболее эффективную компьютерную систему для определения и классификации соответствующих акустических паралингвистических явлений. В статье представлена архитектура данной системы и основные модели и методы, описаны используемые обучающие и тестовые аудиоданные, а также наилучшие полученные результаты по машинной классификации акустических паралингвистических явлений.

Ключевые слова
компьютерная паралингвистика, речевые технологии, акустический анализ, распознавание эмоций, машинное обучение, состояния диктора, акустические паралингвистические явления

Благодарности
Исследование выполнено при финансовой поддержке фонда РФФИ (проект № 16-37-60100) и Совета по грантам Президента РФ (проект № МД-3035.2015.8).

STATE-OF-THE-ART TASKS AND ACHIEVEMENTS OF PARALINGUISTIC SPEECH ANALYSIS SYSTEMS

A.A. Karpov[a,b], H. Kaya[c], A. A. Salah[d]

[a] St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation
[b] ITMO University, Saint Petersburg, 197101, Russian Federation
[c] Namık Kemal University, Çorlu / Tekirdağ, 59860, Turkey
[d] Boğaziçi University, Bebek, Istanbul, 34342, Turkey
Corresponding author: karpov@iias.spb.su

Article info
Received 04.05.16, accepted 14.06.16
doi: 10.17586/2226-1494-2016-16-4-581-592
Article in Russian

For citation: Karpov A.A., Kaya H., Salah A.A.. State-of-the-art tasks and achievements of paralinguistic speech analysis systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 4, pp. 581–592. doi: 10.17586/2226-1494-2016-16-4-581-592

Abstract
We present analytical survey of state-of-the-art actual tasks in the area of computational paralinguistics, as well as the recent achievements of automatic systems for paralinguistic analysis of conversational speech. Paralinguistics studies non-verbal aspects of human communication and speech such as: natural emotions, accents, psycho-physiological states, pronunciation features, speaker's voice parameters, etc. We describe architecture of a baseline computer system for acoustical paralinguistic analysis, its main components and useful speech processing methods. We present some information on an International contest called Computational Paralinguistics Challenge (ComParE), which is held each year since 2009 in the framework of the International conference INTERSPEECH organized by the International Speech Communication Association. We present sub-challenges (tasks) that were proposed at the ComParE Challenges in 2009-2016, and analyze winning computer systems for each sub-challenge and obtained results. The last completed ComParE-2015 Challenge was organized in September 2015 in Germany and proposed 3 sub-challenges: 1) Degree of Nativeness (DN) sub-challenge, determination of nativeness degree of speakers based on acoustics; 2) Parkinson's Condition (PC) sub-challenge, recognition of a degree of Parkinson's condition based on speech analysis; 3) Eating Condition (EC) sub-challenge, determination of the eating condition state during speaking or a dialogue, and classification of consumed food type (one of seven classes of food) by the speaker. In the last sub-challenge (EC), the winner was a joint Turkish-Russian team consisting of the authors of the given paper. We have developed the most efficient computer-based system for detection and classification of the corresponding (EC) acoustical paralinguistic events. The paper deals with the architecture of this system, its main modules and methods, as well as the description of used training and evaluation audio data and the best obtained results on machine classification of these acoustic paralinguistic events.

Keywords
computational paralinguistics, speech technology, acoustical analysis, emotion recognition, machine learning, speaker states, acoustical paralinguistic events

Acknowledgements
This research is financially supported by the Russian Foundation for Basic Research (project No. 16-37-60100) and by the Council for Grants of the President of Russia (project No. MD-3035.2015.8).

Введение

Паралингвистика изучает различные невербальные аспекты в речи и коммуникации, например, эмоции, интонации, психофизиологические состояния, особенности произношения и параметры голоса диктора. Современная паралингвистика касается, в основном, вопросов относительно того, как речь (вербальная информация) произносится, нежели того, что именно произносится. В среднем человек говорит 10–15 мин в день (чистая речь). Информация, передаваемая людьми словами (вербально), составляет лишь около 7% от общего объема информации, получаемой человеком в процессе межчеловеческой коммуникации, тогда как на долю невербальных сигналов приходится до 93%, мимика, позы, жесты, касания, запахи и т.д. составляют свыше половины всего объема информации, а на долю голосовой паралингвистической составляющей приходится не менее трети всей информации [1]. Паралингвистическую информацию также обычно отличают от экстралингвистической, с которой ассоциируют не связанные с речью акустические явления (кашель, смех, вздохи, плач, заикание и другие индивидуальные особенности произношения).

Компьютерная паралингвистика (computational paralinguistics) является одной из самых актуальных и динамично развивающихся областей современных речевых технологий (speech technology). Распознавание эмоций в речи (РЭР) человека является наиболее востребованной областью компьютерной паралингвистики, оно тесно связано с такими направлениями, как распознавание состояния диктора и анализ его особенностей. Общее состояние диктора (человека), как правило, соответствует динамично изменяющимся окружающим условиям и может описываться такими параметрами, как эмоциональное и психофизиологическое состояние, состояние здоровья, усталость, стресс и т.д. Особенности же диктора соответствуют неизменным или относительно постоянным характеристикам человека, таким как пол, возраст, этническая принадлежность, параметры индивидуальности (характер), внешний вид и т.д.

Архитектура базовой системы паралингвистического анализа речи

Общая архитектура базовой автоматической системы, предназначенной для паралингвистического анализа речи [2] и для распознавания состояния диктора, представлена на рис. 1. Эта диаграмма анализа речи содержит методы цифровой обработки сигналов и распознавания образов. В работе [3] было также показано, что, когда система автоматического распознавания речи (АРР) включена в цикл РЭР, применяемые в них модели языка могут помочь улучшить точность распознавания эмоций. Однако такое не всегда возможно, и, соответственно, распознавание эмоциональной (аффективной) речи является независимой значимой задачей [3].

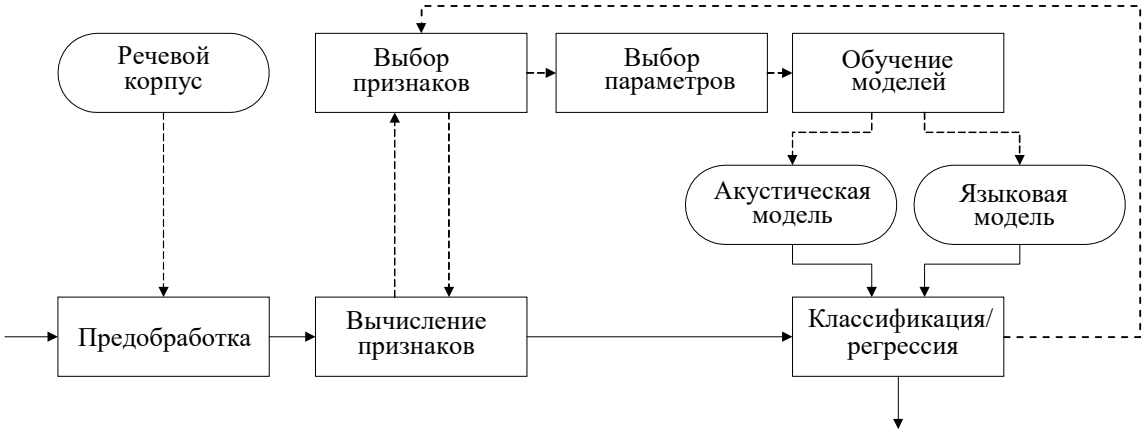

Рис. 1. Общая архитектура автоматической системы компьютерного паралингвистического анализа речи для распознавания состояний диктора

Проблемы обработки эмоциональной речи, в частности, включают в себя (но не ограничиваются) и анализ персональных, культурных и национальных отличий, а также и преднамеренное и непреднамеренное выражение эмоций в повседневной жизни [4–6]. Базы данных (БД) эмоциональной речи зачастую собираются в изолированных от естественной жизни условиях (RUSLANA [7], EMODB [8], BUEMODB [9]), и в них эмоциональная составляющая речи часто оказывается преувеличенной (например, наигранной профессиональными актерами). С другой стороны, исследования в этой области необходимы по двум причинам:

1. современные методы машинного обучения (Machine Learning) и распознавания образов (которые активно используются, например, для задач обработки изображений) пока недостаточно применяются для этой задачи, поэтому их использование позволяет получить значимые результаты;
2. хотя существуют отдельные информативные признаки (параметры) для анализа эмоций в речи, но пока нет хорошего устоявшегося набора признаков для гарантированного распознавания эмоций в разных речевых базах данных [5].

Современные системы для паралингвистического анализа используют пространства признаков огромного размера для интегрального описания целых фраз, а не отдельных слогов или фонем, например, так называемые низкоуровневые описатели (Low Level Descriptors, LLD). Таким является набор признаков ComParE [10], который содержит более 6 тыс. элементов для каждой произнесенной диктором фразы, и вычисляется этот набор при помощи средств программного инструментария openSMILE [11], разработанного немецкой компанией audEERING.

В области математической паралингвистики часто используют вычисление LLD-признаков для описания речевых сигналов. Такой набор параметров может включать в себя, в том числе, частоту основного тона (F0), форманты (резонансные частоты голосового тракта), мел-частотные кепстральные коэффициенты (MFCC), модулированный спектр сигнала, коэффициенты перцептивного линейного предсказания (RASTA-PLP), энергетические признаки сигнала и их вариативности (так называемые джиттер и шиммер) и т.д. MFCC и RASTA-PLP-признаки известны в АРР довольно давно и были заимствованы в РЭР из этой задачи, недавно к ним добавились также частотные признаки речи LSF (Line Spectral Frequency) [12]. Наиболее распространенными и эффективными методами моделирования и классификации в области РЭР, как и в АРР, на сегодняшний день являются смеси гауссовских распределений (Gaussian Mixture Models, GMM), скрытые марковские модели (Hidden Markov Models, HMM), искусственные нейронные сети (Artificial Neural Networks, ANN), метод опорных векторов (Support Vector Machines, SVM) [5]. Практически все современные системы РЭР строятся с использованием методов SVM и ANN, при этом последние получили недавно свою «вторую жизнь» после появления методов глубокого обучения и глубоких нейронных сетей (Deep Neural Networks, DNN) [2]. Семейство DNN-сетей составляют такие их разновидности,как: сверточные нейронные сети (Convolutional Neural Networks, CNN), рекуррентные нейронные сети (Recurrent Networks RNN) и т.д. [13], которые сейчас успешно применяются для задач распознавания речи. Из семейства нейронных сетей вышла также и модель экстремального обучения (Extreme Learning Machines, ELM), которая соединяет возможности быстрого обучения модели и точного предсказания данных. Модель ELM недавно была успешно использована для задачи многомодального (аудиовизуального) распознавания эмоций в реальных условиях (in the wild) [14]. В недавней обзорной статье также отмечалось [5], что некоторые методы, устоявшиеся в области машинного обучения [15], не очень обдуманно применяются к задаче РЭР, например, объединение множественных моделей на уровне позднего объединения информации.

Международные соревнования по компьютерной паралингвистике INTERSPEECH Computational Paralinguistics Challenge

С 2009 года в рамках международной конференции INTERSPEECH проходят ежегодные соревнования автоматических систем паралингвистического анализа Computational Paralinguistics ChallengE (ComParE, http://compare.openaudio.eu), посвященные различным направлениям исследований в области компьютерной паралингвистики и проводимые сообществом AAAC (Association for the Advancement of Affective Computing, она же HUMAINE ассоциация: http://emotion-research.net/sigs/speech-sig).

В 2009 году первые соревнования INTERSPEECH 2009 Emotion Challenge были посвящены анализу эмоций в речи дикторов (базовая статья по этому соревнованию [16]) по двум конкурсам/направлениям (sub-challenges): наиболее точная классификация эмоций в речи (Classifier Performance Sub-Challenge) [17] (здесь и далее приведены ссылки на статьи, описывающие системы, победившие в каждом из конкурсов) и конкурс открытых проектов по автоматическому паралингвистическому анализу речи (Open Performance Sub-Challenge) [18].

В 2010 году соревнования Paralinguistic Challenge проходили по трем направлениям (базовая статья по данному соревнованию [19]): распознавание по речи диктора его возраста (Age Sub-Challenge) [20], пола (Gender Sub-Challenge) [21] и состояния аффекта (Affect Sub-Challenge) [22].

В 2011 году соревнования Speaker State Challenge проходили по двум направлениям (базовая статья по этому соревнованию [23]): распознавание интоксикации человека (состояния опьянения) по речи (Intoxication Sub-Challenge) [24] и состояния сонливости диктора (Sleepiness Sub-Challenge) [25].

В 2012 году соревнования Speaker Trait Challenge проходили по направлениям (базовая статья [26]): распознавание индивидуальности диктора (Personality Sub-Challenge) [27], оценка привлекательности голоса диктора (Likability Sub-Challenge) [28] и определение патологии речи (Pathology Sub-Challenge) [29]. В табл. 1 представлена информация о результатах и работах победителей и призеров данных трех конкурсов, а также сравнение с базовыми результатами организаторов. В данном соревновании все результаты (качество работы автоматических систем классификации акустических паралингвистических явлений) оцениваются и сравниваются с применением основной метрики полноты (Unweighted Average Recall, UAR) [26].

Конкурс	Personality Sub-Challenge		Likability Sub-Challenge		Pathology Sub-Challenge	
Победитель	69,3	[27]	65,8	[28]	76,8	[29]
2-е место	68,4	[28]	64,0	[31]	73,7	[33]
3-е место	68,1	[30]	62,5	[32]	71,9	[34]
Базовая	68,3	[26]	59,0	[26]	68,9	[26]

Таблица 1. Результаты соревнования Computational Paralinguistics Challenge в 2012 году

В 2013 году соревнования ComParE проходили по направлениям (базовые статьи [10] и [35]): распознавание степени заболевания аутизмом (Autism Sub-Challenge) [36], определение наличия конфликта между дикторами (Conflict Sub-Challenge) [37], распознавание эмоционального состояния диктора (Emotion Sub-Challenge) [38], анализ социальных сигналов (Social Signals Sub-Challenge) [39]. Тут стоит также упомянуть, что результаты последнего конкурса были впоследствии превзойдены в работе Х. Кайа и его коллег [40]. В табл. 2 представлена информация о результатах и работах победителей и призеров данных четырех конкурсов, а также сравнение с базовыми результатами.

Конкурс	Autism Sub-Challenge		Conflict Sub-Challenge		Emotion Sub-Challenge		Social Signals Sub-Challenge	
Победитель	69,4	[36]	83,9	[37]	42,3	[38]	91,5	[39]
2-е место	66,1	[41]	83,1	[43]	41,0	[42]	89,8	[45]
3-е место	64,8	[42]	–	–	35,7	[44]	89,7	[38]
Базовая	67,1	[10]	80,8	[10]	40,9	[10]	83,3	[10]

Таблица 2. Результаты соревнования Computational Paralinguistics Challenge в 2013 году

Предпоследние соревнования ComParE в 2014 году проходили по следующим направлениям (базовая статья по этому соревнованию [46]): определение оказания физической нагрузки на диктора (Physical-Load Sub-Challenge) [47] и определение оказания умственной нагрузки на диктора (Cognitive Load Sub-Challenge) [48]. Данный конкурс соревнования был выигран коллективом турецкого Босфорского университета, г. Стамбул. В разработанной ими системе для оценки степени физической нагрузки (контролировалась посредством частоты сердцебиения), накладываемой на диктора, использовались многопоточные признаки аудиосигнала (multi-view discriminative projection), вычисляемые из LLD-наборов признаков [47, 49]. Разработанная система позволила превзойти базовую систему распознавания, показав наилучшее среднее значение полноты 75,4% (Unweighted Average Recall, UAR=75,4%). В табл. 3 представлена ин-

формация о результатах и работах победителей и призеров данных двух конкурсов, а также сравнение с базовыми результатами.

Конкурс	Physical-Load Sub-Challenge		Cognitive Load Sub-Challenge	
Победитель	75,4	[47]	68,9	[48]
2-е место	73,9	[48]	63,7	[50]
3-е место	73,0	[51]	63,1	[51]
Базовая	71,9	[46]	61,6	[46]

Таблица 3. Результаты соревнования Computational Paralinguistics Challenge в 2014 году

Последние завершенные международные соревнования ComParE-2015 (http://emotion-research.net/sigs/speech-sig/is15-compare) проходили в начале сентября 2015 года в Дрездене, Германия, в рамках конференции INTERSPEECH-2015 по следующим трем направлениям (базовая статья [52]): распознавание дикторов, которые говорят на родном для них языке (Degree of Nativeness (DN) Sub-Challenge) [53]; предсказание наличия болезни Паркинсона по речи (Parkinson's Condition (PC) Sub-Challenge) [54]; автоматическое определение, ест ли человек (диктор) во время говорения или диалога, и классификация вида пищи, которую он принимает в это время (Eating Condition (EC) Sub-Challenge) [55]. В табл. 4 представлена информация о результатах и работах победителей и призеров данных трех конкурсов, а также сравнение с базовыми результатами организаторов.

Конкурс	Degree of Nativeness (DN) Sub-Challenge		Parkinson's Condition (PC) Sub-Challenge		Eating Condition (EC) Sub-Challenge	
Победитель	0,745	[53]	0,649	[54]	83,1	[55]
2-е место	0,580	[56]	0,430	[57]	75,9	[58]
3-е место	0,510	[54]	0,349	[59]	74,6	[57]
Базовая	0,425	[52]	0,390	[52]	65,9	[52]

Таблица 4. Результаты соревнования Computational Paralinguistics Challenge в 2015 году

Система автоматической классификации акустических паралингвистических явлений для международного соревнования ComParE-2015

В 2015 г. А.А. Карпов (представляющий СПИИРАН и Университет ИТМО) участвовал в данном международном соревновании совместно с турецкими коллегами из Босфорского университета (Х. Кайа и А. Салах). Турецко-российская команда разработала автоматическую систему паралингвистического анализа речи. Предложенная система классификации паралингвистических событий заняла 1-е место в соревновании INTERSPEECH 2015 Computational Paralinguistics Challenge в номинации «Eating Condition Sub-Challenge», где программно по речи нужно было определить, ест ли человек, в то время как он говорит, и если ест, то какой тип пищи он принимает.

В аудиофайлах из базы данных iHEARu-EAT [52, 60], предоставленной организаторами из Германии, присутствовал один из 6 типов пищи, которую человек ел во время разговора, или она отсутствовала:

1. яблоко (жесткий фрукт), класс «Apple»;
2. банан (мягкий фрукт), класс «Banana»;
3. нектарин (полумягкий фрукт), класс «Nectarine»;
4. бисквит, класс «Biscuit»;
5. чипсы (хрустящая еда), класс «Crisp»;
6. жевательные конфеты, класс «Haribo», а также
7. без пищи во рту, класс «No Food».

В этой базе данных содержатся записи немецкой речи/диалоги 30 анонимных дикторов (по 15 мужчин и женщин); общая длительность речевых данных составляет около 3 часов, она содержит более 1400 файлов-высказываний.

На рис. 2 представлена архитектура разработанной автоматической системы классификации паралингвистических событий, а на рис. 3 детальнее показана схема предложенной трехуровневой каскадной нормализации данных. Система основана на вычислении низкоуровневых акустических признаков с их последующей нормализацией на основе векторов Фишера (ВФ). Для классификации акустических паралингвистических явлений применяются искусственные нейронные сети Extreme Learning Machines (ELM) и метод Partial Least Squares (PLS). Данная система детально описана в предыдущих статьях коллектива авторов [55, 61, 62].

Предложенная турецко-российским коллективом автоматическая система анализа речи позволила достичь наилучшего показателя классификации паралингвистических явлений UAR=83,1%. При этом базовая система, предложенная организаторами соревнования и основанная на низкоуровневых признаках, вычисляемых посредством средства OpenSMILE и машинах опорных векторов (SVM), показала ре-

зультат UAR=65,9%, т.е. разработанная система позволила получить относительное улучшение по сравнению с базовой системой в 26% и намного превзойти конкурентов. На рис. 4 представлена матрица спутывания (Confusion matrix) наилучшей предложенной автоматической системы классификации паралингвистических явлений.

Рис. 2. Архитектура автоматической системы для классификации речевых паралингвистических явлений

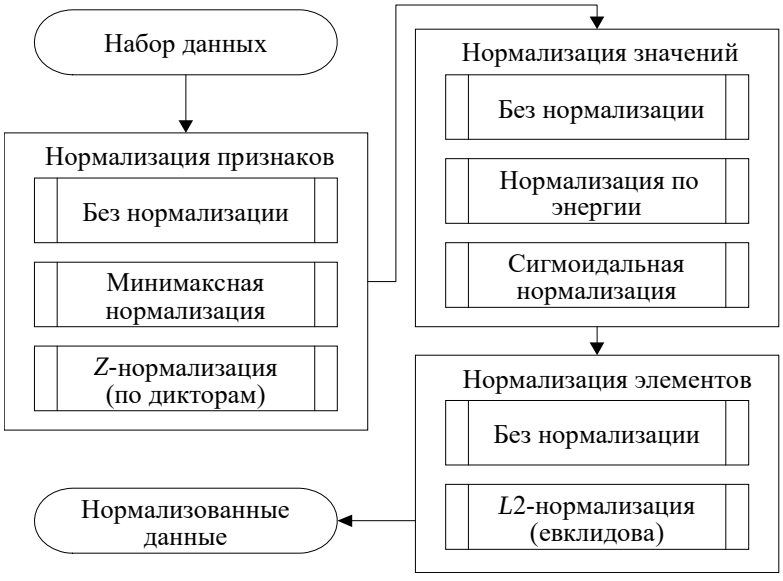

Рис. 3. Схема предложенной трехуровневой каскадной нормализации данных

По информации от организаторов, всего в соревновании INTERSPEECH 2015 Computational Paralinguistics Challenge приняли участие 30 научных коллективов из разных стран мира, а в конкурсе EC участвовали 15 различных международных коллективов. При этом второе место с результатом UAR=75,9% занял немецкий коллектив Технического Университета Дармштадта [58], а третьим стал американский коллектив из Университета Южной Калифорнии [57] с результатом UAR=74,6%.

Стоит сказать, что возможная практическая польза от такой системы заключается в следующем. Актуально ее применение в перспективных системах автоматического распознавания речи, так как известно, что иногда некоторые специалисты едят, в то время как монотонно надиктовывают длинные тексты, протоколы или заключения (например, следователи, врачи и т.д.). Кроме того, так как при принятии пищи изменяются параметры речевого тракта и характеристики голоса человека, то системы идентификации/верификации дикторов перестают надежно работать в таких случаях, поэтому необходимо создавать такие системы, адаптирующиеся к условиям и паралингвистическим явлениям.

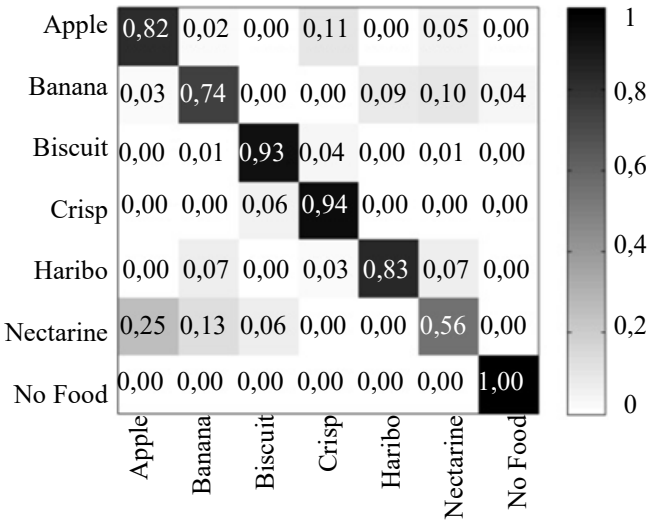

Рис. 4. Матрица спутывания наилучшей системы классификации

Заключение

В статье представлен аналитической обзор современных и актуальных задач, стоящих в области компьютерной паралингвистики, а также последних достижений автоматических систем паралингвистического анализа разговорной речи. Приведена информация о международных соревнованиях по компьютерной паралингвистике Computational Paralinguistics Challenge (ComParE), представлены задачи, которые решались в рамках данного соревнования в период с 2009 по 2016 г.г., а также компьютерные системы, победившие в каждом из проведенных конкурсов и полученные результаты. В последнем соревновании ComParE-2015 участвовала объединенная турецко-российская команда авторов данной статьи, представляющая Босфорский Университет, Турция, СПИИРАН и Университет ИТМО, Российская Федерация. Команда одержала победу в одном из конкурсов («Eating Condition Sub-Challenge», EC) соревнования, задача которого заключалась в том, чтобы автоматически по акустике определить, ест ли человек во время говорения/диалога, и классифицировать вид пищи (одного из 7 типов), которую он принимает в это время. Предложенная система позволила достичь показателя классификации UAR=83,1%. В настоящей статье представлена архитектура данной системы и основные модели и методы, а также описаны используемые обучающие и тестовые аудиоданные, и наилучшие полученные результаты по машинной классификации акустических паралингвистических явлений.

В нынешнем, 2016 году соревнования ComParE-2016 объявлены и проходят в настоящее время по трем направлениям: обнаружение ложности/истинности речевых сообщений (Deception Sub-Challenge); определение степени искренности говорящего (Sincerity Sub-Challenge), а также определение по англоязычной речи иностранцев, какой язык (записаны люди из 11 различных стран) для них является родным (Native Language Sub-Challenge). Условия данного конкурса представлены в статье организаторов [63], а результаты будут подведены в сентябре 2016 г. в ходе 17-й международной конференции INTERSPEECH-2016 в США. Турецко-российский коллектив авторов статьи также принимает участие в соревновании ComParE 2016 года [64].

References

1. Basov O.O., Karpov A.A., Saitov I.A. *Metodologicheskie Osnovy Sinteza Polimodal'nykh Infokommunikatsionnykh Sistem Gosudarstvennogo Upravleniya* [Methodological Bases of Synthesis of Multimodal Infocommunication Governance Systems]. Orel, Russian Academy of SSF, 2015, 271 p.
2. Schuller B. Voice and speech analysis in search of states and traits. In: *Computer Analysis of Human Behavior*. Eds. A.A. Salah, T. Gevers. Springer, 2011, pp. 227–253. doi: 10.1007/978-0-85729-994-9_9
3. Schuller B., Rigoll G., Lang M. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - Belief network architecture. *Proc. IEEE Int. Conf. on Acoustic, Speech and Signal Processing, ICASSP-2004*. Montreal, Canada, 2004, pp. 577–580.
4. Schuller B., Vlasenko B., Eyben F., Wollmer M., Stuhlsatz A., Wendemuth A., Rigoll G. Cross-corpus acoustic emotion recognition: variances and strategies. *IEEE Transactions on Affective Computing*, 2010, vol. 1, no. 2, pp. 119–131. doi: 10.1109/T-AFFC.2010.8
5. El Ayadi M., Kamel M.S., Karray F. Survey on speech emotion recognition: features, classification schemes, and databases. *Pattern Recognition*, 2011, vol. 44, no. 3, pp. 572–587. doi: 10.1016/j.patcog.2010.09.020

6. Dhall A., Goecke R., Lucey S., Gedeon T. Collecting large, richly annotated facial-expression databases from movies. *IEEE MultiMedia*, 2012, vol. 19, no. 3, pp. 34–41. doi: 10.1109/MMUL.2012.26
7. Makarova V., Petrushin V. RUSLANA: a database of Russian emotional utterances. *Proc. ICSLP-2002*. Denver, USA, 2002, pp. 2041–2044.
8. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss B. A database of German emotional speech. *Proc. 9th European Conf. on Speech Communication and Technology*. Lisbon, Portugal, 2005, pp. 1517–1520.
9. Kaya H., Salah A.A., Gurgen S.F., Ekenel H. Protocol and baseline for experiments on Bogazici University Turkish emotional speech corpus. *Proc. 22nd Signal Processing and Communications Applications Conf*. Trabzon, Turkey, 2014, pp. 1698–1701. doi: 10.1109/SIU.2014.6830575
10. Schuller B., Steidl S., Batliner A., Vinciarelli A., Scherer K., Ringeval F., Chetouani M., Weninger F., Eyben F., Marchi E., Mortillaro M., Salamin H., Polychroniou A., Valente F., Kim S. The INTERSEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 148–152.
11. Eyben F., Weninger F., Groß F., Schuller B. Recent developments in OpenSMILE, the Munich open-source multimedia feature extractor. *Proc. 21st ACM Int. Conf. on Multimedia*. Barcelona, Spain, 2013, pp. 835–838. doi: 10.1145/2502081.2502224
12. Bozkurt E., Erzin E., Erdem C.E., Erdem A.T. Formant position based weighted spectral features for emotion recognition. *Speech Communication*, 2011, vol. 53, no. 9–10, pp. 1186–1197. doi: 10.1016/j.specom.2011.04.003
13. Alpaydin E. *Introduction to Machine Learning*. 2nd ed. MIT Press, 2010, 581 p.
14. Kaya H., Salah A.A. Combining modality-specific extreme learning machines for emotion recognition in the wild. *Proc. 16th Int. Conf. on Multimodal Interaction ICMI-2014*. Istanbul, Turkey, 2014, pp. 487–493. doi: 10.1145/2663204.2666273
15. Schuller B., Villar R.J., Rigoll G., Lang M.K. Meta-classifiers in acoustic and linguistic feature fusion-based affect recognition. *Proc. IEEE Int. Conf. ICASSP-2005*. Philadelphia, USA, 2005, pp. 325–328. doi: 10.1109/ICASSP.2005.1415116
16. Schuller B., Steidl S., Batliner A. The INTERSEECH 2009 emotion challenge. *Proc. INTERSEECH-2009*. Brighton, UK, 2009, pp. 312–315.
17. Lee C.-C., Mower E., Busso C., Lee S., Narayanan S. Emotion recognition using a hierarchical binary decision tree approach. *Proc. INTERSEECH-2009*. Brighton, UK, 2009, pp. 320–323.
18. Dumouchel P., Dehak N., Attabi Y., Dehak R., Boufaden N. Cepstral and long-term features for emotion recognition. *Proc. INTERSEECH-2009*. Brighton, UK, 2009, pp. 344–347.
19. Schuller B., Steidl S., Batliner A., Burkhardt F., Devillers L., Mueller C., Narayanan S. The INTERSEECH 2010 paralinguistic challenge. *Proc. INTERSEECH-2010*. Makuhari, Japan, 2010, pp. 2794–2797.
20. Kockmann M., Burget L., Cernocky J. Brno University of Technology system for INTERSEECH 2010 paralinguistic challenge. *Proc. INTERSEECH-2010*. Makuhari, Japan, 2010, pp. 2822–2825.
21. Meinedo H., Trancoso I. Age and gender classification using fusion of acoustic and prosodic features. *Proc. INTERSEECH-2010*. Makuhari, Japan, 2010, pp. 2818–2821.
22. Jeon J.H., Xia R., Liu Y. Level of interest sensing in spoken dialog using multi-level fusion of acoustic and lexical evidence. *Proc. INTERSEECH-2010*. Makuhari, Japan, 2010, pp. 2802–2805
23. Schuller B., Steidl S., Batliner A., Schiel F., Krajewski J. The INTERSEECH 2011 speaker state challenge. *Proc. INTERSEECH-2011*. Florence, Italy, 2011, pp. 3201–3204.
24. Bone D., Black M.P., Li M., Metallinou A., Lee S., Narayanan S.S. Intoxicated speech detection by fusion of speaker normalized Hierarchical features and GMM supervectors. *Proc. INTERSEECH-2011*. Florence, Italy, 2011, pp. 3217–3220.
25. Huang D.Y., Ge S.S., Zhang Z. Speaker state classification based on fusion of asymmetric SIMPLS and support vector machines. *Proc. INTERSEECH-2011*. Florence, Italy, 2011, pp. 3301–3304.
26. Schuller B., Steidl S., Batliner A., Nöth E., Vinciarelli A., Burkhardt F., van Son R., Weninger F., Eyben F., Bocklet T., Mohammadi G., Weiss B. The INTERSEECH 2012 speaker trait challenge. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 254–257.
27. Ivanov A., Chen X. Modulation spectrum analysis for speaker personality trait recognition. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 278–281.
28. Montacie C., Caraty M.-J. Pitch and intonation contribution to speakers' traits classification. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 526–529.
29. Kim J., Kumar N., Tsiartas A., Li M., Narayanan S. Intelligibility classification of pathological speech using fusion of multiple subsystems. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 534–537.
30. Anumanchipalli G.K., Meinedo H., Bugalho M., Trancoso I., Oliveira L.C., Black A.W. Text-dependent pathological voice detection. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 530–533.

31. Brueckner R., Schuller B. Likability classification - a not so deep neural network approach. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 290–293.
32. Buisman H., Postma E. The log-Gabor method: speech classification using spectrogram image analysis. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 518–521.
33. Lu D., Sha F. Predicting likability of speakers with Gaussian processes. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 286–289.
34. Huang D.-Y., Zhu Y., Wu D., Yu R. Detecting intelligibility by linear dimensionality reduction and normalized voice quality hierarchical features. *Proc. INTERSEECH-2012*. Portland, USA, 2012, pp. 546–549.
35. Zhang Z., Coutinho E., Deng J., Schuller B. Cooperative learning and its application to emotion recognition from speech. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015, vol. 23, no. 1, pp. 115–126.
36. Asgari M., Bayestehtashk A., Shafran I. Robust and accurate features for detecting and diagnosing autism spectrum disorders. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 191–194.
37. Rasanen O., Pohjalainen J. Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 210–214.
38. Gosztolya G., Busa-Fekete R., Toth L. Detecting autism, emotions and social signals using Adaboost. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 220–224.
39. Gupta R., Audhkhasi K., Lee S., Narayanan S. Paralinguistic event detection from speech using probabilistic time-series smoothing and masking. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 173–177.
40. Kaya H., Ozkaptan T., Salah A.A., Gürgen F. Random discriminative projection based feature selection with application to conflict recognition. *IEEE Signal Processing Letters*, 2015, vol. 22, no. 6, pp. 671–675. doi: 10.1109/LSP.2014.2365393
41. Martinez D., Ribas D., Lleida E., Ortega A., Miguel A. Suprasegmental information modelling for autism disorder spectrum and specific language impairment classification. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 195–199.
42. Lee H.-Y., Hu T.-Y., Jing H., Chang Y.-F., Tsao Y., Kao Y.-C., Pao T.-L. Ensemble of machine learning and acoustic segment model techniques for speech emotion and autism spectrum disorders recognition. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 215–219.
43. Grezes F., Richards J., Rosenberg A. Let me finish: automatic conflict detection using speaker overlap. *Proc. INTERSPEECH-2013*. Lyon, France, 2013, pp. 200–204.
44. Sethu V., Epps J., Ambikairajah E., Li H. GMM based speaker variability compensated system for interspeech 2013 compare emotion challenge. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 205–209.
45. Janicki A. Non-linguistic vocalisation recognition based on hybrid GMM-SVM approach. *Proc. INTERSEECH-2013*. Lyon, France, 2013, pp. 153–157.
46. Schuller B., Steidl S., Batliner A., Epps J., Eyben F., Ringeval F., Marchi E., Zhang Y. The INTERSEECH 2014 computational paralinguistics challenge: cognitive & physical load. *Proc. INTERSEECH-2014*. Singapore, 2014, pp. 427–431.
47. Kaya H., Ozkaptan T., Salah A.A., Gurgen S.F. Canonical correlation analysis and local Fisher discriminant analysis based multi-view acoustic feature reduction for physical load prediction. *Proc. INTERSEECH-2014*. Singapore, 2014, pp. 442–446.
48. Van Segbroeck M., Travadi R., Vaz C., Kim J., Black M.P., Potamianos A., Narayanan S. Classification of cognitive load from speech using an i-vector framework. *Proc. INTERSEECH-2014*. Singapore, 2014, pp. 751–755.
49. Kaya H., Eyben F., Salah A.A., Schuller B.W. CCA based feature selection with application to continuous depression recognition from acoustic speech features. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP-2014*. Florence, Italy, 2014, pp. 3729–3733.
50. Kua J., Sethu V., Le P., Ambikairajah E. The UNSW submission to INTERSPEECH 2014 compare cognitive load challenge. *Proc. INTERSEECH-2014*. Singapore, 2014, pp. 746–750.
51. Gosztolya G., Grosz T., Busa-Fekete R., Toth L. Detecting the intensity of cognitive and physical load using AdaBoost and deep rectifier neural networks. *Proc. INTERSEECH-2014*. Singapore, 2014, pp. 452–456.
52. Schuller B., Steidl S., Batliner A., Hantke S., Honig F., Orozco-Arroyave J.R., Noth E., Zhang Y., Weninger F. The INTERSEECH 2015 computational paralinguistics challenge: nativeness, Parkinson's & eating condition. *Proc. INTERSEECH-2015*. Dresden, Germany, 2015, pp. 478–482.
53. Black M., Bone D., Skordilis Z., Gupta R., Xia W., Papadopoulos P., Chakravarthula S., Xiao B., Segbroeck M., Kim J., Georgiou P., Narayanan S. Automated evaluation of non-native English pronunciation quality: combining knowledge- and data-driven features at multiple time scales. *Proc. INTERSEECH-2015*. Dresden, Germany, 2015, pp. 493–497.

54. Grosz T., Busa-Fekete R., Gosztolya G., Toth L. Assessing the degree of nativeness and Parkinson's condition using Gaussian processes and deep rectifier neural networks. *Proc. INTERSEECH-2015*. Dresden, Germany, 2015, pp. 919–923.
55. Kaya H., Karpov A., Salah A. Fisher vectors with cascaded normalization for paralinguistic analysis. *Proc. INTERSEECH-2015*. Dresden, Germany, 2015, pp. 909–913.
56. Ribeiro E., Ferreira J., Olcoz J., Abad A., Moniz H., Batista F., Trancoso I. Combining multiple approaches to predict the degree of nativeness. *Proc. INTERSEECH-2015*. Dresden, Germany, 2015, pp. 488–492.
57. Kim J., Nasir M., Gupta R., Segbroeck M., Bone D., Black M., Skordilis Z., Yang Z., Georgiou P., Narayanan S. Automatic estimation of parkinson's disease severity from diverse speech tasks. *Proc. INTERSEECH-2015*. Dresden, Germany, 2015, pp. 914–918.
58. Milde B., Biemann C. Using representation learning and out-of-domain data for a paralinguistic speech task. *Proc. INTERSEECH-2015*. Dresden, Germany, 2015, pp. 904–908.
59. Hahm S., Wang J. Parkinson's condition estimation using speech acoustic and inversely mapped articulatory data. *Proc. INTERSEECH-2015*. Dresden, Germany, 2015, pp. 513–517.
60. Hantke S., Weninger F., Kurle R., Ringeval F., Batliner A., El-Desoky Mousa A., Schuller B. I hear you eat and speak: automatic recognition of eating condition and food type, use-cases, and impact on ASR performance. *PLoS ONE*, 2016, vol. 11(5). doi:10.1371/journal.pone.0154486
61. Kaya H., Karpov A., Salah A.A. Robust acoustic emotion recognition based on cascaded normalization and extreme learning machines. *Lecture Notes in Computer Science*, 2016, vol. 9719. doi:10.1007/978-3-319-40663-3_14
62. Lyakso E., Frolova O., Dmitrieva E., Grigorev A., Kaya H., Salah A.A., Karpov A. EmoChildRu: emotional child Russian speech corpus. *Lecture Notes in Computer Science*, 2015, vol. 9319, pp. 144–152. doi: 10.1007/978-3-319-23132-7_18
63. Schuller B., Steidl S., Batliner A., Hirschberg J., Burgoon J.K., Baird A., Elkins A., Zhang Y., Coutinho E., Evanini K. The INTERSPEECH 2016 computational paralinguistics challenge: deception, sincerity & native language. *Proc. INTERSEECH-2016*. San Francisco, USA, 2016.
64. Kaya H., Karpov A. Fusing acoustic feature representations for computational paralinguistics tasks. *Proc. INTERSEECH-2016*. San Francisco, USA, 2016.

Карпов Алексей Анатольевич – доктор технических наук, доцент, заведующий лабораторией, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация; профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, karpov@iias.spb.su

Кайа Хейсем – PhD, научный сотрудник, Университет Намык Кемаль, Чорлу/Текирдаг 59860, Турция, hkaya@nku.edu.tr

Салах Альберт Али – PhD, доцент, Босфорский Университет, Стамбул, 34342, Турция, salah@boun.edu.tr

Alexey A. Karpov – D.Sc., Associate professor, Head of Laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation; Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, karpov@iias.spb.su

Heysem Kaya – PhD, scientific researcher, Namık Kemal University, Çorlu / Tekirdağ, 59860, Turkey, hkaya@nku.edu.tr

Albert A. Salah – PhD, Associate professor, Boğaziçi University, Bebek, Istanbul, 34342, Turkey, salah@boun.edu.tr

Алексей Анатольевич Карпов – заведующий лабораторией речевых и многомодальных интерфейсов ФГБУН Санкт-Петербургского института информатики и автоматизации Российской Академии наук (СПИИРАН), а также профессор кафедры речевых информационных систем Университета ИТМО (по совместительству), доктор технических наук (2014 г.), доцент по специальности (2012 г.). В 2002 г. окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). С 2002 г. по настоящее время работает в СПИИРАН в лаборатории речевых и многомодальных интерфейсов (до 2008 г. – в группе речевой информатики), с 2015 г. возглавляя данную лабораторию, а также с 2014 г. – в Университете ИТМО. Победил в международном соревновании по компьютерной паралингвистике Computational Paralinguistics Challenge (конкурс «Eating Condition Sub-Challenge») в рамках международной конференции INTERSPEECH-2015 (Германия) и в конкурсе Loco Mummy Contest 2006 (Бельгия). Был приглашенным редактором международных журналов Speech Communication и

Journal of Electrical and Computer Engineering, был председателем семинара SLTU-2014, является сопредседателем оргкомитета серии конференций SPECOM. Автор более 220 статей, опубликованных в международных и отечественных научных журналах и трудах конференций, 3 монографий. Признанный эксперт в области речевых технологий и многомодальных пользовательских интерфейсов. Область научных интересов – речевые технологии, автоматическое распознавание речи, обработка аудиовизуальной речи, многомодальные человеко-машинные интерфейсы, компьютерная паралингвистика.

Alexey A. Karpov is the Head of the Speech and Multimodal Interfaces Laboratory of St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS) and Professor of the Speech Information Systems Department at the ITMO University, Dr.Tech.Sc. (2014), Ph.D. (2007), Associate Professor (2012). He has graduated from St. Petersburg State University of Aerospace Instrumentation (SUAI) in 2002. He has been working since 2002 in the Speech and Multimodal Interfaces Laboratory (until 2008 in the Speech Informatics Group) of SPIIRAS, leading this laboratory since 2015, as well as he is the Professor of ITMO University since 2014. He won the Computational Paralinguistics Challenge Award in «Eating Condition Sub-Challenge» at the International Conference INTERSPEECH-2015 (Germany) and Loco Mummy Contest 2006 (Belgium). He served as a Guest Editor for special issues in Speech Communication, Journal of Electrical and Computer Engineering. He was the chairman of SLTU-2014 and co-chair of organizing committee of SPECOM series. He has published more than 220 articles in International and Russian scientific journals and proceedings, including 3 monographs. He is an expert in the scientific domains of speech technology and multimodal (audio-visual) user interfaces. His research interests are: speech technology, automatic speech recognition, audio-visual speech processing, multimodal human-computer interfaces, and computational paralinguistics.

Хейсем Кайа защитил докторскую диссертацию (Ph.D.) на кафедре компьютерной инженерии Босфорского Университета (Boğaziçi University), Стамбул, Турция в 2015 году по тематике компьютерной паралингвистики и многомодального анализа эмоций. Получил диплом бакалавра в Босфорском Университете (в области компьютерных и образовательных технологий) в 2006 году и диплом магистра в Университете Бахчешехир (Bahçeşehir University) в области компьютерной инженерии в 2009 году соответственно. Был награжден почетными дипломами за выдающиеся достижения в его студенческих исследованиях. Двукратный победитель международных соревнований по компьютерной паралингвистике Computational Paralinguistics Challenge: «Physical Load Sub-Challenge» в 2014 году в рамках международной конференции INTERSPEECH-2014 и «Eating Condition Sub-Challenge» в 2015 году в INTERSPEECH-2015. Возглавляемая им команда также победила в соревновании по распознаванию естественных эмоций на основе видеообработки (EmotiW 2015 в рамках международной конференции ICMI). Область научных интересов – вычислительное моделирование, обработка речи, компьютерная паралингвистика, анализ эмоций. Является рецензентом ряда международных журналов, таких как IEEE Trans. on. Affective Computing; IEEE Trans. on Neural Networks and Learning Systems; Computer, Speech and Language; Digital Signal Processing; Information Fusion; Neurocomputing and IEEE Signal Processing Letters.

Heysem Kaya completed his Ph.D. thesis on computational paralinguistics and multimodal affective computing at Computer Engineering Department, Boğaziçi University, Istanbul, Turkey in 2015. He received BS and MS degrees from Boğaziçi University (Computer and Educational Technology) and Bahçeşehir University (Computer Engineering) in 2006 and 2009, respectively. He was awarded high honor and outstanding achievement certificates for his undergraduate studies. His works won two Computational Paralinguistics Challenge Awards: Physical Load Sub-Challenge at INTERSPEECH 2014, Eating Condition Sub-Challenge at INTERSPEECH 2015. His team also was the first runner up in video based emotion recognition in the wild challenge (EmotiW 2015 at ICMI International conference). His research interests include mixture model selection, speech processing, computational paralinguistics, affective computing. He serves as reviewer in IEEE Trans. on. Affective Computing; IEEE Trans. on Neural Networks and Learning Systems; Computer, Speech and Language; Digital Signal Processing; Information Fusion; Neurocomputing and IEEE Signal Processing Letters.

Альберт Али Салах получил ученую степень доктора (Ph.D.) на кафедре компьютерной инженерии Босфорского Университета (Boğaziçi University), Стамбул, Турция. В 2007–2011 г.г. работал в Институте CWI в Амстердаме и в Институте информатики Амстердамского Университета. В настоящее время работает доцентом в Босфорском Университете на кафедре компьютерной инженерии и руководит программой по когнитивным наукам. Он является признанным экспертом в области компьютерного зрения, многомодальных интерфейсов, распознавания образов, компьютерного анализа поведения людей. Опубликовал свыше 150 публикаций в данных областях, включая монографию по компьютерному анализу человеческого поведения. В 2006 году получил премию «EBF European Biometrics Research Award», а в 2014 и 2015 годах стал победителем международного соревнования «INTERSPEECH Computational Paralinguistics Challenge», также в 2014 году был удостоен премии «Boğaziçi University Foundation's Outstanding Research» Босфорского Университета. Является членом руководящего комитета семинара eNTERFACE и наблюдательного совета конференции ICMI. В 2010 году он основал международный семинар по распознаванию человеческого поведения (HBU) и

являлся его сопредседателем в 2010–2016 г.г. Был приглашенным редактором ряда специальных выпусков различных международных журналов, таких как IEEE Trans. Affective Computing, Journal of Ambient Intelligence and Smart Environments (JAISE), Journal on Multimodal Interfaces, IEEE Trans. Autonomous Mental Development, IEEE Pervasive Computing. Член редакционных коллегий журналов JAISE EAI Endorsed Transactions on Creative Technologies, JAISE and IEEE Trans. Cognitive and Developmental Systems. Является действительным членом ассоциаций ACM, IEEE, технического комитета IEEE AMD Action & Perception и биометрического совета IEEE Biometrics Council.

Albert Ali Salah received the Ph.D. degree from the Computer Engineering Department of Boğaziçi University, Istanbul, Turkey. Between 2007–2011 he worked at the CWI Institute, Amsterdam and the Informatics Institute of the University of Amsterdam. He is currently an assistant professor at Boğaziçi University, Computer Engineering Department and the chair of the Cognitive Science program. He works on computer vision, multimodal interfaces, pattern recognition, and computer analysis of human behavior, with more than 150 publications in related areas, including an edited book on computer analysis of human behavior. He received the inaugural EBF European Biometrics Research Award in 2006, INTERSPEECH Computational Paralinguistics Challenge awards in 2014 and 2015, and Boğaziçi University Foundation's Outstanding Research award in 2014. He is a member of the eNTERFACE Steering Committee and ICMI Advisory Board. He initiated the International Workshop on Human Behavior Understanding in 2010 and acted as a co-chair between 2010-2016. He served as a Guest Editor for special issues in IEEE Trans. Affective Computing, Journal of Ambient Intelligence and Smart Environments (JAISE), Journal on Multimodal Interfaces, IEEE Trans. Autonomous Mental Development, and IEEE Pervasive Computing. He is an editorial board member of JAISE EAI Endorsed Transactions on Creative Technologies, JAISE and IEEE Trans. Cognitive and Developmental Systems. He is a member of ACM, IEEE, the IEEE AMD Technical Committee taskforce on Action & Perception, and the IEEE Biometrics Council.