

УДК 004.89, 004.93

ВЫДЕЛЕНИЕ ИДЕНТИФИКАЦИОННЫХ ХАРАКТЕРИСТИК ДИКТОРА НА ОСНОВЕ ВЕЙВЛЕТ-РАЗЛОЖЕНИЯ ГОЛОСОВОГО СИГНАЛА

Белозерова Я.А.

THE ISOLATION SPEAKER IDENTIFICATION CHARACTERISTICS ON THE BASIS OF WAVELET DECOMPOSITION OF VOICE SIGNAL

Byelozorova Ya.

В статье рассмотрены математические основы выделения уникальных идентификационных признаков диктора, основанных на характерных сингулярностях в структуре голосового сигнала диктора. Идентификация сингулярностей выполняется путем вейвлет-разложения голосового сигнала, с последующей его сегментацией и классификацией особенностей. На основе выделенных сингулярностей голосового сигнала определяется частота основного тона, являющаяся в работе основной характеристикой идентификации диктора. Проведенные исследования эффективности идентификации диктора предлагаемым методом выделения характеристик голосового сигнала и расчета на их основе частоты основного тона показали достаточно высокую степень точности в идентификации диктора.

Ключевые слова: голосовой сигнал, распознавание, идентификация диктора, вейвлет-разложение.

Введение. Слух является одной из важнейших способностей биологических организмов воспринимать звуки органами слуха и имеет большое значение для понимания и коммуникации. В процессе обработки и понимания звуковых данных важной способностью является идентификация, как отдельных звуковых фрагментов, так и говорящего в целом.

Постановка проблемы. Развитие человечества все больше требует создания сложных программно-аппаратных систем, в возможности которых должны закладываться способности идентификации звуков, речи, а также личности самого диктора. Причем учитывая, что задачи идентификации звуков и речи в достаточной степени исследованы, то наибольшее значение имеет изучение аспектов идентификации диктора.

Анализ последних исследований и публикаций. Исследованию идентификации диктора посвящено множество работ, которые используют ряд математических средств, таких как сопоставление с эталонами, алгоритмы линейного выравнивания и

динамического искажения времени, марковские модели, нейронные сети, векторное квантование и построение кодовой книги [1-7]. В работах Х. Найквиста, К.Шеннона, Э.Т. Уиттекера, Л.Р. Рабинера, Р.В. Шафера, В.А. Котельникова, Л.М. Гольденберга, В.П. Яковлева, М.В. Назарова, Ю.Н. Прохорова неоднократно говорилось о том, что основные характеристики звукового сигнала могут быть выделены на основе вейвлет-анализа, а также бандаж-фильтров различного типа. Однако при практической реализации на основании теоремы отсчетов В.А. Котельникова возникают ошибки наложения, вызванные нефинитностью спектра сигнала; усечения, обусловленные конечным числом отсчетов, и округления, связанные с неточностью представления отсчетных значений в цифровом виде.

Таким образом, актуальной является задача разработки и исследования алгоритмов дискретной обработки речевых сигналов на основе вейвлет-представления с целью выделения характеристик, необходимых для идентификации диктора.

Цель статьи. В работе предложена модель построения системы идентификации характерных особенностей голосового сигнала для определения диктора.

Результаты исследований. При выполнении основных операций выделения характерных особенностей голосового сигнала необходимо проведение некоторых дополнительных операций в частности фильтрация и сегментация на вокализованные и невокализованные участки. Для выполнения подобных операций в работе использовались методы и алгоритмы, рассмотренные в работах [8,9].

Для выделения характерных особенностей сигнала не обходимо выполнить его математическое описание. Согласно выполненного анализа в данном случае наиболее эффективным подходом является вейвлет-разложение сигнала по определенному базису. Вейвлет – преобразование раскладывает сиг-

налы по растянутым и сдвинутым вейвлетам ψ . Так как вейвлет ψ имеет нулевое среднее значение, вейвлет – интеграл

$$Wf(a, b) = \int f(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt,$$

измеряет изменение f в окрестности точки b , размер которой пропорционален a . При стремлении масштаба a к нулю вейвлет – коэффициенты характеризуют свойства функции f в окрестности b . Если функция f дифференцируема на $[v-l; v+l]$ а $\rho_v(t)$ – многочлен Тейлора в окрестности v , то

$$\rho_v(t) = \sum_{k=0}^{m-1} \frac{f^{(k)}(v)}{k!} (t-v)^k.$$

Погрешность такой аппроксимации $\varepsilon_v(t) = f(t) - \rho_v(t)$ удовлетворяет условию

$$\forall t \in [v-l; v+l] \quad \left| \varepsilon_v(t) \right| \leq \frac{|t-v|^m}{m!} \sup_{u \in [v-l; v+l]} \left| f^{(m)}(u) \right|.$$

Порядок дифференцируемости f в окрестности v определяет верхнюю границу погрешности $\varepsilon_v(t)$ при t , стремящемся к v .

Естественно, что убывание амплитуды вейвлет – преобразования в зависимости от масштаба связано с равномерной и точечной гладкостью Липшица сигнала. Показатели Липшица могут произвольно меняться от точки к точке. Характеризовать гладкость f в точке v бывает достаточно трудно, потому что f может иметь различные виды особенностей, присутствующих в окрестностях точки v . На основании гладкости Липшица также можно сделать заключение, что когда масштаб убывает, амплитуды вейвлет – коэффициентов имеют быстрое убывание до нуля в областях, где сигнал гладкий.

Для идентификации особенностей в звуковом сигнале будем считать, что функция f в окрестности точки v имеет изолированную характерную особенность, если

$$|Wf(s, u)| > As^{\alpha+1/2}, \tag{1}$$

где A – коэффициент, s – масштаб.

На основании ранее проведенных исследований установлено, что если $|Wf(s, u)|$ не будет иметь локальных максимумов на достаточно малых масштабах, то f – является локально гладкой функцией и операция выделения характерных изолированных

особенностей этой функции f может быть построена путем определения максимальных значений функции $|Wf(s, u)|$ на малых масштабах. При этом необходимо учитывать, что при обработке дискретных данных наименьший масштаб ограничен шагом выборки дискретного сигнала, используемого при вычислениях.

После выделения особенности стоит задача ее классификации, на основании которой и будет приниматься решение о ее учете в задаче идентификации диктора.

Неравенство (1) эквивалентно неравенству

$$\log_2 |Wf(s, u)| \leq \log_2 A + \left(\alpha + \frac{1}{2}\right) \log_2 s, \tag{2}$$

поэтому гладкость Липшица в точке v определяется наклоном $\log_2 |Wf(s, u)|$ как функции $\log_2 s$ вдоль линии максимумов, сходящихся к v . Под линией максимумов здесь понимается любая связная кривая $s(u)$ в пространственно-масштабной плоскости (s, u) , вдоль которой все точки – это точки максимумов модуля.

Построим операцию классификации выделенных особенностей в сигнале на основе неравенства (2) следующим образом.

Обозначим $O_v(s, u)$ – линию максимумов, сходящуюся к точке $u=v$, при $s \rightarrow 0$. Для каждой такой точки v определим угол наклона $\log_2 O_v(s, u)$ как функции $\log_2 s$ при $s \rightarrow 0$:

$$\log_2 O_v(s, u) = \log_2 A + \left(\alpha + \frac{1}{2}\right) \log_2 s. \tag{3}$$

Будем считать, что в точке $u=v$ имеем особенность вида α .

При этом необходимо учитывать, что решение задачи классификации особенности зависит от свойств базисной функции вейвлета ψ .

Важность такого метода состоит в том, что разномасштабные составляющие сигнала $y[n]$ рассортировываются и хранятся в различных пространствах $W_{-j_i, n}$. Операцию выделения изолированных особенностей случайной функции y построим на анализе детализирующих коэффициентов $\bar{d}^{-j_i} = \left\{ d_n^{-j_i} \right\}_{n \in Z}$, которые для малых $-j_i$ определяют мелкомасштабные компоненты сигнала.

В соответствии с теоремой Жаффара при уменьшении масштаба $-j_i$, абсолютные значения коэффициентов \bar{d}^{-j_i} , определяющих компоненту

$g \left[2^{j_i} \right]$, имеют быстрое убывание до нуля в областях сигнала, не содержащих изолированных особенностей его структуры. Процедуру выделения изолированных особенностей построим на проверке условия

$$\left| d_n^{-j_i} \right| = \left| \left\langle f, \psi_{-j, n} \right\rangle \right| \leq T', \quad (4)$$

где T' - пороговое значение, определяющее наличие в сигнале изолированной особенности, на малых масштабах $-j_i$. В случае невыполнения условия (4) в точке, $n = n_0$ будем считать, что в точке $n = n_0$ имеет изолированную особенность.

Выделенный представленным образом набор характерных особенностей голосового сигнала, который позволяет на основе максимумов коэффициентов вейвлет-преобразования выполнить идентификацию частоты основного тона (ОТ).

Анализ показывает, что расположение вейвлет-максимумов по временному параметру строго соответствует локальным максимумам амплитуды звуковой волны во временной области. При этом локальные максимумы соответствуют всплескам амплитуды звуковой волны, обусловленными частотой ОТ.

Анализ представленных зависимостей показал, что при различных значениях величин максимумов по частоте при совмещении графиков частотных зависимостей по положению максимумов они обладают высокой степенью подобия после соответствующих нормировок для одних и тех же характеристик голоса. Окончательное принятие гипотезы об идентичности характеристик голоса двух речевых фрагментов принимается после проверки гипотезы об идентичности распределений частотных зависимостей.

Суммарные результаты испытаний по оценке обобщенной ошибки идентификации диктора, по десяти дикторам мужчинам, десяти дикторам женщинам, для 300 языковых фрагментов для каждого диктора. Обобщенная ошибка вычисляется по нормализованному коэффициенту корреляции с единичной задержкой для расчетов по каждому методом с последующим суммированием. Выполненное исследование показало достаточно высокую степень идентификации диктора с ошибкой, не превышающей 20%, что является очень большим для метода не учитывающего фильтрацию и нормализацию сигнала.

Вывод. Таким образом, предложен алгоритм выделения идентификационных характеристик, которые позволяют определить частоту ОТ, сравнительный анализ которой позволяет выполнить идентификацию диктора на различных временных интервалах на основании проведенного вейвлет-разложения сигнала.

Литература

1. Гольденберг Л.М., Матюшкин Б.Д., Поляк М.П. Цифровая обработка сигналов. - М.: Радио и связь, 1990. - 256с.
2. Пазаров М.В., Прохоров Ю.П. Методы цифровой обработки и передачи речевых сигналов. - М.: Радио и связь, 1985. - 176 с.
3. Фант Г. Акустическая теория речеобразования. - М.: Паука, 1964. - 284 с.
4. Фланган Дж. Анализ, синтез и восприятие речи: Пер. с англ. / Под ред. А.А. Пирогова. - М.: Связь, 1968. - 396с.
5. Маркел Дж.Д., Грей А.Х. Линейное предсказание речи. - М.: Связь, 1980.-308 с.
6. Цвикер Э., Фельдкеллер Р. Ухо как приемник информации. 2-е изд., перераб. и доп. - М.: Связь, 1971. - 255с.
7. Солодовников А.И., Спиваковский А.М. Основы теории и методы спектральной обработки информации. - Л.: Ленинград, 1986. - 272 с
8. Соловьев В.И., Брюханова Я.А. Идентификация заданных фрагментов в звуковых файлах // Вісник Східноукраїнського національного університету ім.В.Даля. – 2008. - №9(127) ч.2 – с. 30-33
9. 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

References

1. Goldenberg L.M., Matyuschkin B.D., Polyak M.P. Cifrovaya obrabotka signalov. - M. - Radio i svyaz - 1990. - 256p.
2. Pazarov M.V., Prohorov Yu.P. Metodi cifrovoi obrabotki i peredachi rechevikh signalov. - M. - Radio i svyaz - 1985. - 176 p.
3. Fant G. Akusticheskaya teoriya recheobrazovaniya. - M.- Pauka - 1964. - 284 p.
4. Flanagan Dj. Analiz sintez i vospriyatie rechi_ Per. s angl. / Pod red. A.A. Pirogova. - M.- Svyaz - 1968. - 396p.
5. Markel Dj.D. Grei A.H. Lineinoe predskazanie rechi. - M. - Svyaz - 1980. - 308 p.
6. Cviker E., Feldkeller R. Uho kak priemnik informacii. 2_e izd.pererab. i dop. - M. - Svyaz - 1971. - 255p.
7. Solodovnikov A.I., Spivakovskii A.M. Osnovi teorii i metodi spektralnoi obrabotki informacii. - L.: Leningrad - 1986. - 272 p
8. Solov'yev V.I., Bryukhanova YA.A. Identifikatsiya zadannykh fragmentov v zvukovykh faylakh // Visnik Shidnoukrains'kogo nacional'nogo universitetu im. V. Dalja. – 2008. - №9(127) part.2 – p. 30-332.
9. Solov'yev V.I., Byelozorova YA.A. Ispol'zovaniye fraktal'noy razmernosti audiofaylov v zadache segmentatsii zvukovogo fayla // Visnik Shidnoukrains'kogo nac. un-tu im. V. Dalja. – 2013 – №5(194). part.2. – p. 165-169

Белозерова Я.А. Виділення ідентифікаційних характеристик диктора на основі вейвлет-розкладання мовного сигналу.

У статті розглянуто математичні основи виділення унікальних ідентифікаційних ознак диктора, заснованих на характерних сингулярностях в структурі голосового сигналу диктора. Ідентифікація сингулярностей виконується шляхом вейвлет-розкладання голосового сигналу.

лу з подальшою його сегментацією і класифікацією особливостей. На основі виділених сингулярностей голосового сигналу визначається частота основного тону, що є в роботі основною характеристикою ідентифікації диктора. Проведені дослідження ефективності ідентифікації диктора пропонуваним методом виділення характеристик голосового сигналу і розрахунку на їх основі частоти основного тону показали досить високий ступінь точності в ідентифікації диктора.

Ключові слова: голосовий сигнал, розпізнавання, ідентифікація диктора, вейвлет-розкладання.

Byelozorova Ya. The isolation speaker identification characteristics on the basis of wavelet decomposition of voice signal.

The article describes the mathematical foundations allocate unique identification signs speaker based on the characteristic structure of the singularities in the voice signal speaker. The identification of singularities is performed by wavelet decomposition of the voice signal with its subsequent segmen-

tation and classification features. On the basis of isolated singularities of the voice signal is determined by the frequency of the pitch, which is in the main characteristic of speaker identification. Studies have suggested efficacy speaker identification method of isolation characteristics of the voice signal and calculating the frequency based on their pitch showed reasonably high degree of accuracy in the speaker identification. As an improvement of this method is proposed to introduce additional criteria to improve the accuracy of classification of singularities.

Keywords: voice signal recognition, speaker identification, wavelet decomposition.

Бєлєзьорова Я.А. – старший викладач кафедри «Кібернетики та комп'ютерних систем» СХУ ім. В. Даля, e-mail: bryukhanova.ya@gmail.com.

Рецензент: д.т.н., проф. Марченко Д.М.

Стаття подана 01.04.2015