

<https://doi.org/10.21869/2223-1536-2026-16-1-93-106>

УДК 615.8

Анализ методов обработки и исследования речевого сигнала для установления психоэмоционального состояния человека

Д. А. Кравчук¹✉, А. С. Котенко¹

¹ Южный федеральный университет

ул. Шевченко, д. 2/Е, Ростовская область, г. Таганрог 347922, Российская Федерация

✉ e-mail: kravchukda@sfedu.ru

Резюме

Целью исследований является всесторонний анализ речевого сигнала для выявления его параметров, наиболее информативных для определения психоэмоционального состояния говорящего. Работа направлена на сравнительное изучение и демонстрацию возможностей различных методов обработки сигналов (на основе анализа функции автокорреляции, спектрального, вейвлет-анализа и аттракторного анализа) в контексте извлечения биомаркеров психоэмоциональных состояний.

Методы. В работе применяется комплекс методов частотно-временного и нелинейного анализа речевого сигнала, реализованных в среде Matlab: автокорреляционный анализ – для оценки периодичности сигнала и определения частоты основного тона, параметры которой (среднее значение, дисперсия, характер изменений) служат биомаркерами эмоций; спектральный анализ (спектрограмма) – для визуализации и анализа изменения спектрального состава сигнала во времени, выявления формант и особенностей артикуляции, связанных с эмоциональной окраской; вейвлет-анализ (с использованием вейвлета Морле) – для исследования локализованных во времени особенностей сигнала на разных частотных масштабах, что позволяет анализировать переходные процессы и распределение энергии по частотным диапазонам; аттракторный анализ (анализ динамических систем) – для реконструкции фазового пространства речевого сигнала и исследования его нелинейной динамики через построение и анализ аттракторов (их формы, сложности, фрактальной размерности).

Результаты. Подтверждена эффективность каждого метода в выявлении специфических биомаркеров эмоционального состояния. Показано, что комплексное применение методов обеспечивает более глубокое и полное понимание речевого сигнала, чем каждый метод в отдельности. Установлено, что аттракторный анализ является наиболее перспективным методом для первичного выявления глубинных изменений, связанных с психоэмоциональным состоянием.

Заключение. Комбинированный подход позволяет всесторонне оценивать речевую динамику. Особенно выделен потенциал аттракторного анализа как метода, способного выявлять тонкие нелинейные паттерны, что открывает новые возможности в разработке систем автоматического определения психоэмоционального состояния, диагностики нарушений и создания более естественных человеко-машинных интерфейсов.

Ключевые слова: методы исследования речевого сигнала; определение психоэмоционального состояния человека; параметры речевого сигнала; обработка речевого сигнала.

Конфликт интересов: Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Кравчук Д. А., Котенко А. С. Анализ методов обработки и исследования речевого сигнала для установления психоэмоционального состояния человека // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2026. Т. 16, № 1. С. 93–106. [https://doi.org/ 10.21869/2223-1536-2026-16-1-93-106](https://doi.org/10.21869/2223-1536-2026-16-1-93-106)

Поступила в редакцию 11.01.2026

Подписана в печать 10.02.2026

Опубликована 31.03.2026

Analysis of speech signal processing and research methods for establishing the psycho-emotional state of a person

Denis A. Kravchuk¹✉, Alexander S. Kotenko¹

¹ Southern Federal University

2/E Shevchenko Str., Rostov region, Taganrog 347922, Russian Federation

✉ e-mail:

Abstract

The purpose of the research is a comprehensive analysis of the speech signal to identify its parameters most informative for determining the speaker's psychoemotional state. The work is aimed at the comparative study and demonstration of the capabilities of various signal processing methods (based on autocorrelation function analysis, spectral, wavelet, and attractor analysis) in the context of extracting biomarkers of psychoemotional states.

Methods. The work employs a set of methods for time-frequency and nonlinear analysis of the speech signal, implemented in the Matlab environment: autocorrelation analysis – to assess signal periodicity and determine fundamental frequency, the parameters of which (mean value, variance, nature of changes) serve as emotion biomarkers; spectral analysis (spectrogram) – for visualization and analysis of changes in the spectral composition of the signal over time, identifying formants and articulation features associated with emotional coloring; wavelet analysis (using the Morlet wavelet) – for investigating time-localized features of the signal at different frequency scales, allowing the analysis of transient processes and energy distribution across frequency bands; attractor analysis (dynamic systems analysis) – for reconstructing the phase space of the speech signal and investigating its nonlinear dynamics through the construction and analysis of attractors (their shape, complexity, fractal dimension).

Results. The effectiveness of each method in revealing specific biomarkers of emotional state was confirmed. It was shown that the combined application of the methods provides a deeper and more complete understanding of the speech signal than each method individually. It was concluded that attractor analysis is the most promising method for the primary detection of profound changes related to psychoemotional state, due to its ability to reveal nonlinear and hidden patterns inaccessible to traditional linear methods.

Conclusion. The research demonstrates that modern speech signal analysis methods, from classical ones (autocorrelation function, spectrogram) to advanced ones (wavelet and attractor analysis), provide a powerful toolkit for extracting objective emotion biomarkers. A combined approach allows for a comprehensive assessment of speech dynamics. The potential of attractor analysis is particularly highlighted as a method capable of detecting subtle nonlinear patterns, which opens new opportunities in developing systems for automatic determination of psychoemotional state, diagnosis of disorders, and creating more natural human-machine interfaces.

Keywords: speech signal research methods; human psychoemotional state determination; speech signal parameters; speech signal processing.

Conflict of interest: The Authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

For citation: Kravchuk D.A., Kotenko A.S. Analysis of speech signal processing and research methods for establishing the psycho-emotional state of a person. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Serija: Upravlenie, vychislitel'naja tekhnika, informatika. Meditsinskoe priborostroenie = Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering*. 2026;16(1):93–106. (In Russ.) <https://doi.org/10.21869/2223-1536-2026-16-1-93-106>

Received 11.01.2026

Accepted 10.02.2026

Published 31.03.2026

Введение

Актуальность исследования методов анализа речевого сигнала обусловлена их широким потенциалом применения в различных областях, включая медицинские учреждения, авиацию, диспетчерские службы, службы общественной безопасности и экстренного реагирования [1]. Методы анализа речевого сигнала применяются во многих системах, например: системы распознавания и синтеза речи, системы идентификации личности по голосу или выявления признаков стресса или лжи [2], системы диагностики заболеваний по изменениям в голосе [3], системы улучшения качества связи, кодирования [4] и сжатия речи [5]. Инновации в области обработки речи и нейронных сетей значительно повысили точность и надежность этих методов, делая их все более востребованными при разработке систем анализа речи.

Материалы и методы

Постановка задачи

Функция автокорреляции измеряет степень сходства сигнала с самим собой, сдвинутым по времени. Для периодических сигналов (а речь в определенной степени периодична, особенно гласные звуки) [6], автокорреляция по-

казывает пики на временах, соответствующих периодам сигнала [7], что позволяет определять частоту основного тона речевого сигнала [8]. Частота основного тона (F_0), представляющая собой скорость вибрации голосовых связок, является важной характеристикой речи [9], определяющей высоту голоса [10].

Оценка F_0 применяется в различных задачах, включая анализ эмоциональной окраски речи, гендерную идентификацию [11], сегментацию многоголосых аудиозаписей и в клинической фонологии для диагностики дисфоний и других нарушений голосовой функции [12].

Анализ частоты основного тона по функции автокорреляции

Функция автокорреляции вычисляется для определенного фрагмента речевого сигнала согласно следующему уравнению:

$$R(\tau) = \sum_{i=0}^n [x(t_n) \cdot x(t_n + \tau)], \quad (1)$$

где $R(\tau)$ — значение автокорреляции при сдвиге τ ; n — количество всех точек выбранного фрагмента, по которым ведется суммирование; $x(t_n + \tau)$ — значение сигнала в момент времени $t_n + \tau$; $x(t_n)$ — значение сигнала в момент времени t_n .

Представление частоты основного тона речевого сигнала с помощью функции автокорреляции в Matlab 2017b приведено ниже (рис. 1). В качестве речевого сигнала использовался фрагмент речи диктора (слово «положиться»), испытывающего злость.

Основные биомаркеры эмоционального состояния на графике частоты основного тона речевого сигнала:

- высокая средняя F0: часто указывает на возбуждение, радость, страх или гнев (активные эмоции часто связаны с повышением F0);
- низкая средняя F0: может указывать на грусть, усталость или депрессию (пассивные эмоции часто связаны с понижением F0);

– большая дисперсия F0 (широкий диапазон колебаний): может указывать на возбуждение, интерес или нестабильность эмоционального состояния (повышенная эмоциональная вовлеченность часто приводит к большему колебанию высоты голоса);

– маленькая дисперсия F0 (узкий диапазон колебаний): может указывать на спокойствие, безразличие или подавленность (сниженная эмоциональная вовлеченность может приводить к монотонности голоса);

– резкие скачки F0: могут указывать на удивление, страх или гнев;

– плавные изменения F0: могут указывать на спокойствие, грусть или задумчивость [9].

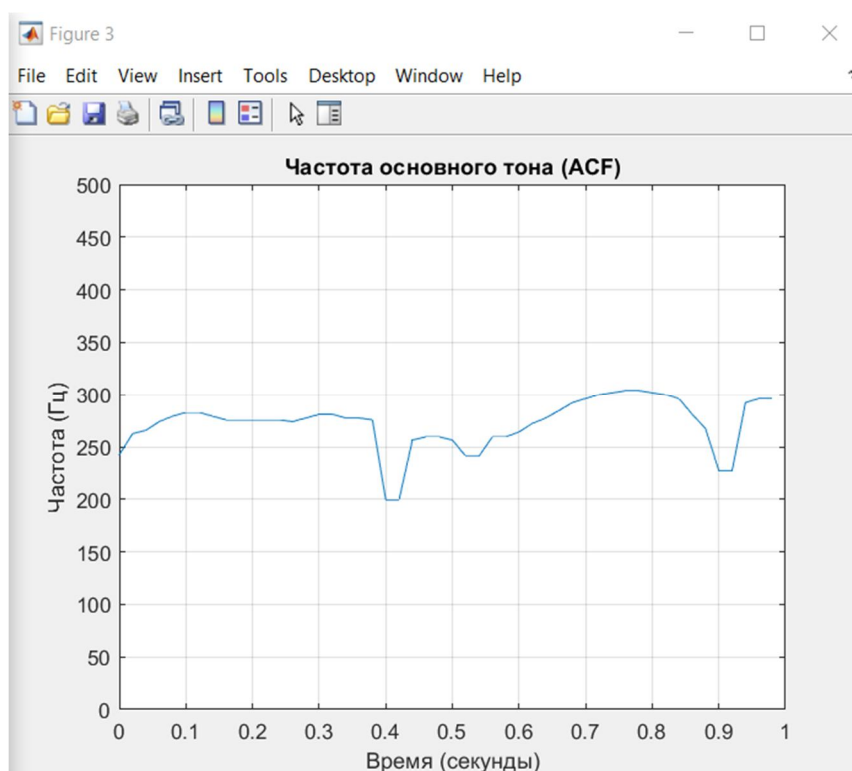


Рис. 1. Представление частоты основного тона речевого сигнала с помощью функции автокорреляции в Matlab

Fig. 1. Representation of the fundamental frequency of a speech signal using the autocorrelation function in Matlab

Результаты и их обсуждение

Обработка экспериментальных данных на основе различных моделей речевого сигнала. Частотный анализ речевого сигнала

Спектрограмма – это визуальное представление частотного состава сигнала во времени [1]. Она отображает, какие частоты присутствуют в сигнале в каждый момент времени, и их относительную интенсивность. Обычно отображается в виде изображения, где:

- ось X: время;
- ось Y: частота;
- цвет (или оттенок серого): интенсивность частоты (чем ярче цвет, тем больше интенсивность).

Применение:

- спектрограммы позволяют визуально анализировать речевой сигнал, различать фонемы, определять границы слов и фраз, выявлять особенности произношения;

- речевая диагностика включает использование спектрограмм для анализа речи у людей с различными нарушениями речевого аппарата;

- спектрограммы также применяются в системах автоматического распознавания речи [9].

Представление сигнала во временно-частотной области выполняется с помощью Фурье-анализа (FT view): отсчёты сигнала $X[k]$ на отдельных кадрах умножаются на оконную функцию и затем преобразуются с помощью дискретного преобразования Фурье, в результате чего получается последова-

тельность кратковременных комплексных спектров [1]:

$$\{X[k]\} \rightarrow DFT \rightarrow \{X(n, k)\}, \quad (2)$$

где k – индекс кадра; n – индекс частоты n -й полосы спектра f_n .

Последовательность отсчётов спектра $\{X(n, k)\}$ можно рассматривать как комплексный сигнал, расположенный в частотной полосе f_n и дискретизированный с частотой, соответствующей переключению кадров.

Рассмотрим спектральное представление речевого сигнала в Matlab (рис. 2). В качестве речевого сигнала использовался фрагмент речи диктора (слово «удерживать»), находящегося в нейтральном эмоциональном состоянии.

Основные биомаркеры эмоционального состояния на спектрограмме речевого сигнала свидетельствуют о том, что:

- более яркие и широкие форманты (резонансные частоты): могут указывать на более громкую речь, связанную с возбуждением, гневом или радостью (форманты формируются речевым трактом и отражают артикуляцию звуков);

- увеличение энергии на высоких частотах: может указывать на возбуждение, страх или агрессию;

- уменьшение энергии на высоких частотах: может указывать на грусть, усталость или депрессию;

- изменение длительности и четкости гласных и согласных: эмоциональ-

ное состояние может влиять на то, как человек произносит звуки (например, при гневe гласные могут быть более короткими и резкими);

– наличие «вокальных дрожаний» (вокализаций, не относящихся к речи): могут указывать на сильные эмоции, такие как плач или смех [9].

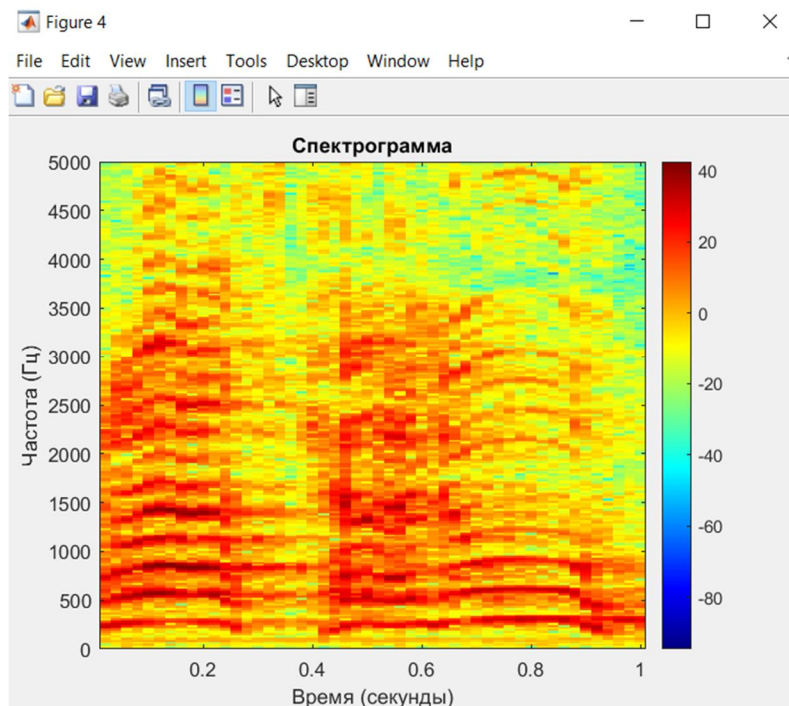


Рис. 2. Спектральное представление речевого сигнала в Matlab

Fig. 2. Spectral representation of a speech signal in Matlab

Расчет модели на основе частотно-временного вейвлет-анализа

Вейвлет-анализ, как и спектрограмма, является методом частотно-временного анализа, но использует вейвлеты вместо синусоид, как в преобразовании Фурье. Вейвлеты – это короткие колебания, локализованные во времени и частоте. Разные вейвлеты хорошо подходят для анализа разных типов сигналов [12].

Вейвлет-анализ обеспечивает лучшее разрешение по времени для высокочастотных компонентов сигнала и лучшее разрешение по частоте для низ-

кочастотных компонентов. Это позволяет более точно анализировать переходные процессы и низкочастотные характеристики речевого сигнала. Это особенно важно для анализа речевых сигналов, которые содержат как быстроизменяющиеся звуки (например, взрывные согласные), так и медленно меняющиеся звуки (например, гласные). Также существенным преимуществом вейвлет-анализа является то, что он менее чувствителен к шуму, чем традиционный Фурье-анализ [10].

Вейвлет-преобразование описывается следующей формулой [1]:

$$\Psi_{ab} = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right), \quad (3)$$

где Ψ_{ab} – функция (вейвлет) порождаемая базовым вейвлетом; $\Psi\left(\frac{t-b}{a}\right)$ – базовый вейвлет; b – сдвиг во времени; a – коэффициент масштабирования.

Выбирается базовый вейвлет, который является «материнским» вейвлетом. Существуют различные типы вейвлетов (например, вейвлет Морле или вейвлет Хаара), каждый из которых имеет свои особенности. Базовый вейвлет масштабируется (сжимается / растягивается) и сдвигается по времени, чтобы создать набор вейвлетов разного масштаба и положения.

Масштабирование позволяет анализировать сигнал в разных частот-

ных диапазонах, а сдвиг – в разных временных точках. Далее сигнал свертывается с каждым вейвлетом из набора. Результат свертки показывает степень соответствия сигнала и вейвлета в данной временной точке и частотном диапазоне. Результат вейвлет-преобразования обычно представляется в виде масштаб-граммы, которая аналогична спектрограмме, но отображает коэффициенты вейвлет-преобразования вместо спектральных амплитуд.

Вейвлет-анализ речевого сигнала в Matlab приведен ниже (рис. 3). В качестве речевого сигнала использовался фрагмент речи диктора (слово «устойчиво»), находящегося в нейтральном эмоциональном состоянии.

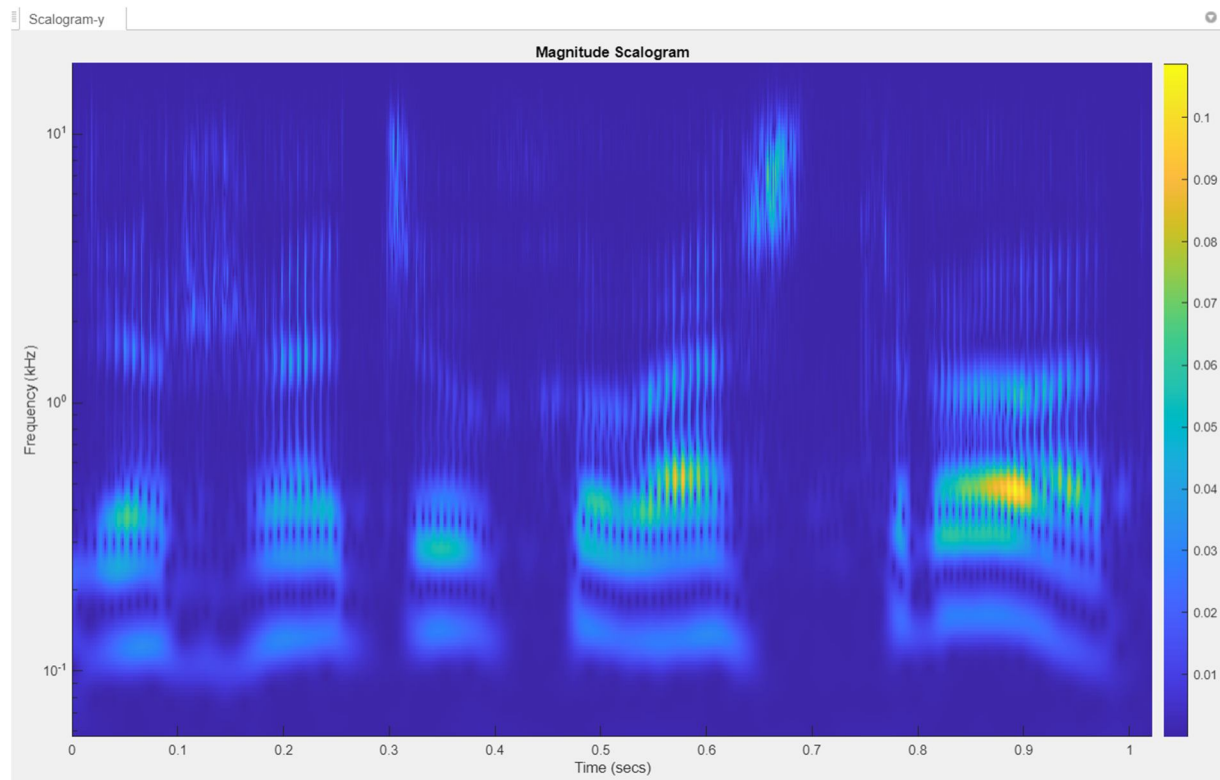


Рис. 3. Вейвлет-анализ (вейвлет Морле) речевого сигнала в Matlab

Fig. 3. Wavelet analysis (Morlet wavelet) of a speech signal in Matlab

Основные биомаркеры эмоционального состояния на вейвлет-масштабограмме речевого сигнала:

- общая вейвлет-энергия: изменения общей вейвлет-энергии могут быть связаны с уровнем возбуждения и общей выразительностью эмоций (более высокие значения часто соответствуют более энергичным эмоциям, таким как радость или гнев, а более низкие – грусти или спокойствию);

- вейвлет-энергия в низкочастотном диапазоне (обычно ниже 500 Гц): изменения вейвлет-энергии на низких частотах могут указывать на изменения в ритме речи, интонации и тембре голоса, которые являются важными индикаторами эмоций (например, увеличение энергии на низких частотах может быть связано с усилением напряжения или стрессом, а снижение – с подавленностью или грустью);

- вейвлет-энергия в среднечастотном диапазоне (обычно от 500 Гц до 2кГц): изменения вейвлет-энергии в среднечастотном диапазоне могут быть связаны с изменениями в форме вокального тракта, которые, в свою очередь, могут быть вызваны эмоциональным напряжением (например, при страхе или стрессе наблюдается смещение формант вверх);

- вейвлет-энергия в высокочастотном диапазоне (обычно выше 2кГц): изменения вейвлет-энергии в высокочастотном диапазоне могут быть связаны с артикуляционными усилиями и уров-

нем напряженности, что отражает эмоциональное возбуждение (например, усиление энергии на высоких частотах может наблюдаться при гневе или страхе) [10].

Анализ речевого сигнала на основе построения аттракторов

При проведении аттракторного анализа речевой сигнал рассматривается как динамическая система. При этом полагают, что речевой сигнал генерируется сложной, но детерминированной (хотя и, возможно, хаотической) системой. В результате проведения исследований требуется восстановить фазовое пространство этой системы и идентифицировать аттракторы. Аттрактор – это состояние, к которому система стремится с течением времени [13]. Форма и структура аттрактора предоставляет исследователю информацию о динамике речевого сигнала и его характеристиках [14].

Построение аттрактора речевого сигнала:

1. Исходный одномерный речевой сигнал преобразуется в многомерное фазовое пространство с использованием метода задержек (теорема Флориса Такенса о вложении с задержкой). Это означает, что создаются новые временные ряды, состоящие из значений исходного сигнала, взятых с определенной задержкой:

$$\{x(t), x(t + \tau), x(t + 2\tau), \dots, x(t + (d - 1)\tau)\}, \quad (4)$$

где τ – временная задержка (delay time); d – размерность вложения (embedding dimension). Выбор оптимальных значений τ и d является важной задачей при моделировании аттрактора [15].

2. Идентификация аттрактора: в реконструированном фазовом пространстве выполняется поиск областей, к которым траектории системы (представленные точками в фазовом пространстве) стремятся. Эти области и являются аттракторами¹.

3. Анализ характеристик аттрактора: рассчитываются различные ха-

рактеристики аттрактора, например, размерность аттрактора отражает сложность динамической системы. Фрактальная размерность хаотического аттрактора свидетельствует о наличии хаотических процессов в генерации речи [16].

Представление речевого сигнала с помощью аттрактора в Matlab приведено ниже (рис. 4).

В качестве речевого сигнала использовался фрагмент речи диктора (слово «королю»), испытывающего злость.

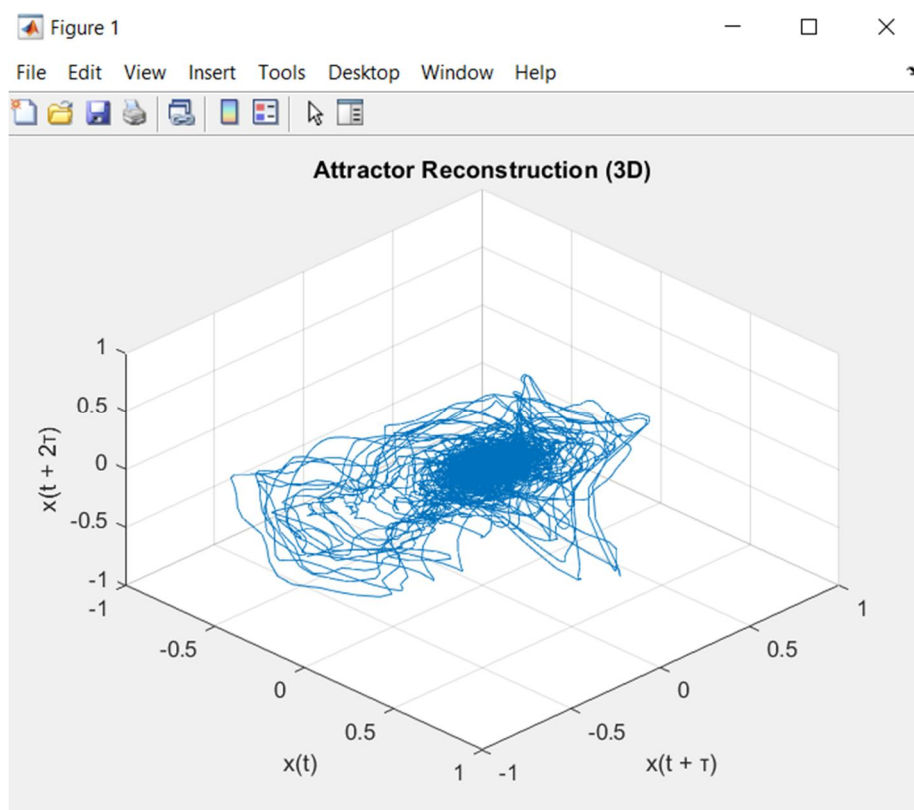


Рис. 4. Представление речевого сигнала с помощью аттрактора в Matlab

Fig. 4. Representation of a speech signal using an attractor in Matlab

¹ Новиков Е. Д. Система анализа психофизиологического состояния человека по паралингвистическим признакам речи на основе методов explainable AI: дис. ... канд. техн. наук. Казань, 2022. 167 с.

Биомаркеры эмоционального состояния на аттракторе речевого сигнала:

- размер и форма аттрактора: аттрактор, соответствующий спокойному состоянию, может быть меньше и более регулярным, в то время как аттрактор, соответствующий возбужденному или стрессовому состоянию, может быть больше и более хаотичным [17];

- сложность аттрактора: более сложные и нерегулярные аттракторы могут указывать на более нестабильные или интенсивные эмоции;

- фрактальная размерность аттрактора: фрактальная размерность может быть использована для количественной оценки сложности аттрактора (более высокая фрактальная размерность может указывать на более сложную динамику, связанную с определенными эмоциями) [18];

- переходы между аттракторами: анализ того, как система переходит между разными аттракторами¹, может дать информацию о переходе между разными эмоциональными состояниями [19].

Выводы

Современные исследования в области обработки сигналов открывают новые горизонты для изучения сложных явлений, и речевой сигнал человека является одним из наиболее информативных источников. Для глубокого и все-

стороннего анализа речевых данных, особенно в контексте определения психоэмоционального состояния человека, активно применяются разнообразные методы обработки сигналов.

Применение таких методов анализа речевого сигнала в Matlab, как анализ частоты основного тона, спектральный анализ, частотно-временной вейвлет-анализ и аттракторный анализ, позволяет получить разностороннюю информацию о речевом сигнале и его характеристиках.

При этом следует отметить, что каждый метод выделяет определенные аспекты сигнала, а комбинирование этих методов позволяет перейти от фрагментарного понимания к целостному анализу сигнала. Такой комплексный подход обеспечивает получение максимально полного представления о параметрах речевого сигнала.

Считаем, что наиболее информативным для первичного анализа речи, особенно с точки зрения выявления глубинных изменений, связанных с психоэмоциональным состоянием, является аттракторный анализ речевого сигнала. Этот метод позволяет выявлять тонкие, нелинейные особенности, которые могут быть упущены при использовании только традиционных частотных или временных характеристик.

Аттракторный анализ может служить мощным инструментом для первичной дифференциации состояний, после чего другие методы могут быть применены для уточнения полученных результатов.

¹ Новиков Е. Д. Система анализа психофизиологического состояния человека по паралингвистическим признакам речи на основе методов explainable AI: дис. ... канд. техн. наук. Казань, 2022. 167 с.

Список литературы

1. Bhatt S., Jain A., Dev A. Continuous Speech Recognition Technologies – A Review // Recent Developments in Acoustics. Lecture Notes in Mechanical Engineering / ed. by M. Singh, Y. Rafat. Singapore: Springer, 2021. P. 85–94. https://doi.org/10.1007/978-981-15-5776-7_8
2. Иванов А. А., Смирнова К. Д. Классификация дикторов по полу и возрасту на основе сверточных нейронных сетей и мел-частотных кепстральных коэффициентов // Искусственный интеллект и принятие решений. 2022. № 3. С. 45-56.
3. Mohammed A. A., Nagy A. Fundamental Frequency and Jitter Percent in MDVP and PRAAT // J. Voice. 2023. Vol. 37, N 4. P. 496–503.
4. Панченко А. А., Кравчук Д. А. Диагностика психофизиологического состояния человека на основе речевого сигнала // XXXV сессия Российского акустического общества: сборник трудов конференции. М.: ГЕОС, 2023. С. 112-117. <https://doi.org/10.34756/GEOS.2023.17.38436>
5. Кравчук Д. А. Ультразвуковая система контроля психофизиологического состояния машиниста поезда // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2020. Т. 10, № 1. С. 134-142.
6. Мырадов М. Т., Назарова С. Н. Современные технологии распознавания речи // Наука и мировоззрение. 2024. № 1 (23). С. 301–305.
7. Катаев М. Ю. Методика распознавания речевых команд в школьных информационных системах // Речевые технологии. 2024. № 1. С. 18–34.
8. Кравчук Д. А. Разработка метода распознавания эмоций по речевому сигналу // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2024. Т. 14, № 2. С. 72-80. <https://doi.org/10.21869/2223-1536-2024-14-2-72-80>
9. Петров А. С., Козлов А. В. Алгоритм оценки частоты основного тона на основе адаптивной спектральной обработки для задач биометрической идентификации // Информатика, управление и системы связи. 2023. № 1(15). С. 45-54.
10. Семенова Е. А., Иванов П. К. Акустические корреляты дисфонии у профессиональных пользователей голоса: результаты скринингового исследования с использованием мобильных приложений // Вестник оториноларингологии. 2022. Т. 87, № 3. С. 33-40.
11. Калинина М. А., Белова А. Н. Анализ параметров стабильности голоса у детей с двигательными нарушениями с использованием методов машинного обучения // Детская и подростковая реабилитация. 2021. № 4(47). С. 28-35.
12. Методы предварительной обработки речевых сигналов / В. В. Козлов, А. А. Трофимов, Е. А. Фокина, В. Н. Пономарев, Т. О. Жуков // Измерение. Монито-

ринг. Управление. Контроль. 2023. № 4. С. 43–49. <https://doi.org/10.21685/2307-5538-2023-4-5>

13. Горшков Ю. Г. Визуализация эмоциональной напряженности человека по речевому сигналу // Научная визуализация. 2023. № 2. С. 102–112.

14. Смирнов И. П., Орлова Ю. С. Применение методов реконструкции фазового пространства для анализа нелинейных динамических систем в биоакустике // Нелинейный мир. 2021. Т. 19, № 4. С. 215–225.

15. Федоров Р. А., Соколова Д. И. Классификация эмоциональных состояний по речи с использованием ансамбля признаков нелинейной динамики и глубоких нейронных сетей // Искусственный интеллект и принятие решений. 2023. № 2. С. 74–85.

16. Qin Ji. Emotional state analysis based on speech recognition signals // Highlights in Science Engineering and Technology. 2022. N 23. P. 58–64.

17. Multi-Modal Speech Emotion Recognition with Temporal-Perceptual Features and Hybrid Deep Learning Models / Y. Zhao, Z. Wang, J. Zhang, B. Schuller // IEEE Transactions on Affective Computing. 2024. Vol. 15, is. 1. P. 234–247.

18. Khan M., Kwon S. A CNN-Assisted Enhanced Audio Signal Processing for Speech Emotion Recognition // Sensors. 2020. Vol. 20, is. 1. P. 183.

19. Pepino L., Riera P., Ferrer L. Emotion Recognition from Speech Using wav2vec 2.0 Embeddings. URL: <https://arxiv.org/abs/2104.03502> (дата обращения: 17.12.2025).

References

1. Bhatt S., Jain A., Dev A. Continuous Speech Recognition Technologies – A Review. Recent Developments in Acoustics. Lecture Notes in Mechanical Engineering. Singapore: Springer; 2021. P. 85–94. https://doi.org/10.1007/978-981-15-5776-7_8

2. Ivanov A.A., Smirnova K.D. Speaker gender and age classification based on convolutional neural networks and mel-frequency cepstral coefficients. *Iskusstvenny`j intellekt i prinyatie reshenij = Artificial Intelligence and Decision Making*. 2022;(3):45–56. (In Russ.)

3. Mohammed A.A., Nagy A. Fundamental Frequency and Jitter Percent in MDVP and PRAAT. *J. Voice*. 2023;37(4):496–503.

4. Panchenko A.A., Kravchuk D.A. Diagnosis of a person's psychophysiological state based on the speech signal. In: *XXXV sessiya Rossiiskogo akusticheskogo obshchestva: sbornik trudov konferentsii = XXXV session of the Russian Acoustic Society: Proceedings of the conference*. Moscow: GEOS; 2023. P. 112–117. (In Russ.) <https://doi.org/10.34756/GEOS.2023.17.38436>

5. Kravchuk D.A. Ultrasonic system for monitoring the psychophysiological state of a train driver. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naja tekhnika, informatika. Meditsinskoe priborostroenie = Proceedings of the*

Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering. 2020;10(1):134-142. (In Russ.)

6. Myradov M.T., Nazarova S.N. Modern speech recognition technologies. *Nauka i mirovozzrenie = Science and Worldview.* 2024;(1):301–305. (In Russ.)

7. Kataev M.Yu. Methodology for recognizing speech commands in school information systems. *Rechevy'e texnologii = Speech Technologies.* 2024;(1):18–34. (In Russ.)

8. Kravchuk D.A. Development of a method for emotion recognition from a speech signal. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naja tekhnika, informatika. Meditsinskoe priborostroenie = Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering.* 2024;14(2):72-80. (In Russ.) <https://doi.org/10.21869/2223-1536-2024-14-2-72-80>

9. Petrov A.S., Kozlov A.V. A fundamental frequency estimation algorithm based on adaptive spectral processing for biometric identification tasks. *Informatics. Informatika, upravlenie i sistemy` svyazi = Computer Science, Control and Communication Systems.* 2023;(1):45-54. (In Russ.)

10. Semenova E.A., Ivanov P.K. Acoustic correlates of dysphonia in professional voice users: results of a screening study using mobile applications. *Vestnik otorinolaringologii = Bulletin of Otolaryngology.* 2022;87(3):33-40. (In Russ.)

11. Kalinina M.A., Belova A.N. Analysis of voice stability parameters in children with motor disorders using machine learning methods. *Detskaya i podrostkovaya reabilitaciya = Children's and Adolescent Rehabilitation.* 2021;(4):28-35. (In Russ.)

12. Kozlov V.V., Trofimov A.A., Fokina E.A., Ponomarev V.N., Zhukov T.O. Methods of preliminary processing of speech signals. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measurement. Monitoring. Management. Control.* 2023;(4):43–49. (In Russ.) <https://doi.org/10.21685/2307-5538-2023-4-5>

13. Gorshkov Yu.G. Visualization of human emotional stress from a speech signal. *Nauchnaya vizualizaciya = Scientific Visualization.* 2023;(2):102-112. (In Russ.)

14. Smirnov I.P., Orlova Yu.S. Application of phase space reconstruction methods for the analysis of nonlinear dynamic systems in bioacoustics. *Nelinejny`j mir = Nonlinear World.* 2021;19(4):215-225. (In Russ.)

15. Fedorov R.A., Sokolova D.I. Classification of emotional states from speech using an ensemble of nonlinear dynamics features and deep neural networks. *Iskusstvenny`j intellekt i prinyatie reshenij = Artificial Intelligence and Decision Making.* 2023;(2):74-85. (In Russ.)

16. Qin Ji. Emotional state analysis based on speech recognition signals. *Highlights in Science Engineering and Technology.* 2022;(23):58-64.

17. Zhao Y., Wang Z., Zhang J., Schuller B. Multi-Modal Speech Emotion Recognition with Temporal-Perceptual Features and Hybrid Deep Learning Models. *IEEE Transactions on Affective Computing*. 2024;15:234-247.
18. Khan M., Kwon S. A CNN-Assisted Enhanced Audio Signal Processing for Speech Emotion Recognition. *Sensors*. 2020;20:183.
19. Pepino L., Riera P., Ferrer L. Emotion Recognition from Speech Using wav2vec 2.0 Embeddings. Available at: <https://arxiv.org/abs/2104.03502> (accessed 17.12.2025).

Информация об авторах / Information about the Authors

Кравчук Денис Александрович, доктор технических наук, профессор, Институт нанотехнологий, электроники и приборостроения, Южный федеральный университет, г. Таганрог, Российская Федерация, e-mail: kravchukda@sfedu.ru, ORCID: 0000-0002-4220-3928

Denis A. Kravchuk, Doctor of Sciences (Engineering), Professor, Institute of Nanotechnology, Electronics and Instrumentation, Southern Federal University, Taganrog, Russian Federation, e-mail: kravchukda@sfedu.ru, ORCID: 0000-0002-4220-3928

Котенко Александр Сергеевич, студент, Институт нанотехнологий, электроники и приборостроения, Южный федеральный университет, г. Таганрог, Российская Федерация, e-mail: akotenko@sfedu.ru

Alexander S. Kotenko, Student, Institute of Nanotechnology, Electronics and Instrumentation, Southern Federal University, Taganrog, Russian Federation, e-mail: akotenko@sfedu.ru