

Структура системы распознавания речи на основе вейвлет-преобразования



*Гапочкин Артём Владимирович,
Северо-Кавказский федеральный университет,
г. Ставрополь*

E-mail: warrior_555@rambler.ru

Аннотация: В данной работе рассматривается структура системы распознавания речи на основе вейвлет-преобразования. Использование новых методов предобработки речевых сигналов на основе вейвлет-преобразования позволяет существенно повысить точность системы распознавания речевого сигнала.

Ключевые слова: речь, сигналы, вейвлет-преобразования.

В связи со стремительным развитием информационных технологий в течение последних десятилетий значительно расширился круг задач, решаемых с помощью вычислительной техники. Также более разнообразными стали способы взаимодействия человека с различного рода электронными информационными системами. В частности речевые пользовательские интерфейсы внедряются в разнообразные автоматизированные системы.

Речь является наиболее естественной формой обмена информацией между человеком и машиной. Преимущества использования устной речи для общения с машинами сказываются также в снятии некоторых ограничений, которые присущи традиционным устройствам ввода-вывода, например, в освобождении рук и фиксированного положения тела оператора, необходимости использования клавиатуры и дисплея. Для многих миниатюрных устройств, таких как мобильные телефоны и ладонные компьютеры наличие встроенных систем распознавания и синтеза речи является единственным удобным способом обмена информацией. Несмотря на то, что реализация речевого диалога с компьютерами на естественном разговорном языке до сих пор остается нерешенной задачей, современные методы цифровой обработки речевых сигналов успешно позволяют частные прикладные задачи из этой области. [3].

Распознавание речи – это многоуровневая задача распознавания образов, в которой акустические сигналы анализируются и структурируются в иерархию

структурных элементов (например, фонем), слов, фраз и предложений. Каждый уровень иерархии может предусматривать некоторые временные константы, например, возможные последовательности слов или известные виды произношения, которые позволяют уменьшить количество ошибок распознавания на более низком уровне. Чем больше мы знаем (или предполагаем) априорной информации о входном сигнале, тем качественнее мы можем его обработать и распознать. Обобщенная структурная схема системы распознавания речи представлена на рисунке 1.

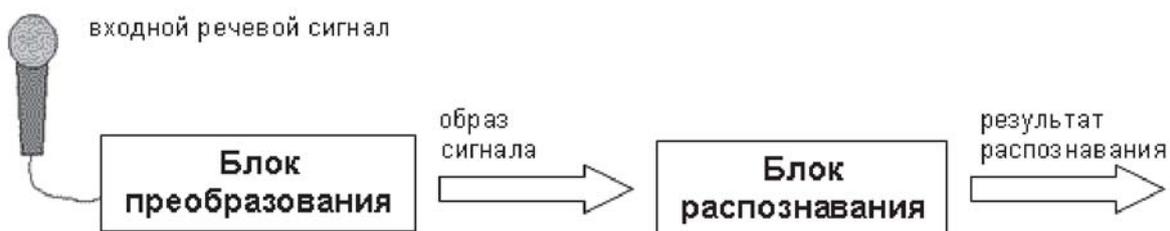


Рис. 1 Общая структурная схема системы распознавания речи

Обычно, речевой сигнал, записанный с высокой дискретизацией (20 КГц при записи с микрофона либо 8 КГц при записи с телефонной линии) имеет вид представленный на рисунке 2 [2].

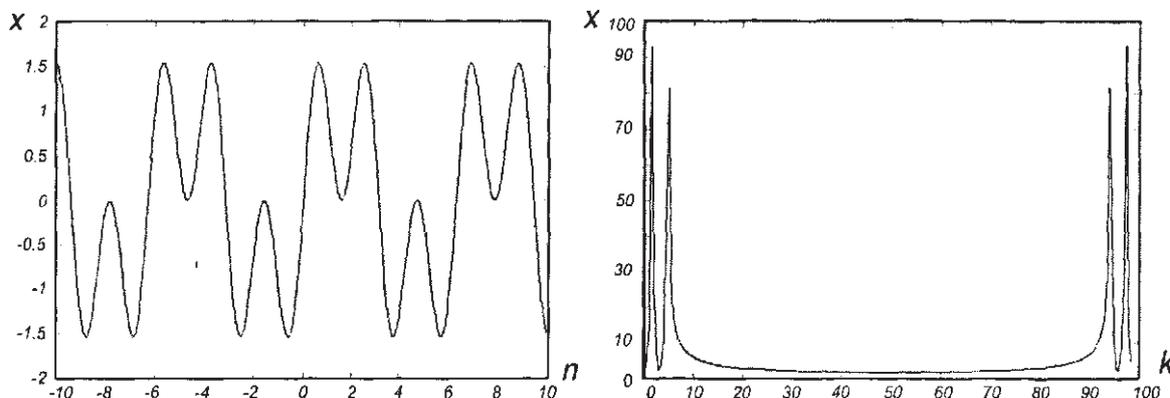


Рис. 2 Речевой сигнал и его спектр

Основная задача анализа речевого сигнала состоит в вычислении совокупности параметров (признаков), которые содержат информацию, используемую при распознавании. Есть различные методы для извлечения полезных параметров и сжатия исходных данных в десятки раз без потери полезной информации. Наиболее эффективным и популярным является метод вейвлет - преобразования речевого сигнала: Вейвлеты и вейвлетное преобразование появились как разви-

тие Фурье-преобразования для решения парадокса всегда существующих волн. Вейвлеты так же, как синусы и косинусы, относятся к ортогональным функциям. В остальном вейвлеты отличаются от них. Во-первых, вейвлеты конечны, во-вторых, вместо амплитудного масштабирования в вейвлет-преобразовании используется временная «растяжка», в-третьих, операция сложения в Фурье-преобразовании заменяется временными сдвигами и масштабированием.

Вейвлет-анализ определяется выражением:

$$\gamma(s, \tau) = \int f(t) \Psi_{s, \tau}^*(t) dt \quad (1)$$

где $f(t)$ – исходный сигнал, а $\Psi_{s, \tau}(t)$ – базовая функция, называемая вейвлетом. Вейвлеты получают от одного основного вейвлета $\Psi(t)$, называемого материнским, через масштабирование и сдвиги:

$$\Psi_{s, \tau}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-\tau}{s}\right), \quad (2)$$

где s – коэффициент масштабирования, а τ – временной сдвиг.

Обратное вейвлет-преобразование имеет вид:

$$f(t) = \iint \gamma(s, \tau) \Psi_{s, \tau}(t) dt ds . \quad (3)$$

Из данных выражений следует возможность выбора в качестве материнского вейвлета функции, которая в наибольшей степени удовлетворяет нуждам задачи, если только она удовлетворяет определенным условиям. Одним из таких условий является условие допустимости и непрерывности[4]:

$$\int \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < +\infty . \quad (4)$$

Из выражения (4) следует, что вейвлет должен иметь представление, подобное спектру. Так же важно, чтобы среднее значение вейвлета было равно нулю, т.е.

$$\int \Psi(t) dt = 0 . \quad (5)$$

Данные преобразования предназначены для аналоговой обработки. Вместе с тем компьютерные системы имеют дело с дискретными представлениями непрерывных сигналов. В выражении (1) вейвлетное преобразование вычисляется непрерывным сдвигом и масштабированием функции и вычислением корреляции между ними. Но основной проблемой является невозможность аналитически вычислить результат вейвлет-преобразования и их вычисляют численными методами. Из всего вышесказанного следует, что в дискретном преобразовании сигнал масштабируется и сдвигается не непрерывно, а только с определенным дискретным шагом:

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \Psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right), \quad (6)$$

где j и k – целые числа и $s_0 > 1$ – фиксированный шаг. Коэффициент сдвига τ_0 определяется шагом. Обычно s_0 выбирается равным 2 как наиболее естественный для компьютерных реализаций. Значение для τ_0 обычно выбирается равным 1.

Для улучшения качества распознавания, в кадры может быть добавлена информация о первой или второй производной значений их коэффициентов для описания динамики изменения речи (рис.2).

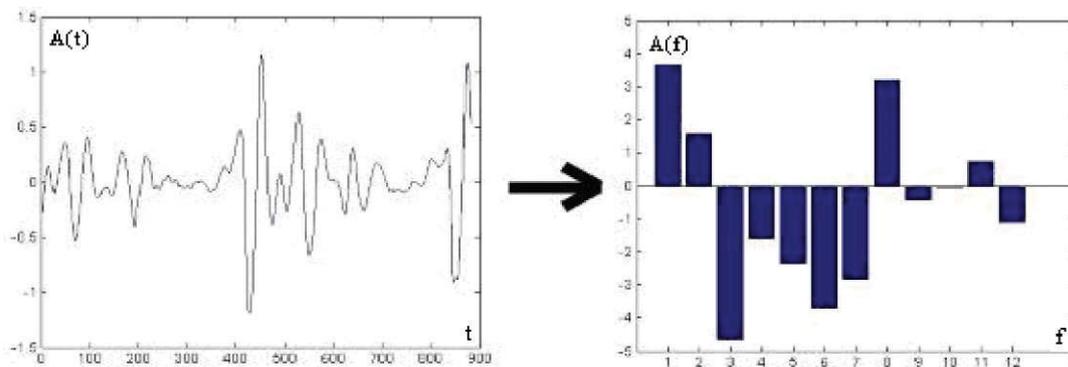


Рис. 2. Вейвлет-преобразование речевого сигнала

Когда дискретные вейвлеты используются для преобразования непрерывного сигнала, результатом будет серия коэффициентов вейвлета, называемая последовательной декомпозицией вейвлета. Важным в схеме декомпозиции есть вопрос восстановления. В [5] доказывается, что необходимое и достаточное условие для восстановления состоит в том, что энергия

коэффициентов вейвлета должна лежать между двумя положительными границами, то есть:

$$A\|f\|^2 \leq \sum_{j,k} \left| \langle f, \Psi_{j,k} \rangle \right|^2 \leq B\|f\|^2, \quad (7)$$

где $\|f\|^2$ – энергия сигнала $f(t)$; $A > 0$; $B < \infty$ и A, B – независимы от $f(t)$.

Когда условие (7) выполняется, семейство функций $\Psi_{j,k}(t)$ с $j, k \in Z$ называется фреймом, а A и B – границами фрейма. Когда $A = B$, фрейм сжат и дискретные вейвлеты ведут себя подобно ортонормальной основе. Когда $A \neq B$, точное восстановление возможно за счет двойного фрейма. В двойном фрейме дискретного вейвлета преобразующий вейвлет отличается от восстанавливающего вейвлета.

На последнем шаге нужно сделать дискретные вейвлеты ортогональными. Дискретные вейвлеты могут быть сделаны ортогональными в своих растяжениях и сдвигах специальным выбором «материнского» вейвлета:

$$\int \Psi_{j,k}(t) \Psi_{m,n}^*(t) dt = \begin{cases} 1, & \text{если } j = m \text{ и } k = n, \\ 0, & \text{в других случаях} \end{cases} \quad (8)$$

Произвольный сигнал может восстанавливаться суммированием ортогональных базисных функций, с весом коэффициентов вейвлетного преобразования:

$$f(t) = \sum_{j,k} \gamma(j,k) \Psi_{j,k}(t) \quad (9)$$

Можно сделать вывод, что использование новых методов предобработки речевых сигналов на основе вейвлет-преобразования позволяет существенно повысить точность системы распознавания речевого сигнала. Из указанного следует необходимость проведения дальнейшего исследования о применении теории вейвлет-преобразований в качестве возможной альтернативы существующим методам.

Литература:

1. Блаттер, К. Вейвлет-анализ. Основы теории/К. Блаттер. - М.: Техносфера, 2004. - 280 с.
2. Винцюк, Т.К. Анализ, распознавание и интерпретация речевых сигналов/Т.К. Винцюк. - Киев: Нав. думка, 1987. - 264 с.
3. Жожикашвили В.А. и др. Применение распознавания речи в автоматизированных системах массового обслуживания// Автоматизация и современные технологии, 2003, №11, с. 23-29.
4. Чуи К. Введение в вейвлеты: Пер. с англ. М.: Мир, 2001. 412 с.
5. Daubechies, I. Ten Lectures on Wavelets. 2nd ed. Philadelphia: SIAM, 1992.