

УДК 681.327.12

МОДЕЛЬ СЛУХОВОЙ СИСТЕМЫ ЧЕЛОВЕКА ДЛЯ ЗАДАЧ ИДЕНТИФИКАЦИИ ДИКТОРА

В. И. Соловьев, Я. А. Белозерова

Восточнoукраинский национальный университет им. В. Даля
Украина, г. Северодонецк
bryukhanova.ya@gmail.com

Предложена модель слуховой системы человека, которая позволила сформулировать принципы обработки информативных признаков распознавания голоса и обобщенную последовательность преобразований информационного канала голосового общения. На основе такой модели предложены математические основы сегментации голосового сигнала на вокализованные сегменты и паузы, а также выделения частоты основного тона (ОТ). Полученные результаты позволяют усовершенствовать имеющиеся алгоритмы распознавания диктора.

Ключевые слова: диктор, распознавание, голос, частота основного тона, вейвлет-анализ.

Создание алгоритмов и программно-аппаратных средств для систем компьютерного распознавания речевой информации является частью области информационных технологий и систем обработки данных. Одним из важных направлений исследований является разработка интеллектуальных систем образного восприятия речевой информации, среди которых значительную роль играют системы распознавания дикторов.

Исследованию языкового аппарата и математическому обоснованию частотных характеристик звуков речи были посвящены работы А. Белла, Г. Фанта и Д. Фланагана, Р. Якобсона [1—3]. Появление ЭВМ способствовало развитию методов цифровой обработки языкового голосового сигнала. Важную роль в этой области сыграли работы Б. Гоулда, Д. Маркела, Л. Рабинера, Д. Рэйди, Р. Шафера, Б. Янга, Х. Сако, С. Чиба, Ф. Итакура, В. М. Величко, Н. Г. Загоруйко, В. М. Сорокина. Разработанные методы базировались, в основном, на статистическом подходе с использованием скрытых Марковских цепей, критерия максимального правдоподобия и Байесовских правил. Несмотря на значительное количество публикаций в области распознавания диктора, разработанные методы и алгоритмы не имеют достаточной точности идентификации, в связи с чем исследования в этой отрасли все еще остаются актуальными.

Учитывая присущее слуховой системе человека свойство распознавать голоса, в качестве модели потенциальной системы для распознавания диктора предложена упрощенная модель слуховой системы человека.

Модель состоит из следующих блоков: блока, описывающего работу барабанной перепонки, молоточка и наковальни; блока, представляющего работу слуховой улитки, и блока восприятия.

Блок предварительной обработки сигнала отражает работу барабанной перепонки, молоточка и наковальни, выполняет функции фильтрации шума речевого сигнала, поступающего на его вход.

Сегмент, реализующий работу улитки уха, представляется двумя взаимосвязанными блоками: блоком фрактального анализа и блоком спектрального анализа. Блок фрактального анализа отражает действие улитки уха, сам по себе представляет собой фрактальную самоподобную структуру, которая выполняет сегментацию сигнала на вокализованные сегменты и паузы. Блок спектрального анализа реализует поиск самоподобных структур с целью уплотнения информации в сигнале и определения основных особенностей голоса диктора.

Блок восприятия отражает действие слуховых нейронов, соединенных с волосковыми клетками базилярной мембраны уха. Она учитывает такие слуховые эффекты, как однополупериодное выпрямление и регулирование усиления.

Рассмотрим математическое представление основных блоков, выполняющих обработку речевого сигнала ТС.

Учитывая значительную проработанность вопросов фильтрации голосового сигнала и системы восприятия голоса, не будем останавливаться на аспектах работы этих блоков рассматриваемой модели.

Определим принципы необходимости использования блока фрактального анализа при исследовании голосового сигнала.

Согласно теореме вложения вектор, составленный из образцов исходного сигнала и замедляемый кратно постоянной задержке по времени, определяет движение в реконструированном многомерном пространстве, которое разделяет общие аспекты с исходной фазой — пространством. В частности, инвариантные величины допустимой динамической системы, такие как фрактальные размерности, сохраняются в реконструированном пространстве. Таким образом, изучая конструктивную динамическую систему, мы можем раскрыть полезную информацию о сложности оригинала неизвестной динамической системы, связанной с этими инвариантными величинами. Указанное возможно при условии, что раскрытие динамики успешно, например, размерность вложения достаточно велика. Однако теорема вложения не указывает методы определения необходимых параметров, а только устанавливает ограничения на их значения. Например, размерность вложения должна быть в два раза больше, чем окно подсчета размерности многомерного набора. Таким образом, важнейшим инструментом преобразования звукового сигнала является оценка фрактальной размерности фрагментов голосового сигнала, которая сохраняет основные его характеристики при любом преобразовании. В предлагаемой модели фрактальный анализ используется для сегментации фрагментов сигнала на вокализованные и невокализованные участки [3].

Предположим, что возможность узнавать человека по голосу является способностью выделять в фрагменте голоса характерные признаки, присущие диктору. При этом, идентификация личности является действием по выделению признаков, повторяющихся в сигнале (то есть постоянно присутствующих в сигнале), а значит можно сделать вывод, что в сигнале должны быть постоянные самоподобные структуры, формирующиеся при речеобразовании каждого конкретного человека. Подобие структур самим себе возможно за счет их масштабирования в языковых фрагментах. Поэтому наиболее удачным инструментом выявления самоподобных структур можно считать методы, основанные на спектральном разложении сигнала, в частности вейвлет-анализ. В связи с этим в создаваемой модели слуховой системы человека присутствует блок именно такого анализа. В работе предлагается алгоритм определения частоты основного тона на основе значения коэффициентов трехмерного вейвлета Морле речевого сигнала на отрезках, где наблюдаются экстремумы корреляции частоты основного тона.

Таким образом, в работе на основе анализа структуры слуховой системы человека предложен набор математических методов преобразования голосового сигнала с целью последующей идентификации диктора. Учитывая физическое подобие алгоритмов преобразования сигнала в модели и в реальной слуховой системе человека, ожидается значительное улучшение качества распознавания дикторов в сравнении с имеющимися методами.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Фланаган Дж. Л. Анализ, синтез и восприятие речи.— Москва: Связь, 1968.
2. Фант Г. Акустическая теория речеобразования.— Москва: Связь, 1964.
3. Винцюк Т. К. Анализ, распознавание и интерпретация речевых сигналов.— Киев: Наукова Думка, 1987.
4. Соловьев В. И., Белозерова Я. А. Использование фрактальной размерности аудиофайлов в задаче сегментации звукового файла // Вісник Східноукраїнського національного університету ім. В. Даля.— 2013.— № 5 (194). Ч. 2.— С. 165–169.

V. I. Solovjov, Ya. A. Byelozorova

A model of human auditory system for speaker identification.

A model of the human auditory system was proposed, which allowed formulating the principles of treatment of informative features of voice recognition and synthesis of a sequence of transformations of the information channel of voice communication. Based on the model, the mathematical foundations were proposed for the segmentation of the voice signal to voiced segments and pauses, as well as for allocation of the pitch frequency. The obtained results make it possible to improve existing speaker recognition algorithms.

Keywords: *speaker, recognition, voice frequency, wavelet analysis.*