

УДК 004.934

DOI: 10.18413/2518-1092-2026-11-1-0-4

**Балабанова Т.Н.<sup>1</sup>**  
**Белов А.С.<sup>1</sup>**  
**Пашков А.С.<sup>2</sup>**  
**Маматов Е.М.<sup>1</sup>****АНАЛИЗ ПРОСОДИЧЕСКИХ ПАРАМЕТРОВ  
ЭМОЦИОНАЛЬНО ОКРАШЕННОЙ РЕЧИ**

<sup>1</sup> Белгородский государственный национальный исследовательский университет,  
ул. Победы, 85, г. Белгород, 308015, Россия

<sup>2</sup> Белгородский университет кооперации, экономики и права,  
ул. Садовая, 116а, г. Белгород, 308023, Россия

*e-mail: sozonova@bsuedu.ru, belov\_as@bsuedu.ru, Pogosad@yandex.ru, mamatov@bsuedu.ru*

**Аннотация**

В работе представлено исследование просодических параметров эмоционально окрашенной речи на русском языке. Целью исследования является выявление наиболее информативных акустических признаков, позволяющих различать эмоциональные состояния говорящего. В качестве экспериментальных данных использовались аудиозаписи из корпуса эмоциональной речи Dusha, включающие четыре эмоциональных состояния: злость, радость, грусть и нейтральную речь. Всего было проанализировано 240 аудиофайлов, содержащих записи мужской и женской речи.

В работе были извлечены и исследованы просодические характеристики речевого сигнала, включающие параметры высоты основного тона, энергетические, темпоральные и фонационные признаки. Для анализа данных применялся комплекс статистических методов и методов машинного обучения, включающий корреляционный анализ, оценку важности признаков с использованием алгоритма Random Forest, а также анализ главных компонент (Principal Component Analysis (PCA)).

Результаты эксперимента показали, что наибольшую информативность для распознавания эмоций в речи имеют энергетические и интонационные характеристики сигнала, в частности средняя энергия речи, вариативность частоты основного тона, темп речи и среднее значение F0. Проведённый анализ позволил выделить компактное пространство признаков и выявить характерные акустические профили для различных эмоциональных состояний. Полученные результаты могут быть использованы при разработке систем автоматического распознавания эмоций в речевых сигналах и интеллектуальных речевых интерфейсов.

**Ключевые слова:** просодические параметры речи; эмоциональная речь; распознавание эмоций; анализ речевых сигналов; частота основного тона; машинное обучение; Random Forest; анализ главных компонент; акустические признаки

**Для цитирования:** Балабанова Т.Н., Белов А.С., Пашков А.С., Маматов Е.М. Анализ просодических параметров эмоционально окрашенной речи // Научный результат. Информационные технологии. – Т.11, №1, 2026. – С. 29-39. DOI: 10.18413/2518-1092-2026-11-1-0-4

Balabanova T.N.<sup>1</sup>  
Belov A.S.<sup>1</sup>  
Pashkov A.S.<sup>2</sup>  
Mamatov E.M.<sup>1</sup>

ANALYSIS OF PROSODIC PARAMETERS  
OF EMOTIONALLY COLORED SPEECH

<sup>1)</sup> Belgorod State National Research University,  
85 Pobedy St., Belgorod, 308015, Russia

<sup>2)</sup> Belgorod University of Cooperation, Economics and Law,  
116a Sadovaya St., Belgorod, 308023, Russia

*e-mail: sozonova@bsuedu.ru, belov\_as@bsuedu.ru, Pogosad@yandex.ru, mamatov@bsuedu.ru*

**Abstract**

This paper presents a study of prosodic parameters of emotionally colored speech in the Russian language. The aim of the study is to identify the most informative acoustic features that allow distinguishing the emotional state of a speaker. The experimental data consisted of audio recordings from the Dusha emotional speech dataset, including four emotional states: anger, joy, sadness, and neutral speech. In total, 240 audio recordings of both male and female speakers were analyzed.

The study focused on extracting and analyzing prosodic characteristics of speech signals, including pitch-related, energy, temporal, and phonation features. A combination of statistical analysis and machine learning methods was applied, including correlation analysis, feature importance estimation using the Random Forest algorithm, and Principal Component Analysis (PCA).

The experimental results demonstrate that energy and pitch-related characteristics of speech are the most informative features for emotion recognition. In particular, mean signal energy, variability of the fundamental frequency, speech rate, and mean F0 showed the highest contribution to emotion classification. The analysis allowed identifying a compact feature space and revealing characteristic acoustic profiles for different emotional states. The obtained results can be used in the development of automatic speech emotion recognition systems and intelligent speech-based human–computer interaction technologies.

**Keywords:** prosodic parameters; emotional speech; speech emotion recognition; speech signal analysis; fundamental frequency; machine learning; Random Forest; principal component analysis; acoustic features

**For citation:** Balabanova T.N., Belov A.S., Pashkov A.S., Mamatov E.M. Analysis of Prosodic Parameters of Emotionally Colored Speech // Research result. Information technologies. – Т. 11, № 1, 2026. – P. 29-39. DOI: 10.18413/2518-1092-2026-11-1-0-4

Одну из ключевых ролей при выражении эмоционального состояния человека играет речь. В частности, просодические параметры речевого сигнала, к которым можно отнести интонацию, темп речи, динамические характеристики, являются хорошим индикатором для определения эмоционального состояния.

Эмоциональное состояние человека при общении выражается в различных аспектах: жестах, позе, мимике и т.д., однако выражение эмоций через речевую коммуникацию является наиболее сильным и показательным фактором. При общении человек не только передает информацию собеседнику, но и выражает свое отношение как к информации, так и к собеседнику. Такое психоэмоциональное состояние отражается на различных параметрах речевого сигнала, например, таких как высота основного тона, интенсивность звучания, темп и ритм речи, а также длительность пауз. Все эти элементы в совокупности формируют интонационную структуру высказывания, что позволяет слушателю не только воспринимать сказанное, но и интерпретировать эмоциональную окраску речевого сигнала.

В последние годы наблюдается значительный рост интереса к исследованию эмоциональной речи, что обусловлено развитием систем автоматического распознавания речи, речевых интерфейсов, интеллектуальных ассистентов и технологий человеко-машинного взаимодействия. Для эффективного функционирования подобных систем одним из важных аспектов является автоматический анализ эмоционального состояния человека по его речи. В свою очередь

акустические и просодические характеристики речевого сигнала являются ключевыми параметрами, по которым можно провести такую оценку.

Одними из первых систематических исследований эмоциональной экспрессии в голосе стали работы, посвящённые изучению вокальной передачи эмоций. В частности, в исследованиях показано, что изменения высоты основного тона, амплитуды и темпа речи являются важными индикаторами эмоционального состояния говорящего [1]. В дальнейшем было установлено, что различные эмоции характеризуются специфическими акустическими профилями, отражающими физиологическое состояние человека и уровень эмоционального возбуждения [2].

Анализ интонационной структуры речевого сигнала так же привнес значимый вклад в исследования эмоциональной речи. В исследованиях рассматривались такие акустические характеристики речи, как вариативность высоты основного тона и диапазон частот. Было показано, что данные параметры являются значимыми для определения эмоционального состояния человека [3].

В дальнейших исследованиях были использованы методы машинного обучения для автоматического извлечения большого числа параметров речевого сигнала для дальнейшего их использования для решения задач паралингвистического анализа [4]. Результаты исследований показали, что объединение просодических и спектральных характеристик речи позволяют увеличить качество анализа.

Ряд работ показал, что динамические характеристики интонации и интенсивности речевого сигнала так же являются значимыми параметрами для распознавания эмоционально окрашенной речи [5]. В работах [6, 7] показано, что для систем автоматического распознавания эмоционального состояния по речи значимыми параметрами являются высота основного тона, интенсивность и темп речи.

Важным аспектом исследования эмоционально окрашенной речи являются исследования, посвященные теоретическим основам описания эмоций с точки зрения психологических аспектов [8]. Эти исследования играют важную роль для формирования корпусов эмоционально окрашенной речи, которые в дальнейшем используются для тренировки автоматических алгоритмов паралингвистического анализа. Следует отметить, что выражение естественных эмоций в речи в значительной степени отличается от выражения эмоций актерской игры. Это создает определенные сложности при формировании корпусов эмоциональной речи для тренировки алгоритмов [9].

В работах [10, 14], посвященных проблеме распознавания эмоций по речи сделаны выводы о том, что качество систем распознавания эмоций зависит, в основном, от двух факторов: выбора параметров речевого сигнала для анализа и методов обработки. Методы обработки, в свою очередь, сильно зависят от корпуса эмоциональной речи, на котором осуществлялось обучение [13].

Современные исследования в области распознавания эмоций в речи активно используют методы глубокого обучения и нейронные сети. В обзорных работах последних лет отмечается значительный прогресс в данной области благодаря использованию больших речевых корпусов и методов представления признаков на основе нейронных моделей [16].

Современные методы представления речевых сигналов, основанные на самообучающихся моделях, позволяют извлекать информативные признаки напрямую из аудиосигнала. С одной стороны это позволяет повысить точность распознавания эмоций по речи, с другой стороны, извлекаемые параметры часто невозможно интерпретировать с точки зрения физических характеристик сигнала [17, 18].

Так же, в научных исследованиях последних лет, например, материалах ведущей международной конференции, посвященной науке и технологиям обработки устной речи, довольно часто встречаются работы, посвященные использованию глубокого обучения при решении задачи распознавания различных невербальных аспектов поведения человека по его речи. В том числе, определение эмоционального состояния. Многие работы демонстрируют увеличение эффективности систем автоматического анализа речи при использовании нейросетевого подхода [19].

Однако, большое число исследований, представленных в последнее время, отмечают, что несмотря на наличие успешных решений задачи распознавания невербальных аспектов поведения человека по речи, остается проблема выбора набора параметров речевого сигнала, который бы позволил адекватно отнести речевой сигнал к той или иной эмоции и был интерпретируем с точки зрения физической характеристики сигнала. Такими параметрами могут выступать акустические и просодические параметры речевого сигнала [20].

Существует множество работ, посвящённых анализу эмоциональной речи, однако вопрос о том, какие именно просодические характеристики являются наиболее информативными, остаётся открытым и требует дальнейшего внимания. Одной из причин такой сложности является высокая вариативность проявления эмоций в речевом сигнале, которая зависит от множества факторов.

В данной работе представлено исследование просодических параметров эмоционально окрашенной речи на русском языке на основе анализа акустических характеристик аудиосигналов дикторов обоих полов.

В качестве экспериментальных данных были использованы 240 аудиофайлов речевого сигнала из датасета эмоционально окрашенной речи от Сбера Dusha. В отобранных файлах содержалось 120 аудио женского голоса, 120 – мужского. При этом анализировались аудиофайлы 4 эмоции: злость, нейтраль, радость, грусть (по 60 аудио каждой эмоции: 30 женской речи, 30 мужской).

Исследуемые группы просодических параметров речи представлены в таблице 1. Так же в таблице представлены усредненные значения исследуемых параметров по всем аудиофайлам.

Таблица 1

Группы просодических параметров

Table 1

Groups of prosodic parameters

	Группа параметров	Параметр	Обозначение	Значение
1	Интонационные	Средняя частота основного тона	F0_mean	219,692547
		Стандартное отклонение частоты основного тона	F0_std	87,490058
		Диапазон изменения частоты	F0_range	280,092888
		Среднее изменение F0 между соседними временными отсчётами (общая динамика интонационного контура речи)	$\Delta F0\_mean$	-0,638419
		Степень вариативности изменения F0 (насколько сильно изменяется высота тона в процессе речи)	$\Delta F0\_std$	26,333538
		Наклон интонационного контура	F0_slope	-0,755006
		Регрессия F0	F0_intercept	239,963501
2	Энергетические	Средняя энергия сигнала	Energy_mean	0,029578
		Вариативность интенсивности	Energy_std	0,034774
3	Темпоральные	Темп речи	Speech_rate	17,662191
		Отношение длительности пауз к общей длительности сигнала	Pause_ratio	0,437258
4	Фонационные	Микровариации частоты (нестабильность периода колебаний голосовых связок)	jitter	0,025268
		Микровариации амплитуды (нестабильность амплитуды речевого сигнала)	shimmer	0,103951

С целью анализа и интерпретации полученных просодических параметров эмоциональной речи использовался комплекс статистических методов анализа и методов машинного обучения, что позволило выявить взаимосвязи между параметрами, определить статистическую значимость различий между эмоциями и выделить наиболее информативные признаки.

Одной из ключевых задач анализа рассматриваемых просодических параметров речи является выявление корреляционных зависимостей между ними. Это позволит определить наличие и степень зависимости между рассматриваемыми параметрами, а, так же, установить есть ли избыточность у данного набора признаков, так как нет смысла использовать для распознавания эмоционального состояния человека параметры с сильной корреляционной зависимостью. Таким образом, первоначально был проведен корреляционный анализ величин, представленных в таблице 1. Результаты корреляционного анализа представлены в таблице 2.

Таблица 2

Выявленные сильные корреляционные зависимости ( $|r|>0,7$ )

Table 2

Revealed strong correlations ( $|r|>0,7$ )

№	Параметр 1	Параметр 2	r
1	F0_mean	F0_intercept	0,739
2	F0_std	F0_range	0,982
3	F0_std	$\Delta F0\_std$	0,942
4	F0_range	$\Delta F0\_std$	0,931
5	$\Delta F0\_mean$	F0_slope	0,931
6	$\Delta F0\_mean$	F0_intercept	-0,769
7	F0_slope	F0_intercept	-0,806
8	Energy_mean	Energy_std	0,886
9	Speech_rate	Pause_ratio	-1,000

По итогам корреляционного анализа были выявлены зависимости между рядом рассматриваемых параметров речевого сигнала. Причем преимущественно корреляционная зависимость наблюдалась внутри групп параметров.

1. Группа интонационных параметров. В этой группе представлены характеристики, связанные с частотой основного тона. Так, наиболее сильная корреляционная зависимость наблюдается между стандартным отклонением частоты основного тона (F0\_std) и диапазоном изменения частоты основного тона (F0\_range)  $r=0,982$ , а, так же, между стандартным отклонением частоты основного тона (F0\_std) и степенью вариативности изменения F0 ( $\Delta F0\_std$ )  $r=0,942$ . Этот факт свидетельствует о том, что все три интонационных параметра описывают степень интонационной изменчивости речи и использование их всех влечет за собой избыточность при анализе эмоционального состояния человека, что может отрицательно сказываться на работе моделей классификации. В связи с этим, из дальнейшего анализа были исключены параметры F0\_range и  $\Delta F0\_std$ , а оставлен параметр F0\_std.

Так же анализ показал высокую обратную корреляционную зависимость между наклоном интонационного контура (F0\_slope) и регрессией частоты основного тона (F0\_intercept)  $r=-0,806$ . Оба эти параметра характеризуют один и тот же тренд. В связи с этим из дальнейшего анализа исключается параметр F0\_slope.

2. Группа энергетических параметров. Эта группа содержит два параметра: среднюю энергию сигнала и вариативность интенсивности. Эти два параметра имеют высокую корреляцию  $r=0,886$ . Данный факт свидетельствует о том, что один из этих параметров можно исключить из дальнейшего рассмотрения. В дальнейшем исследовании будет рассматриваться средняя энергия сигнала и исключен параметр вариативности интенсивности. Это обусловлено тем, что параметр средней энергии сигнала является основным.

3. Группа темпоральных признаков. В данной группе рассматривались два параметра: отношение длительности пауз к общей длительности сигнала и темп речи. Между ними наблюдается высокая отрицательная корреляционная зависимость. Поэтому для дальнейшего анализа был оставлен только один параметр, а именно, темп речи.

В результате проведения корреляционного анализа были выбраны 8 параметров из 13. Результирующий набор параметров представлен в таблице 3.

Таблица 3

Параметры с небольшой корреляционной зависимостью

Table 3

Parameters with a small correlation dependence

№	Параметр	Обозначение
1	Средняя частота основного тона	F0_mean
2	Стандартное отклонение частоты основного тона	F0_std
3	Среднее изменение F0 между соседними временными отсчётами	$\Delta F0\_mean$
4	Регрессия F0	F0_intercept
5	Средняя энергия сигнала	Energy_mean
6	Темп речи	Speech_rate
7	Микровариации частоты	Jitter
8	Микровариации амплитуды	Shimmer

Проведенный корреляционный анализ позволил уменьшить количество анализируемых параметров для определения эмоционального состояния по речи, не уменьшая информативности рассматриваемого набора параметров.

На следующем этапе эксперимента был применен метод Random Forest. Целью исследования было определение значимости каждого из рассматриваемых параметров речи. В результате была получена информация о том, какие из параметров оказывают большее влияние на распознавание эмоционального состояния по речи, а какие наименьшее.

Для оценки важности признаков использовался набор просодических параметров, полученный после корреляционного анализа. На основе данного набора признаков была обучена модель Random Forest. При построении деревьев осуществлялось последовательное разбиение обучающей выборки по рассматриваемым признакам. На каждом шаге выбирался тот признак, который максимально уменьшает неопределённость (критерий Джини) в данных. Важность признака определялось как суммарное уменьшение критерия нечистоты. Чем сильнее признак снижает неопределённость при разделении данных, тем выше его информативность. После обучения модели была вычислена важность каждого признака, характеризующая его вклад в процесс классификации эмоциональных состояний. Результаты эксперимента представлены в таблице 4.

Таблица 4

Результаты оценки важности признаков

Table 4

Results of the assessment of the importance of the signs

Energy_mean	F0_std	Speech_rate	F0_mean	jitter	F0_intercept	$\Delta F0\_mean$	shimmer
0,140	0,139	0,129	0,126	0,122	0,122	0,117	0,106

По результату эксперимента видно, что наибольшую важность при оценке эмоционального состояния по речевому сигналу имеет средняя энергия сигнала (Energy\_mean), поскольку эмоциональная речь имеет больше экспрессии. Например, при выражении радости и злости человек говорит громче, и средняя энергия сигнала увеличивается, при грусти, наоборот, речь становится более тихой. Однако, следует заметить, что этот факт является показательным только в том случае, если при анализе известно значение средней энергии сигнала в предыдущие моменты времени. Если этой информации нет, то основываться только на средней энергии сигнала необоснованно.

Следующим по значимости параметром является стандартное отклонение частоты основного тона (F0\_std). Это объясняется тем, что речь, содержащая эмоции радости и злости обладает большой динамикой изменения высоты голоса, а содержащая грусть или нейтральная речь является менее динамичной. Так же, эмоциональная речь, как правило является либо более быстрой (радость, злость), либо более медленной (грусть) по отношению к нейтральной, что отражается в важности Speech\_rate, который характеризует темп речи. Следует отметить, что важность рассматриваемых признаков по итогам анализа имеет отличия во втором знаке после запятой, то есть нет признака с ярко выраженной важностью.

Таким образом, можно заключить, что рассматриваемые признаки речевого сигнала позволяют определить эмоциональное состояние человека, однако для этого необходимо использовать анализ нескольких признаков в совокупности, а не одного из них.

На следующем этапе анализа был использован метод анализа главных компонент (Principal Component Analysis (PCA)), который позволил выявить скрытые факторы, которые определяют структуру параметров эмоциональной речи. Данный метод предполагает преобразование исходного набора параметров в набор ортогональных (некоррелированных) компонент. Полученные компоненты показывают наиболее значимые направления вариации. Первая компонента показывает максимальную дисперсию данных, вторая – максимальную долю оставшейся дисперсии и так далее.

Важно отметить, что в процессе анализа использовались все восемь отобранных компонент, что позволяет получить более полное представление о данных. Перед проведением анализа главных компонент все признаки были нормированы. Нормализация необходима для того, чтобы исключить влияние различий в масштабах измерений, что может привести к искажению результатов анализа. Результаты PCA анализа представлены в таблице 5.

Таблица 5

Результаты PCA анализа

Table 5

PCA analysis results

Компонента	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
% общей дисперсии	34,5	18,6	16,1	12,4	10,1	5,0	2,4	0,9
Сумма	53,1							
	69,2							
	91,7							

По результатам эксперимента видно, что 91,7 % общей дисперсии приходится на пять компонент. Это говорит о том, что описание эмоционально окрашенной речи может быть представлено в пространстве меньшей размерности. Так же, следует отметить, что всего первыми двумя компонентами объясняется 53,1 % дисперсии. Что подтверждает возможность использовать меньшую размерность, которая будет состоять из первых нескольких компонент.

Для более развернутого анализа были рассмотрены веса исходных параметров в первых трех компонентах. Данные представлены в таблице 6.

Таблица 6

Анализ трех первых компонент

Table 6

Analysis of the first three components

№	Компонента	Параметры с наибольшим весом	Обозначение параметра
1	PC1	Средняя частота основного тона	F0_mean
		Стандартное отклонение частоты основного тона	F0_std
		Степень вариативности изменения F0	$\Delta F0$ mean
2	PC2	Средняя энергия сигнала	Energy_mean
		Темп речи	Speech_rate
3	PC3	Микровариации частоты	Jitter
		Микровариации амплитуды	Shimmer

Таким образом, первая компонента в основном определяется интонационными параметрами речи. Вторая компонента определяется динамикой речи. Третья компонента имеет наибольшие веса для параметров микровариации частоты и амплитуды основного тона. Большие значения данной компоненты указывают на нестабильность работы голосовых связок, что свидетельствует о таких эмоциях, как радость и злость, маленькие значения свидетельствуют о таких эмоциях, как грусть или нейтральная речь.

На рисунке представлена PCA-диаграмма для первых двух компонент, на которой четыре эмоции отображены в пространстве просодических признаков. Первая компонента (PC1) в основном отражает общую акустическую активность речи (высоту тона, вариативность F0 и энергию), вторая компонента (PC2) связана с микровариациями голоса и динамикой речи. Каждая точка на диаграмме представляет собой эмоцию, описанную набором из 8 параметров, оставленных после корреляционного анализа.

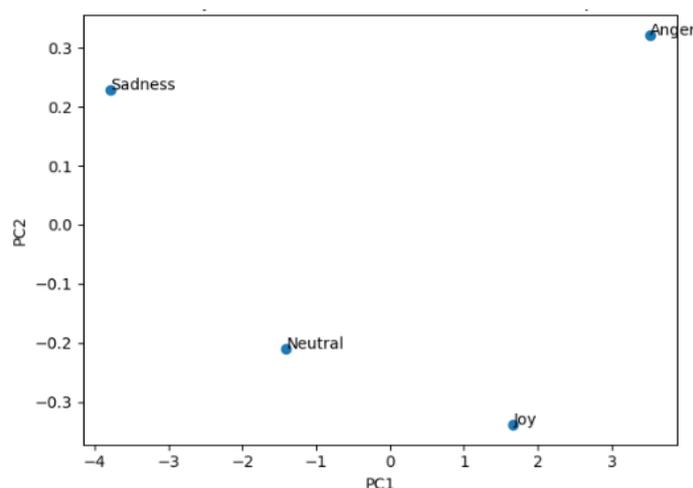


Рис. PCA-диаграмма для первых двух компонент  
Fig. PCA-diagram for the first two components

На данной диаграмме эмоция «злость» (Anger) находится в правой части пространства PC1. Это означает, что у речевого сигнала данного класса высокая энергия, большое значение F0 и высокая вариативность интонации. То есть злость – это самая акустически активная эмоция.

Эмоция «радость» (Joy) расположена близко к злости по PC1, но отличается по PC2. Это означает, что уровень энергии и F0 имеют высокие значения, но структура микровариаций голоса отличается.

Эмоция «грусть» (Sadness) находится в противоположной части пространства PC1. Это указывает на низкую энергию, низкое значение F0 и слабую интонационную вариативность.

Отображение нейтрального состояния (Neutral) представлено ближе к центру пространства признаков, что соответствует умеренным значениям всех параметров.

Таким образом, на основе проведенного анализа эмоциональной речи можно привести описание четырех эмоциональных состояний человека (злость, радость, грусть и нейтральная речь) на основе восьми значимых просодических признаков речевого сигнала.

1. Речь, выражающая злость, характеризуется высокой энергетикой и значительной вариативностью интонации.

- параметр F0\_mean имеет высокие значения, что отражает повышенную высоту голоса. При этом показатель F0\_std также высокий, что указывает на сильные колебания частоты основного тона;

- средняя энергия сигнала (Energy\_mean) достигает максимальных значений среди рассматриваемых эмоций, поскольку злость часто сопровождается усилением громкости речи;

- параметр Speech\_rate также повышен, так как говорящий может произносить фразы быстрее, что связано с высоким уровнем эмоционального возбуждения;

- показатели Jitter и Shimmer, характеризующие микровариации частоты и амплитуды, находятся на среднем уровне. Несмотря на напряжённость речи, голос обычно остаётся достаточно устойчивым;

- параметр  $\Delta F0\_mean$  демонстрирует высокие значения, отражая резкие изменения высоты голоса внутри фраз;

- значение F0\_intercept, характеризующее общий уровень тонального контура, также повышено.

2. Речь, выражающая радость, характеризуется повышенной высотой голоса и выраженной интонационной динамикой, однако отличается большей плавностью по сравнению со злостью:

- среднее значение частоты основного тона (F0\_mean) является высоким, что связано с эмоциональным подъёмом и повышением тональности речи;

- параметр F0\_std также имеет высокие значения, отражая богатую интонационную вариативность;

- средняя энергия сигнала (Energy\_mean) повышена, однако обычно несколько ниже, чем при выражении злости, поскольку радостная речь реже сопровождается криком или резким усилением громкости;

- темп речи (Speech\_rate) может быть ускоренным, так как положительные эмоции часто сопровождаются более активной речевой динамикой;

- параметры Jitter и Shimmer имеют относительно низкие значения, что указывает на стабильность голосообразования и отсутствие напряжения голосовых связок;

- показатель  $\Delta F0\_mean$  высок, что отражает активные изменения высоты тона внутри высказывания;

- параметр F0\_intercept также принимает высокие значения, соответствующие общему повышенному уровню тонального контура.

3. Речь, выражающая грусть, характеризуется пониженной энергетикой, медленным темпом и слабой интонационной вариативностью:

- параметр F0\_mean имеет низкие значения, что отражает пониженную высоту голоса;

- стандартное отклонение частоты основного тона (F0\_std) также низкое, что указывает на монотонность интонационного рисунка;

- средняя энергия сигнала (Energy\_mean) значительно ниже, чем при других эмоциональных состояниях, поскольку грустная речь обычно произносится тихо и спокойно;

- темп речи (Speech\_rate) замедлен, что связано с общим снижением эмоциональной активности;

- показатели Jitter и Shimmer имеют относительно высокие значения. Это может отражать нестабильность голосообразования и эмоциональную подавленность;

- параметр  $\Delta F0$ \_mean демонстрирует низкие значения, поскольку изменения высоты тона происходят медленно и слабо выражены;

- показатель  $F0$ \_intercept также является низким, что соответствует общему пониженному уровню тонального контура.

4. Нейтральная речь характеризуется умеренными значениями всех просодических параметров и отсутствием выраженных эмоциональных проявлений.

В качестве дальнейших исследований представляется целесообразным провести исследование просодических параметров эмоциональной речи с учетом пола диктора. Мужской и женский голос имеют свои особенности, которые отражаются на просодических параметрах речи и учет пола диктора может повысить качество распознавания эмоции.

### Список литературы

#### References

1. Scherer K.R. Vocal communication of emotion: A review of research paradigms // *Speech Communication*. – 2003. – Vol. 40, № 1–2. – P. 227-256.
2. Scherer K.R., Wallbott H.G. Evidence for universality and cultural variation of emotional expression in voice // *Journal of Cross-Cultural Psychology*. – 1994. – Vol. 25, № 1. – P. 92-110.
3. Bänziger T., Scherer K.R. The role of intonation in emotional expressions // *Speech Communication*. – 2005. – Vol. 46. – P. 252-267.
4. Schuller B., Batliner A. *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. – Chichester: Wiley, 2014. – 324 p.
5. Schröder M. Emotional speech synthesis: A review // *Proceedings of the European Conference on Speech Communication and Technology*. – Geneva, 2003. – P. 561-564.
6. Busso C., Bulut M., Narayanan S. Toward effective automatic recognition systems of emotion in speech // *IEEE Transactions on Audio, Speech, and Language Processing*. – 2009. – Vol. 17, № 5. – P. 846-859.
7. Narayanan S., Busso C. Analysis of emotional speech: A review // *IEEE Signal Processing Magazine*. – 2011. – Vol. 28, № 5. – P. 98-112.
8. Ekman P. An argument for basic emotions // *Cognition and Emotion*. – 1992. – Vol. 6, № 3–4. – P. 169-200.
9. Cowie R., Douglas-Cowie E. Emotion recognition in human-computer interaction // *IEEE Signal Processing Magazine*. – 2001. – Vol. 18, № 1. – P. 32-80.
10. Ververidis D., Kotropoulos C. Emotional speech recognition: Resources, features and methods // *Speech Communication*. – 2006. – Vol. 48, № 9. – P. 1162-1181.
11. Rabiner L., Juang B.-H. *Fundamentals of Speech Recognition*. – New Jersey: Prentice Hall, 1993. – 507 p.
12. Murray I., Arnott J. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion // *Journal of the Acoustical Society of America*. – 1993. – Vol. 93, № 2. – P. 1097–1108.
13. Banse R., Scherer K.R. Acoustic profiles in vocal emotion expression // *Journal of Personality and Social Psychology*. – 1996. – Vol. 70, № 3. – P. 614-636.
14. El Ayadi M., Kamel M., Karray F. Survey on speech emotion recognition: Features, classification schemes and databases // *Pattern Recognition*. – 2011. – Vol. 44, № 3. – P. 572-587.
15. Schuller B. Speech emotion recognition: Two decades in a nutshell, benchmarks and ongoing trends // *Communications of the ACM*. – 2018. – Vol. 61, № 5. – P. 90-99.
16. Latif S., Rana R., Qadir J., Epps J. Speech emotion recognition: State-of-the-art review // *IEEE Access*. – 2021. – Vol. 9. – P. 114509-114539.
17. Neumann M., Vu N.T. Improving speech emotion recognition with unsupervised representation learning on unlabeled speech // *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. – 2021. – Vol. 29. – P. 2388-2399.
18. Pepino L., Riera P., Ferrer L. Emotion recognition from speech using wav2vec 2.0 embeddings // *Proceedings of the Interspeech Conference*. – 2021. – P. 3400-3404.

19. Wagner J., Triantafyllopoulos A., Schuller B. Deep learning in paralinguistics: Recent trends and perspectives // IEEE Signal Processing Magazine. – 2023. – Vol. 40, № 3. – P. 104-118.

20. Zhang Z., Deng J., Schuller B. Advances in speech emotion recognition: A survey // IEEE Transactions on Affective Computing. – 2024. – Vol. 15, № 1. – P. 123-139.

**Балабанова Татьяна Николаевна**, кандидат технических наук, доцент, заведующий кафедрой автоматизированных систем и технологий, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

**Белов Александр Сергеевич**, кандидат технических наук, доцент, доцент кафедры автоматизированных систем и технологий, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

**Пашков Александр Сергеевич**, аспирант кафедры информационной безопасности, Белгородский университет кооперации, экономики и права, г. Белгород, Россия

**Маматов Евгений Михайлович**, кандидат технических наук, доцент, доцент кафедры автоматизированных систем и технологий, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

**Balabanova Tatyana Nikolaevna**, Candidate of Technical Sciences, Associate Professor, Head of the Department of Automated Systems and Technologies, Belgorod State National Research University, Belgorod, Russia

**Belov Alexander Sergeevich**, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Automated Systems and Technologies, Belgorod State National Research University, Belgorod, Russia

**Pashkov Alexander Sergeevich**, Graduate student of the Information Security Department, Belgorod University of Cooperation, Economics and Law, Belgorod, Russia

**Mamatov Evgeny Mikhailovich**, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Automated Systems and Technologies, Belgorod State National Research University, Belgorod, Russia