

СИСТЕМА ГОЛОСОВОГО МОНИТОРИНГА VOICE SCREENING MODEL (VSM) ДЛЯ АНАЛИЗА РУССКОЯЗЫЧНОЙ РЕЧИ И ПОМОЩИ В ДИАГНОСТИКЕ ПСИХИЧЕСКИХ ЗАБОЛЕВАНИЙ

Ковриго А. В.

Научный руководитель: канд. мед. наук, доц. Обьедков В. Г.

Белорусский государственный медицинский университет, г. Минск

Резюме. Диагностика психических заболеваний, таких как депрессия и шизофрения, традиционно основана на субъективных методах оценки, что ограничивает раннее выявление и объективный мониторинг состояния пациентов. Голосовые аудиомаркеры – объективные характеристики речи (высота голоса, темп, качество звучания) – отражают изменения психоэмоционального состояния и могут служить инструментом для неинвазивного скрининга. В работе представлена система голосового мониторинга Voice Screening Model (VSM), предназначенная для автоматизированного анализа русскоязычной речи с целью выявления признаков психических расстройств. Система объединяет два подхода: извлечение акустических параметров голоса (высота тона, вариативность, энергия, темп речи, качество голоса) и анализ глубинных речевых характеристик с помощью современных методов искусственного интеллекта. Разработанная система готова к пилотному клиническому тестированию и позволяет создать базу голосовых аудиомаркеров для диагностики депрессии, шизофрении и других психических заболеваний. Предварительные результаты демонстрируют способность системы выявлять отклонения голосовых параметров от нормы с указанием конкретных признаков (повышенная напряженность речи, снижение выразительности, изменение энергетика голоса).

Ключевые слова: голосовые аудиомаркеры, психические заболевания, депрессия, шизофрения, автоматизированная диагностика, анализ речи.

Актуальность. Психические расстройства – ведущая проблема здоровья, затрагивающая более 450 млн человек. Традиционные методы диагностики являются клинические интервью, психометрические шкалы ограничены субъективностью, зависимостью от самооценки пациента и высокой стоимостью. Голосовые аудиомаркеры – акустические и интонационные параметры позволяют объективизировать диагностику. При депрессии и шизофрении изменения в нервной системе влияют на речевые центры мозга, что проявляется в снижении тона, темпа, выразительности

и изменении звучания голоса. Автоматизированный анализ выявляет депрессию с точностью 78–93%, сопоставимой со стандартными опросниками.

Цель: разработка системы голосового мониторинга Voice Screening Model (VSM) для автоматизированного анализа русскоязычной речи и создания базы голосовых аудиомаркеров, способствующих диагностике психических заболеваний.

Задачи:

1. Разработать систему VSM для выявления признаков психических заболеваний

2. Собрать и проанализировать аудиозаписи, сравнив акустические параметры с нормативами

3. Провести пилотное тестирование и сформировать базу аудио-маркеров

Материалы и методы. Система VSM представляет собой программный комплекс, предназначенный для анализа аудиозаписей речи пациентов. Запись голоса производится с помощью стандартных устройств (смартфон, компьютер с микрофоном) длительностью от 30 секунд до 5 минут. Система обрабатывает аудиофайлы и автоматически извлекает набор голосовых характеристик, которые затем сравниваются с нормативными значениями для соответствующей возрастной и половой группы.

Анализ речи в системе VSM включает два основных компонента:

1. Акустические параметры голоса. Система автоматически измеряет следующие характеристики речи:

2. Высота голоса – отражает колебания голосовых связок; при депрессии часто наблюдается снижение высоты и уменьшение ее изменчивости.

3. Качество голоса – измеряется через джиттер и нестабильность амплитуды голосовых связок; повышенные значения указывают на напряженность, эмоциональный дистресс или телесные проявления тревоги.

4. Энергия и громкость речи – отражают интенсивность и силу голоса. Темп речи и паузы – замедление речи и увеличение длительности пауз

ассоциируются с психомоторным торможением при депрессии, в то время как ускоренная речь может наблюдаться при мании.

Спектральные характеристики – описывают частотное распределение звука и отражают артикуляционные особенности; измеряются через мел-частотные кепстральные коэффициенты.

Для извлечения этих параметров используется специализированный инструментальный OpenSMILE – программное обеспечение, разработанное в Мюнхенском техническом университете и широко применяемое в исследованиях эмоционального состояния и психических расстройств.

Глубинный анализ речевых паттернов. Помимо явных акустических параметров, система использует метод искусственного интеллекта для выявления скрытых закономерностей в речи, которые не очевидны при обычном прослушивании. Применяется самообучающаяся модель wav2vec2-large-ru-golos, специально адаптированная для русского языка. Эта модель анализирует речь на глубинном уровне и выделяет сложные комбинации признаков, характерные для различных психоэмоциональных состояний.

Для каждого измеренного параметра система вычисляет степень отклонения от нормы (z-оценка), учитывая пол и возраст пациента.

Для каждого измеренного параметра система вычисляет степень отклонения от нормы (z-оценка), учитывая пол и возраст пациента. Отклонения классифицируются как:

1. Норма – параметр находится в пределах типичных значений для данной группы.

2. Умеренное отклонение – параметр незначительно отклоняется от нормы, что может потребовать клинического внимания.

3. Значительное отклонение – параметр существенно отличается от нормы, что указывает на возможные нарушения.

Система генерирует отчет, включающий: Таблицу всех измеренных параметров с указанием их значений и степени отклонения от нормы. Описывающее общее состояние голосовых характеристик и выделяющие параметры, требующие внимания специалиста. Рекомендации для клинициста по дальнейшему наблюдению или углубленной диагностике.

Результаты и их обсуждение. Разработана и реализована система Voice Screening Model (VSM), готовая к пилотному клиническому тестированию. Система способна обрабатывать аудиозаписи русскоязычной речи и извлекать 15 основных голосовых параметров, включая высоту голоса, качество звучания, энергетiku, темп речи и спектральные характеристики. Дополнительно система использует глубинный анализ речи для выявления более 800 скрытых признаков, связанных с психоэмоциональным состоянием. Предварительное тестирование системы проведено на образцах речи добровольцев. Пример анализа голоса мужчины 21 года продемонстрировал следующие результаты:

“Параметры в пределах нормы:

Высота голоса: 123,5 Гц (нормальный диапазон для мужчин молодого возраста). Основные показатели качества голоса (джиттер, RAP): отклонения незначительны. Темп речи: в пределах нормы. Спектральные характеристики: соответствуют средне-статистическим значениям.

Параметры, требующие внимания: Частота пересечений нулевой линии технический показатель, отражающий шумность и напряженность речи: значительно выше среднего (z-оценка +1,08), что может указывать на повышенное напряжение во время речи или недостаточную расслабленность голосовых связок. Вариативность амплитуды голоса: ниже среднего (z-оценка –1,29), что может свидетельствовать о снижении выразительности речи – признак, характерный для эмоциональной притупленности. Энергия голоса: несколько ниже среднего, что может указывать на относительно тихую речь или сниженную активность. Показатель PPQ5: умеренно повышен (z-оценка +1,76), что может отражать нестабильность голосовых колебаний.

Общий вывод системы: голос демонстрирует параметры, близкие к норме, однако выявлены признаки повышенной напряженности речи и снижения выразительности, требующие дополнительной клинической оценки с учетом контекста записи и общего психического статуса обследуемого.”

Данный пример иллюстрирует способность системы VSM выявлять тонкие отклонения голосовых параметров, которые могут быть не оче-

видны при обычном клиническом наблюдении, но потенциально связаны с психоэмоциональным состоянием.

Разработанная система Voice Screening Model представляет собой перспективный инструмент для объективизации диагностики психических расстройств. Основные преимущества системы включают:

1. **Объективность.** В отличие от субъективных опросников и клинических интервью, голосовые аудиомаркеры представляют собой измеримые физические характеристики, которые не зависят от желания пациента сообщать о своих симптомах.

2. **Неинвазивность и простота сбора данных.** Запись голоса не требует специализированного медицинского оборудования, может выполняться дистанционно через телефон или компьютер, что особенно важно для мониторинга пациентов в амбулаторных условиях или на дому.

3. **Возможность непрерывного мониторинга.** Система может использоваться для регулярной оценки состояния пациента в процессе терапии, что позволяет отслеживать динамику симптомов и своевременно выявлять ухудшение состояния или недостаточный ответ на лечение.

4. **Адаптация для русскоязычной популяции.** Применение модели, специально обученной на русском языке (wav2vec2-large-ru-golos), обеспечивает учет фонетических, просодических и лингвистических особенностей русской речи, что повышает точность анализа по сравнению с универсальными англоязычными си-

стемами.

Важным аспектом является также учет вариативности условий записи (качество микрофона, фоновый шум, акустическая обстановка), которая может влиять на измеряемые параметры. Система включает модуль подавления шума и фильтрации речевых сегментов, однако требуется дополнительная оценка робастности системы в реальных клинических условиях.

Система VSM может найти применение в следующих клинических сценариях:

1. **Скрининг психических расстройств в группах риска** (пациенты первичного звена здравоохранения, студенты, военнослужащие).

2. **Мониторинг эффективности антидепрессантной терапии** через регулярную оценку изменений голосовых параметров в процессе лечения.

3. **Раннее выявление рецидивов** у пациентов с рекуррентной депрессией или биполярным расстройством.

4. **Телемедицинская оценка психического состояния** пациентов, находящихся в удаленных регионах или не имеющих доступа к специализированной психиатрической помощи.

Дальнейшее развитие системы предполагает создание базы голосовых аудиомаркеров для различных нозологических форм психических расстройств. Интеграция голосовых аудиомаркеров с другими источниками клинической информации (данные психометрических шкал, биологические маркеры, нейровизуализация) может обеспечить комплексный под-

ход к диагностике и персонализации терапии психических расстройств.

Выводы. Разработана система голосового мониторинга Voice Screening Model (VSM), предназначенная для автоматизированного анализа русскоязычной речи и выявления голосовых аудиомаркеров психических заболеваний. Система объединяет анализ акустических параметров голоса (высота, качество, энергия, темп речи) с глубинным анализом речевых паттернов на основе искусственного интеллекта. Предварительное тестирование демонстрирует способность системы выявлять отклоне-

ния голосовых характеристик от нормы, включая признаки повышенной напряженности речи и снижения выразительности. Система готова к пилотному клиническому тестированию для создания базы голосовых аудиомаркеров депрессии, шизофрении и других психических расстройств. Внедрение системы VSM в клиническую практику может способствовать объективизации диагностики, непрерывному мониторингу терапии и расширению доступности психиатрической помощи через телемедицинские технологии.

Литература

1. Zaremba L., Dłotko P., Piasecki J. Listening to the Mind: Integrating Vocal Biomarkers into Digital Mental Health // *Frontiers in Psychiatry*. 2025. Vol. 16. P. 1-15.
2. Gan J., Wang H., Chen Y. Bio-acoustic features of depression: A review // *Biomedical Signal Processing and Control*. 2024. Vol. 91. P. 105-118.
3. Cummins N., Scherer S., Krajewski J. et al. A review of depression and suicide risk assessment using speech analysis // *Speech Communication*. 2015. Vol. 71. P. 10-49.
4. Eyben F., Wöllmer M., Schuller B. openSMILE: the Munich versatile and fast open-source audio feature extractor // *Proceedings of the 18th ACM International Conference on Multimedia*. 2010. P. 1459-1462.
5. Baevski A., Zhou Y., Mohamed A., Auli M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // *Advances in Neural Information Processing Systems*. 2020. Vol. 33. P. 12449-12460.
6. Williamson J.R., Quatieri T.F., Helfer B.S. et al. Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice // *Digital Biomarkers*. 2021. Vol. 5(1). P. 78-88.
7. Cummins N., Baird A., Schuller B.W. Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning // *Methods*. 2018. Vol. 151. P. 41-54.

VOICE SCREENING MODEL (VSM) FOR MONITORING AND ANALYSIS OF RUSSIAN SPEECH TO ASSIST IN THE DIAGNOSIS OF MENTAL DISORDERS

Kovrigo A. V.

Tutor: PhD, associate professor Obyedkov V. G.

Belarusian State Medical University, Minsk

Resume. Diagnosis of mental disorders such as depression and schizophrenia has traditionally been based on subjective assessment methods, which limits early detection and objective monitoring of patients' conditions. Voice biomarkers—objective speech characteristics (pitch, speech rate, quality of phonation)—reflect changes in psycho-emotional status and can serve as a tool for non-invasive screening. This work presents the Voice Screening Model (VSM), a voice monitoring system for automated analysis of Russian-language speech to identify signs of mental disorders. The system combines two approaches: extraction of acoustic voice parameters (pitch, variability, energy, speech rate, voice quality) and the analysis of deep speech features using modern artificial intelligence methods. The developed system is ready for pilot clinical testing and enables the creation of a database of voice audio biomarkers for the diagnosis of depression, schizophrenia, and other mental illnesses. Preliminary results demonstrate the system's ability to detect deviations of voice parameters from the norm, indicating specific features (increased vocal tension, reduced expressiveness, changes in vocal energy).

Keywords: voice biomarkers, mental disorders, depression, schizophrenia, automated diagnosis, speech analysis.