ИЗДАТЕЛЬСТВО грамота

Филологические науки. Вопросы теории и практики
Philology. Theory & Practice

ISSN 2782-4543 (online)
ISSN 1997-2911 (print)

2024. Том 17. Выпуск 6 | 2024. Volume 17. Issue 6
Материалы журнала доступны на сайте (articles and issues available at): philology-journal.ru

RU

Анализ эффективности ML-алгоритмов распознавания эмоций с учетом просодических и спектральных признаков

Заврумов З. А., Гончарова О. В., Левит А. А.

Аннотация. Цель исследования – определить оптимальный классификатор для идентификации эмоционального состояния на основании результатов сравнительного анализа эффективности различных алгоритмов машинного обучения, основанных на комбинации просодических и спектральных признаков. Научная новизна состоит в применении ML-алгоритмов в распознавании эмоционально-маркированной речи северокавказских билингвов в задаче бинарной классификации наличия или отсутствия акцента с определением оптимальной комбинации универсальных просодических и спектральных признаков. В ходе исследования создан экспериментальный корпус речи представителей трех этногрупп (русских, кабардинцев и армян) с аннотацией степени выраженности акцента, извлечены просодические (94 признака) и спектральные (74 признака) характеристики из речевых сигналов, проведен сравнительный анализ эффективности алгоритмов машинного обучения (логистическая регрессия, k-ближайших соседей, метод опорных векторов, деревья решений) в задаче бинарной классификации наличия/отсутствия акцента. Результаты исследования показали, что на слоговом уровне наиболее эффективной является модель дерева решений с комбинированными признаками, а на фразовом уровне – модель k-ближайших соседей с просодическими признаками. Были выявлены универсальные просодические признаки, составляющие основу «языковой модели эмоций», а также типологические различия в их реализации, отражающие влияние родного языка на эмоциональную речь билингвов.

EN

Analysis of the effectiveness of ML algorithms for emotion recognition, taking into account prosodic and spectral features

Z. A. Zavrumov, O. V. Goncharova, A. A. Levit

Abstract. The aim of the study is to determine the optimal classifier for identifying an emotional state based on the results of a comparative analysis of the effectiveness of various machine learning algorithms based on a combination of prosodic and spectral features. The scientific novelty consists in the application of ML algorithms in the recognition of emotionally marked speech of North Caucasian bilinguals in the problem of binary classification of the presence or absence of an accent with the determination of the optimal combination of universal prosodic and spectral features. During the study, an experimental corpus of speech of representatives of three ethnic groups (Russians, Kabardians and Armenians) was created with an annotation of the degree of accent, prosodic (94 signs) and spectral (74 signs) characteristics were extracted from speech signals, a comparative analysis of the effectiveness of machine learning algorithms (logistic regression, k-nearest neighbors, the method of support vectors, decision trees) in the problem of binary classification of the presence/absence of emphasis. The results of the study showed that at the syllabic level, the most effective is the decision tree model with combined features, and at the phrasal level, the k-nearest neighbor model with prosodic features. Universal prosodic features that form the basis of the "language model of emotions" were identified, as well as typological differences in their implementation, reflecting the influence of the native language on the emotional speech of bilinguals.

Введение

Несмотря на широкое использование продуктов распознавания эмоций в повседневной жизни (Cowie, Douglas-Cowie, Tsapatsoulis, Votsis, Fellenz, Taylor, 2001, p. 32-37), их обнаружение в речи все еще является сложной задачей, поскольку эмоции не только субъективны, но и культурно обусловлены (Ekman, 1971).

Научная статья (original research article) | https://doi.org/10.30853/phil20240264

Эмоциональное содержание произносимых высказываний очевидно закодировано в речевом сигнале, точное определение специфических особенностей, которые способствуют передаче эмоций, остается открытым вопросом. В большинстве работ по автоматическому распознаванию речи используются низкочастотные кепстральные коэффициенты (MFCC) (McGilloway, Cowie, Douglas-Cowie, Gielen, Westerdijk, Stroeve, 2000, p. 200-205), однако их роль как в описании качественных структур звучащей речи, так и для задачи получения наглядных взвешенных параметров диагностики вариантов неочевидна, так как, как правило, они представляют собой числовые массивы без какой-либо отсылки к значению (Liu, Wei, Morris, Zhuang, 2023), в то время как, несомненно, важно понять, «что именно в составе звука несет определенную функцию в системе языка» (Трубецкой, 2012, с. 19). Таким образом, становится очевидной актуальность проведения анализа алгоритмов машинного обучения (основанных на просодических и спектральных признаках) с последующим выбором оптимального классификатора идентификации эмоционального состояния.

В рамках настоящего исследования предполагалось решить следующие задачи:

- создать экспериментальный корпус речи представителей трех этногрупп с аннотацией степени выраженности акцента;
- извлечь просодические и спектральные признаки из речевых сигналов;
- проанализировать просодические и спектральные признаки в терминах просодических подсистем тона, длительности и интенсивности, разработанных в ходе исследований интерференции в рамках Пятигорской фонетической школы (Кипа, 2003; Садовая, 2003; Шишимер, 2003; Лукова, 2004; Ермакова, 2006; Мартьянова, 2006; Гончарова, 2008; Воробьева, 2008; Дубовский, Воробьева, Гончарова и др., 2008);
- провести сравнительный анализ эффективности различных алгоритмов машинного обучения в задаче бинарной классификации наличия/отсутствия акцента;
- определить оптимальную комбинацию признаков, обеспечивающую наилучшие результаты в распознавании акцента;
- проанализировать важность полученных признаков для задачи распознавания акцента;
- получить предварительный список универсалий «языковой модели эмоций» и «базовых этнокультурных комбинаций признаков».

В качестве материала исследования были использованы аудиозаписи квазиспонтанной реализации эмоционально-маркированных диалогов на русском языке (состояниями «радость» и «гнев») между представителями русской, кабардинской и армянской этногруппами, а также их эмоционально-нейтральных вариантов, необходимых для работы классификатора.

С точки зрения машинного обучения, распознавание эмоций в речи – это задача классификации, в которой входной образец (аудиозапись) нужно отнести к нескольким заранее определенным эмоциям, то есть на заранее размеченных и обработанных данных показать модели – вот здесь «не радость», а здесь «радость». Очевидно, что сложность этой задачи выходит за рамки технической проблемы: как вообще определить эмоцию и ее класс, когда это может быть неоднозначными даже для человека, особенно учитывая тот факт, что записи были сделаны квазиспонтанно. *Как правило, большинство специалистов, работающих со звуком, признают проблему «эмоциональных датасетов», поэтому для валидации нашего массива данных был использован метод аудитивного анализа, в ходе которого лингвисты и «наивные носители» языка давали оценку естественности звучания записи и пытались определить, с какой эмоцией была произнесена фраза. Для чистоты эксперимента было решено не предъявлять аудиторам список эмоций. Поэтому в исходном отчете, помимо искомых «гнева» и «радости», встречались «печаль», «счастье» и даже «тоска». В итоге, из более чем 5000 тыс. записей для последующего обучения модели было выбрано по 500 реализаций в каждом этно-варианте.*

Для анализа было использовано программное обеспечение Praat и скрипт на языке Python, специально разработанный для целей исследования. Мы сравнили и объединили два типа признаков: просодические и спектральные, полученные из показателей частоты основного тона, интенсивности и длительности фраз, а также их соотношений, частот первой и второй формант, статистик среднечастотных кепстральных коэффициентов (MFCC). Для анализа информативности и спектральных, и просодических признаков нами было получено несколько различных наборов с вариативными комбинациями типа признака (спектральный или просодический) и области высказывания (уровень фразы и уровень слога). Для каждого набора признаков были нормализованы динамические, темпоральные и тональные значения, контуры частоты основного тона, значения F2 и F1, а также получены статистические данные, включающие средние и медианные значения признаков, стандартное отклонение и дисперсия, минимум и максимум частоты основного тона, интенсивности и длительности, как на уровне фразы, так и послогово. Общая численность набора просодических признаков составила 94 единицы.

Спектральные характеристики представляли собой значения стандартных отклонений и средних частот MFCC, коэффициентов дельты и ускорения как первой, так и второй производных MFCC, используя конечные разности данных производных (26 признаков). Общее количество спектральных характеристик составило 74 единицы.

После первичной обработки был создан пайплайн моделей для распознавания, включающий в себя машины опорных векторов, деревья принятия решений, логистическую регрессию и алгоритм k-ближайших соседей. Далее мы использовали метод кодирования и токенизации признаков, получили два набора данных: обучающий – 80% и тестовый 20%. В нашей работе мы сравнили и объединили два типа признаков: просодические признаки и спектральные признаки, полученные из показателей частоты основного тона, интенсивности и длительности фраз, а также их соотношений, частот первой и второй формант, статистики среднечастотных

кепстральных коэффициентов (MFCC). Чтобы проанализировать полезность спектральных и просодических признаков и сравнить их эффективность с существующими подходами, мы вычислили несколько различных наборов данных, варьируя тип признака (спектральный / просодический) и область высказывания, для которой они были вычислены (уровень фразы / уровень слога). Используя данные по частотным уровням, интервалам, диапазонам и значениям формант, мы перевели первичные акустические показатели в относительные при помощи StandardScaler и MinMaxScaler библиотеки scikit-learn, далее все данные были нормализованы посредством метода нормализации z-score, который позволяет преобразовать значения переменных таким образом, чтобы они имели среднее значение 0 и стандартное отклонение 1. Принцип работы метода заключается в следующем: для каждого значения переменной вычисляется z-оценка, которая показывает, насколько это значение отклоняется от среднего значения переменной в единицах стандартного отклонения. Категориальные признаки были также преобразованы в числа в виде набора бинарных признаков со значениями 0 и 1 (см. Таблицу 1).

Таблица 1. *Фрагмент базового набора признаков для классификации до и после кодирования*

id	level_ start	level_ stressed	level_ end	тональный диапазон	регистр	int_ rising	int_ falling
R41	СН	СН	СВ	Малый	ВШвв	Малый	Узкий
R44	СВ	СВ	СВ	Ср.	ВУ	Узкий	Суж.
R42	СВ	СВ	СВ	Малый	ВШвв	Суж.	Узкий
id	level_ start	level_ stressed	level_ end	тональный диапазон	регистр	int_ rising	int_ falling
R41	1	1	2	1	4	1	2
R44	2	2	2	3	6	2	3
R42	2	2	2	1	4	3	2

Теоретической базой настоящего исследования служат работы, посвященные суперсегментным единицам речи (Анашкина, 1998; Кантер, 1988; Соколова, Гинтовт, Тихонова и др., 1991; Bolinger, 1958; Pike, 1979), интерференции и ее влиянию на речь билингвов (Вишневская, 1985; Фомиченко, 2005; Светозарова, 2006; Лаврентьева, 2008; Гончарова, 2008; Дубовский, 2008); использованию алгоритмов машинного обучения для распознавания эмоций по речевому сигналу (Богданова, Акушев, 2021).

Проведенное исследование имеет практическую значимость, так как выявленные нами универсальные просодические признаки, составляющие основу «языковой модели эмоций», и типологические различия в их реализации билингвами могут быть использованы для улучшения алгоритмов распознавания эмоций в речи вообще и в разработке более чувствительных к национальным типам эмоционально-маркированной коммуникации голосовых ассистентов и чат-ботов.

Обсуждение и результаты

Поиски методов и инструментов идентификации различных эмоциональных состояний человека привлекают ученых уже не одно десятилетие, но за последние годы возможности исследователей, безусловно, значительно расширились, в первую очередь благодаря развитию современных алгоритмов машинного обучения. Уже предложены инструментальные методы распознавания динамических жестов человека (Девятков, Алфимцев, 2007), алгоритм детекции эмоций на основе локальных бинарных паттернов (Shan, Gong, McOwan, 2009), алгоритм распознавания выражения лица (FER), использующий нейронную сеть с радиальной базисной функцией в качестве классификатора (Yi, Mao, Chen et al., 2014). Д. А. Астахов и А. В. Катаев пришли к выводу, что в решении задач классификации эмоций наибольшую эффективность среди методов ML-обучения демонстрируют ANN и SVM, однако «искусственные нейронные сети показывают более точный показатель распознавания» и лучшие результаты (2018, с. 678).

В нашем исследовании гораздо более важным представлялось проведение сравнительного анализа эффективности ML-алгоритмов в распознавании эмоционально-маркированной речи в задаче бинарной классификации наличия или отсутствия акцента с последующим определением оптимальной комбинации универсальных просодических и спектральных признаков. Как мы уже упоминали ранее, на первом этапе исследования мы провели аудитивный анализ экспериментального корпуса записей, в результате которого была получена окончательная выборка по 500 реализаций в каждом этно-варианте (русской, кабардинской и армянской этногрупп) для последующего обучения модели. Далее мы работали с помощью программы Praat и специально разработанного Python-скрипта, объединили просодические и спектральные признаки, полученные из показателей ЧОТ, интенсивности и длительности фраз, их соотношений, частот F1 / F2 и статистик среднечастотных кепстральных коэффициентов (MFCC). Таким образом, мы получили несколько наборов с вариативными комбинациями типа признака и области высказывания, для каждого из которых была осуществлена нормализация динамических, темпоральных и тональных значений, контуров ЧОТ, значений F1 / F2, получены статистические данные по средним и медианным значениям признаков, стандартному отклонению и дисперсии, минимуму/максимуму ЧОТ, интенсивности и длительности. Как мы уже отмечали, спектральные характеристики представляли собой значения стандартных отклонений и средних частот MFCC, коэффициентов дельты

и ускорения первой и второй производных MFCC (26 признаков). В итоге набор просодических признаков включил 94 единицы, спектральных характеристик – 74 единицы. Наконец, мы создали пайплайн моделей для распознавания (см. Таблицу 2), включающий в себя машины опорных векторов, деревья принятия решений, логистическую регрессию и алгоритм k-ближайших соседей.

Таблица 2. *Результаты метрик моделей на тестовой выборке*

Оценка	Precision						Recall					
Призн.	*Spectral*		*Prosodic*		*Comb*		*Spectral*		*Prosodic*		*Comb*	
Модель	*f*	*s*	*f*	*s*	*f*	*s*	*f*	*s*	*f*	*s*	*f*	*s*
LogReg	0.71	0.68	0.69	0.63	0.77	0.73	0.68	0.74	0.71	0.72	0.61	0.69
KNN	0.60	0.64	0.79	0.74	0.71	0.74	0.70	0.59	0.78	0.63	0.76	0.77
SVM	0.69	0.75	0.61	0.67	0.72	0.77	0.73	0.71	0.68	0.73	0.69	0.75
DT	0.62	0.76	0.74	0.69	0.71	0.80	0.76	0.69	0.74	0.71	0.71	0.82

Обозначения Таблицы 2:
Precision – доля правильно классифицированных положительных классов среди предсказанных положительными; Recall – доля правильно классифицированных положительных классов среди реально положительных; Spectral – спектральные признаки; Prosodic – просодические признаки; Combination – комбинация двух видов признаков; LogReg – логистическая регрессия; KNN – метод ближайших соседей; SVM – метод опорных векторов; DT – деревья решений; f – фразовый уровень; s – слоговой уровень.

Согласно результатам, представленным в Таблице 2, можно сделать следующие выводы:

1. На слоговом уровне:

- Наилучшие результаты по метрике Precision (точность) показала модель дерева решений (DT) с комбинированными признаками (спектральными и просодическими) – 0.80.
- Наилучшие результаты по метрике Recall (полнота) также показала модель дерева решений (DT) с комбинированными признаками – 0.82.
- Модели с комбинированными признаками в целом демонстрируют более высокие показатели точности и полноты по сравнению с моделями, использующими только спектральные или только просодические признаки.

2. На фразовом уровне:

- Наилучшие результаты по метрике Precision (точность) показала модель k-ближайших соседей (KNN) с просодическими признаками – 0.79.
- Наилучшие результаты по метрике Recall (полнота) показала модель дерева решений (DT) с комбинированными признаками – 0.76.
- Модели с просодическими признаками в целом демонстрируют более высокие показатели точности по сравнению с моделями, использующими только спектральные или комбинированные признаки.

Таким образом, можно сделать вывод, что на слоговом уровне наиболее эффективной является модель дерева решений (DT) с комбинированными признаками, а на фразовом уровне – модель k-ближайших соседей (KNN) с просодическими признаками. Это может быть связано с тем, что на слоговом уровне комбинация спектральных и просодических признаков позволяет лучше дифференцировать эмоциональные состояния, в то время как на фразовом уровне просодические характеристики оказываются более информативными.

Этногруппа «кабардинцы»

Тональные характеристики:

1. Начало фраз. Русские реализации характеризуются более высоким начальным уровнем ЧОТ (преимущественно среднеповышенный уровень), в то время как интерферентные фразы начинаются на среднепониженном или среднеповышенном уровнях.
2. Терминальная часть. В русских репликах наблюдается нисходящее движение тона от среднеповышенного к среднепониженному и низкому уровням. В интерферентных фразах отмечается восходяще-нисходящая конфигурация тона с пиком на среднеповышенном уровне.
3. Локализация тонального максимума. В русских высказываниях максимум ЧОТ чаще всего приходится на первый ударный слог шкалы или на стык предъядерного и ядерного слогов. В интерферентных репликах тональный пик локализуется преимущественно на ядерном слоге.
4. Частотный диапазон. Русские реализации характеризуются более широким частотным диапазоном (средний и суженный типы), в то время как интерферентные фразы обладают преимущественно суженным диапазоном.
5. Частотные регистры. Русские высказывания отличаются большей регистровой разнородностью, включая средний широкий, средний узкий и полный регистры. Интерферентные реплики реализуются в основном в среднем узком регистре.

Динамические характеристики:

1. Среднеслоговая интенсивность. Русские фразы характеризуются большей интенсивностью произнесения, реализуясь в основном в уменьшенной и средней степенях контраста. Интерферентные реплики произносятся с меньшей интенсивностью, преимущественно в малой и уменьшенной зонах.

2. Динамический максимум. В русских высказываниях пик интенсивности чаще всего приходится на ядерный слог, реже – на первый ударный слог шкалы. В интерферентных репликах максимум интенсивности локализуется преимущественно на ядерном слоге.

3. Динамический диапазон. Русские фразы обладают более широким диапазоном значений интенсивности (преимущественно средний и увеличенный типы), в то время как интерферентные реализации характеризуются меньшим динамическим диапазоном (малый и уменьшенный типы).

Темпоральные характеристики:

1. Среднеслоговая длительность. Русские высказывания произносятся с меньшей средней длительностью слогов (преимущественно уменьшенный тип), в то время как интерферентные реплики характеризуются более растянутым темпом (средний и увеличенный типы).

2. Длительность ударных слогов. Ударные слоги в русских фразах обладают меньшей длительностью по сравнению с интерферентными репликами, где ударные слоги произносятся более растянуто.

3. Длительность безударных слогов. Безударные слоги в русских высказываниях характеризуются меньшей длительностью (преимущественно малый и уменьшенный типы), в то время как в интерферентных репликах безударные слоги произносятся более растянуто (уменьшенный и средний типы).

4. Соотношение длительности слогов. В русских фразах ядерный слог характеризуется большей длительностью по сравнению с предъядерным и заядерным слогами, а также первым ударным слогом шкалы. В интерферентных репликах соотношение длительности слогов более вариативно, и ядерный слог может быть как длиннее, так и короче сравниваемых с ним слогов.

Наиболее важными признаками, позволяющими дифференцировать русский и интерферированный (с кабардинским акцентом) варианты речи с эмоцией «радость», являются тональные и темпоральные характеристики. Русские реализации характеризуются более высоким начальным уровнем ЧОТ, нисходящим движением тона в терминальной части, локализацией тонального максимума на ударных слогах шкалы, более широким частотным диапазоном и разнообразием частотных регистров. Интерферентные фразы отличаются восходяще-нисходящей конфигурацией тона, локализацией тонального пика на ядерном слоге, суженным частотным диапазоном и преобладанием среднего узкого регистра.

В динамическом плане русские высказывания произносятся с большей интенсивностью, с локализацией динамического максимума на ядерном или первом ударном слоге, а также с более широким динамическим диапазоном.

Этногруппа «армяне»

Для дифференциации русского и интерферированного варианта наиболее важными признаками являются следующие:

Тональные характеристики:

1. Уровень начала фразы: в русском варианте преобладает СН и СВ уровни, в интерферированном – СВ и высокий уровни.

2. Уровень первого ударного слога: в русском варианте доминирует СВ уровень, в интерферированном – высокий и СВ уровни.

3. Тональный максимум: в русском варианте локализуется преимущественно в шкале, в интерферированном – в шкале и ядерном слоге.

4. Тональный минимум: в русском варианте локализуется в предшкале или заядерных слогах, в интерферированном – в шкале или заядерных слогах.

5. Направление движения тона: в русском варианте дугообразно-нисходящее, в интерферированном – восходяще-нисходящее.

Динамические характеристики:

1. Среднеслоговая интенсивность: в русском варианте преобладает средняя и повышенная интенсивность, в интерферированном – средняя и пониженная.

2. Максимальная интенсивность: в русском варианте локализуется в шкале, в интерферированном – в предшкале (39%) или шкале (44,5%).

3. Динамический минимум: в русском варианте локализуется в предшкале или заядерных слогах, в интерферированном – в шкале или заядерных слогах.

4. Диапазон интенсивности: в русском варианте преобладает средний диапазон, в интерферированном – суженный.

5. Соотношение интенсивности шкалы и шкалы/заядерных слогов: в русском варианте контраст больше, в интерферированном – меньше.

Темпоральные характеристики:

1. Средняя длительность слога: в русском варианте преобладает средняя и уменьшенная длительность, в интерферированном – уменьшенная.

2. Средняя длительность ударного слога: в русском варианте больше, чем в предшкале и заядерных слогах, в интерферированном – разница меньше.

3. Соотношение длительности ядерного слогоносителя и гласных шкалы/заядерных слогов: в русском варианте контраст больше, в интерферированном – меньше.

Наиболее важными признаками для дифференциации русского и интерферированного вариантов с эмоцией «радость» являются тональные характеристики (направление движения тона, локализация тонального максимума и минимума), а также соотношение динамических и темпоральных характеристик шкалы с предшкалой и заядерными слогами.

Проанализировав просодические признаки, выражающие эмоцию радости в речи билингвов-кабардинцев и билингвов-армян, можно выделить следующие общие и различные характеристики, которые могут составить основу для «языковой модели эмоций» и типологических особенностей конкретных языков.

Общие просодические признаки:

1. Тональные характеристики:

Локализация тонального максимума преимущественно на ядерном слоге.

Наличие восходяще-нисходящей конфигурации тона.

Суженный частотный диапазон.

2. Динамические характеристики:

Локализация динамического максимума на ядерном слоге.

Меньший динамический диапазон.

3. Темпоральные характеристики:

Более растянутый темп речи.

Большая длительность ударных слогов по сравнению с безударными.

Типологические различия в просодическом выражении эмоции радости:

Этногруппа «кабардинцы»

1. Тональные характеристики:
- начало фраз на среднепониженном или среднеповышенном уровнях;
- нисходяще-восходящая конфигурация тона в терминальной части;
- локализация тонального максимума преимущественно на ядерном слоге;
- преобладание среднего узкого частотного регистра.
2. Динамические характеристики:
- произнесение с меньшей интенсивностью, преимущественно в малой и уменьшенной зонах;
- локализация динамического максимума на ядерном слоге;
- меньший динамический диапазон.
3. Темпоральные характеристики:
- более растянутый темп речи (средний и увеличенный типы);
- большая длительность ударных и безударных слогов.

Этногруппа «армяне»

1. Тональные характеристики:
- начало предшкалы и шкалы на высоком и среднеповышенном уровнях;
- завершение шкалы на высоком, среднеповышенном и среднепониженном уровнях;
- локализация тонального максимума в шкале и заядерных слогах;
- преобладание среднего узкого частотного регистра.
2. Динамические характеристики:
- произнесение со средней и пониженной интенсивностью;
- локализация динамического максимума в предшкале или шкале;
- суженный динамический диапазон.
3. Темпоральные характеристики:
- преобладание уменьшенной длительности слогов;
- меньший контраст длительности ядерного слога и шкалы/заядерных слогов.

Таким образом, универсальные просодические признаки, составляющие основу «языковой модели эмоций», включают тональные, динамические и темпоральные характеристики, связанные с локализацией максимумов/минимумов, конфигурацией движения тона, диапазонами частоты и интенсивности, а также соотношением длительности слогов. Типологические различия проявляются в конкретных реализациях этих просодических параметров, отражающих влияние родного языка на эмоциональную речь билингвов.

Заключение

В результате проведенного исследования нами были сделаны выводы относительно как общих, так и типологических различий в просодическом выражении эмоций у представителей разных этногрупп. В частности, универсальными просодическими признаками, составляющими основу «языковой модели эмоций» при выражении «радости» являются: 1) локализация тонального максимума преимущественно на ядерном слоге; 2) наличие восходяще-нисходящей конфигурации тона; 3) суженный частотный диапазон; 4) локализация динамического максимума на ядерном слоге; 5) меньший динамический диапазон; 6) более растянутый темп речи; 7) большая длительность ударных слогов по сравнению с безударными.

Типологическими различиями в просодическом выражении «радости» являются:

– для этногруппы «кабардинцы»: 1) начало фраз на среднепониженном или среднеповышенном уровнях; 2) нисходяще-восходящая конфигурация тона в терминальной части; 3) локализация тонального максимума преимущественно на ядерном слоге; 4) преобладание среднего узкого частотного регистра; 5) меньшая интенсивность, преимущественно в малой и уменьшенной зонах; 6) локализация динамического максимума на ядерном слоге; 7) меньший динамический диапазон; 8) более растянутый темп речи (средний и увеличенный типы); 9) большая длительность ударных и безударных слогов;

– для этногруппы «армяне»: 1) начало предшкалы и шкалы на высоком и среднеповышенном уровнях; 2) завершение шкалы на высоком, среднеповышенном и среднепониженном уровнях; 3) локализация тонального максимума в шкале и заядерных слогах; 4) преобладание среднего узкого частотного регистра; 5) средняя или пониженная среднеслоговая интенсивность; 6) локализация динамического максимума в предшкале или шкале; 7) суженный динамический диапазон; 8) преобладание уменьшенной длительности слогов; 9) меньший контраст длительности ядерного слога и шкалы / заядерных слогов.

В результате сравнительного анализа эффективности различных алгоритмов машинного обучения (деревья решений, k-ближайших соседей, логистическая регрессия) в задаче бинарной классификации наличия/отсутствия акцента было установлено, что на слоговом уровне наиболее эффективна модель деревьев решений с комбинированными признаками, а на фразовом уровне – модель k-ближайших соседей с просодическими признаками. Определены оптимальные комбинации признаков, обеспечивающие наилучшие результаты в распознавании акцента. Показано, что использование комбинированных (просодических и спектральных) признаков повышает точность и полноту классификации по сравнению с моделями, использующими только один тип признаков.

Перспективы дальнейших исследований связаны с расширением экспериментальной базы за счет включения других этногрупп, а также с углубленным изучением просодических механизмов выражения эмоций в речи билингвов с учетом влияния родного языка. Кроме того, представляется перспективным применение методов глубокого обучения для автоматического распознавания эмоционального состояния говорящего на основе комплексного анализа просодических и спектральных характеристик речевого сигнала.

Источники | References

1. Анашкина И. А. Звучащий текст в аспекте культурной аксиологии / М-во общ. и проф. образования РФ. Морд. гос. пед. ин-т им. М. Е. Евсевьева. Саранск: Морд. гос. пед. ин-т им. М. Е. Евсевьева, 1998.
2. Астахов Д. А., Катаев А. В. Использование современных алгоритмов машинного обучения для задачи распознавания эмоций // Cloud of science. 2018. № 4.
3. Богданова Д. Р., Акушев А. Т. Распознавание эмоций по речевому сигналу // E-Scio. 2021. № 6 (57).
4. Вишневская Г. М. Английская интонация (в условиях русской интерференции): учебное пособие / Иван. гос. ун-т. Иваново, 1985.
5. Воробьева О. В. Просодия имплицитного несогласия в русской речи северокавказских армянских билингвов: экспериментально-фонетическое исследование: дисс. ... к. филол. н. Пятигорск, 2008.
6. Гончарова О. В. Просодия русского побуждения в условиях кабардино-черкесской интерференции: экспериментально-фонетическое исследование: дисс. ... к. филол. н. Пятигорск, 2008.
7. Девятков В. В., Алфимцев А. Н. Распознавание манипулятивных жестов // Вестник Московского государственного технического университета им. Н. Э. Баумана. Серия: Приборостроение. 2007. Т. 68. Вып. 3.
8. Дубовский Ю. А., Воробьева О. В., Гончарова О. В., Мартьянова Е. О., Садовая А. Е., Шишимер Л. Ф. Русская просодия на Северном Кавказе: в 2-х т. / под общ. ред. Ю. А. Дубовского; Федеральное агентство по образованию; Пятигорский государственный лингвистический университет. Пятигорск, 2008. Т. 1.
9. Ермакова Н. А. Просодия русского восклицания в условиях осетинской интерференции: Экспериментально-фонетическое исследование: дисс.. к. филол. н. Пятигорск, 2006.
10. Кантер Л. А. Системный анализ речевой интонации. М.: Высшая школа, 1988.
11. Кипа Е. В. Просодия русского общего вопроса в условиях кабардино-черкесской интерференции: экспериментально-фонетическое исследование: дисс. ... к. филол. н. Пятигорск, 2003.
12. Лаврентьева Н. Г. Особенности русско-английской интерференции применительно к акценто-ритмической организации английской речи // Современный билингвизм: теоретические и прикладные аспекты: межвуз. сб. науч. тр. / под ред. Г. М. Вишневской. Иваново, 2008.
13. Лукова Н. В. Просодия русского специального вопроса в условиях греческой интерференции: Экспериментально-фонетическое исследование: дисс. ... к. филол. н. Пятигорск, 2004.
14. Мартьянова Е. О. Просодия русского восклицания в условиях Карачаево-Балкарской интерференции: экспериментально-фонетическое исследование на материале реплик с модальностью восхищения: дисс.. к. Филол. н. Пятигорск, 2006.
15. Садовая А. Е. Просодические черты обращения в русской речи северокавказских армянских билингвов: экспериментально-фонетическое исследование: дисс. ... к. филол. н. Пятигорск, 2003.
16. Светозарова Н. Д. Интонационная система русского языка. СПб.: Изд-во Санкт-Петербургского ун-та, 2006.
17. Соколова М. А., Гинтовт К. П., Тихонова И. С., Тихонова Р. М. Теоретическая фонетика английского языка. М.: Высш. шк., 1991.
18. Трубецкой Н. С. Основы фонологии. М.: URSS, 2012.
19. Фомиченко Л. Г. Когнитивные основы просодической интерфренции: монография. Волгоград: Изд-во Волгоградского ун-та, 2005.
20. Шишимер Л. Ф. Просодия русской ответной реплики в условиях кабардино-черкесской интерференции: экспериментально-фонетическое исследование: дисс. ... к. филол. н. Пятигорск, 2003.
21. Bolinger D. A theory of pitch accent in English // Word. 1958. Vol. 14.
22. Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W., and Taylor J. G. Emotion recognition in human-computer interaction // IEEE Signal Processing Magazine. 2001. Vol. 18. № 1.

23. Ekman P. Universals and cultural differences in facial expressions of emotion. Nebraska symposium on motivation, University of Nebraska Press, 1971.
24. Liu L., Wei L., Morris Sh., Zhuang M. Knowledge-Based Features for Speech Analysis and Classification: Pronunciation Diagnoses // Electronics. 2023. № 12 (9): 2055. URL: https://doi.org/10.3390/electronics12092055.
25. McGilloway S., Cowie S., Douglas-Cowie E., Gielen S., Westerdijk M., Stroeve S. Approaching automatic recognition of emotion from voice: A Rough benchmark // Proc. ISCA Workshop on Speech and Emotion. 2000. January.
26. Pike K. The intonation of American English // University of Michigan Publications. Linguistics, 1. Greenwood Press, 1979.
27. Shan C., Gong Sh., McOwan Peter W. Facial expression recognition based on Local Binary Patterns: A Comprehensive study // Image and Vision Computing. 2009. № 27.
28. Yi J., Mao X., Chen L., Xue Y., Compare A. Facial expression recognition considering individual differences in facial structure and texture // IET Computer Vision. 2014. Vol. 8. Iss. 5. DOI: 10.1049/iet-cvi.2013.0171.

Финансирование | Funding

RU Публикация подготовлена в рамках поддержанного РНФ и Министерством образования Ставропольского края научного проекта № 23-28-10124 «Квантитативно-статистическая модель анализа эмоционально-маркированной коммуникации в условиях межэтнических взаимодействий в регионе Кавказские Минеральные Воды».

EN The publication was prepared within the framework of the scientific project No. 23-28-10124 supported by the Russian Academy of Sciences and the Ministry of Education of the Stavropol Territory "Quantitative statistical model for the analysis of emotionally marked communication in the context of interethnic interactions in the Caucasian Mineral Waters region".

Информация об авторах | Author information

RU **Заврумов Заур Асланович**[1]
Гончарова Оксана Владимировна[2], к. филол. н., доц.
Левит Алина Александровна[3]
[1, 2, 3] Пятигорский государственный университет

EN **Zaur Aslanovich Zavrumov**[1]
Oksana Vladimirovna Goncharova[2], PhD
Alina Aleksandrovna Levit[3]
[1, 2, 3] Pyatigorsk State University

[1] *nauka@pgu.ru,* [2] *goncharovaov@pgu.ru,* [3] *levitaa@pgu.ru*

Информация о статье | About this article

Дата поступления рукописи (received): 01.05.2024; опубликовано online (published online): 13.06.2024.

Ключевые слова (keywords): языковая модель эмоций; идентификация эмоционального состояния; алгоритмы машинного обучения; просодические и спектральные признаки в речи билингва; распознавание акцента в речи билингва; language model of emotions; identification of emotional state; machine learning algorithms; prosodic and spectral features in bilingual speech; accent recognition in bilingual speech.