**Методы распознавания дефектной речи по речевому сигналу**

Беленко М.В.

Факультет ПИиКТ, Университет ИТМО, Россия, Санкт-Петербург

**Аннотация**

В работе проведено исследование наиболее распространенных типов расстройств речи и современных методов их обнаружения и классификации. Произведена оценка применимости рассмотренных методов для обнаружения каждого из расстройств.

**Ключевые слова** – распознавание речи, расстройства речи, определение, классификация, методы.

**Введение**

Люди все чаще подвергаются риску патологических проблем с голосом. Около 25% населения земного шара, чья профессия вынуждает их говорить громче, чем нормальный уровень, страдают от такого рода проблем. Например, певцы, актеры, юристы, учителя, аукционисты, инструкторы по аэробике и руководители производства - все они, как считается, работают в профессиях, которые предъявляют большие требования к голосу. Использование цифровой обработкой речевых сигналов обеспечивает неинвазивный аналитический метод, который может считаться эффективным вспомогательным инструментом для врачей при выявлении нарушений голоса, особенно на ранних стадиях.

Для построения системы определения, классификации и исправления дефектной речи были рассмотрены и проанализированы наиболее распространенные речевые расстройства.

**Апраксия речи.** Апраксия речи происходит, когда нарушается нервное сообщение между мозгом и речевыми мышцами человека. Человек знает, что он хочет сказать, и даже может записать то, что он хочет сказать на бумаге, – однако мозг не может послать правильные сигналы речевым мышцам, чтобы сформулировать то, что он хочет сказать, даже если сами речевые мышцы работают нормально. Существуют различные уровни выраженности апраксии, начиная от преимущественно функционального, до бессвязной речи. Апраксия может быть вызвана как повреждением мозга, например, у взрослого человека, перенесшего инсульт, так и иметь генетическую природу, проявляясь у детей.

Симптомы легких форм апраксии являются общими для целого ряда различных речевых расстройств и включают неправильное произношение слов и нарушения в тоне, ритме или акценте (просодия). Тяжелые случаи заболевания диагностируются легче, и проявляются такими симптомами как неспособность артикулировать слова, нащупывание звуковых позиций, нецелевые движения, искажающие звуки, и непоследовательность в произношении.

**Заикание.** Заикание, настолько распространено, что каждый знает, как оно звучит, и может легко распознать его. У каждого, вероятно, были моменты заикания, по крайней мере, один раз в жизни. Самое раннее нарушение беглости речи может проявиться, когда ребенок учится говорить, может также проявиться позже, в детстве. Гораздо реже оно развивается у взрослых, хотя многие взрослые сохранили заикание с детства.

Причины заикания по большей части остаются загадкой. Существует корреляция с историей семьи, указывающая на генетическую связь. Другая теория состоит в том, что заикание - это форма непроизвольного или полувольного тика. Большинство исследований заикания сходятся во мнении, что здесь задействовано много факторов.

**Дизартрия.** Дизартрия - это симптом поражения нервов или мышц. Она проявляется в виде невнятной речи, замедленной речи, ограниченного движения языка, челюсти или губ, ненормального ритма и высоты тона при разговоре, изменения качества голоса, трудности артикуляции, затрудненной речи и других сопутствующих симптомов, в связи с повреждением мышц или нервов, участвующих в процессе речи, таких как диафрагма, губы, язык и голосовые связки. Поскольку дизартрия это симптом повреждения нервов и/или мышц, он может быть вызван широким спектром явлений, которые влияют на людей всех возрастов. Она может начаться во время развития в утробе матери или вскоре после рождения в результате таких состояний, как мышечная дистрофия и детский церебральный паралич. У взрослых одними из наиболее распространенных причин дизартрии являются инсульт, опухоли и др.

**Шепелявость.** Существует четыре различных типа шепелявости.

Фронтальная шепелявость - при фронтальной шепелявости люди, слишком сильно выталкивают свой язык вперед, так что звуки /z/ и /s/ заменяются на звук /th/. Фронтальная шепелявость-самый распространенный тип шепелявости.

Латеральная шепелявость - при боковых шепелявостях воздух скользит по бокам языка так, что/s/ и /z/ звуки выходят “влажными”. Люди с латеральной шепелявостью часто говорят так, как будто у них во рту слишком много слюны.

Небная шепелявость - при небной шепелявости говорящий касается языком неба при произнесении /s/, /z/, а иногда /r/. В результате звуки выходят размытыми и гнусавыми.

Зубная шепелявость – человек, страдающий зубной шепелявостью, вместо того, чтобы просунуть язык между передними зубами, прижимает свой язык к зубам. По звучанию очень похожа на фронтальную щепелявость.

**Спазматическая дисфония.** Спазматическая дисфония (СД) - хроническое длительное расстройство, влияющее на голос. Он характеризуется спазмами голосовых связок, когда человек пытается говорить. Голос может быть описан как дрожащий, хриплый, стонущий, напряженный или дрожащий. Это может привести к значительному изменению акцента речи.

**Захламление.** Нарушение беглости речи, загромождение характеризуется тем, что речь человека слишком быстрая, слишком отрывистая или и то и другое вместе. Чтобы квалифицироваться как загромождение, речь человека должна также иметь чрезмерное количество “ну",” гм“,” как“,” хм “или” так " (речевые нарушения), чрезмерное исключение или сжатие слогов, или ненормальные слоговые ударения или ритмы.

**Немота, Избирательный мутизм.** Избирательный мутизм – речевое расстройство, при котором человек не говорит в некоторых или большинстве ситуаций, однако этот человек физически способен говорить. Чаще всего это происходит у детей, и обычно примером может служить ребенок, говорящий дома, но не в школе.Избирательный мутизм связан с психологией. Он появляется у детей, которые очень застенчивы, у которых есть тревожное расстройство или которые переживают период социальной изоляции.

**Афазия.** Афазия - это нарушение коммуникации, вызванное повреждением языковых способностей мозга. Афазия отличается от апраксии речи и дизартрии тем, что она относится исключительно к речевому и языковому центру мозга. Таким образом, любой человек может страдать от афазии, потому что повреждение мозга может быть вызвано рядом факторов. Распространенными причинами афазии являются инсульт, опухоли головного мозга, черепно-мозговые травмы и дегенеративные заболевания головного мозга.

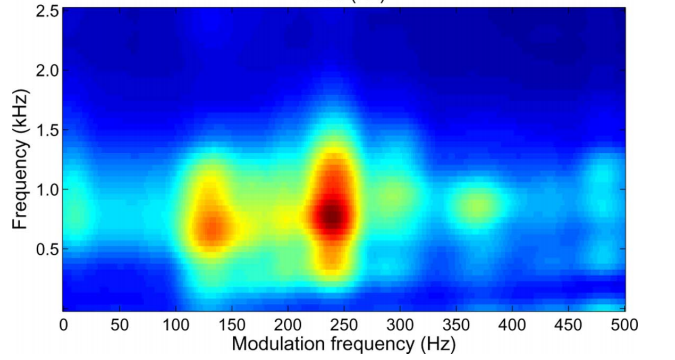
**Задержка Речи, Алалия.** Задержка речи, известная специалистам как алалия, проявляется у детей и выражается в том, что ребенок не делает нормальных попыток вербально общаться. Это может быть вызвано целым рядом факторов.

Также необходимо рассмотреть существующие методы детектирования и классификации расстройств речи, и оценить их применимость по отношению к вышеперечисленным заболеваниям.

**Методы**

**Обнаружение и распознавание патологии голоса на основе Модуляционных спектральных признаков**. Начальное представление сначала преобразуется в область более низкой размерности с использованием сингулярной декомпозиции более высокого порядка (HOSVD). Исходя из этого представления с уменьшенной размерностью, предлагается процесс выбора признаков с использованием теоретико-информационного критерия, основанного на взаимной связи между классами голоса (т. е. нормофоническими/дисфоническими) и признаками [1].

Рис. 1 Усредненная спектрограмма звука A здорового человека



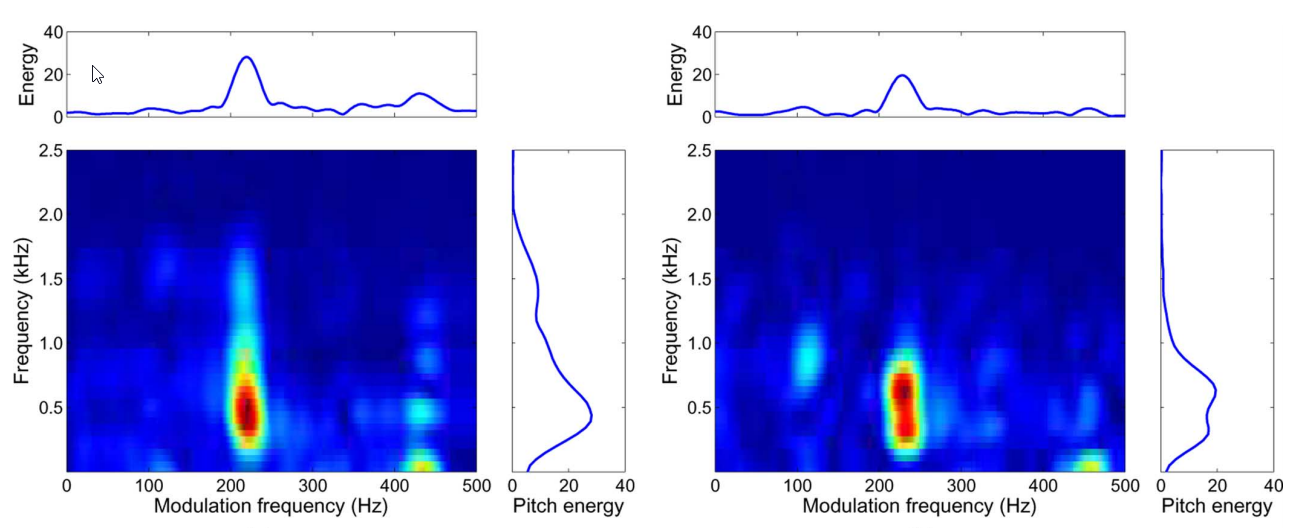
****

Рис. 2 Спектрограммы пациентов с полипами и спазмотической дисфонией

**Автоматическое обнаружение патологических голосов с использованием мер сложности, шумовых параметров и мел-кепстральных коэффициентов**. В статье, посвященной данному методу[2], рассматриваются возможности распознавания 11 признаков, выделенных с помощью нелинейного анализа временных рядов. Две из этих характеристик основаны на традиционной нелинейной статистике (наибольший показатель Ляпунова и корреляционная размерность), две - на рекуррентном и фрактально-масштабном анализе, а остальные-на различных оценках энтропии. Кроме того, в данной работе используется стратегия, основанная на объединении классификаторов для слияния нелинейного анализа с информацией, представленной классическими подходами параметризации, найденными в литературе (шумовые параметры и мел-частотные кепстральные коэффициенты).В данной работе не представляется возможности классификации расстройств речи.

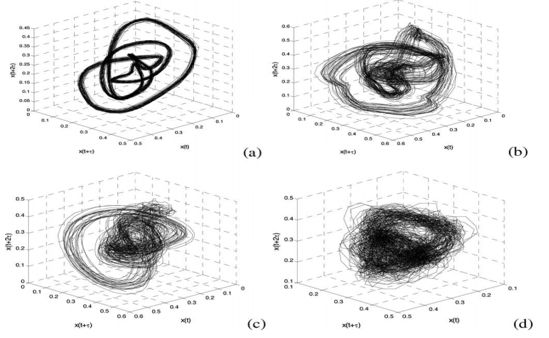
****

Рис. 3 Визуализация звука ‘a’ с различными растройствами

**Автоматическое обнаружение и классификация патологии голоса с использованием нерегулярности области голосового тракта.** Патология голоса связана с проблемой голосовых связок, и поэтому область голосового сигнала, которая связана с голосовыми связками или голосовым трактом, должна демонстрировать нерегулярные паттерны во фреймах в случае устойчивого гласного для патологического голоса[3]. Этот нерегулярный паттерн количественно выражается в числе фреймов, и может использоваться чтобы различать нормальные и патологические голоса. Данный метод может применяться только для выявления патологий вызванных голосовым трактом.

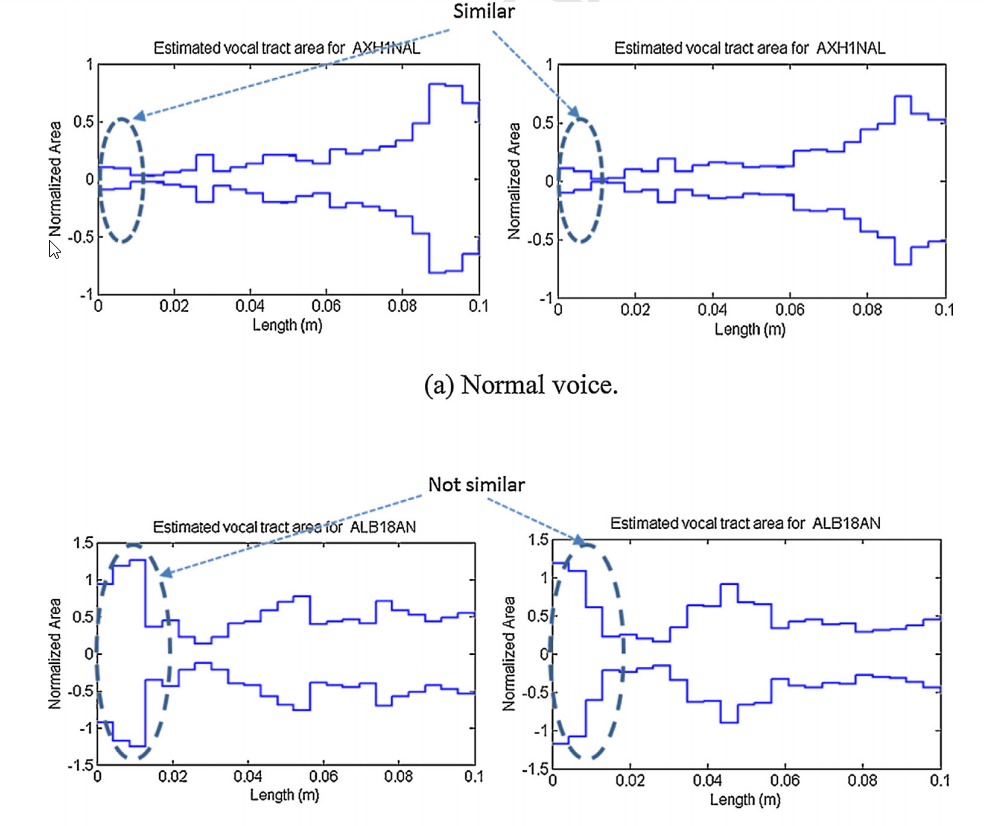


Рис. 4 Нерегулярные паттерны в голосовом сигнале

**Выявление и классификация патологии голоса с использованием автокорреляционных и энтропийных признаков в различных частотных областях.** Извлекаются максимальные пиковые значения и соответствующие им значения задержек из каждого кадра озвученного сигнала, используя автокорреляцию в качестве признаков для обнаружения и классификации патологических образцов[4]. Также извлекается энтропия для каждого кадра речевого сигнала после того, как его значения нормализуются, для использования в качестве признаков. Эти особенности были исследованы в различных частотных диапазонах для оценки вклада каждого диапазона в процессы обнаружения и классификации.

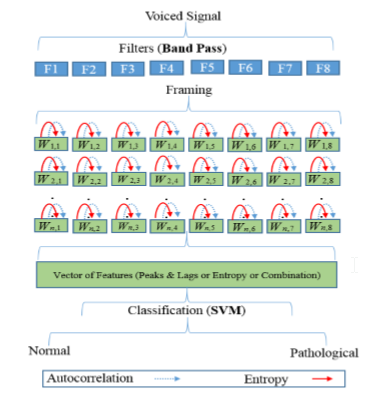


Рис 5. Архитектура предложенной системы

**Обнаружение патологии голоса с помощью глубокого обучения.** Идея этого подхода заключается в использовании сверточных слоев в сочетании с рекуррентными слоями Long-Short-Term-Memory (LSTM) на необработанном звуковом сигнале [5]. На рисунке представлена архитектура используемой сети. Преимуществом этого подхода является универсальность, так как с его помощью можно классифицировать любое расстройство речи, необходима только достаточная обучающая выборка. Однако создание такой выборки является трудоемкой задачей, что усложняет применение данного подхода.

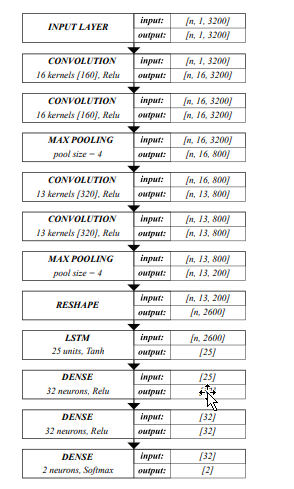


Рис. 6 Архитектура предложенной сети

**Результаты**

Рассмотрев основные подходы к детектированию дефектной речи и результаты их работы можно сделать следующие выводы.

* В настоящее время подходы к решению данной проблемы делятся на 2 группы – выделение из сигнала признаков, на основании которых происходит классификация, и обработка всего сигнала нейронной сетью.
* Первая группа подходов обеспечивает большую точность, но только для определенных групп заболеваний, обычно вызванных дефектами голосового тракта или связок.
* Вторая группа потенциально более универсальна и может использоваться для детектирования любого расстройства из перечисленных выше, однако, на данный момент, обеспечивает посредственную точность.

**Литература**

1. Markaki M., Stylianou Y. Voice pathology detection and discrimination based on modulation spectral features //IEEE Transactions on audio, speech, and language processing. – 2011. – Т. 19. – №. 7. – С. 1938-1948.

2. Arias-Londoño J. D. et al. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients //IEEE Transactions on biomedical engineering. – 2010. – Т. 58. – №. 2. – С. 370-379.

3. Muhammad G. et al. Automatic voice pathology detection and classification using vocal tract area irregularity //Biocybernetics and Biomedical Engineering. – 2016. – Т. 36. – №. 2. – С. 309-317.

4. Al-Nasheri A. et al. Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions //IEEE Access. – 2017. – Т. 6. – С. 6961-6974.

5. Harar P. et al. Voice pathology detection using deep learning: a preliminary study //2017 international conference and workshop on bioinspired intelligence (IWOBI). – IEEE, 2017. – С. 1-4.