



DOI: <https://doi.org/10.15688/jvolsu2.2024.4.10>

UDC 81'33
LBC 81.1

Submitted: 13.11.2023
Accepted: 09.04.2024

INTERFERED RUSSIAN SPEECH OF NATIVE SPEAKERS OF THE KABARDIAN-CIRCASSIAN LANGUAGE: AN EXPERIMENTAL STUDY

Irina A. Gurtueva

Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences, Nalchik, Russia

Michael V. Kamensky

North-Caucasus Federal University, Stavropol, Russia

Abstract. The mismatch between the speech of non-native's and the native language resources used in training, as well as acoustic and linguistic modelling of pronunciation based on a data-driven approach induce the challenging task of automatic speech recognition for non-native's. The existing approaches to overcoming these difficulties are based on the use of non-native speech resources or multilingual corpora and are proved to be ineffective. It is necessary to study the specifics of speech psychophysiological mechanisms of bilinguals who use two or more language systems in communication. This paper shows the results of the experimental phonetic study of interlingual interference and accent in Russian speech of the native Kabardian-Circassian speakers. The experimental materials include audio recordings of phonatory reading of a previously prepared list of words. The preliminary acoustic analysis of the initial formant measurements executed on the non-native Russian speech of the Kabardian-Circassian speakers showed a systematic deviation of the acoustic patterns of [A] and [O] vowel allophones in the F_1 - F_2 space. The crucial identification feature of the accent vowel [A] in the speech of male and female native speakers of the Kabardian-Circassian language is the systematic deviation of its formant characteristics in space (F_1 - F_2) to the right-up direction from the standard values of the orthoepic norm of the Russian language. The identification features of vowel [O] in the speech of male speakers are the high variability of the F_2 second formant values and a significant shift of the average value of F_2 upwards from the gender-neutral and gender-specific reference values. For female speakers, the distinctive feature of the Kabardian-Circassian accent is the localisation of observations to the right by F_1 and down by F_2 relative to the gender-specific reference in space (F_1 - F_2). Comparison of vocal systems based on acoustic characteristics allows the authors to assess objectively the mutual influence of language systems, identify systematic errors in real speech activity, and reduce the inaccuracy of categorisation based on rather conventional phonetic features of vowels. This research is aimed at developing methods for modelling the speech of non-native speakers for subsequent use in speech recognition systems, as well as accent and language identification in conditions of limited linguistic resources. The experimental data obtained can be useful for further development of contrastive analysis theory, typical interference pronunciation errors identification, and also contribute to elaboration of theoretical models of foreign language acquisition.

Key words: accent identification, phonetic interference, automatic speech recognition, speech databases, contrastive analysis, language contact theory, non-standard speech databases, acoustic analysis.

Citation. Gurtueva I.A., Kamensky M.V. Interfered Russian Speech of Native Speakers of the Kabardian-Circassian Language: An Experimental Study. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2024, vol. 23, no. 4, pp. 119-141. (in Russian). DOI: <https://doi.org/10.15688/jvolsu2.2024.4.10>

**ИНТЕРФЕРИРОВАННАЯ РУССКАЯ РЕЧЬ
НОСИТЕЛЕЙ КАБАРДИНО-ЧЕРКЕССКОГО ЯЗЫКА:
ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ****Ирина Асланбековна Гуртуева**

Кабардино-Балкарский научный центр РАН, г. Нальчик, Россия

Михаил Васильевич Каменский

Северо-Кавказский федеральный университет, г. Ставрополь, Россия

Аннотация. Сложность задачи распознавания ненативной речи возникает как следствие несоответствия между речью носителей и нативными языковыми ресурсами, используемыми при обучении, а также акустическом и языковом моделировании произношения на основе data-driven approach. Существующие подходы для решения этой задачи, эксплуатирующие ресурсы ненативной речи или мультилингвальные корпуса, малоэффективны. Необходимо исследование специфики речевых психофизиологических механизмов билингов, оперирующих в коммуникации двумя и более языковыми системами. В статье показаны результаты экспериментально-фонетического исследования межязыковой интерференции и акцента в русской речи носителей кабардино-черкесского языка. Материалы эксперимента включают аудиозаписи фонационного чтения предварительно подготовленного списка слов. Предварительный акустический анализ исходных формантных измерений ненативной русской речи носителей кабардино-черкесского языка на фонетическом уровне показал систематическое отклонение акустических паттернов аллофонов гласных [А] и [О] в пространстве F_1 - F_2 . Установлено, что идентификационно значимым признаком акцентного гласного [А] в речи мужчин и женщин носителей кабардино-черкесского языка является систематическое отклонение его формантных характеристик в пространстве F_1 - F_2 вправо-вверх от эталонных значений орфоэпической нормы русского языка. Классифицирующим признаком акцентного [О] в речи мужчин – это высокая вариативность значений второй форманты F_2 и выраженное отклонение общего среднего по F_2 вверх относительно гендерно нейтрального и гендерно специфицированного эталонов. Определено, что в речи женщин отличительным признаком кабардино-черкесского акцента служит локализация наблюдений правее по F_1 и ниже по F_2 относительно гендерно специфицированного эталона в пространстве F_1 - F_2 . Сопоставление вокалических систем на основе акустических характеристик позволяет объективно оценить взаимовлияние языковых систем, обнаружить систематические ошибки реальной речевой деятельности, снизить неточность категоризации на основе довольно условных фонетических признаков гласных. Данное исследование нацелено на разработку методов моделирования речи носителей для последующего применения в системах распознавания речи, идентификации языка и акцентов в условиях ограниченности лингвистических ресурсов. Полученные экспериментальные данные могут быть полезны для продолжения разработки теории контрастивного анализа, выявления типичных интерферированных произносительных ошибок, а также способствовать дальнейшему развитию теоретических моделей усвоения иностранных языков.

Ключевые слова: идентификация акцентов, фонетическая интерференция, распознавание речи, речевые базы данных, контрастивный анализ, теория языковых контактов, базы данных нестандартной речи, акустический анализ.

Цитирование. Гуртуева И. А., Каменский М. В. Интерферированная русская речь носителей кабардино-черкесского языка: экспериментальное исследование // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2024. – Т. 23, № 4. – С. 119–141. – DOI: <https://doi.org/10.15688/jvolsu2.2024.4.10>

Введение

Востребованность речевых систем как удобного интерфейса значительно возросла с достижением высокой эффективности автоматических систем при решении отдельных задач

распознавания речи [Stolcke et al., 2006; Seide, Li, Yu, 2011]. Речевые приложения встраиваются в самые разные продукты – от систем естественно-языкового управления программным и аппаратным обеспечением до систем биометрии и помощи людям с ограниченными возмож-

ностями здоровья. В то же время, хотя разработчики IBM и Microsoft заявили о достижении человеческого уровня точности [Chen et al., 2006; Matsoukas et al., 2006; Evermann et al., 2004], однозначно оценить современное состояние исследований в области автоматического распознавания речи довольно сложно [Lippman, 1997; Pallett, 2003]. Условия оценки и эффективность любой системы могут сильно различаться в зависимости от параметров, определяющих производительность, быстродействие и конфигурацию системы распознавания речи. Почти любая система при специально подобранных условиях достигает точности, сравнимой с человеческой.

Прорывные успехи в автоматическом транскрибировании речи связаны, главным образом, с переходом от использования скрытых моделей Маркова и гауссовых смесей к применению двунаправленных рекуррентных сетей [Stolcke et al., 2006]. Однако данный алгоритм характеризуется довольно высоким временем задержки, определяемым длиной высказывания. При этом, поскольку алгоритмы, понижающие показатели времени задержки, повышают количество требуемой вычислительной мощности, необходимо учитывать целесообразность улучшения точности распознавания речи.

Кроме того, хотя при решении проблемы автоматической сегментации смешанного аудиосигнала [Гуртуева, 2020; Das et al., 2020; Weng et al., 2014; Kristjanson et al., 2006] алгоритмы глубокого обучения успешно расщепляют источники звука, они не справляются с задачей масштабирования до большего числа говорящих, а также страдают от переобучения. Обнадеживают недавние открытия свойств слуховой репрезентации, открывающие возможности для разработки ассистивных средств для людей с нарушениями слуха [Han et al., 2019; Straetsmans et al., 2022]. Алгоритмы декодирования слухового внимания разрабатываются с использованием сетей глубоких аттракторов (DAN) [Straetmans et al., 2022].

Остается нерешенной и проблема распознавания акцентной речи. По результатам сравнительного тестирования работы приложений S Voice (Samsung) и Dragon Dictation (Nuance) с использованием образцов акцентной английской речи носителей китайского языка интернет-архива The Speech Accent Archive из общего числа сбоев для первого из указанных программ-

ных продуктов 6,0 % были вызваны речевыми сбоями, а 75,7 % – акцентной речью, для второго данное соотношение составило 22 % против 54 % соответственно. Для сравнительного анализа производительности Deep Speech 2 (Baidu) и человека при решении задачи транскрипции авторы часто цитируемой работы [Amodei et al., 2016] создали тестовый набор на широкой коллекции данных из общедоступного источника VoxForge (<http://www.voxforge.org>) с образцами из четырех акцентных групп: американско-канадской, европейской, индийской и акцентов стран Содружества наций (британский, ирландский, южноафриканский, австралийский и новозеландский акценты). Эффективность на человеческом уровне заметно лучше, чем у моделей Deep Speech 2, для всех акцентов, кроме индийского (22,15 % и 22,44 % соответственно). Необходимо отметить, что первичной лингвистической компетенцией стенографов, принимавших участие в эксперименте, было владение американским английским. Возможно, стенографы из соответствующих регионов допустили бы меньше ошибок при распознавании акцентов родных стран. Кроме того, человек быстро адаптируется к нестандартной или атипичной речи, в отличие от предварительно обученной нейронной сети.

Сложность задачи распознавания ненативной речи возникает как следствие несоответствия между речью носителей и нативными языковыми ресурсами, используемыми при обучении, а также акустическом, языковом и моделировании произношения на основе data-driven approach [Гуртуева, 2020]. Очевидный способ преодоления указанных трудностей заключается в создании информационных ресурсов речи носителей. Однако простое увеличение обучающих баз данных затратно и малоэффективно. По разным оценкам в настоящее время насчитывается около 6–7 тыс. языков. Записать речь носителей для каждого из них сложно, если не невозможно. Например, разработка речевой системы для английского языка с учетом вариатива американских акцентов требует создания речевого корпуса объемом в пять тысяч часов. Альтернативным подходом к решению проблемы учета акцентов является применение мультилингвальных ресурсов для адаптации речи носителей с использованием результатов исследований феномена «межъязыкового переноса» [Weinreich, 1979].

На основе многоязычного ресурса и информации о межъязыковом переносе создается новое лингвистическое пространство, выровненное с целевым пространством, которое можно использовать для оценки пространства речи носителей. В зависимости от типа лингвистических ресурсов, таких как многоязычные акустические модели или корпуса, предлагаются разные методы. Наиболее перспективными разработчики считают методы самоконтролируемого предварительного обучения, а также методы на основе i -вектора и глубокой нейронной сети, основанной на скрытой марковской модели (DNN-NMM) [Deng, Cao, Ma, 2021; Najafian, Russell, 2020]. Однако эффективность данных методов вне лабораторных условий остается пока нерешенной задачей.

В настоящей работе представлены результаты экспериментально-фонетического исследования межъязыковой интерференции и акцента в русской речи носителей кабардино-черкесского языка. Был проведен первичный анализ исходных формантных измерений ненативной русской речи носителей кабардино-черкесского языка на фонетическом уровне. Предварительный анализ возможности идентификации акцентов с использованием специально созданного ресурса ненативной русской речи [Фонетико-акустический...; Nagoev et al., 2022] показало систематическое отклонение акустических паттернов аллофонов гласных в пространстве F_1 - F_2 . Данная работа нацелена на развитие методов моделирования речи носителей для последующего внедрения в системах распознавания речи, идентификации языка и акцентов в условиях ограниченности лингвистических ресурсов. Полученные экспериментальные данные могут быть полезны для продолжения разработки теории контрастивного акустического анализа вокалических систем разноструктурных языков, выявлении типичных интерферированных произносительных ошибок, а также способствовать дальнейшему развитию теоретических моделей усвоения иностранных языков.

Материалы и методы

Информанты

Для выявления универсальных и специфических признаков в интерферированной русской

речи билингвов-носителей кабардино-черкесского и русского языков к записи были приглашены десять информантов – пять мужчин и пять женщин в возрасте от 21 до 49 лет (средний возраст 33,1 года). Все они – жители Кабардино-Балкарской Республики. По роду деятельности все информанты относятся к работникам сферы высшего образования и науки. 9 дикторов из 10 получили высшее образование в г. Нальчике (КБР), 1 – в г. Пятигорске (Ставропольский край). Большинство участников билингвы – носители кабардино-черкесского языка с высоким уровнем знания русского ($n = 2$). Одна из участниц владеет арабским, кабардинским, русским и английским языками ($n = 4$). В записи также участвовала испытуемая, владеющая английским языком ($n = 3$).

К участию в эксперименте привлекались информанты с разными уровнями владения родной речью, поскольку это дает возможность пронаблюдать эволюцию фонетических ошибок в речи билингвов. Уровень владения кабардино-черкесским языком участники оценивали самостоятельно при анкетировании перед началом эксперимента. Усредненная самооценка составила 3,7 балла по пятибалльной шкале.

Методика подготовки и сбора языкового материала для базы данных акустико-фонетического корпуса акцентной русской речи

Материалы эксперимента включают аудиозаписи фонационного чтения предварительно подготовленного списка слов. Лексический материал при составлении произносительного словаря ранее опубликованного проекта базы данных [Гуртуева, 2020; Программа для анализа...; Nagoev et al., 2022; SpecApp] подбирались с учетом изменения акустических паттернов аллофонов под влиянием позиционных и комбинаторных факторов, а также принимался во внимание фонетический закон оглушения согласных в конце слов русского языка. Выбранные лексемы позволяют проанализировать контактную аккомодацию, ассимиляцию и диссимиляцию шести гласных русского языка в пяти левых (абсолютное начало слова, после твердых билабиальных / переднеязычных, среднеязычных / заднеязычных, некоторых гласных / мягких согласных,

гласного *и*) и четырех правых контекстах (перед паузой, перед твердыми билабиальными / передне-, заднеязычными, гласными [О], [А], [У], [Э], [Ы] / мягкими согласными и гласным [И]). Для исследования позиционной аллофонии гласных в ненативной русской речи словарный состав сформирован таким образом, что в него включены следующие позиционные аллофоны: абсолютная сильная позиция – изолированное произнесение; первая сильная позиция – ударная позиция в начале слова перед твердым согласным; вторая сильная позиция – любая ударная позиция; первая слабая позиция – первый предударный слог или безударная позиция в абсолютном начале слова; вторая слабая позиция – любое безударное (не первый предударный или заударный) положение. Таким образом, стимульные слова произносительного словаря представляют 480 аллофонов шести гласных русского языка.

Для учета позиционной аллофонии согласных при проектировании словаря рассматривались два позиционных аллофона: сильная позиция – в ударном слоге; слабая позиция – в безударном слоге. Фонетическая комбинаторика согласных представлена в пяти правых контекстах (конец слова, перед глухими / звонкими согласными, безударными / ударными гласными). Таким образом, тридцать шесть согласных фонем представлены 180 аллофонами.

Общее количество аллофонов составляет 640. По разным оценкам, общее количество аллофонов, в зависимости от степени детализации, может колебаться от нескольких сотен до нескольких десятков тысяч [Lobanov, Tsurulnik, 2006]. Данное количество контекстов, определившее число аллофонов, было выбрано, поскольку, как показали экспериментальные исследования [Lobanov, Tsurulnik, 2006], это – минимально необходимый набор для синтеза речи, удовлетворяющего критериям разборчивости и естественности.

Объем произносительного словаря составляет 461 слово (288 слов для репрезентации гласных аллофонов и 173 слова – для согласных). Общий объем реализаций для 10 дикторов-носителей кабардино-черкесского языка составляет 4610 слов.

Продолжительность записи для каждого диктора составляет около двадцати минут в зависимости от индивидуального темпа речи.

Обработка полученных фонограмм для подготовки к последующему статистическому, психолингвистическому, аудиторскому анализу включала лексическую сегментацию, аннотацию и сплошную акустическую разметку.

Методика проведения эксперимента

Эксперимент состоял из чтения предложенного списка слов. В целях устранения лексико-грамматической интерференции чтение было подготовленным. Таким образом, описанию и анализу подлежали фонетические отклонения, возникающие в акцентной русской речи кабардинцев в процессе фонационного чтения. Перед участниками эксперимента была поставлена задача как можно более естественно прочитать слова. Считалось допустимым вносить поправки, повторы, если информант не был уверен в правильности произнесения. Все реализации записывались с помощью встроенного микрофона Realtek High-Definition Audio. Характеристики качества записи в итоговых аудиофайлах – 2 канала, 16 bit, 44100 Hz. Запись проводилась в офисных условиях. Зашумленные записи исключались из последующего анализа.

Результаты и обсуждение

Предварительный акустический анализ

Для измерения акустических характеристик аллофонов гласных [А] и [О] при проведении первичного анализа лингвистического материала, представленного аудиозаписями русской речи 10 билингов – носителей кабардино-черкесского (L1) и русского языков (L2), было выбрано по 22 слова. Отобранные слова представляют варианты каждой из указанных фонем в ударных позициях (изолированное произнесение, ударная позиция в начале слова, любая ударная позиция) в окружении непалатализованных и палатализованных смычных, сонорных, фрикативных согласных, а также некоторых гласных [Nagoev et al., 2022]. Таким образом, для каждой фонемы было проанализировано по 218 токенов. Поскольку основная акустическая информация о качестве гласных передается начальными формантами гласных [Ladefoged, Johnson, 2014], для каждого диктора с помощью программы для ана-

лиза звучащей речи Praat version 6.3.08 [Boersma, Weenink, 2003] вычислялись средние значения F_1 и F_2 в анализируемых речевых сегментах с использованием функций 'Get first formant' и 'Get second formant' (параметры настройки: метод Бурга, временное окно 25 мс, частотный диапазон 5500 Гц, число формант 5). Результаты измерений F_1 и F_2 в герцах сопоставлялись с известными в специальной научной литературе акустическими характеристиками тех же гласных нормативной русской речи [Козлачков, Дворянкин, Бонч-Бруевич, 2016; Сорокин, Цыплихин, 2004]. Далее канонические усредненные формантные значения F_1 и F_2 , полученные в работе [Козлачков, Дворянкин, Бонч-Бруевич, 2016], обозначены Reference1, а эталонные гендерно дифференцированные значения из работы [Сорокин, Цыплихин, 2004] – Reference2. Формантные измерения приведены в таблицах 1–4 и на рисунках 1, 3, 5, 7. На всех диаграммах по оси абсцисс отмечены значения первой форманты, F_1 , Гц.; на оси ординат показаны измерения значений второй форманты, F_2 , Гц; km1-km5 указывают отдельные наблюдения дикторов-мужчин; kf1-kf5 указывают индивидуальные наблюдения дикторов-женщин; km1(mean)-km5(mean)/kf1(mean)-kf5(mean) – индивидуальные средние дикторов; Mean – общее среднее по всем дикторам. На рисунках 2, 4, 6, 8 показано распределение формантных значений.

Так как гендерные различия в основном выражаются в характеристиках гласных [Леонов, Макаров, Сорокин, 2009], в данном исследовании женские и мужские голоса рассматриваются отдельно.

Анализ исходных формантных значений F_1 и F_2 аллофонов [A] для мужских голосов

Наблюдаемое распределение аллофонов [A] в пространстве F_1 - F_2 для мужских голосов характеризуется следующими признаками.

1. Средние значения координат (F_1 - F_2) для всех дикторов близки, для дикторов km1, km4 и km5 практически совпадают, для информантов km2 и km3 совпадают по F_1 .

2. Наибольшее расстояние между индивидуальными средними дикторов составляет около 50 Гц по F_1 и 150 Гц по F_2 .

3. Расстояние между Reference2 и общим средним значением по F_1 в 2 раза превышает максимальное расстояние между индивидуальными средними дикторов.

4. Все индивидуальные средние дикторов находятся выше Reference2 как по F_1 , так и по F_2 ; также все средние значения находятся выше Reference1 по F_2 и не ниже по F_1 . Более того, все наблюдения были выше Reference2 по F_1 и всего 11 % (13 из 109) были ниже по F_2 .

5. Разброс формантных значений для всех дикторов характеризуется небольшим стандартным отклонением, равным 48,01 для F_1 , и 127,64 для F_2 . Коэффициенты вариации по F_1 и F_2 равны 7,8 и 9,6 % соответственно.

6. 61,5 % наблюдений находятся выше Reference2 на расстоянии 100 Гц и более по F_2 ; 95,4 % – выше на 50 Гц и более по F_1 . Средние значения всех дикторов находятся на расстоянии более 2 стандартных отклонений по F_1 , более 1 стандартного отклонения по F_2 .

Акустический анализ приведенных на рисунке 1 формантных измерений интерферированного гласного [A] в речи мужчин позволяет сделать вывод о том, что отклонение F_1 и F_2 от значений гендерно дифференцированного эталона Reference2 устойчиво и значительно выражено. Характерной особенностью кабардино-черкесского акцента является систематическое отклонение формантных характеристик фонемы [A] в пространстве (F_1 - F_2) вправо-вверх. При сопоставлении с каноническим усредненным отклонение по F_2 выражено значительно, по F_1 – менее значимо. Для определения акцента удобнее использовать Reference2.

Анализ исходных формантных значений F_1 и F_2 аллофонов [A] для женских голосов

Наблюдаемое распределение аллофонов [A] в пространстве F_1 - F_2 для женских голосов характеризуется следующими признаками.

1. Значения отдельных наблюдений (F_1 - F_2) для всех дикторов-женщин довольно близки, варьируются около собственных индивидуальных средних в диапазоне 100 Гц (от 41,11 до 141,32 Гц) по F_1 , по F_2 – в диапазоне 150 Гц (от 67,21 до 206,15 Гц).

2. Наибольшее расстояние между индивидуальными средними информанток составляет около 140 Гц по F_1 и 200 Гц по F_2 .

3. Максимальное отклонение индивидуального среднего (kf_1) от среднего по всем наблюдениям по F_1 составляет 80 Гц. При этом Reference2 находится на расстоянии 193 Гц от общего среднего, то есть в 2,4 раза дальше.

4. Все индивидуальные средние дикторов-женщин находятся выше и эталона Reference2, и Reference1 как по F_1 , так и по F_2 . Более того, все наблюдения были выше эталона Reference2 по F_1 и всего 2 из 110 были ниже по F_2 .

5. Разброс значений формант для всех билингов-женщин характеризуется незначительным стандартным отклонением, равным 87 Гц для F_1 , и 133 Гц – для F_2 . Коэффициенты вариации по F_1 и F_2 равны 11,9 и 8,9 % соответственно.

6. 86,4 % наблюдений находятся выше эталона Reference2 на расстоянии 100 Гц и более по F_1 , 98,2 % – выше на 100 Гц и более по F_2 . По F_1 средние значения всех дикторов находятся на расстоянии более 1 стандартного отклонения от эталона, причем для четырех из пяти дикторов – на расстоянии более 2 стандартных отклонений. По F_2 все средние значения находятся на расстоянии от эталона, превышающем два стандартных отклонения.

Формантные измерения акцентного гласного [А] в речи женщин, показанные на рисунке 3, демонстрируют отклонение аналогичное наблюдениям, полученным для мужских голосов. Акустические признаки локализуются правее и выше обоих эталонных значений. При сопоставлении с Reference2 наблюдается смещение, обладающее большей идентификационной значимостью.

Анализ исходных данных F_1 и F_2 аллофонов [О] для мужских голосов

Наблюдаемое распределение аллофонов [О] в пространстве F_1 - F_2 для мужских голосов характеризуется следующими признаками.

1. Индивидуальные средние формантных значений F_1 смещены вправо, но находятся в пределах одного стандартного отклонения от обоих эталонов.

2. Наибольшее расстояние между индивидуальными средними дикторов составляет около 70 Гц по F_1 (то есть более одного стандартного отклонения) и более 300 Гц по F_2 .

Заметим, что значение коэффициента вариации для F_2 очень велико (32 %), что подтверждает визуально наблюдаемый большой разброс значений по F_2 .

3. Максимальное отклонение индивидуального среднего от общего среднего по F_2 составляет 202,06 Гц. При этом эталонное значение Reference1 находится на расстоянии 409,52 Гц от общего среднего, то есть значительно дальше (более чем в два раза). Таким образом, данное отклонение можно считать устойчивым и выраженным признаком акцента.

Таким образом, характерным признаком исследуемого акцентного [О] в речи мужчин является высокая дисперсия по значениям второй форманты, F_2 . Отклонение общего среднего по F_2 вверх устойчиво и достаточно выражено. Для решения задачи обнаружения акцента эффективнее использовать значение Reference1 в качестве эталона нормативного гласного [О]. Наблюдаемое среднее отклонение интерферирующего [О] от обоих эталонных значений по F_1 не превышает стандартного отклонения, поэтому может использоваться лишь как вспомогательный индикатор при идентификации исследуемого акцента.

Анализ исходных формантных значений F_1 и F_2 аллофонов [О] для женских голосов

Наблюдаемое распределение аллофонов [О] в пространстве F_1 - F_2 для женских голосов характеризуется следующими признаками.

1. Средние значения координат (F_1 - F_2) для всех дикторов-женщин довольно близки, варьируются в диапазоне около 40 Гц (отклонения индивидуальных средних от общего среднего по всем дикторам составляет от 0,07 до 39,04 Гц, то есть от 0 до 8,0 %) по F_1 , по F_2 – в диапазоне 158,65 Гц (от 888,53 до 1076,79 Гц).

2. Наибольшее расстояние между индивидуальными средними дикторов-женщин составляет около 65 Гц по F_1 и 200 Гц по F_2 , то есть находится в пределах одного стандартного отклонения как по F_1 , так и по F_2 . В то же время соотношение между стандартным отклонением и средним значением для F_1 (12 %) значительно меньше, чем для F_2 (21 %), что подтверждает визуально наблюдаемый больший разброс значений по F_2 .

3. Все индивидуальные средние дикторов-женщин находятся выше эталона Reference2 по F_1 и ниже по F_2 .

4. Максимальное отклонение индивидуального среднего ($kf1$) от среднего по всем наблюдениям по F_1 составляет 39,74 Гц. При этом эталонное значение находится на расстоянии 57,76 Гц от общего среднего, то есть незначительно дальше. Следует заметить, что следующее по значению отклонение индивидуального среднего составляет всего 25,28 Гц, то есть более чем в два раза ниже.

5. Что касается F_2 , то максимальное отклонение индивидуального среднего ($kf1$) от среднего по всем наблюдениям составляет 108,69 Гц. При этом эталонное значение F_2 находится на расстоянии 231,9 Гц от общего среднего, то есть значительно дальше, более, чем в два раза. Это позволяет сделать вывод, что отклонение по F_2 является более надежной характеристикой для идентификации акцентного гласного [O].

6. 83,6 % наблюдений находятся правее эталона по F_1 , 88,2 % – ниже по F_2 . По F_1 средние значения четырех из пяти дикторов находятся на расстоянии около 1 стандартного отклонения от эталона, по F_2 средние значения трех из пяти дикторов находятся на расстоянии значительно более 1 стандартного отклонения.

Таким образом, индивидуальные средние дикторов-женщин локализованы в пространстве (F_1 - F_2) правее F_1 и ниже F_2 относительно эталона Reference2 на расстоянии, позволяющем использовать данное положение в качестве идентифицирующего признака. Отклонение общего среднего по F_1 от эталонного значения Reference2 можно рассматривать в качестве значимого идентифицирующего признака. Однако более надежной характеристикой для обнаружения акцентного [O] в речи женщин является отклонение по F_2 .

Заключение

Первичный акустический анализ лингвистического материала продемонстрировал

возможность обнаружения отличительных признаков интерферирующей русской речи билингвов, носителей кабардино-черкесского языка.

Дискриминативным признаком акцентного гласного [A] в речи носителей кабардино-черкесского языка является систематическое отклонение его формантных характеристик в пространстве (F_1 , F_2) вправо-вверх от эталонных значений орфоэпической нормы. Наблюдаемое отклонение усредненных формантных наблюдений устойчиво и обладает высокой степенью идентификационной значимости. Для обнаружения акцента эффективнее использовать Reference2.

Идентифицирующим признаком исследуемого акцентного [O] в речи билингвов-мужчин являются, во-первых, высокая вариативность значений второй форманты, F_2 ; во-вторых, выраженное отклонение общего среднего по F_2 вверх относительно обоих эталонов. Отклонение по F_1 может служить лишь в качестве второстепенного индикатора принадлежности данному акценту. При проведении лингвистической экспертизы более оптимально использование значения Reference1 в качестве эталона гласного [O]. В речи билингвов-женщин идентифицирующим признаком кабардино-черкесского акцента является локализация наблюдений в IV квадранте относительно эталона Reference2 в пространстве (F_1 - F_2), то есть правее по F_1 и ниже по F_2 .

Сопоставление вокалических систем на основе формантных характеристик позволяет объективно оценить степень и качество интерферирующего взаимовлияния языковых систем, выявить систематические ошибки реальной речевой деятельности, снизить неточность категоризации на основе довольно условных фонетических признаков гласных. Объективный анализ с использованием акустических признаков позволяет эмпирически обосновать и дополнить перцептивные теории усвоения языка, избавиться от экспертной субъективности в оценках акцентной речи, что принципиально важно при проведении лингвистической экспертизы.

ПРИЛОЖЕНИЕ

Таблица 1. Канонические, гендерно-специфические эталонные (Reference1, Reference2 соответственно) и измеренные формантные значения нормативного и ненормативного гласного [A] русской речи билингов-мужчин *, Гц

Table 1. Canonical, gender-specific reference (Reference1, Reference2, respectively) and measured formant values of the normative and non-normative vowel [A] in the Russian speech of male bilinguals *, Hz

Показатель	F ₁	F ₂
Индивидуальные измерения		
km1 медиана	592,09	1263,31
km1 среднее	592,14	1288,13
km1 SD	40,39	108,22
km2 медиана	648,94	1437,00
km2 среднее	647,64	1439,60
km2 SD	35,98	110,92
km3 медиана	649,40	1277,97
km3 среднее	644,34	1264,44
km3 SD	54,81	120,02
km4 медиана	577,44	1342,02
km4 среднее	599,43	1349,82
km4 SD	47,63	111,06
km5 медиана	600,39	1339,80
km5 среднее	600,13	1328,11
km5 SD	26,98	120,65
Общие измерения		
Медиана	610,50	1330,51
Общее среднее	616,73	1334,02
SD	48,01	127,64
Махнаблюдение	736,04	1731,22
Миннаблюдение	503,31	1024,73
Reference1	600	1200
Reference2	500	1180

Примечание. Количество наблюдений – 109. * – L1 – кабардино-черкесский язык; L2 – русский язык.
 Note. Number of observations – 109. * – L1 – Kabardian-Circassian language; L2 – Russian language.

Таблица 2. Канонические, гендерно специфические эталонные (Reference1, Reference2 соответственно) и измеренные формантные значения нормативного и ненормативного гласного [А] русской речи билингвов-женщин *, Гц

Table 2. Canonical, gender-specific reference (Reference1, Reference2, respectively) and measured formant values of the normative and non-normative vowel [A] in the Russian speech of female bilinguals *, Hz

Показатель	F ₁	F ₂
Индивидуальные измерения		
kf1 медиана	647,13	1437,52
kf1 среднее	653,67	1448,23
kf1 SD	113,38	143,56
kf2 медиана	794,39	1417,34
kf2 среднее	794,99	1430,12
kf2 SD	40,84	110,30
kf3 медиана	746,01	1548,16
kf3 среднее	753,88	1549,88
kf3 SD	61,08	78,65
kf4 медиана	704,89	1609,09
kf4 среднее	715,09	1617,09
kf4 SD	91,24	91,44
kf5 медиана	748,30	1401,59
kf5 среднее	748,51	1410,94
kf5 SD	37,23	108,13
Общие измерения		
Медиана	742,56	1496,95
Общее среднее	733,23	1491,25
SD	87,38	133,00
Махнаблюдение	943,40	1815,95
Миннаблюдение	415,53	1210,78
Reference1	600	1200
Reference2	540	1140

Примечание. Количество наблюдений – 110. * – L1 – кабардино-черкесский язык; L2 – русский язык.
 Note. Number of observations – 110. * – L1 – Kabardian-Circassian language; L2 – Russian language.

Таблица 3. Канонические, гендерно специфические эталонные (Reference1, Reference2 соответственно) и измеренные формантные значения нормативного и ненормативного гласного [O] русской речи билингвов-мужчин *, Гц

Table 3. Canonical, gender-specific reference (Reference1, Reference2, respectively) and measured formant values of the normative and non-normative vowel [O] in the Russian speech of male bilinguals *, Hz

Показатель	F ₁	F ₂
Индивидуальные измерения		
km1 медиана	483,06	1370,20
km1 среднее	484,64	1351,09
km1 SD	43,43	261,31
km2 медиана	430,04	994,48
km2 среднее	437,41	1123,38
km2 SD	52,88	335,1147
km3 медиана	455,14	992,28
km3 среднее	485,20	1074,10
km3 SD	72,13	480,85
km4 медиана	491,08	1380,55
km4 среднее	507,12	1411,58
km4 SD	56,90	358,02
km5 медиана	465,38	986,35
km5 среднее	451,00	1087,44
km5 SD	71,25	371,52
Общие измерения		
Медиана	466,77	1142,23
Среднее	473,08	1209,52
SD	64,44	388,74
Махнаблюдение	681,71	2270,77
Миннаблюдение	353,93	646,37
Reference1	450	800
Reference2	380	1090

Примечание. Количество наблюдений – 109. * – L1 – кабардино-черкесский язык; L2 – русский язык.
Note. Number of observations – 109. * – L1 – Kabardian-Circassian language; L2 – Russian language.

Таблица 4. Канонические, гендерно специфические эталонные (Reference1, Reference2 соответственно) и измеренные формантные значения нормативного и ненормативного гласного [O] русской речи билингвов-женщин *, Гц

Table 4. Canonical, gender-specific reference (Reference1, Reference2, respectively) and measured formant values of the normative and non-normative vowel [O] in the Russian speech of females-bilinguals *, Hz

Показатель	F ₁	F ₂
Индивидуальные измерения		
kf1 медиана	457,48	1042,98
kf1 среднее	448,02	1076,79
kf1 SD	49,34	177,06
kf2 медиана	494,91	886,38
kf2 среднее	513,04	946,31
kf2 SD	76,86	194,05
kf3 медиана	505,04	848,77
kf3 среднее	502,43	888,53
kf3 SD	49,06	103,88
kf4 медиана	482,18	959,47
kf4 среднее	487,64	1030,23
kf4 SD	51,67	307,14
kf5 медиана	472,88	864,37
kf5 среднее	487,69	898,63
kf5 SD	63,22	125,32
Общие измерения		
Медиана	487,42	907,36
Среднее	487,76	968,10
SD	62,02	205,15
Махнаблюдение	766,24	2105,79
Миннаблюдение	305,72	755,19
Reference1	450	800
Reference2	430	1200

Примечание. Количество наблюдений – 110. * – L1 – кабардино-черкесский язык; L2 – русский язык.
 Note. Number of observations – 110. * – L1 – Kabardian-Circassian language; L2 – Russian language.

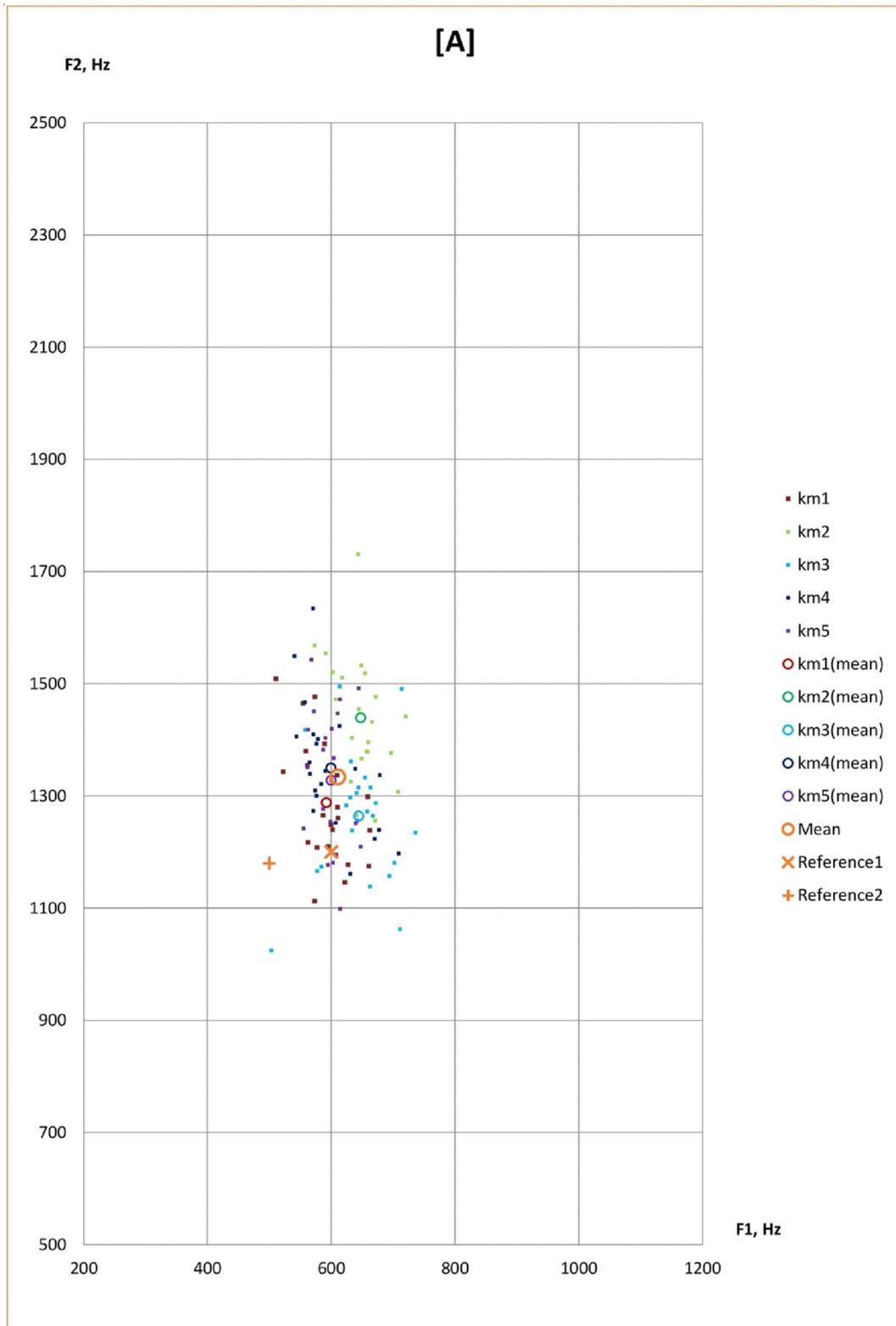
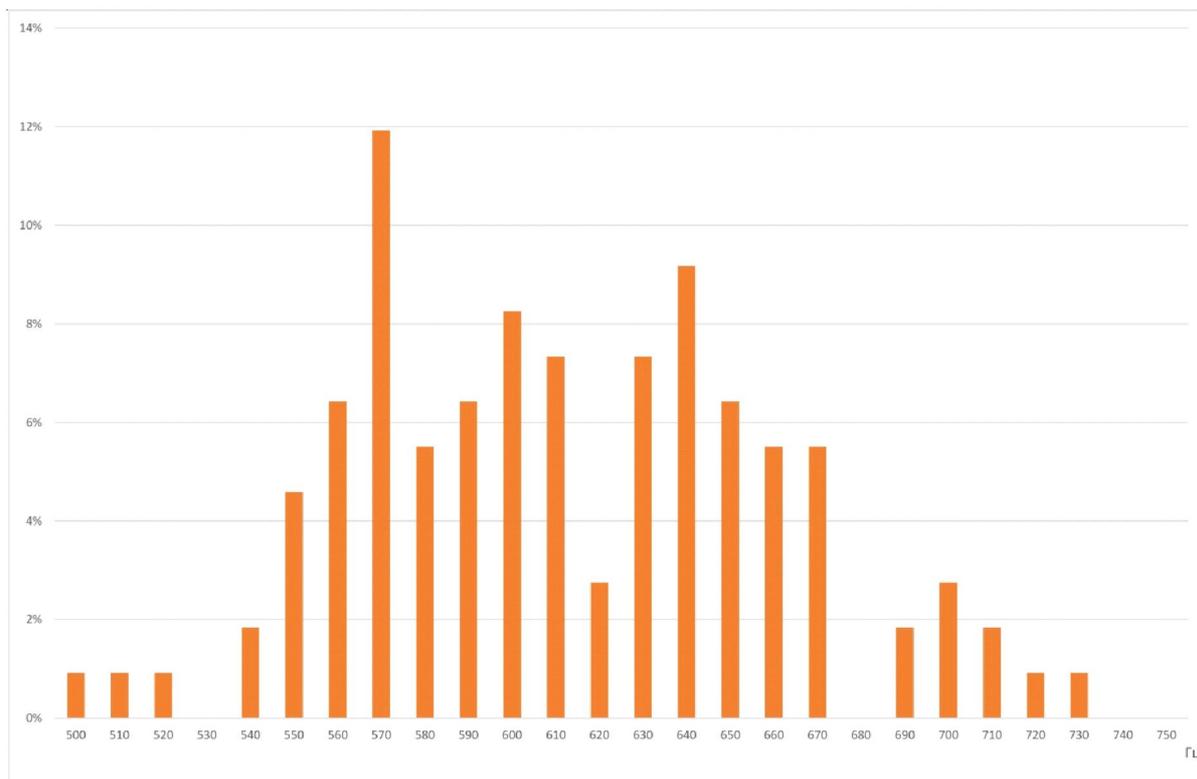
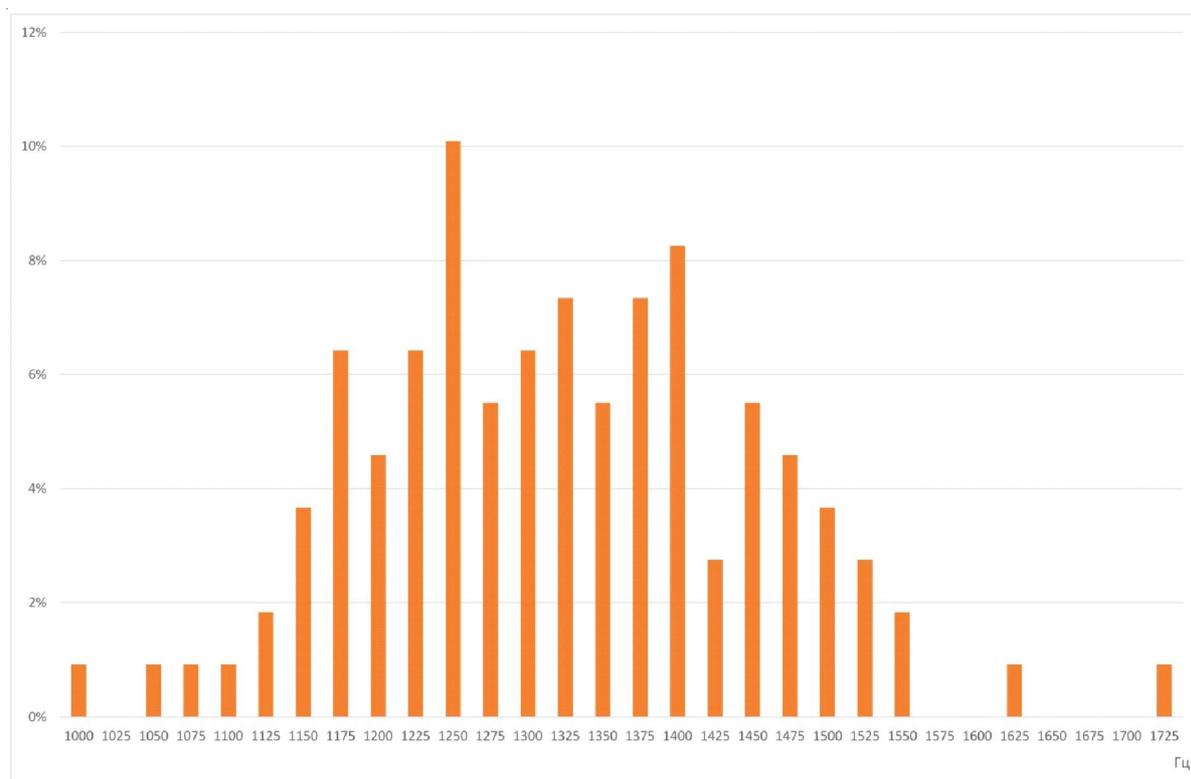


Рис. 1. Формантные значения аллофонов гласного [А] в речи мужчин, владеющих кабардино-черкесским (L1) и русским (L2) языками

Fig. 1. Formants values of the vowel [A] in the male bilinguals' speech (L1 – Kabardian-Circassian language, L2 – Russian language)



а



б

Рис. 2. Распределение значений первой форманты (а) и второй форманты (б) для акцентного [А] речи дикторов-мужчин (L1 – кабардино-черкесский язык, L2 – русский язык)

Fig. 2. Distribution for values of the first (a) and second (б) formants for the accented [A] in the speech of male speakers (L1 – Kabardian-Circassian language, L2 – Russian language)

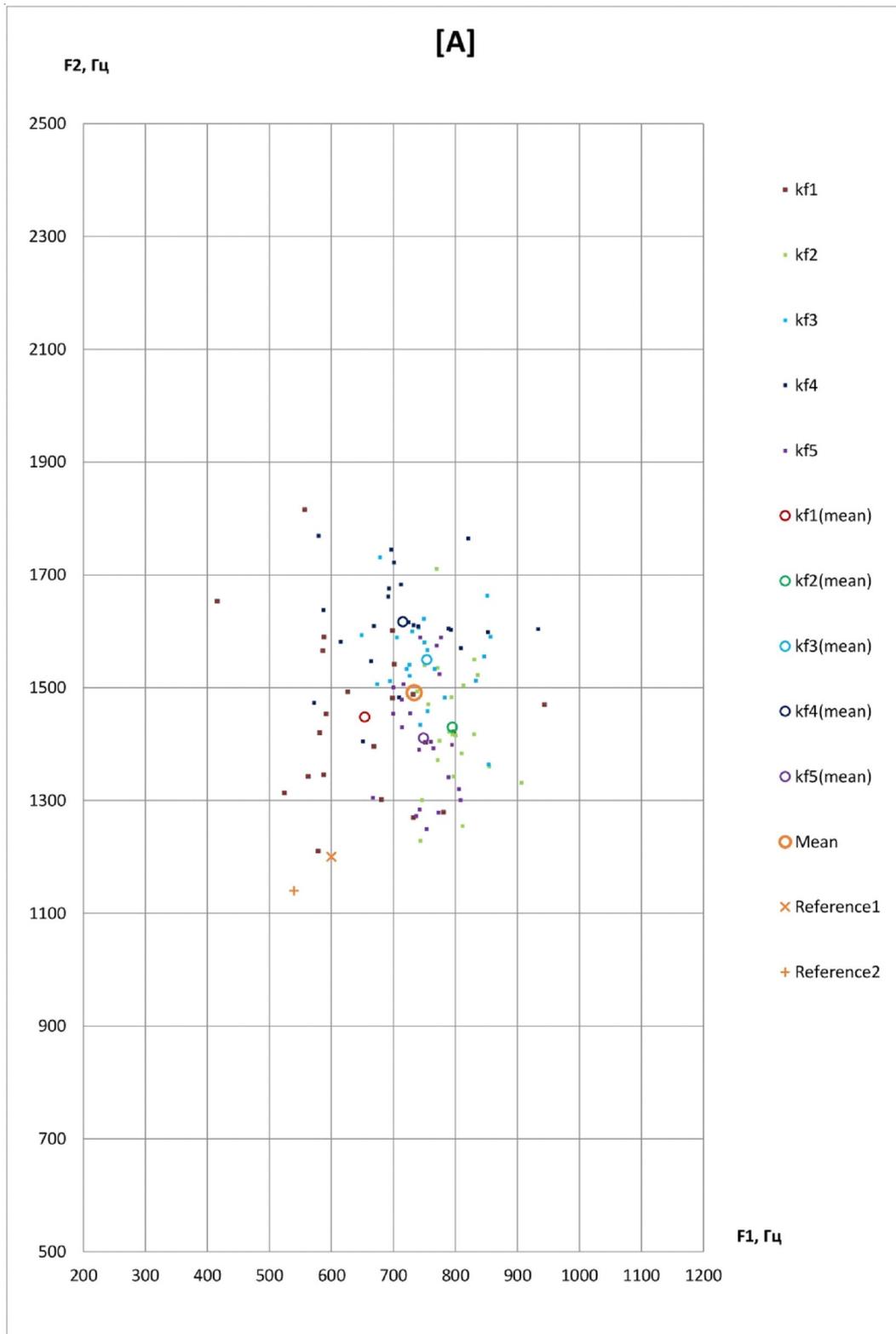
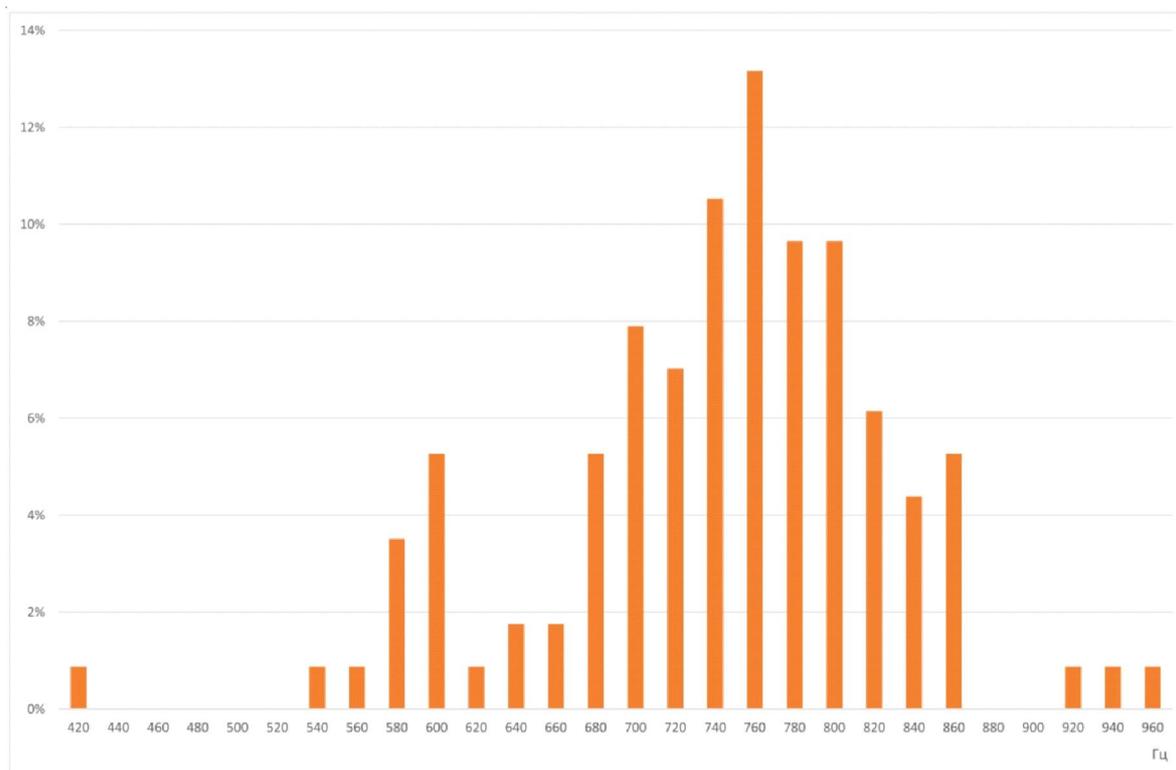
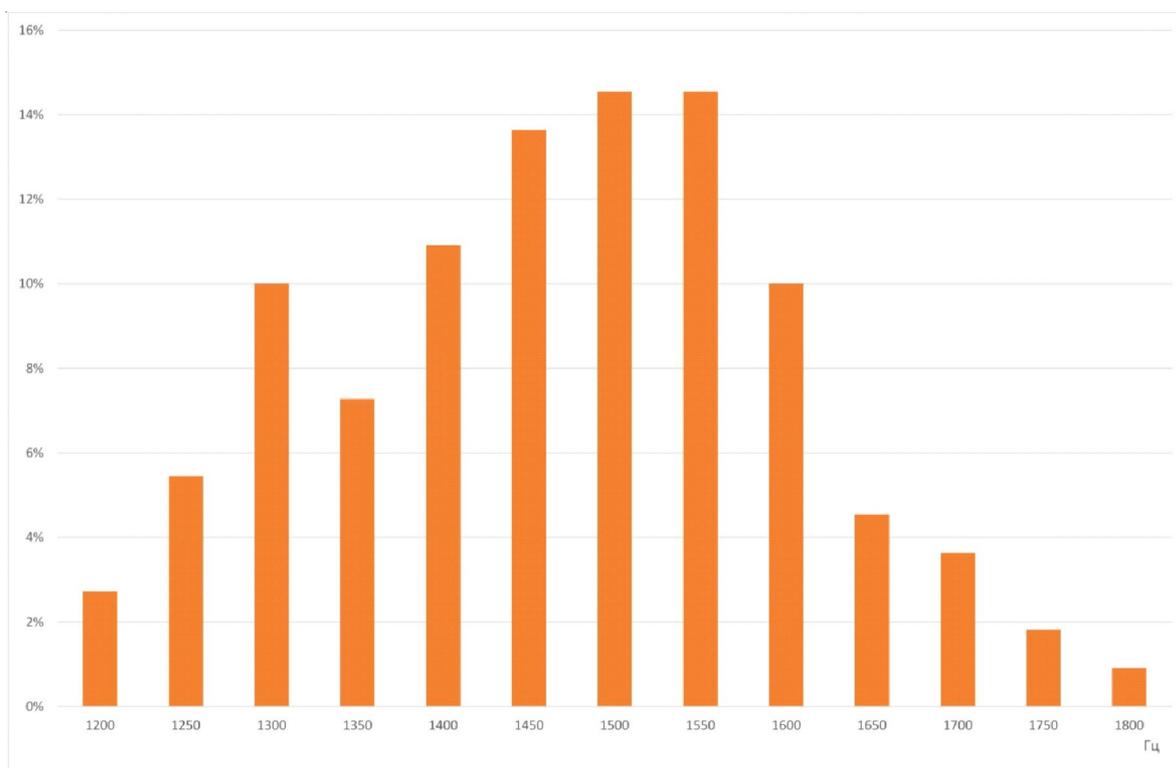


Рис. 3. Формантные значения аллофонов гласного [А] в речи женщин, владеющих кабардино-черкесским (L1) и русским (L2) языками

Fig. 3. Formants values of the vowel [A] in the females-bilinguals speech (L1 – Kabardian-Circassian language, L2 – Russian language)



a



б

Рис. 4. Распределение значений первой форманты (а) и второй форманты для акцентного (б) [А] в речи дикторов-женщин (L1 – кабардино-черкесский язык, L2 – русский язык)

Fig. 4. Distribution for values of the first (a) and second (б) formants for the accented [A] in the speech of female speakers (L1 – Kabardian-Circassian language, L2 – Russian language)

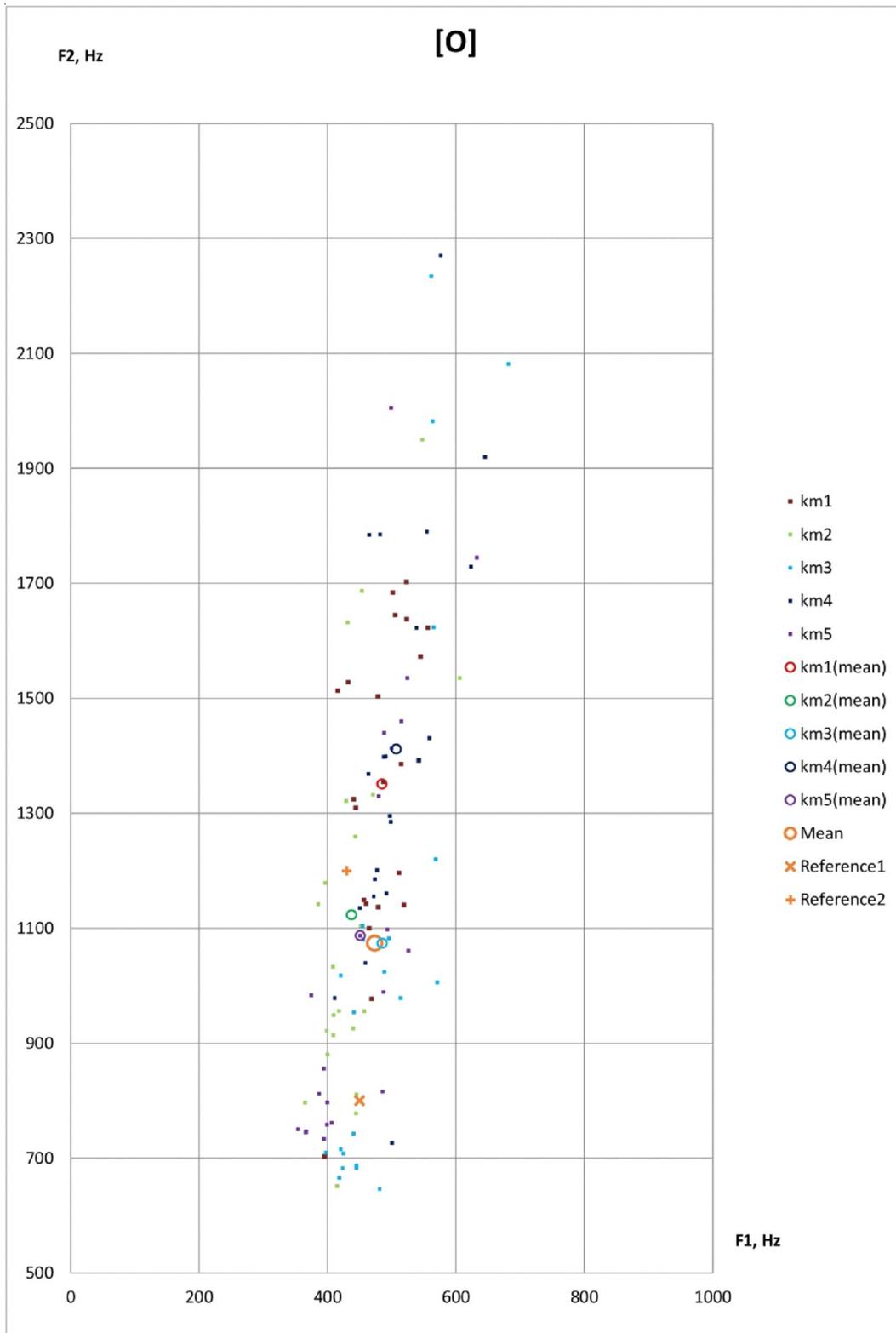
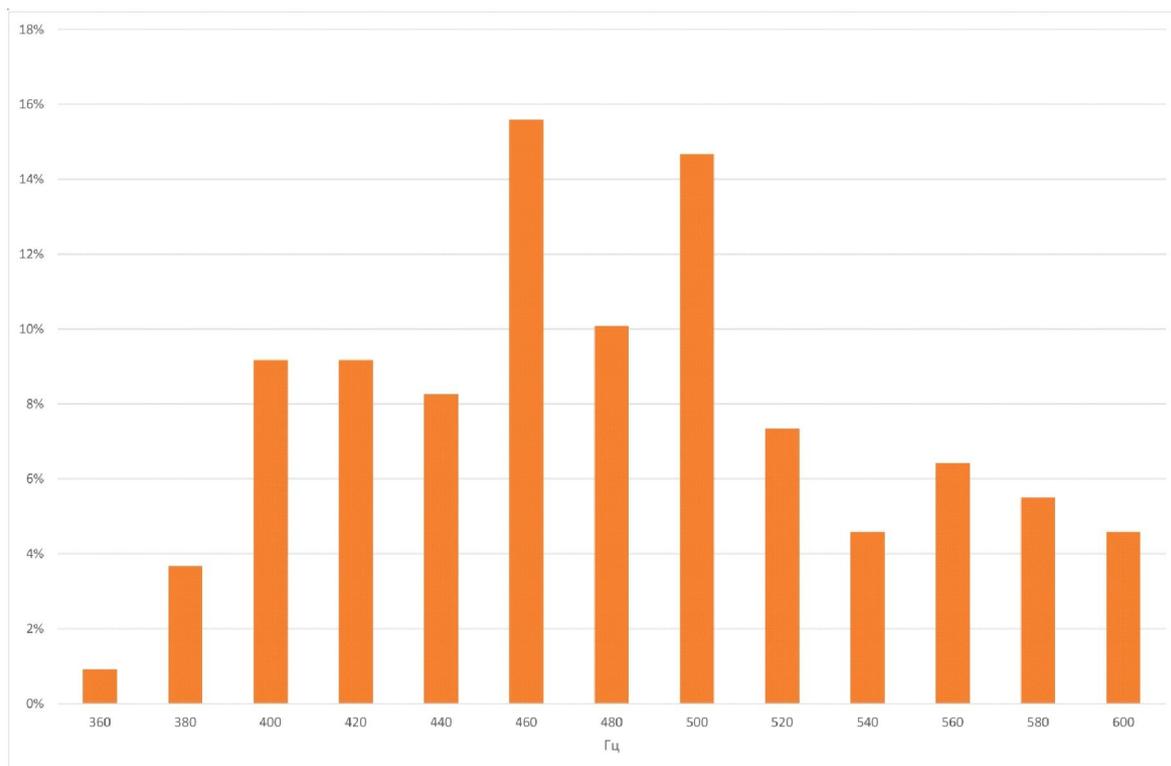
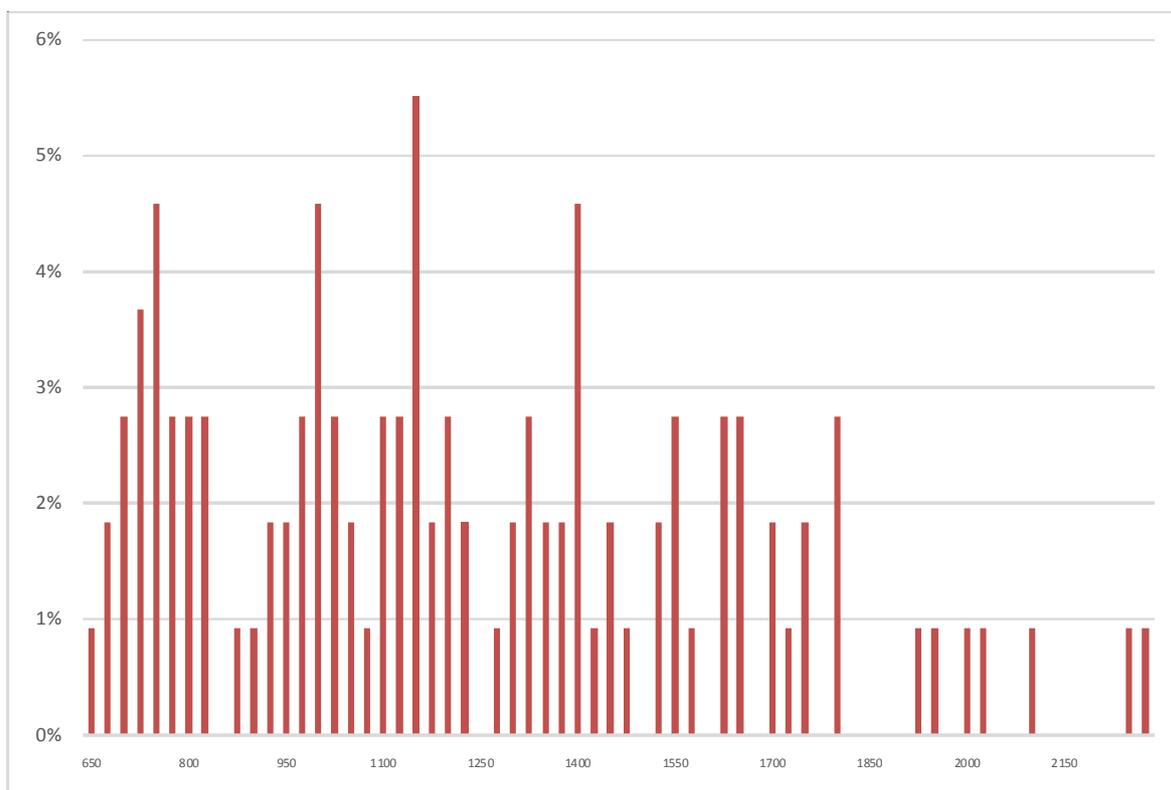


Рис. 5. Формантные значения аллофонов гласного [O] в речи мужчин, владеющих кабардино-черкесским (L1) и русским (L2) языками.

Fig. 5. Formants values of the vowel [O] in the male bilinguals' speech (L1 – Kabardian-Circassian language, L2 – Russian language)



a



б

Рис. 6. Распределение значений первой форманты (а) и второй форманты (б) для акцентного [O] в речи дикторов-мужчин (L1 – кабардино-черкесский язык, L2 – русский язык)

Fig. 6. Distribution for values of the first (a) and second (б) formants for the accented [O] in the speech of male speakers (L1 – Kabardian-Circassian language, L2 – Russian language)

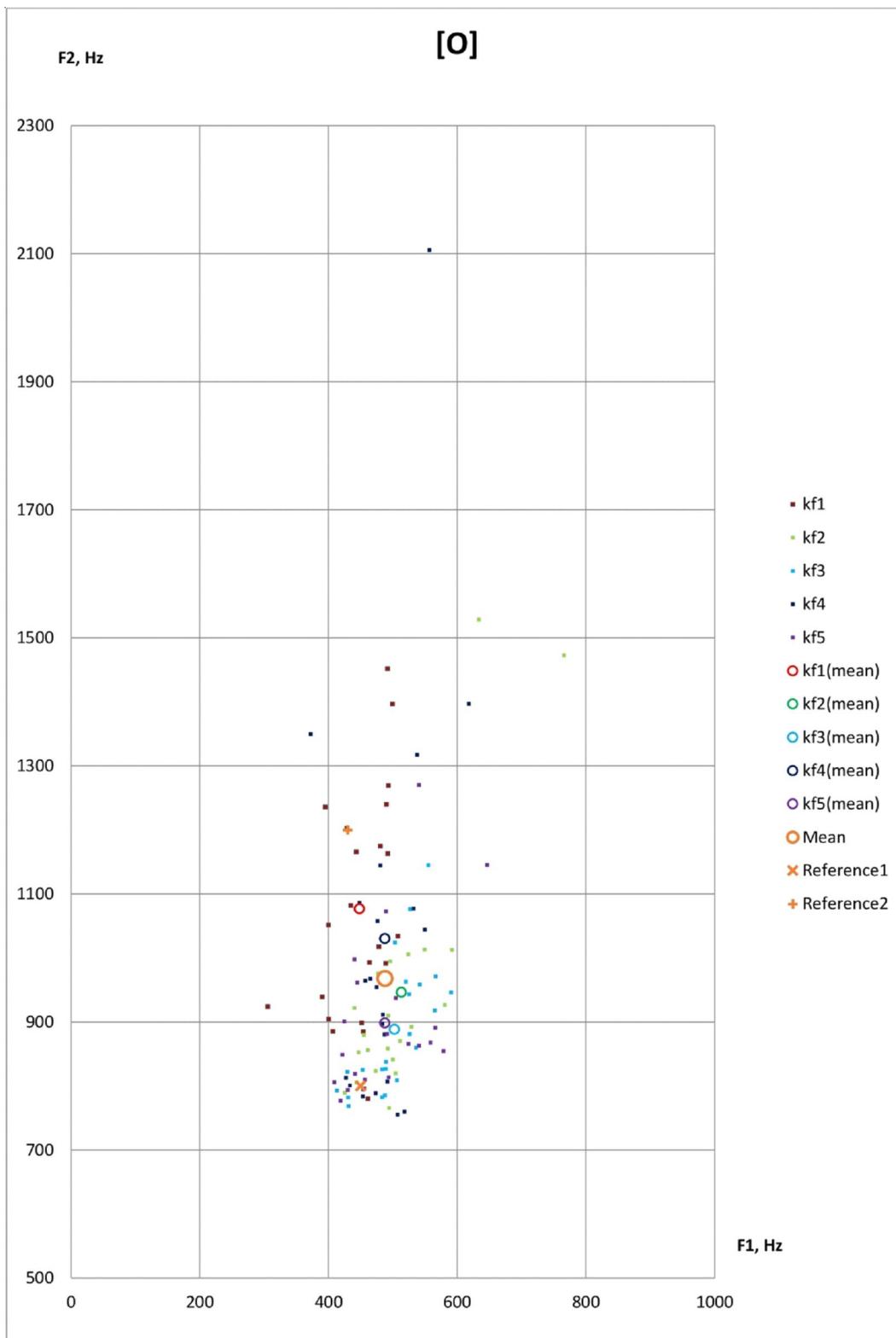
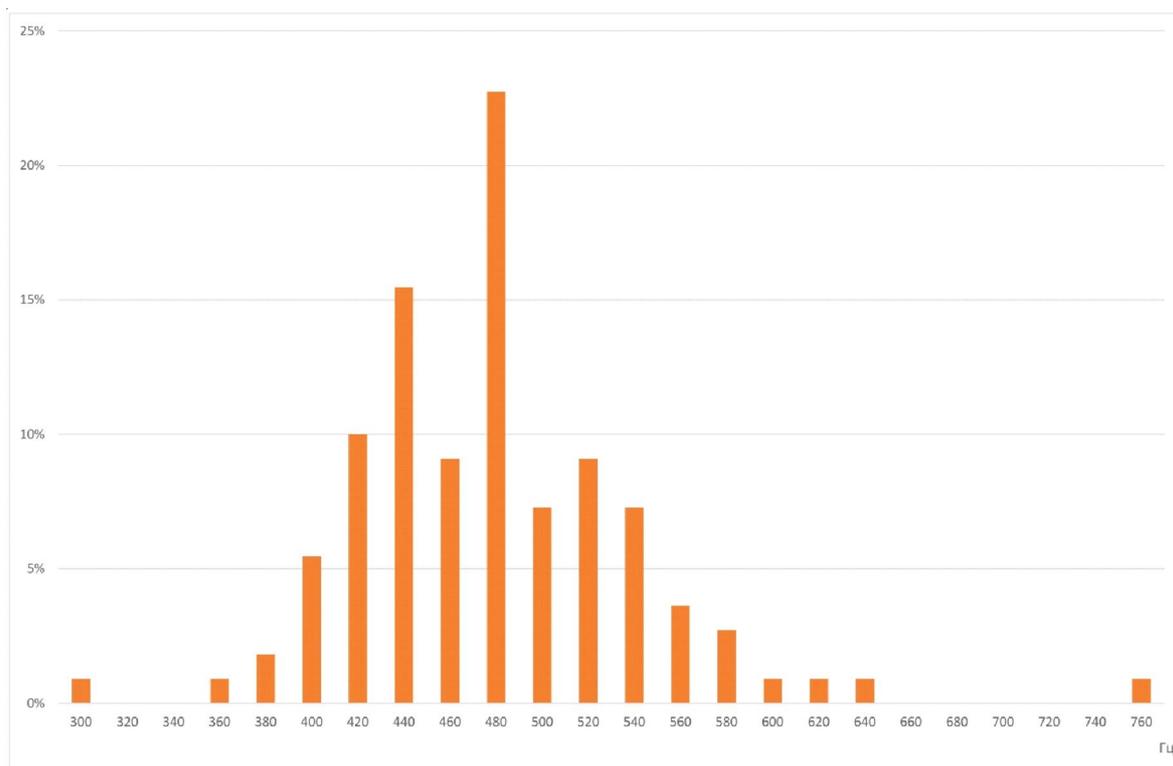
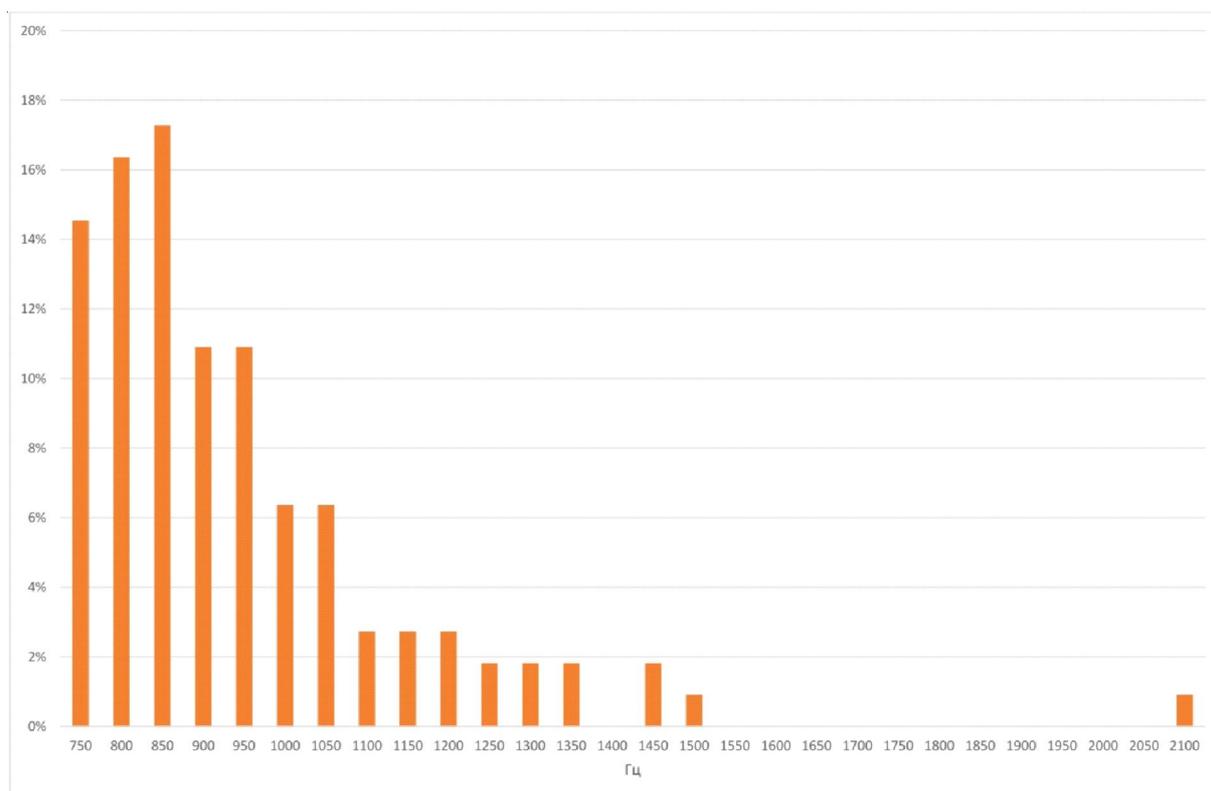


Рис. 7. Формантные значения аллофонов гласного [O] в речи женщин, владеющих кабардино-черкесским (L1) и русским (L2) языками

Fig. 7. Formants values of the vowel [O] in the females-bilinguals speech (L1 – Kabardian-Circassian language, L2 – Russian language)



a



б

Рис. 8. Распределение значений первой форманты (а) и второй форманты (б) для акцентного [O] в речи дикторов-женщин (L1 – кабардино-черкесский язык, L2 – русский язык)

Fig. 8. Distribution for values of the first (a) and second (б) formants for the accented [O] in the speech of female speakers (L1 – Kabardian-Circassian language, L2 – Russian language)

СПИСОК ЛИТЕРАТУРЫ

- Гуртуева И. А., 2020. Современные проблемы автоматического распознавания речи // Известия Кабардино-Балкарского научного центра РАН. № 6(98). С. 20–32. DOI: 10.35330/1991-6639-2020-6-98-20-33
- Козлачков С. Б., Дворянkin С. В., Бонч-Бруевич А. М., 2016. Ограничения формантной теории разборчивости речи в приложениях защиты речевой информации // Вопросы кибербезопасности. № 5 (18). С. 28–35.
- Леонов А. С., Макаров И. С., Сорокин В. Н., 2009. Частотные модуляции в речевом сигнале // Акустический журнал. Т. 55, № 6. С. 809–821.
- Программа для анализа и визуализации спектрального состава аудиосообщений : свидетельство о гос. регистрации программы для ЭВМ № 2020618637 от 22.07.2020 / Бжихатлов К. Ч., Гуртуева И. А.
- Сорокин В. Н., Цыплихин А. И., 2004. Сегментация и распознавание гласных // Информационные процессы. Т. 4, № 2. С. 202–220.
- Фонетико-акустический корпус акцентной русской речи : свидетельство о гос. регистрации базы данных № 2023620331 от 23.01.2023 : заявка № 2022623155 / Гуртуева И. А.
- Amodei D., Anubhai R., Battenberg E., Case C., Casper J., Catanzaro B. et al., 2016. Deep Speech 2: End-to-End Speech Recognition in English and Mandarin // ICML'16 : Proceedings of the 33rd International Conference on International Conference on Machine Learning. Vol. 48. P. 173–182. DOI: 10.48550/arXiv.1512.02595
- Chen S. F., Kingsbury B., Mangu L., Povey D., Saon G., Soltau H., Zweig G., 2006. Advances in Speech Transcription at IBM Under the DARPA EARS program // IEEE Trans. Audio, Speech, and Language Processing. Vol. 14. P. 1596–1608.
- Das N., Zegers J., Van Hamme H., Francart T., Bertrand A., 2020. Linear Versus Deep Learning Methods for Noisy Speech Separation for EEG-Informed Attention Decoding // Journal of Neural Engineering. Vol. 17, № 4. DOI: 10.1088/1741-2552/aba6f8
- Deng K., Cao S., Ma L., 2021. Improving Accent Identification and Accented Speech Recognition Under a Framework of Self-Supervised Learning // Proc. Interspeech. P. 1504–1508. DOI: 10.48550/arXiv.2109.07349
- Evermann G., Chan H. Y., Gales M. J. F., Hain T., Liu X., Mrva D., Wang L., Woodland P. C., 2004. Development of the 2003 CU-HTK Conversational Telephone Speech Transcription System // Proc. IEEE ICASSP. Vol. 1. Monreal : [s.n.]. P. 1–249.
- Boersma P., Weenink D., 2003. Praat 6.3.08: Computer Soft Package for Speech Analysis in Phonetics. URL: <https://www.fon.hum.uva.nl/praat/>
- Han C., O'Sullivan J., Luo Y., Herrero J., Mehta A. D., Mesgarani N., 2019. Speaker-Independent Auditory Attention Decoding Without Access to Clean Speech Sources // Sci Adv. Vol. 15, iss. 5. DOI: 10.1126/sciadv.aav6134
- Kristjansson T. T., Hershey J. R., Olsen P. A., Rennie S. J., Gopinath R. A., 2006. Super-Human Multi-Talker Speech Recognition: The IBM 2006 Speech Separation Challenge System // Proc. Interspeech. Vol. 12. Pittsburg : [s.n.]. P. 155.
- Ladefoged P., Johnson K. A., 2014. Course in Phonetics. Boston : [s. n.]. 352 p.
- Lippmann R. P., 1997. Speech Recognition by Machines and Humans // Speech Communication. Vol. 22, № 1. P. 1–15.
- Lobanov B., Tsurulnik L., 2006. Development of Multi-Voice and Multi-Language TTS Synthesizer (Languages: Belarussian. Polish. Russian) // SPECOM'2006. P. 274–283.
- Matsoukas S., Gauvain J.-L., Adda G., Colthurst T., Kao C.-L., Kimball O., Lamel L., Lefevre F., Ma J. Z., Makhoul J. et al., 2006. Advances in Transcription of Broadcast News and Conversational Telephone Speech Within the Combined Ears bbn/limsi System // IEEE Transactions on Audio, Speech, and Language Processing. Vol. 14. P. 1541–1556.
- Nagoev Z., Gurtueva I., Bzhikhatlov K., Anchekov M., 2022. Phonetic-Acoustic Database of High-Accent Russian Speech // Procedia Computer Science. Vol. 213. P. 518–522. DOI: 10.1016/j.procs.2022.11.099
- Najafian M., Russell M., 2020. Automatic Accent Identification as an Analytical Tool for Accent Robust Automatic Speech Recognition // Speech Communication. Vol. 122. P. 44–55. DOI: 10.1016/j.specom.2020.05.003
- Pallett D. S., 2003. A Look at NIST's Benchmark ASR Tests: Past, Present, and Future // IEEE Automatic Speech Recognition and Understanding Workshop. P. 483–488. DOI: 10.1109/ASRU.2003.1318488
- Seide F., Li G., Yu D., 2011. Conversational Speech Transcription Using Context-Dependent Deep Neural Networks // Proceedings Interspeech. Florence. P. 437–440. DOI: 10.21437/Interspeech.2011-169
- SpecApp : свидетельство о гос. регистрации программы для ЭВМ №2021615667 от 21.04.2021 / Гуртуева И. А., Бжихатлов К. Ч.
- Stolcke A., Chen B., Franco H., Gadde V. R. R., Graciarena M., Hwang M.-Y., Kirchhoff K., Mandal A., Morgan N., Lei X. et al., 2006. Recent Innovations in Speech-to-Text Transcription at SRI-ICSI-UW // IEEE Transactions on Audio, Speech, and Language Processing. Vol. 14. P. 1729–1744. DOI: 10.1109/TASL.2006.879807

- Straetmans L., Holtze B., Debener S., Jaeger M., Mirkovic B., 2022. Neural Tracking to Go: Auditory Attention Decoding and Saliency Detection with Mobile EEG // *J. Neural Eng.* Vol. 18 (6). DOI: 10.1088/1741-2552/ac42b5
- Weinreich U., 1979. *Languages in Contact: Findings and Problems.* The Hague : Mouton Publishers. 148 p.
- Weng C., Yu D., Seltzer M. L., Droppo J., 2014. Single-Channel Mixed Speech Recognition Using Deep Neural Networks // *Proc. IEEE ICASSP.* Florence. P. 5632–5636. DOI: 10.1109/ICASSP.2014.6854681

REFERENCES

- Gurtueva I.A., 2020. Sovremennye problemy avtomaticheskogo raspoznavaniya rechi [State-of-the-Art of Automatic Speech Recognition]. *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [Bulletin of Kabardino-Balkarian Scientific Center RAS], no. 6 (98), pp. 20-32. DOI: 10.35330/1991-6639-2020-6-98-20-33
- Kozlachkov S.B., Dvoryankin S.V., Bonch-Bruevich A.M., 2016. Ogranicheniya formantnoy teorii razborchivosti rechi v prilozheniyakh zashchity rechevoy informatsii [Limitations of Formant Theory of Speech Intelligibility in Speech Security Applications]. *Voprosy kiberbezopasnosti* [Cybersecurity Issues], no. 5 (18), pp. 28-35.
- Leonov A.S., Makarov I.S., Sorokin V.N., 2009. Chastotnye modulyatsii v rechevom signale [Frequency Modulations in a Speech Signal]. *Akusticheskiy zhurnal* [Acoustic Magazine], vol. 55, no. 6, pp. 809-821.
- Bzhikhatlov K.Ch., Gurtueva I.A., 2020 *Programma dlya analiza i vizualizatsii spektralnogo sostava audioobshcheniy: svidetelstvo o gos. registratsii programmy dlya EVM № 2020618637 ot 22.07.2020* [Program for Analysis and Visualization of Spectral Composition of Audio Messages. Certificate of State Registration of Computer Program No. 2020618637 Dated July 22, 2020].
- Sorokin V.N., Tsyplikhin A.I., 2004. Segmentatsiya i raspoznavanie glasnykh [Vowel Segmentation and Recognition]. *Informatsionnye protsessy* [Information Processes], vol. 4, no. 2, pp. 202-220.
- Gurtueva I.A., 2023. *Fonetiko-akusticheskiy korpus aktsentnoy russkoy rechi: svidetelstvo o gos. registratsii bazy dannykh № 2023620331 ot 23.01.2023: zayavka № 2022623155* [Phonetic-Acoustic Corpus of Accented Russian Speech. Certificate of State Registration of Databases No. 2023620331, January 23, 2023. Application No. 2022623155].
- Amodei D., Anubhai R., Battenberg E., Case C., Casper J., Catanzaro B. et al., 2016. Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. *ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning*, vol. 48, pp. 173-182. DOI: 10.48550/arXiv.1512.02595
- Chen S.F., Kingsbury B., Mangu L., Povey D., Saon G., Soltau H., Zweig G., 2006. Advances in Speech Transcription at IBM Under the DARPA EARS Program. *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, pp. 1596-1608.
- Das N., Zegers J., Van Hamme H., Francart T., Bertrand A., 2020. Linear Versus Deep Learning Methods for Noisy Speech Separation for EEG-Informed Attention Decoding. *Journal of Neural Engineering*, vol. 17, no. 4. DOI: 10.1088/1741-2552/aba6f8
- Deng K., Cao S., Ma L., 2021. Improving Accent Identification and Accented Speech Recognition Under a Framework of Self-Supervised Learning. *Proc. Interspeech*, pp. 1504-1508. DOI: 10.48550/arXiv.2109.07349
- Evermann G., Chan H.Y., Gales M.J.F., Hain T., Liu X., Mrva D., Wang L., Woodland P.C., 2004. Development of the 2003 CU-HTK Conversational Telephone Speech Transcription System. *Proc. IEEE ICASSP. Vol. 1.* Montreal, s. n., pp. 1-249.
- Boersma P., Weenink D., 2023. *Praat 6.3.08: Computer Soft Package for Speech Analysis in Phonetics.* URL: <https://www.fon.hum.uva.nl/praat/>
- Han C., O'Sullivan J., Luo Y., Herrero J., Mehta A.D., Mesgarani N., 2019. Speaker-Independent Auditory Attention Decoding Without Access to Clean Speech Sources. *Sci Adv*, vol. 15, iss. 5. DOI: 10.1126/sciadv.aav6134
- Kristjansson T.T., Hershey J.R., Olsen P.A., Rennie S.J., Gopinath R.A., 2006. Super-Human Multi-Talker Speech Recognition: The IBM 2006 Speech Separation Challenge System. *Proc. Interspeech. Vol. 12.* Pittsburg, s. n., p. 155.
- Ladefoged P., Johnson K.A., 2014. *Course in Phonetics.* Boston, s.n. 352 p.
- Lippmann R.P., 1997. Speech Recognition by Machines and Humans. *Speech Communication*, vol. 22, no. 1, pp. 1-15.
- Lobanov B., Tsurulnik L., 2006. Development of Multi-Voice and Multi-Language TTS Synthesizer (Languages: Belarussian. Polish. Russian). *SPECOM'2006*, pp. 274-283.
- Matsoukas S., Gauvain J.-L., Adda G., Colthurst T., Kao C.-L., Kimball O., Lamel L., Lefevre F., Ma J.Z., Makhoul J. et al., 2006. Advances in Transcription of Broadcast News and Conversational Telephone Speech Within the Combined Ears bbn/limsi System. *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 1541-1556.

- Nagoev Z., Gurtueva I., Bzhikhatlov K., Anchekov M., 2022. Phonetic-Acoustic Database of High-Accent Russian Speech. *Procedia Computer Science*, vol. 213, pp. 518-522. DOI: 10.1016/j.procs.2022.11.099
- Najafian M., Russell M., 2020. Automatic Accent Identification as an Analytical Tool for Accent Robust Automatic Speech Recognition. *Speech Communication*, vol. 122, pp. 44-55. DOI: 10.1016/j.specom.2020.05.003
- Pallett D.S., 2003. A Look at NIST's Benchmark ASR Tests: Past, Present, and Future. *IEEE Automatic Speech Recognition and Understanding Workshop*, pp. 483-488. DOI: 10.1109/ASRU.2003.1318488
- Seide F., Li G., Yu D., 2011. Conversational Speech Transcription Using Context-Dependent Deep Neural Networks. *Proceedings Interspeech*. Florence, pp. 437-440. DOI: 10.21437/Interspeech.2011-169
- Gurtueva I.A., Bzhikhatlov K.Ch., 2021. *SpecApp: svidetelstvo o gos. registratsii programmy dlya EVM №2021615667 ot 21.04.2021* [SpecApp. Certificate of State Registration of Computer Programs No. 2021615667 Dated April 21, 2021].
- Stolcke A., Chen B., Franco H., Gadde V.R.R., Graciarena M., Hwang M.-Y., Kirchhoff K., Mandal A., Morgan N., Lei X. et al., 2006. Recent Innovations in Speech-To-Text Transcription at SRI-ICSI-UW. *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 1729-1744. DOI: 10.1109/TASL.2006.879807
- Straetmans L., Holtze B., Debener S., Jaeger M., Mirkovic B., 2022. Neural Tracking to Go: Auditory Attention Decoding and Saliency Detection with Mobile EEG. *J. Neural Eng.*, vol. 18 (6). DOI: 10.1088/1741-2552/ac42b5
- Weinreich U., 1979. *Languages in Contact: Findings and Problems*. The Hague, Mouton Publishers. 148 p.
- Weng C., Yu D., Seltzer M.L., Droppo J., 2014. Single-Channel Mixed Speech Recognition Using Deep Neural Networks. *Proc. IEEE ICASSP*. Florence, pp. 5632-5636. DOI: 10.1109/ICASSP.2014.6854681

Information About the Authors

Irina A. Gurtueva, Researcher, Laboratory of Computational Linguistics, Institute of Computer Science and Problems of Regional Management, Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences, I. Armand St, 37a, 360000 Nalchik, Russia, gurtueva-i@yandex.ru, <https://orcid.org/0000-0003-4945-5682>

Michael V. Kamensky, Doctor of Sciences (Philology), Associate Professor, Head of the Department of Linguistics, Linguodidactics and Intercultural Communication, North-Caucasus Federal University, Pushkina St, 1, Bld. 20, 355009 Stavropol, Russia, mkamenskii@ncfu.ru, <https://orcid.org/0000-0001-8358-9516>

Информация об авторах

Ирина Асланбековна Гуртуева, научный сотрудник лаборатории компьютерной лингвистики, Институт информатики и проблем регионального управления, Кабардино-Балкарский научный центр РАН, ул. И. Арманд, 37а, 360000 г. Нальчик, Россия, gurtueva-i@yandex.ru, <https://orcid.org/0000-0003-4945-5682>

Михаил Васильевич Каменский, доктор филологических наук, доцент, заведующий кафедрой лингвистики, лингводидактики и межкультурной коммуникации, Северо-Кавказский федеральный университет, ул. Пушкина, 1, корп. 20, 355009 г. Ставрополь, Россия, mkamenskii@ncfu.ru, <https://orcid.org/0000-0001-8358-9516>