

УДК 004.89

DOI: [10.26102/2310-6018/2026.60.9.011](https://doi.org/10.26102/2310-6018/2026.60.9.011)

Применение нейросетевых методов идентификации характеристик речевого потока на примере построения цифровых профилей пользователей финансовых услуг

В.Ю. Кузнецова, К.Д. Кузовлев✉

*Астраханский государственный технический университет, Астрахань,
Российская Федерация*

Резюме. В статье рассматривается применение нейросетевых методов для идентификации характеристик речевого потока с целью построения расширенных цифровых профилей пользователей финансовых услуг. Актуальность исследования обусловлена развитием дистанционных форм банковского и микрофинансового обслуживания, при которых речевая коммуникация становится одним из значимых источников информации о клиенте. Описаны основные просодические характеристики речи, включая темп, громкость, высоту тона, паузы и ритмику, а также рассмотрена их связь с психологическими типами личности по методике «7 радикалов» В.В. Пономаренко. Особое внимание уделено нейросетевому подходу к анализу речевого сигнала: предварительной цифровой обработке аудиозаписи, выделению акустических признаков, формированию вектора просодических параметров и последующей классификации психологических радикалов с использованием современных архитектур CNN, LSTM и механизмов внимания. Представлена информационная технология принятия решений в банковской и микрофинансовой сферах, предусматривающая включение модуля анализа речевой информации в систему построения цифрового профиля клиента. Показано, что использование вероятностного подхода и многокритериального анализа позволяет повысить объективность оценки личности, снизить субъективную неопределенность при взаимодействии с заемщиком и повысить точность кредитного скоринга. Сделан вывод о перспективности применения нейросетевого анализа речевых характеристик для управления финансовыми рисками и персонализации финансовых услуг.

Ключевые слова: нейросетевые методы, речевой поток, просодика, цифровой профиль, финансовые услуги, кредитный скоринг, психологические радикалы, анализ речи, акустические признаки, поддержка принятия решений.

Для цитирования: Кузнецова В.Ю., Кузовлев К.Д. Применение нейросетевых методов идентификации характеристик речевого потока на примере построения цифровых профилей пользователей финансовых услуг. *Моделирование, оптимизация и информационные технологии*. 2026;14(9). URL: <https://moitvvt.ru/ru/journal/article?id=2463> DOI: 10.26102/2310-6018/2026.60.9.011

Application of neural network methods for identifying speech flow characteristics in building digital profiles of financial services users

V.Yu. Kuznetsova, K.D. Kuzovlev✉

Astrakhan State Technical University, Astrakhan, the Russian Federation

Abstract. The article examines the application of neural network methods for identifying speech flow characteristics in order to build enhanced digital profiles of financial services users. The relevance of the study is driven by the development of remote forms of banking and microfinance services, in which speech communication becomes one of the significant sources of information about the client. The main prosodic characteristics of speech are described, including speech rate, loudness, pitch, pauses, and rhythm, as well as their relationship with psychological personality types according to V.V. Ponomarenko's "7 radicals" methodology. Special attention is paid to the neural network approach

to speech signal analysis: preliminary digital processing of the audio recording, extraction of acoustic features, formation of a prosodic parameter vector, and subsequent classification of psychological radicals using modern architectures such as CNN, LSTM, and attention mechanisms. An information technology for decision-making in the banking and microfinance sectors is presented, which involves incorporating a speech information analysis module into the client's digital profile building system. It is shown that the use of a probabilistic approach and multicriteria analysis makes it possible to improve the objectivity of personality assessment, reduce subjective uncertainty in interaction with the borrower, and increase the accuracy of credit scoring. A conclusion is made about the prospects of using neural network analysis of speech characteristics for financial risk management and personalization of financial services.

Keywords: neural network methods, speech stream, prosody, digital profile, financial services, credit scoring, psychological radicals, speech analysis, acoustic features, decision support.

For citation: Kuznetsova V.Yu., Kuzovlev K.D. Application of neural network methods for identifying speech flow characteristics in building digital profiles of financial services users. *Modeling, Optimization and Information Technology*. 2026;14(9). (In Russ.). URL: <https://moitvvt.ru/ru/journal/article?id=2463> DOI: 10.26102/2310-6018/2026.60.9.011

Введение

Цифровизация в России демонстрирует устойчивый рост: вклад ИКТ-сектора в ВВП страны в 2023 году составил 3,5 %, а оборот ИТ-отрасли в 2024 г. достиг 4,5 трлн руб. Активно развиваются госуслуги (109 млн пользователей) и региональная цифровая трансформация (лидеры – Белгородская обл., Татарстан, ЯНАО). В 2024 г. затраты на цифровую экономику превысили 6,7 трлн рублей 88 регионов утвердили программы цифровой трансформации на 2026–2028 годы, при этом лидерами по уровню цифровизации в 2024 году стали Белгородская область (94,22 %), Татарстан (91,35 %) и ЯНАО (90,79 %). Ключевым звеном процесса цифровизации выступает персонализация оказания услуг, которая обеспечивается за счет построения цифровых профилей пользователей систем.

Процесс принятия решений в системах, частью которых является человек, основывается на сборе, обработке и анализе большого количества неструктурированных и зачастую противоречивых данных, что вносит в процесс принятия решения субъективную неопределенность, связанную с особенностями поведения людей. Научная задача формализации поведения и предсказания поведения человека неоднократно претерпевала попытки к решению в части моделирования частных ситуаций принятия решений, где учитываются отдельные характеристики субъектов, в том числе без использования средств автоматизации. Например, достаточно большое количество работ в криминалистике посвящено построению психологических профилей преступника для раскрытия преступлений [1], цифровые профили пациентов строятся с помощью нейросетевых цифровых профилей на базе текстовой информации в медицине для предсказания негативных последствий для пациента [2], в сфере принудительного взыскания на базе набора данных о должниках прогнозируется эффективность возврата задолженностей в коллекторском бизнесе [3]. Большой пул задач с помощью расширенных цифровых профилей решаются в сфере микрофинансирования и кредитного скоринга, когда от полноты информации о заявителе зависит результативность кредитного портфеля банка или микрофинансовой организации (МФО) [4].

Субъективная неопределенность при принятии решения в отношении субъекта, в частности, при выдаче кредита, возникает из-за особенностей восприятия, знаний, опыта и психологических характеристик человека – как того, кто принимает решение со

стороны банка, так и того, в отношении которого принимается решение. В коммуникации субъективная неопределенность проявляется как когнитивная (незнание взглядов и установок исследуемого субъекта) и поведенческая (неспособность предсказать его поведение).

Большинство предлагаемых современными исследователями подходов предполагает сбор разрозненных данных, которые в совокупности позволяют построить некий цифровой профиль актора, на основе которого прогнозируется его предполагаемое поведение. К таким данным могут быть отнесены как его вербальные (внешний вид), так и невербальные характеристики (манера речи), а также его поведение в социальных сетях. К таким работам относятся статьи авторов В. Евсюкова, А. Масютина, С. Скибы, P. Yildirim, E. Aghasian, S. Garg, J. Montgomery, M. Óskarsdóttira, C. Sarrautes, J. Vanthienen, B. Baesens и др. Имеется отдельный пласт академических работ и статьи практической направленности, посвященные анализу просодических характеристик речи для установления кредитного риска, к ним относятся работы Chen X., Yu X., Yang Y., Qin Y., в рамках которых анализируются отдельные признаки, такие как тон, тембр, интонация. При этом ранее описанные методы достаточно полно отражают определение психологического типа актора с точки зрения его внешнего вида в автоматизированном режиме, но практически не позволяют произвести оценку речи и ее особенности, в том числе при дистанционном взаимодействии посредством телефонного разговора, в то время как такой вид коммуникации является наиболее распространенным в организациях, в частности, при кредитном и банковском обслуживании.

Материалы и методы

В современной психологии и смежных дисциплинах все большее внимание уделяется не только содержанию речи, но и ее просодическим характеристикам – темпу, громкости, интонации, паузам. Просодика – это совокупность ритмико-интонационных и динамических особенностей устной речи, которые не связаны напрямую со значением слов, но несут важную информацию о говорящем. В последние десятилетия просодические методы активно используются для диагностики и распознавания психологического типа личности, а также для выявления эмоционального состояния, темперамента, которые в совокупности оказывают существенное влияние на принятие решений.

Просодика включает в себя следующие характеристики речи:

- темп речи (скорость произнесения слов и фраз);
- громкость (интенсивность звукового сигнала);
- высоту тона (мелодику, интонацию);
- паузы (длительность и распределение молчания);
- ритмику (равномерность или неравномерность речи).

Эти параметры не зависят от лексики, но тесно связаны с физиологией, эмоциональным состоянием и индивидуальными особенностями говорящего. Данные характеристики речи достаточно полно отражаются в методике «7 радикалов» В.В. Пономаренко для определения психологического типа личности субъекта [5] (Таблица 1).

Просодические методы достаточно точны, однако важно учитывать, что на речь субъекта могут оказывать влияние усталость, болезнь, культурные особенности среды, в которой происходит коммуникация, поэтому применение исключительно просодических методов рекомендуется в совокупности с другими характеристиками психологического профиля субъекта.

Просодические характеристики речи могут изменяться под влиянием усталости, заболевания, эмоционального состояния и индивидуальных речевых особенностей, поэтому их абсолютные значения не рассматриваются как достаточное основание для классификации. Для снижения влияния этих факторов используется индивидуальная калибровка: при первичном взаимодействии показатели сопоставляются с эталонными диапазонами, а при наличии истории взаимодействия – со средними значениями ранее обработанных речевых фрагментов пользователя. Это позволяет оценивать отклонение темпа, интенсивности, основного тона, паузации и ритмики относительно индивидуального речевого профиля.

Таблица 1 – Характеристики речи, определяющие психологический тип личности

Table 1 – Characteristics of speech that determine the psychological type of personality

Преобладающий радикал	Просодические особенности речи	Темп	Громкость	Высота тона	Паузы	Ритмика
паранойяльный	уверенная, чёткая, часто менторский тон, жёсткие формулировки, последовательное изложение («во-первых», «в-третьих»)	средний, быстрый	средняя, высокая	властная интонация	редкие	равномерная
истероидный	эмоциональная, экспрессивная, речь театрализована	переменный, часто ускоряется	высокая	яркая, с преувеличенными модуляциями	драматические паузы	равномерная
эпилептоидный	речь вялая	медленная	умеренная или низкая	однообразная	длинные	монотонная
шизоидный	высокоинтеллектуальная, с обилием терминологии	любой	средняя	«отстранённая», без эмоциональных акцентов	неожиданные паузы	неравномерная
гипертимный	увлечённая, быстрая, энергичная	быстрый	изменяющаяся	интонация живая, эмоционально окрашенная	короткие	сбивчивая
эмотивный	речь осторожная	медленный или средний	тихая	сочувствующая, без резких акцентов	частые	мягкая, плавная речь
тревожный	тихая, неуверенная, сбивчивая речь	замедленный	низкая	интонация робкая, с извиняющимися нотками	частые и нерешительные	сбивчивая

Дополнительно учитывается качество исходного сигнала: при высоком уровне шума, низкой разборчивости или недостаточной продолжительности фрагмента достоверность просодической оценки снижается, а неоднозначные случаи требуют дополнительного речевого материала либо экспертной проверки. Влияние культурных и

речевых особенностей учитывается при формировании обучающей выборки за счет включения разнообразных речевых образцов и последующей калибровки диапазонов просодических характеристик.

Распознавание просодических характеристик речи с помощью нейросетевого подхода. Нейросетевой подход стал стандартом в распознавании просодических характеристик речи благодаря своей способности выявлять сложные, неформализуемые паттерны, обучаться на реальных данных, адаптироваться к разным условиям и обеспечивать высокую точность анализа¹. Это связано с тем, что характеристики речи – это не просто набор параметров, а сложные, часто неявные признаки, которые трудно описать строгими правилами. Традиционные алгоритмы (например, на основе правил или статистических моделей) плохо справляются с выявлением индивидуальных особенностей речи [6]. Нейросети же способны выявлять сложные зависимости и паттерны, которые невозможно формализовать вручную [7]. Кроме того, нейросетевые модели легко адаптируются к новым языкам, акцентам и индивидуальным особенностям речи, что делает их универсальным инструментом для анализа просодики в мультикультурной среде.

С целью распознавания просодических характеристик речи, представленной в Таблице 1, целесообразно использовать нейросетевой подход, базирующийся на анализе акустических параметров речевого сигнала. Современные методики машинного обучения позволяют автоматически выделять скрытые закономерности в человеческой речи и привязывать их к определенным психологическим характеристикам его личности, в отличие от классических алгоритмов обработки сигнала, которые работают с отдельными, не зависимыми друг от друга значениями. Так, например, нейронные сети способны учитывать сложные взаимосвязи между темпом речи, интонацией, громкостью, длительностью пауз и ритмикой, что особенно важно для построения цифрового профиля субъекта [8].

С программно-технической точки зрения нейросеть не «слышит» интонацию так, как человек. Она получает на вход числовое представление аудиосигнала и обучается находить в нем закономерности, соответствующие просодическим признакам: высоте голоса, темпу, громкости, паузам, ритму, ударениям и мелодическому контуру речи. Поэтому распознавание просодии строится как последовательный вычислительный процесс: от цифровой обработки сигнала к извлечению признаков и далее к их интерпретации нейросетевой моделью.

Распознавание просодических особенностей речи с помощью нейросетевых технологий осуществляется путем автоматического разложения акустических параметров речевого сигнала. Нейронная сеть получает на вход цифровую аудиозапись $x(t)$, представляющую собой изменение амплитуды звуковой волны во времени. На начальном этапе предварительной обработки сигнал дискретизируется и разбивается на короткие (20–40 мс) временные окна, что позволяет анализировать речь в виде последовательности акустических признаков. Алгоритм определения просодических характеристик речи приведен на Рисунке 1 [9].

¹ Abudheen S.K. *AI voice analysis startup BizBaz aims to fix broken credit scoring with 80% accuracy*. LinkedIn. URL: https://www.linkedin.com/posts/sainul-abudheen-k-a9430328_ai-voice-analysis-startup-bizbaz-aims-to-activity-7201526093336465408-QMA- (дата обращения: 19.08.2026).

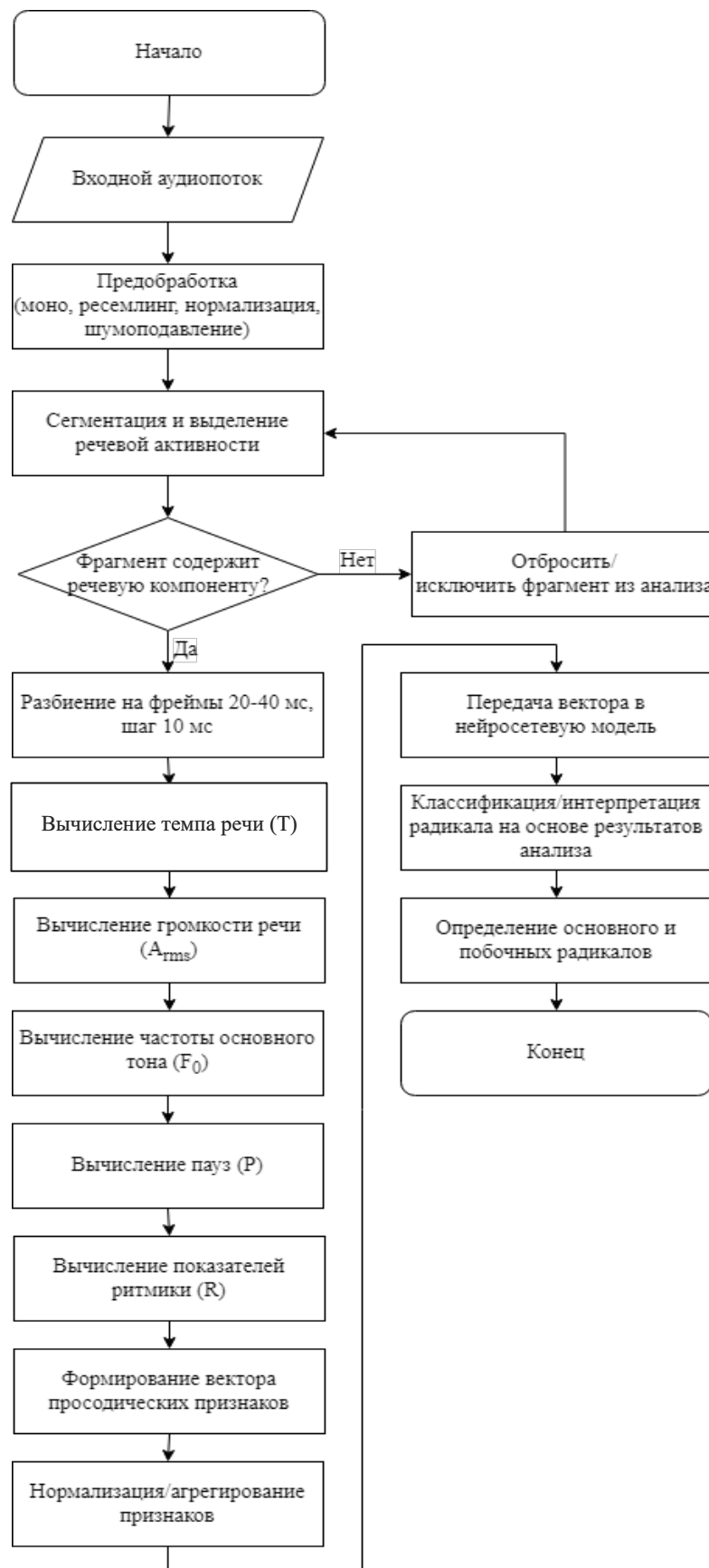


Рисунок 1 – Алгоритм определения просодических характеристик речи
Figure 1 – Algorithm for determining prosodic characteristics of speech

На первом этапе аудиосигнал приводится к единому техническому виду. Запись переводится в моноформат, нормализуется по громкости, очищается от шумов, при необходимости ресемплируется, например, до 16 или 22,05 кГц, и делится на короткие перекрывающиеся окна. Обычно анализ ведется не по всей записи сразу, а по фреймам длительностью примерно 20–40 мс с шагом около 10 мс. Формализуется такой фрейм формулой:

$$x_m[n] = x[n + mH]w[n], \quad (1)$$

где $x[n]$ – исходный аудиосигнал; m – номер фрейма; H – шаг окна; $w[n]$ – оконная функция. Такое разбиение нужно потому, что просодические характеристики разворачиваются во времени: высота голоса, интенсивность и темп меняются от фрагмента к фрагменту.

Затем аудио преобразуется во временно-частотное представление. Для этого применяются спектрограммы, мел-спектрограммы или MFCC-признаки. Спектрограмма показывает, как энергия сигнала распределяется по частотам в каждый момент времени, а мел-спектрограмма дополнительно приближает шкалу частот к особенностям человеческого слуха. Математически такое преобразование обычно строится на кратковременном преобразовании Фурье, представленном в формуле:

$$X(m, k) = \sum_{n=0}^{L-1} x_m[n] e^{-j2\pi kn/L}, \quad (2)$$

где $X(m, k)$ – частотное представление m -го фрейма; k – номер частотной компоненты; L – длина окна. Такие представления удобны для нейросетей, поскольку их можно анализировать почти как изображение: по горизонтали располагается время, по вертикали – частоты, а значения отражают энергию сигнала [10].

Среди характеристик речи первым вычисляется темп речи, задаваемый формулой (3), как отношение количества речевых единиц, обработанных нейросетью к длительности сигнала:

$$T = \frac{N}{t}, \quad (3)$$

где T – темп речи, N – количество полученных в результате обработки слогов, фонем или слов и t – длительность речевого фрагмента.

Далее вычисляется громкость речи при помощи среднеквадратичного значения амплитуды сигнала (A_{RMS}), задаваемого формулой:

$$A_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}, \quad (4)$$

где x_i – амплитуда сигнала в момент времени i , N – количество сегментов сигнала.

Высота тона определяется при помощи частоты основного тона (F_0), задаваемого автокорреляционной функцией, представленной на формуле:

$$R(\tau) = \sum_{n=0}^{N-1} x(n)x(n + \tau), \quad (5)$$

где τ – временной сдвиг сигнала. Максимум функции автокорреляции позволяет определить период основного тона, после чего вычисляется частота при помощи формулы:

$$F_0 = \frac{1}{R_{max}(\tau)}. \quad (6)$$

Паузы в речи определяются как интервал времени, в которых амплитуда сигнала ниже порогового значения. Пороговое значение задается либо экспертом, либо предусмотрено моделью нейросети.

Ритмика речи определяется через дисперсию временных интервалов между ударными речевыми сегментами, задаваемых следующей формулой:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (t_i - \bar{t})^2, \quad (7)$$

где t_i – длительность интервала между сегментами речи, $\bar{t}$ – среднее значение интервалов. Чем выше дисперсия, тем менее равномерной является ритмика речи. После вычисления просодических параметров формируется вектор признаков $X = (T, A, F_0, P, R)$, где T – вычисленное значение темпа, A – громкость, F_0 – высота тона, P – параметры пауз, R – показатели ритмики.

В основе обучения модели лежит заранее размеченный набор входных данных: записи речи людей с заранее определенными психологическими типами и соответствующими просодическими характеристиками для каждого радикала индивидуальности [11]. Так, гипертимный тип проявляется в быстром темпе речи, выраженной эмоциональной интонации и коротких паузах, тогда как тревожный – в приглушенном голосе, замедленном ритме и частых нерешительных остановках. По мере обучения нейросеть выстраивает многомерное признаковое пространство, которое затем позволяет с высокой точностью классифицировать речь новых пользователей. При этом конкретные числовые пороги и зависимости между признаками не задаются вручную – они формируются автоматически на основе того массива данных, на котором проходило обучение.

В задачах анализа речевого сигнала могут применяться гибридные архитектуры CNN–LSTM и механизмы внимания, обеспечивающие совместный учет локальных акустических особенностей и их временной динамики [12, 13]. В настоящей работе для обработки просодических характеристик используется более компактная архитектура, ориентированная на заранее выделенный набор интерпретируемых признаков.

Извлечение темпа речи, интенсивности, характеристик основного тона, паузации и ритмики выполняется средствами openSMILE, используемого как фиксированный необучаемый модуль. После нормализации полученный вектор поступает в обучаемый полносвязный блок Dense–ReLU–Dropout, формирующий 64-мерное скрытое представление просодических характеристик. Далее оно передается в классификационный слой с семью выходными нейронами, соответствующими психологическим радикалам, а функция Softmax формирует вероятности их выраженности.

В процессе обучения параметры openSMILE не изменяются, тогда как веса полносвязного и классификационного блоков обновляются методом обратного распространения ошибки. В качестве функции потерь используется категориальная кросс-энтропия, оптимизация выполняется алгоритмом Adam. Для снижения переобучения применяется dropout. Качество модели оценивается по показателям Accuracy, Precision, Recall, F1 и ROC-AUC.

Заключительным результатом работы нейросетевой системы является вектор вероятностей принадлежности речи индивидуальности к конкретным радикалам. Например, система определяет, что речь содержит 43 % признаков первого радикала, 31 % второго и 19 % третьего. Выделяются три основных радикала, остальные же – признаются несущественными и обнуляются. Такой подход позволяет избавиться от жесткой классификации индивидуальности и обеспечить учет смешанного характера психологического профиля человека. Вычисление процентной доли преобладающего радикала, может происходить на основе выходных значений нейронной сети.

Пусть нейросетевая модель сформировала следующие значения активности выходного слоя $z_1 - z_7$. Тогда вероятность принадлежности к каждому радикалу вычисляется по формуле:

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}, \quad (8)$$

где P_i – процентное соотношение i -го радикала. В качестве примера обработки речи возьмем ситуацию, где сеть выдала значения: 3,1, 1,1, 0,6, то после применения многопеременной логистической (софтмакс) функции система определит, что преобладающим является первый радикал с вероятностью около 58 %, второй – 22 %, третий – 12 %.

Результаты

Для реализации информационной технологии принятия решения предлагается к разработке программный комплекс, автоматизирующий указанный процесс. Для желающих получить заем лично обрабатываются вербальные признаки заявителя, для онлайн-клиентов данные с их страниц в социальных сетях [14]. В дополнение к ранее описанным методикам предлагается к реализации дополнительный модуль обработки речевой информации, результаты обработки данных которого дополняют цифровой профиль клиента, расширяя его за счет интерпретируемых сведений о его манере речи, что является удачным решением при дистанционном телефонном обслуживании заявителей. С помощью прецедентного подхода за счет дополнительных параметров о заявителе на получение финансовых услуг СППР банковского скоринга может принять обоснованные решения по вопросу факта и условий выдачи заемных средств [4].

Для обучения модели использовалась размеченная выборка из 1250 аудиофрагментов, разделенная на обучающую и валидационную части в пропорции 70 на 30. Обучение модели проводилось в течение 50 эпох с контролем функции потерь и показателей качества на валидационной выборке. Для предотвращения переобучения использовалась ранняя остановка при отсутствии улучшения валидационной функции потерь в течение 7 последовательных эпох. Дополнительно применялась калибровка просодических характеристик: при первичном взаимодействии значения темпа речи, интенсивности, основного тона, паузации и ритмики сопоставлялись с эталонными диапазонами, а при наличии истории взаимодействия – со средними значениями ранее обработанных речевых фрагментов пользователя, что позволяло учитывать индивидуальную вариативность речи при последующей классификации.

Экспериментальная проверка нейросетевого подхода проводилась на независимой контрольной выборке, не использовавшейся при обучении модели. Контрольная выборка включала 192 аудиофрагмента трех классов: 64 нейтральных, 62 пограничных и 66 фрагментов с выраженными диагностическими признаками. Средняя продолжительность аудиофрагментов составляла соответственно 23,4, 27,1 и 24,8 с. Дополнительно были сформированы стрессовые подвыборки, включавшие 48 аудиофрагментов с высоким уровнем фонового шума и 24 аудиофрагмента с несколькими говорящими. Тестирование предусматривало оценку корректности классификации, устойчивости к шумовым искажениям, работы при наличии нескольких голосов и воспроизводимости результата при повторной обработке одних и тех же аудиофрагментов. По результатам экспериментальной проверки из 64 нейтральных аудиофрагментов правильно классифицировано 58, из 62 пограничных – 49, из 66 фрагментов с выраженными диагностическими признаками – 59. Общая точность классификации (Accuracy) составила 0,865. Для нейтральных аудиоданных значения

Precision, Recall и F1 составили соответственно 0,879, 0,906 и 0,892; для пограничных – 0,845, 0,790 и 0,817; для аудиоданных с выраженными признаками – 0,868, 0,894 и 0,881. Макроусредненное значение F1 составило 0,863, а ROC-AUC macro – 0,931. Полученные результаты показывают, что наиболее устойчиво распознаются крайние классы, тогда как основная зона неопределенности приходится на пограничные аудиофрагменты, характеризующиеся смешанным сочетанием просодических и акустических признаков.

Дополнительная проверка устойчивости показала, что при умеренном ухудшении качества записи модель сохраняет основное качество классификации. Коэффициент шумоустойчивости составил 0,928. Средний коэффициент реального времени RTF составил 0,045, что соответствует обработке аудиоданных приблизительно в 22 раза быстрее естественного времени их звучания. Таким образом, эксперимент подтверждает не только возможность классификации речевых характеристик нейросетевыми методами, но и применимость предложенного подхода к обработке аудиоданных в условиях, приближенных к реальным условиям дистанционного взаимодействия с пользователями.

Сравнительный анализ показал, что при использовании одинакового набора входных просодических признаков точность классификации методом SVM составила 81,2 %, методом Random Forest – 83,3 %, тогда как разработанная нейросетевая модель обеспечила точность 86,5 %. Прирост точности составил соответственно 5,3 и 3,2 процентного пункта. Аналогичная тенденция наблюдается по показателю F1 macro: 0,811 для SVM, 0,832 для Random Forest и 0,863 для нейросетевой модели, что свидетельствует о более устойчивой классификации при использовании обучаемого нейросетевого представления речевых характеристик.

Усовершенствованная технология процесса принятия решения в организациях финансового обслуживания приведена на Рисунке 2.

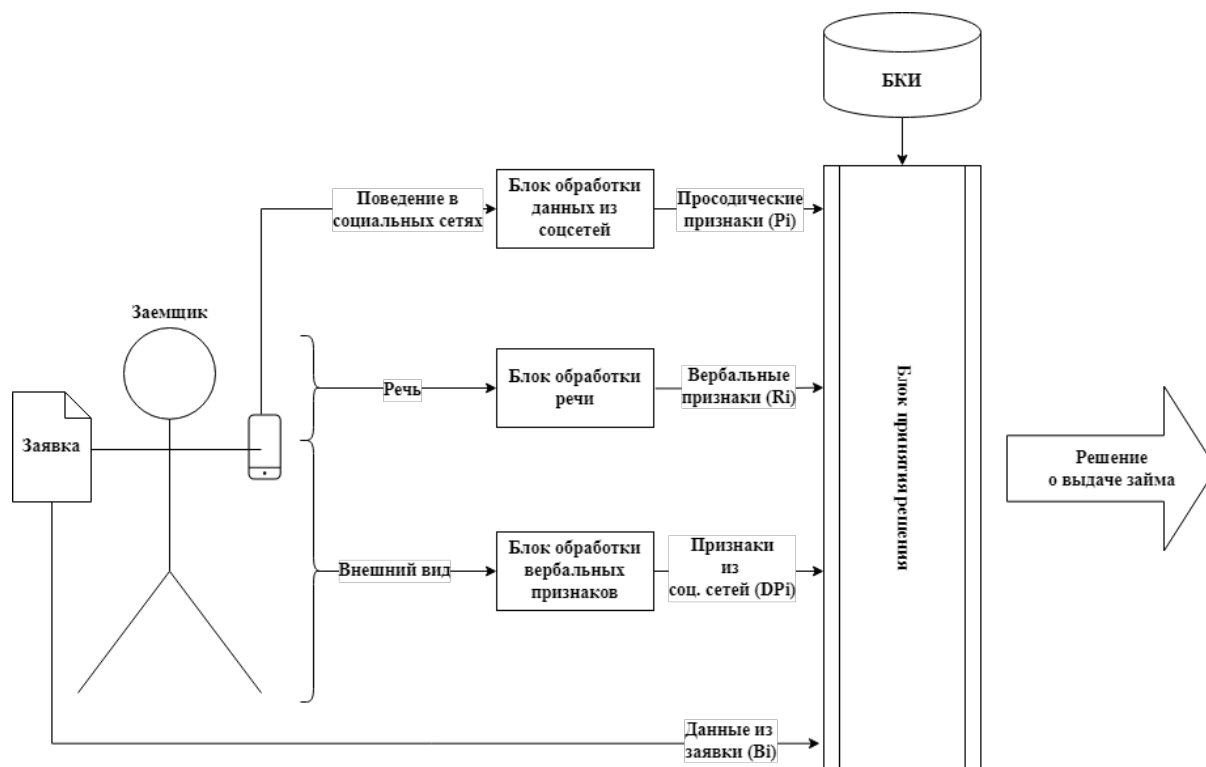


Рисунок 2 – Усовершенствованная технология процесса принятия решения в организациях финансового обслуживания

Figure 2 – Improved technology for decision-making in financial services organizations

Применение данной методики при внедрении в кредитный конвейер одной из российских микрокредитных организаций показало следующие результаты: в сравнении с предыдущими исследованиями, проведенными в 2019–2020 году [4], точность определения предполагаемого срока просрочки с учетом добавления модуля определения речевых признаков возросла на 13,1 %, а общая доля просрочек снизилась на 2,7 %.

Обсуждение

Полученные результаты показывают, что использование в нейросетях методов для анализа характеристик речевого потока может существенно дополнить существующие подходы построения цифровых профилей пользователей финансовых услуг. Напротив традиционных моделей кредитного скоринга, построенных преимущественно на анкетных данных, кредитной истории, социально-демографических и вербальных признаках, предложенный подход позволяет учитывать дополнительные невербальные характеристики клиента в его речевом поведении.

К особой практической значимости можно отнести анализ просодического анализа речи при дистанционном взаимодействии с клиентами банков и микрофинансовых организаций. В условиях телефонного или голосового обслуживания отсутствует возможность непосредственного наблюдения за внешним видом, мимикой, а также иными визуальными признаками субъекта, однако существует возможность задержки к речевому потоку. Темп речи, громкость, высота тона, распределение пауз и ритмика в речи могут дополнительно служить индикаторами эмоционального состояния клиента, коммуникативной уверенности и индивидуальных психологических особенностях клиента.

Интеграция модуля анализа речевой информации в систему поддержки принятия решений делает возможным расширить цифровой профиль пользователя и снизить уровень субъективной неопределенности при оценке заемщика. При этом самым обоснованным является не изолированное применение просодического анализа, а использование его в составе многокритериальной модели, объединяющей вербальные, невербальные, социально-сетевые, анкетные и финансовые данные. Такой подход увеличивает устойчивость итогового решения и уменьшает риск ошибочной классификации клиента, а полученные результаты доказывают его применимость в реальном секторе экономики.

Заключение

Рассмотренные технологии открывают путь к созданию интеллектуальных систем, которые при оценке заемщика выходят за рамки стандартной анкеты – они учитывают поведенческие паттерны и психологические особенности человека, что существенно снижает субъективность при кредитном скоринге и помогает финансовым организациям точнее управлять рисками. Возможность встроить речевой анализ непосредственно в банковские и МФО-платформы позволяет принимать решение о кредите прямо в ходе дистанционного звонка, без личного визита клиента. Для цифровых сервисов это означает принципиально новый уровень – более удобное обслуживание и переход к по-настоящему персонализированным финансовым продуктам.

Полученные результаты показывают, что нейросетевые подходы вполне применимы для анализа речевых характеристик и построения цифровых профилей клиентов – и в нынешних экономических реалиях у них есть реальный потенциал.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Сафуанов Ф.С., Назарова Е.А. Сравнительный анализ различных методов составления психологического портрета предполагаемого преступника. *Психология и право*. 2011;(3):33–43.
Safuanov S.F., Nazarova E.A. Different methods of psychological portrait preparation for an alleged offender: a comparative analysis. *Psychology and Law*. 2011;(3):33–43. (In Russ.).
2. Петраевский В.А., Кравец А.Г. Метод интеллектуального анализа текстовой информации для психиатрической диагностики. *Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика*. 2024;(2):95–104.
Petraevskiy V.A., Kravets A.G. The textual information intellectual analysis method for psychiatric diagnosis. *Vestnik of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*. 2024;(2):95–104. (In Russ.).
3. Демина Р.Ю., Ашимов М.М., Лепехин Д.А., Есауленко В.Н. Формирование обучающего множества на основе данных должников, хранимых коллекторским агентством. *Прикаспийский журнал: управление и высокие технологии*. 2025;(3):38–45.
Demina R.Y., Ashimov M.M., Lepekhin D.A., Esaulenko V.N. Formation of a training set based on debtors' data stored by a collection agency. *Caspian Journal: Management and High Technologies*. 2025;(3):38–45. (In Russ.).
4. Кузнецова В.Ю. Информационная технология принятия решений в микрофинансовой организации. *Системная инженерия и информационные технологии*. 2023;5(3):27–41.
Kuznetsova V.Yu. Information technology for decision making in a microfinance organization. *Systems Engineering and Information Technologies*. 2023;5(3):27–41. (In Russ.).
5. Пономаренко В.В. *Практическая характерология с элементами прогнозирования и управления поведением: методика «7 радикалов»*. Москва: Принт Про; 2017. 206 с.
6. Yang Y., Qin Y., Fan Y., Zhang Zh. Unlocking the Power of Voice for Financial Risk Prediction: A Theory-Driven Deep Learning Design Approach. *MIS Quarterly*. 2023;47(1):63–96. <https://doi.org/10.25300/MISQ/2022/17062>
7. Акай О.М. Вычислительное моделирование просодико-синтаксических корреляций в русской устной речи: методы машинного обучения и лингвистический анализ. В сборнике: *Русская грамматика: полипарадигмальность как методологический принцип современных научных исследований: Материалы IX Международного научного симпозиума, 23–27 сентября 2025 года, Иркутск, Россия*. Иркутск: Издательство ИГУ; 2025. С. 416–422.
Akay O.M. Computational Modeling of Prosody-Syntax Correlations in Russian Spoken Speech: Machine Learning Methods and Linguistic Analysis. In: *Russian Grammar: polyparadigmality as a methodological principle of modern scientific research: Proceedings of the IX International Scientific Symposium, 23–27 September 2025, Irkutsk, Russia*. Irkutsk: Izdatel'stvo IGU; 2025. P. 416–422. (In Russ.).
8. Chen X., Yu X., Chang L., et al. *The Sound of Risk: A Multimodal Physics-Informed Acoustic Model for Forecasting Market Volatility and Enhancing Market Interpretability*. arXiv. URL: <https://arxiv.org/abs/2508.18653> [Accessed 19th August 2026].
9. Кузовлев К.Д., Кузнецова В.Ю. Методика выявления психоэмоционального воздействия в аудиопотоке на основе нейросетевого анализа. *Инженерный вестник Дона*. 2026;(6). URL: <https://ivdon.ru/ru/magazine/archive/n6y2026/11269>

- Kuzovlev K.D., Kuznetsova V.Yu. A method for identifying psycho-emotional impact in an audio stream based on neural network analysis. *Engineering Journal of Don*. 2026;(6). (In Russ.). URL: <https://ivdon.ru/en/magazine/archive/n6y2026/11269>
10. Trigeorgis G., Ringeval F., Brueckner R., et al. Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. In: *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, 20–25 March 2016, Shanghai, China*. IEEE; 2016. P. 5200–5204. <https://doi.org/10.1109/ICASSP.2016.7472669>
 11. Schuller B.W., Steidl S., Batliner A., et al. The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In: *Proceedings of the 14th Annual Conference of the International Speech Communication Association, 25–29 August 2013, Lyon, France*. ISCA; 2013. P. 148–152. <https://doi.org/10.21437/Interspeech.2013-56>
 12. Zhao J., Mao X., Chen L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomedical Signal Processing and Control*. 2019;47(4):312–323. <https://doi.org/10.1016/j.bspc.2018.08.035>
 13. Mirsamadi S., Barsoum E., Zhang Ch. Automatic speech emotion recognition using recurrent neural networks with local attention. In: *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, 05–09 March 2017, New Orleans, LA, USA*. IEEE; 2017. P. 2227–2231. <https://doi.org/10.1109/ICASSP.2017.7952552>
 14. Марьенков А.Н., Кривенко А.И. Сбор и обработка текстовых данных в контексте оценки социальных настроений: методологические аспекты. *Инженерный вестник Дона*. 2022;(7). URL: <https://ivdon.ru/en/magazine/archive/n7y2022/7797>
Marienkov A.N., Krivenko A.I. Collection and processing of textual data in the context of social sentiment assessment: methodological aspects. *Engineering Journal of Don*. 2022;(7). (In Russ.). URL: <https://ivdon.ru/en/magazine/archive/n7y2022/7797>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Кузнецова Валентина Юрьевна, кандидат технических наук, доцент, заведующая кафедрой «Автоматизированные системы обработки информации и управления», Астраханский государственный технический университет, Астрахань, Российская Федерация.
e-mail: v.kuznetsova@astu.ru
ORCID: [0000-0002-6954-5020](https://orcid.org/0000-0002-6954-5020)

Valentina Yu. Kuznetsova, Candidate of Engineering Sciences, Docent, Head of the Department of Automated Information Processing and Control Systems, Astrakhan State Technical University, Astrakhan, the Russian Federation.

Кузовлев Кирилл Дмитриевич, ассистент, Астраханский государственный технический университет, Астрахань, Российская Федерация.
e-mail: k.kuzovlev@astu.ru
ORCID: [0009-0003-3769-0594](https://orcid.org/0009-0003-3769-0594)

Kirill D. Kuzovlev, Assistant, Astrakhan State Technical University, Astrakhan, the Russian Federation.

Статья поступила в редакцию 15.06.2026; одобрена после рецензирования 17.09.2026; принята к публикации 23.09.2026.

The article was submitted 15.06.2026; approved after reviewing 17.09.2026; accepted for publication 23.09.2026.