

ОБНАРУЖЕНИЯ ПАТОЛОГИЙ ГОЛОСОВОГО ТРАКТА В ОПЕРАЦИОННОЙ СИСТЕМЕ ANDROID

И. Э. Хейдоров, А. М. Сорока, М. Н. Лахвич

Белорусский государственный университет,

Минск, Беларусь

e-mail: igorhmm@mail.ru, soroka.a.m@gmail.com, m.lakhvich@gmail.com

Системы распознавания и классификации голоса – это вычислительные системы, которые могут определить речь и выделить набор акустических характеристик. Целью данной работы является демонстрация пользователю его характеристик голоса и предупреждение о наличии риска патологии голосового тракта. В статье приведены результаты исследования различных моделей классификации для нахождения эффективных параметров обнаружения патологии. Классификация патологий голосового тракта реализована на основе метода опорных векторов и многослойной искусственной нейронной сети. Результат эксперимента свидетельствуют о высокой точности данного метода для обнаружений патологий голосового тракта.

Ключевые слова: метод опорных векторов; искусственная нейронная сеть; машинное обучение; мел-частотные кепстральные коэффициенты; речевой сигнал; акустический анализ голоса; предварительная обработка сигнала.

DETECT PATHOLOGIES VOCAL TRACT IN THE OPERATING SYSTEM ANDROID

I. E. Kheidorov, A. M. Soroka, M. N. Lakhvich

Belarusian State University

Minsk, Belarus

Systems of recognition and classification of voices – a computer system that can identify it and to allocate a set of acoustic characteristics. Recognition and classification of pathologies by voice is still a difficult research area in speech processing. The aim of this work is to show the user's voice parameters and the presence of risk prevention vocal tract pathology. This article explores and compares different classification models to find effective detection of healthy and pathological voice parameters. Classification of vocal tract pathologies is realized with a help of Support Vector Machine (SVM) and Artificial Neural Network (ANN). Experiment results allow to make a conclusion about high accuracy of a developed method for detecting pathologies.

Keywords: support vector machines; artificial neural network; machine learning; mel-frequency cepstral coefficients; speech signal; acoustic voice analysis; signal pre-processing.

ВВЕДЕНИЕ

Выявление заболеваний на ранней стадии является одним из ключевых факторов успешного их лечения, особенно в случаях, связанных с социально важными органами. При выявлении заболевания на ранней стадии значительно повышается вероятность его излечения и сокращения периода восстановления голосовой функции. В настоящее время существующие программы обнаружения патологий голосового тракта привязаны к специальному оборудованию, что препятствует возможности сделать экспресс-диагностику самостоятельно.

Акустический сигнал является одним из основных средств связи человека с окружающей средой и людей между собой. Процесс генерации речевого сигнала начинается с преобразования абстрактного сигнала в мозге человека в последовательность нервных импульсов, под воздействием которых происходит изменение физических характеристик голосового тракта, что в свою очередь при прохождении через голосовой тракт потока воздуха, приводит к образованию различных звуков. А так же изменения характеристик голосового тракта могут быть вызваны реальными факторами, которые могут свидетельствовать о наличии заболеваний у пациента. Например, органические образования в гортани или функциональное состояние голосовых связок, в результате чего может происходить изменение основных характеристик речевого сигнала.

МЕТОДИКА

Разработанная методика, с использованием вычислений мел-частотных кепстральных коэффициентов (МЧКК), базируется на применении полученного вектора признаков в обучении многослойной искусственной нейронной сети (ИНС), и метода, основанного на опорных векторах (МОВ). Общий алгоритм системы обнаружения патологий проиллюстрирован на рис. 1.



Рис. 1. Система обнаружения патологий голосового тракта

ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ

Процесс записи фонограммы осуществляется в реальном времени через звуковую карту устройства, с максимально возможной частотой дискретизации в диапазоне от 16 – 48 кГц. Для извлечения акустических характеристик, оцифрованная запись предварительно обрабатывается для выделения границ речевого сигнала, нормировки и сегментации [1].

Для получения акустических характеристик речевого сигнала, используем функцию, основанную на вычислении мел-частотных кепстральных коэффициентов [2]. С

помощью данных коэффициентов влияние высокочастотных составляющих, обычно содержащих посторонний шум, на результат распознавания уменьшается, в этом случае более тщательно анализируется информация, получаемая из низкочастотного диапазона.

$$y_t(k) = \sum_{m=1}^M \log(Y_t(m)) \cos\left(k\left(m - \frac{1}{2}\right) \frac{\pi}{M}\right), \quad (2)$$

где $y_t(k)$ – полученные мел-частотные кепстральные коэффициенты, k – число мел-коэффициентов, $Y_t(m)$ – набор мел-коэффициентов.

Введем в вектор признаков информацию о временных изменениях сигналов. Для этого, помимо самих кепстральных коэффициентов добавляются их первые и вторые производные (дельта и дельта-дельта характеристики). Таким образом вектор признаков состоит из 12 мел-частотных кепстральных коэффициентов, 24 характеристик первой и второй производной и по одной энергетической характеристике каждого поднабора; в общей сложности имеем 39 характеристик на окно. Включение производных в вектор признаков позволяет также снизить влияние сверточных искажений сигнала, в силу того, что эти искажения обычно медленно изменяются во времени и аддитивны в кепстральной области.

МЕТОД ОПОРНЫХ ВЕКТОРОВ

Метод опорных векторов основан на принципе минимизации структурного риска, принадлежит к семейству линейных классификаторов, может быть применен для задач классификации и регрессионного анализа [3, 7].

МОВ строит линейную модель для оценки функции принятия решения с использованием границ нелинейных классов на основе опорных векторов. Принимая на вход данные, метод предсказывает для каждого данного входа, какой из двух возможных классов формирует выходной сигнал. В соответствии с данным методом классифицируемые объекты представляются в виде векторов в некотором пространстве признаков, значения целевой зависимости известны только на объектах обучающей выборки. Требуется построить алгоритм, аппроксимирующий целевую зависимость на всем пространстве. Для классификации необходимо построение такой разделяющей гиперплоскости в этом пространстве, чтобы расстояние от нее до ближайшей точки было максимальным.

Эффективность метода классификации на основе МОВ сильно зависит от вида функции ядра. В данной работе применяется *ядро Гаусса*:

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right). \quad (3)$$

МЕТОД НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

Реализованный алгоритм многослойной ИНС [4] базируется на радиальной базисной функции Гаусса, как функция активации активационной функции и механизма

обратного распространения ошибки [5]. Наиболее часто в качестве активационной функции используется так называемый сигмоид, который имеет следующий вид:

$$f(x) = \frac{1}{1 + e^{-\alpha x}}, \quad (4)$$

где α – параметр наклона сигмоидальной функции.

Принцип работы многослойной ИНС с механизмом обратного распространения ошибки включает в себя три этапа: подачу на вход данных, где каждый из входных нейронов соответствует одному мел-частотному кепстральному коэффициенту; последующее распространение данных в направлении выходов, нейроны которых соответствует желаемому значению распознавания сигнала; вычисление и обратное распространение соответствующей ошибки, корректировку весов.

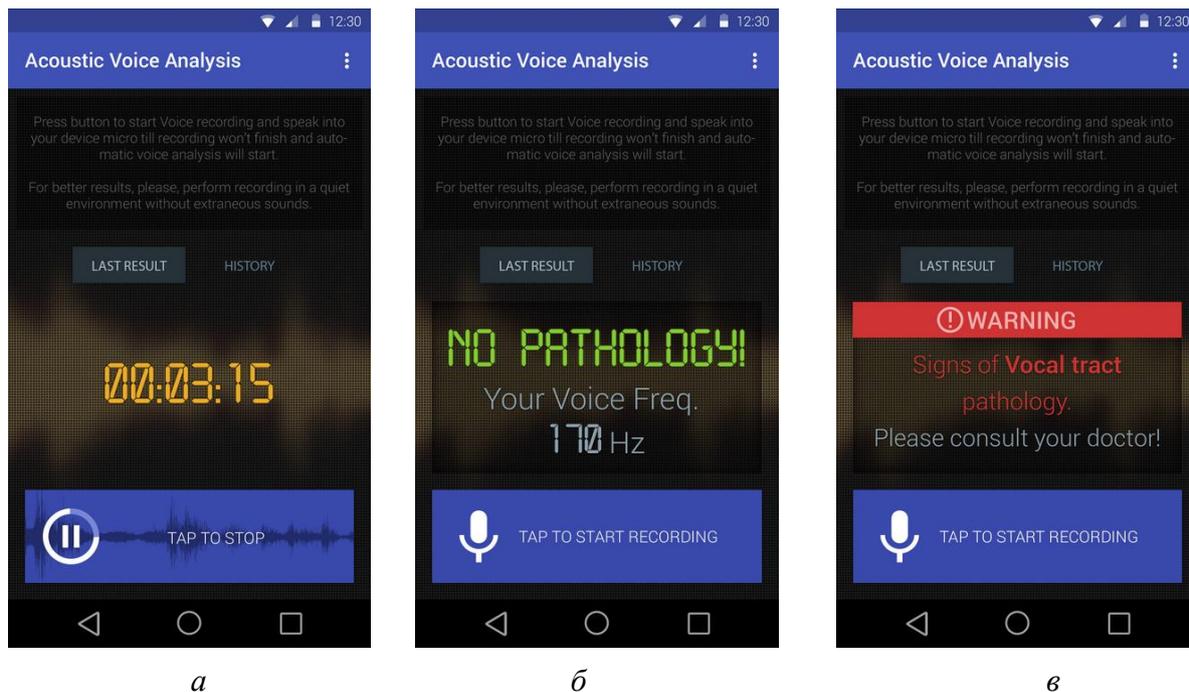
ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

Для осуществления тестирования предложенного алгоритма обнаружения патологий голосового тракта была использована база речевых сигналов, подготовленная на кафедре радиофизики и цифровых медиатехнологий совместно со специалистами Республиканского центра патологий слуха, голоса и речи. В процессе записи пациенты читали специально подобранный текст в течение нескольких минут. Каких-либо требований к произношению или отчетливости артикуляции отдельных звуков не предъявлялись. После записи каждому файлу сопоставлялся соответствующий диагноз, поставленный специалистом-фономатром в процессе медицинского осмотра пациента с помощью специализированного оборудования. Записи были отсортированы специалистами-медиками по группам «здоровый голос» и «голос с патологиями». Экспериментально полученные результаты классификации патологии голосового тракта приведены в таблице. Полученные данные свидетельствуют об эффективности и высокой точности предложенной модели обнаружения патологий.

Экспериментально полученная точность используемых классификаторов

Результат	Точность классификации, %
Метод опорных векторов	82,6
Искусственная нейронная сеть	76,1

Рассмотренный в данной работе алгоритм обнаружения патологий голосового тракта нацелен на использование в мобильных платформах как персональное приложение, что значительно расширяет круг пользователей (рис. 2). Основная задача данного приложения заключается в отображении пользователю его характеристик голоса и предупреждении о наличии риска патологии голосового тракта. Конечной точкой приложения является положительный или отрицательный результат обработанного сигнала. Цель мобильного приложения – повышение осведомленности пользователей и обращение внимания на здоровье своего голосового тракта благодаря возможности сделать экспресс-диагностику до консультации с врачом.



а

б

в

Рис. 2. Приложения для мобильной платформы Android:

а – процесс записи речевого сигнала, анализ и обработка входного сигнала;
б – положительный результат программы; *в* – отрицательный результат программы

ЗАКЛЮЧЕНИЕ

Разработанная модель автоматического обнаружения патологий голосового тракта, включающая этапы предварительной обработки и сегментации речевых сигналов, обеспечивает максимально удобный и быстрый способ анализа речи пациента в реальном времени. Предложенные алгоритм построения признаковового описания речевого сигнала как функции мел-частотных кепстральных коэффициентов и алгоритм классификации на основе метода опорных векторов позволили получить точность обнаружения патологий голосового тракта равную 82,6 %, что на 6,5 % выше точности обнаружения на основе многослойной интеллектуальной нейронной сети. Полученные данные свидетельствуют об эффективности предложенной модели и о хорошей перспективе улучшения используемых методов.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Панченко Д. П., Ле Н. В. Предварительная обработка речевых сигналов для системы распознавания речи // Молодой ученый. 2011. № 5. Т.1. С. 74.
2. Computing mel-frequency cepstral coefficients on the power spectrum / S. Molau [et al.] // IEEE International Conference on Acoustics, Speech, and Signal Processing. 2001. Vol. 1. С.73.
3. Dhanalakshmi P., Palanivel S. and Ramalingam V. Classification of Audio Signals Using SVM and RBFNN, Expert Systems with Applications. 2009. Vol. 36, № 3. P. 6069-6075.
4. Salhi L, Talbi M, Cherif A. Voice Disorders Identification Using Hybrid Approach: Wavelet Analysis and Multilayer Neural Networks. WASET 2008. P. 330.
5. Laurene V. Fausett Fundamentals of Neural Networks: Architectures, Algorithms And Applications. 1994. С. 11, 26.
6. Nicola's S., Juan I. G. Biomedical Signal Processing and Control. 2006. Vol. 46. P. 120.

7. Solera-Urena R., Martín-Iglesias D., Gallardo-Antolín A. Robust ASR using Support Vector Machines // *Speech Communication*. 2007. P. 253
8. Widrow B., Lehr M. A. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation // *Proceedings of the IEEE*. 1990. Vol. 78, № 9, September. P. 1415–1442.