

Исследование речевых расстройств при помощи голосовой идентификации на основе нейронных сетей

П.А. МЕНЬШАКОВ, И.А. МУРАШКО

Предлагается рассмотреть способ анализа состояния голоса и выявления его проблем, таких как усталость голосовых связок, повреждения или воспаления речевого тракта при помощи ранее записанного голоса диктора и нейросетевого анализа голосовых изменений.

Ключевые слова: голосовая идентификация, биометрия, система контроля доступа, нейронная сеть, быстрое преобразование Фурье.

It is proposed to consider a method for analyzing the state of the voice and identifying its problems, such as fatigue of the vocal cords, damage or inflammation of the vocal tract with the help of the previously recorded voice of the speaker and neural network analysis of voice changes.

Keywords: voice recognition, biometrics, access control systems, neural network, fast Fourier transform.

Введение. Оценка нарушения голоса включает использование оценки множественных характеристик речи (степень дыхания, напряжение, шероховатость). Одна из проблем, связанных с использованием многомерных данных – их сравнение. Для выполнения сравнения и классификации предлагается использовать самоорганизующуюся карту Кохонена. Ввиду возможности обучения без учителя, она не нуждается в целевом векторе для выходов и, следовательно, не требует сравнения с предопределенными идеальными ответами, а обучающее множество состоит лишь из входных векторов. Процесс обучения, следовательно, выделяет статистические свойства обучающего множества и группирует сходные векторы в классы. Предъявление на вход вектора из данного класса даст определенный выходной вектор.

Дисфония. Дисфония четко определяется как основное расстройство фонации, является следствием патологии голосовых связок. Есть два вида дисфонии – функциональная и повреждающая. Повреждающие дисфонии делятся на два вида: врожденная и приобретенная дисфония. Повреждающая дисфония в некоторых случаях может появиться после функциональной дисфонии.

Возможны функциональные дисфонии:

- хрипота;
- ларингит;
- воспаление гортани;
- гиперкинез;
- слизистая киста или киста оболочки;
- гипокинез;
- инфекция горла;
- глоссоплегия;
- ринофарингит.

Указанные заболевания имеют различные степени тяжести, что существенно отражается на качестве жизни человека. Таким образом, диагностирование нарушений фонации помогут определить уровень качества жизни. Шкала голосовых нарушений приведена в таблице 1.

На данный момент диагностика нарушения голосовой функции заключается в опросе и осмотре больного. При сборе анамнеза особое внимание обращают на длительность патологии, наличие сопутствующих болезней и факторов, способствующих развитию дисфонии.

Таблица 1 – Шкала голосовых нарушений

Градация нарушений	Степень нарушений	Описание	Рекомендации
Нет нарушений	0	–	–
Незначительные нарушения	1	Лингвистическое нарушение едва чувствуется или ощущается только самим пациентом	Рекомендуется речевая терапия
Умеренные нарушения	2	Снижение легкости и скорости говорения	Необходима речевая терапия
Тяжелые нарушения	3	Говорящий нуждается в помощи слушателя. Пациент часто не может быть понятым, но понимает сам	Необходима речевая терапия и помощь со стороны слушателя
Глубокие повреждения	4	Говорение фрагментарными выражениями. Слушатель должен о многом догадываться, задавать дополнительные вопросы	Необходима речевая терапия и изучение языка жестов, консультации или синтез голоса.

Основными способами диагностики дисфонии на данный момент являются физикальное и инструментальное обследование:

- исследование основных акустических параметров голоса;
- глоттография и электромиография;
- прямая или непрямая ларингоскопия;
- исследование ФВД (функции внешнего дыхания);
- эндофибрларингоскопия;
- микроларингостробоскопия;
- трахеоскопия при необходимости;
- рентгенография гортани;
- КТ;
- бактериологическое исследование отделяемого ротоглотки;
- биопсия при подозрении на онкопатологию.

Частично данные способы обследования возможно заменить при помощи сравнительного анализа векторов голосовых характеристик больного. Первый вектор характеристик голоса выступает в качестве эталонного образца, взятого в момент отсутствия заболевания, второй вектор является текущим и сравнивается на предмет отклонений от нормы.

Главным плюсом данного подхода к анализу является возможность неинвазивной диагностики заболеваний артикуляторного тракта с использованием искусственных нейронных сетей, что позволяет провести диагностику удаленно и частично автоматизировать процесс.

Аппаратная реализация. Для аппаратной реализации необходимы микрофон, фильтр и аналого-цифровой преобразователь для дальнейшей работы с цифровой записью голоса.

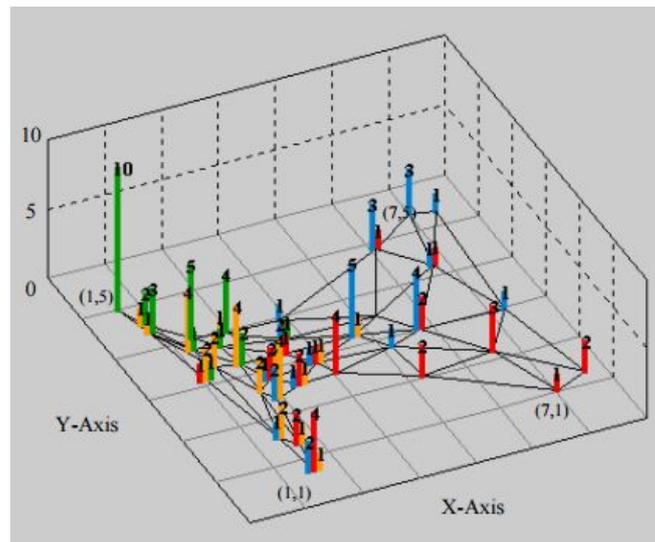
С выхода микрофона сигнал подается на вход блока фильтрации. Следующим этапом является прохождение АЦП [1]. Далее оцифрованный сигнал попадает в блок цифровой обработки. В блоке цифровой обработки сигнал фильтруется и преобразуется в вектор, с которым в дальнейшем будет работать микропроцессор и нейросетевой обработчик. Для последующего сравнения с сохраненным ранее вектором мел-кепстральных коэффициентов полученный вектор заносится в энергонезависимую память. После сравнения вектора в памяти с полученным вектором микроконтроллер подает команду на блок управления внешним устройством, к примеру, на магнитный дверной замок [2].

Сам процесс голосовой идентификации не требователен к ресурсам и состоит из двух этапов. Первым этапом является получение речевых признаков диктора и преобразование к виду, в котором его можно будет сравнить с другими. Вторым шагом является их сравнение при помощи обученной нейронной сети [3].



Рисунок 1 – Архитектура программно-аппаратных средств системы контроля доступа в помещение

Анализ нарушения речи. В исследовании, проведенном Leinonen [4], для замены непосредственного прослушивания голоса была создана шкала оценок различных степеней и форм дисфонии. Для сравнения критериев использовалась нейронная сеть без учителя, обучение которой было проведено с использованием перцепционной карты оценок нормального и дисфонического голоса [4]. Результаты эксперимента представлены на рисунке 2.



(зеленый – нормальный голос, желтый – гипотонусная, красный – гипертонусная, синий – спазматическая)

Рисунок 2 – Результаты определения дисфонии голоса

У данного подхода имеется несколько недостатков:

- Отсутствие сравнительной характеристики с предыдущим состоянием голоса. Ввиду данного недостатка невозможно отделить врожденные нарушения от приобретенных.
- Отсутствие возможности диагностирования нескольких расстройств одновременно.

Определение формы, сложности и степени нарушения речевого тракта может быть произведена не только оценкой явных факторов – ухудшение слышимости, шероховатости и прерывистости речи, но и сверкой вектора голосовых характеристик нарушенного голоса с оригинальными показателями путем включения голосового отпечатка с симуляцией той или иной степени заболевания в качестве входных параметров.

Таким образом, необходимо произвести следующий ряд действий:

- получить голосовую запись диктора;
- коррекция голоса в соответствии с таблицей оценок [5], [6];
- получение голосовых отпечатков из скорректированных голосовых записей;

- обучение нейронной сети с использованием модифицированных голосовых отпечатков [7];
- ввод текущего голоса и нейросетевое сравнение с имеющимися отпечатками [8].

Заключение. Голосовой отпечаток может быть использован не только для средств контроля доступа, но и для оценки голосовых отклонений диктора. Очевидное преимущество данного подхода заключается в отсутствии непосредственного контакта с диктором и возможность удаленной диагностики, включая использование электронных средств связи. Так же данная процедура имеет возможность полной автоматизации и нетребовательна к ресурсам.

Литература

1. Universal developing of persons identification based on RFID / G.I. Raho [et al.] // Journal of emerging trends in computing and information sciences. – 2015. – Vol. 6, № 10. – P. 592–597.
2. Меньшаков, П.А. Методика голосовой идентификации на основе нейронных сетей / П.А. Меньшаков, И.А. Мурашко // Доклады БГУИР. – 2017. – № 4 (106). – С. 12–18.
3. Adeyemo, Z.K. Development of hybrid radio frequency identification and biometric security attendance system / Z.K. Adeyemo, O.J. Oyeyemi, I.A Akanbi // International journal of applied science and technology. – 2014. – Vol. 4, № 5. – P. 190–197.
4. Leinonen, L. Dysphonia detected by pattern recognition of spectral composition / L. Leinonen // Journal of speech, language, and hearing research. – 1992. – Vol. 35. – P. 287–295.
5. Biometrics, computer security systems and artificial intelligence applications / Ed. : K. Saeed, J. Pejas, R. Mosdorf. – Springer, 2006. – 345 p.
6. You, Y. Audio coding: theory and applications / Y. You. – NY : Springer, 2010. – 349 p.
7. Herbig, T. Self-learning speaker identification: a system for enhanced speech recognition / T. Herbig, F. Gerl, W. Minker. – Berlin : Springer Verlag GmbH, 2011. – 185 p.
8. Al-Shayea, Q.K. Speaker identification: a novel fusion samples approach / Q.K. Al-Shayea, M.S. Al-Ani // International journal of computer science and information security (IJCSIS). – 2016. – Vol. 14, № 7. – P. 423–427.