

Н. М. Быков, к. т. н., доц.; В. В. Ковтун, к. т. н.; Н. Г. Савинова, студ.

НАДЕЖНЫЙ МЕТОД ВЫДЕЛЕНИЯ СЛОГОВЫХ СЕГМЕНТОВ В РЕЧЕВОМ СИГНАЛЕ

Предложен новый метод, который повышает надежность выделения слоговых сегментов в речевом сигнале, и разработаны алгоритм и устройство, которые реализуют предложенный метод.

Ключевые слова: *распознавание речи, речевой сигнал, сегментация сигнала, признаки образов, частотный диапазон, устройство выделения слогов, алгоритм выделения признаков слогов.*

Вступление

Использование информации о слогах в иерархических системах распознавания речи позволяет повысить дикторонезависимость и контекстнезависимость этих систем [1], и, как следствие, повысить точность распознавания. В данной работе предложен новый метод, который повышает надежность выделения слоговых сегментов в речевом сигнале, и разработаны алгоритм и устройство, реализующие предложенный метод.

Постановка задачи

Невзирая на то, что сегодня уже существуют коммерческие системы автоматического распознавания языка, актуальной остается проблема разработки таких методов и средств распознавания, которые не требуют настройки под индивидуальные речевые особенности диктора. Одним из подходов к построению дикторонезависимых систем распознавания является использование на верхних уровнях иерархической системы в качестве элементов распознавания фонотипов, характеристики которых мало зависят от диктора и от контекста, например, слогов, полуслогов, звонких, фрикативных, пауз и др. В работе [2] исследовано, что использование, например, только слоговой информации позволяет уже на верхнем уровне распознавания сократить количество кандидатов на классификацию в 2 – 4 раза. Такой информацией является длительность слогов и их количество в высказывании. Одним из основных параметров, которые используются для разграничения слогов в речевом сигнале, является его энергия [3, 4]. При этом ядро слога определяется в месте локализации максимумов энергии, ограниченных существенными (на 40 или 50 дБ) спадами энергии. Однако в ряде работ, например [5], отмечается частое выделение по этому признаку ошибочных слогов, сформированных высокоэнергетическими фрикативными или сонорными звуками. Это подтверждают и экспериментально снятые записи. В работе [6] в качестве параметра для выделения признака слога используется функция "громкости", получаемая как взвешенная сумма амплитуд сигналов 22-х частотных каналов, размещенных в критических полосах. Очевидный недостаток такого метода формирования признака слога – большие аппаратные или вычислительные затраты и недостаточная надежность.

Для исключения недостатков, свойственных упомянутым методам, в нашей работе ставится **цель** разработки нового метода, который позволил бы повысить надежность выделения сегментов языкового сигнала, соответствующих слогам речи, а также алгоритма и устройства, реализующих этот метод. В следующем разделе рассмотрена модель, позволяющая определить информативные признаки слоговых сегментов, на основе которой разрабатываются метод, алгоритм и устройство для выделения этих сегментов.

Математическая модель и метод выделения признаков слогов из языкового сигнала

Раскрытие механизмов отображения спектральных параметров речевого сигнала в пространстве инвариантных признаков возможно путем дополнения процессов его обработки периферийной слуховой системы процессами в центральной слуховой системе.

Структурная схема, дополненной таким образом модели слуха, представлена на рис. 1.



Рис. 1. Обобщенная модель слуховой системы

В приведенной на рис. 1 обобщенной слуховой модели блок спектрального анализа (БСА) отображает частотно-выборочные свойства модели и представляет собой набор фильтров, на вход которых подается речевой сигнал $S(t)$. Обычно фильтры перекрывают частотный диапазон от 250 до 6400 Гц, в котором сосредоточена основная энергия речевого сигнала.

Модель сенсорных слуховых нейронов (МССН) отображает действие слуховых нейронов, которые соединены с волосковыми клетками базилярной мембраны уха. Она учитывает такие слуховые эффекты, как динамическое сжатие входного сигнала, однополупериодное его выпрямление и регулировка усиления сигнала. Механизмы слухового восприятия, которые представлены данными моделями, достаточно хорошо изучены, однако многочисленные попытки их применения в распознающих устройствах не дали желаемых результатов. Поэтому в работе рассматривается уточненная модель слуховой системы с учетом работы нейронной сети, представленной ее моделью (МНС).

Особенности формирования речевых образов моделью нейронной сети с сигналами $\{y(t)\}$ на выходах нейронов проанализированы в работах [7, 8]. Проведенный анализ показал, что признаки речевого сигнала необходимо искать среди элементов автокорреляционной матрицы спектральных параметров данного сигнала:

$$\|y_x\| = x \cdot x^T = \begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & \dots & x_1 \cdot x_n \\ x_2 \cdot x_1 & x_2 \cdot x_2 & \dots & x_2 \cdot x_n \\ \dots & \dots & x_i \cdot x_j & \dots \\ x_n \cdot x_1 & x_n \cdot x_2 & \dots & x_n \cdot x_n \end{bmatrix}. \quad (1)$$

С учетом полученных результатов моделирования разработан новый метод выделения признаков слоговых сегментов, для формирования которых в качестве первичных параметров используются огибающие сигналов в частотных диапазонах $\Delta_1 = 800 - 2500$ Гц и $\Delta_2 = 250 - 540$ Гц. Результирующий параметр, который в дальнейшем используется для выделения признаков слогов, получается корреляционным методом и записывается так:

$$U_c(t) = U_{\Delta_1}(t) \cdot U_{\Delta_2}(t), \quad (2)$$

где $U_{\Delta_1}(t)$ – огибающая энергии в полосе частот Δ_1 , а $U_{\Delta_2}(t)$ – огибающая энергии в полосе Δ_2 . Диапазон частот первого полосового фильтра, равный 250 – 540 Гц, выбран в виду того, что в нем отсутствует энергия высокоэнергетических фрикативных звуков типа /ш/ и /ч/, которые создают ошибочные слоговые ядра, а также сосредоточена значительная часть энергии всех звонких звуков, в том числе и гласных. Однако в этом диапазоне энергия сонорных звуков типа /л/, /м/, /н/ сравнима с энергией гласных, из-за чего определение слоговых сегментов только с учетом огибающей речевого сигнала в этом диапазоне будет сопровождаться многими ошибками. Поэтому диапазон частот второго полосового фильтра, выбран в пределах 800 – 2500 Гц, в котором энергия гласных звуков минимум в два раза превышает энергию сонорных звуков.

Благодаря операции умножения огибающих $U_{\Delta_1}(t)$ и $U_{\Delta_2}(t)$ в результирующей временной функции, происходит усиление участков кривой в области гласных звуков из-за корреляции их энергий в обоих диапазонах. Кроме того ошибочные максимумы энергии,

предопределенные наличием в диапазоне 800 – 2500 Гц значительной части энергии фрикативных звуков, устраняются путём их умножения на практически нулевое значение амплитуды фрикативных звуков в диапазоне 250 – 540 Гц.

Алгоритм и устройство выделения признаков составов

Соответственно с описанием функционирования устройства выделения слоговых сегментов алгоритм его работы состоит из таких шагов:

1. Ввод речевого сигнала с помощью микрофона.
2. Фильтрация сигнала двумя полосовыми фильтрами Баттерворта четвертого порядка в диапазонах 250 – 540 Гц и 800 – 2500 Гц соответственно.
3. Детектирование исходных сигналов фильтров для получения огибающих.
4. Перемножение огибающих исходных сигналов фильтров.
5. Дифференцирование результирующего сигнала.
6. Сравнение полученного сигнала с положительным и отрицательным пороговыми напряжениями и выделение логического сигнала.
7. Формирование из полученного сигнала логических сигналов для положительных и отрицательных полупериодов дифференцированного сигнала.
8. Выделение сегментов слогов и центров слогов путем логического добавления и умножения полученных логических сигналов соответственно.

Схема алгоритма изображена на рис. 2.

В алгоритме использованы такие обозначения:

$ff_1 = filter(p1, p2, s)$ и $ff_2 = filter(p3, p4, s)$ – функции фильтрации в полосах частот $\Delta_1 = 800 – 2500$ Гц и $\Delta_2 = 250 – 540$ Гц соответственно; $U_{g1} = abs(ff_1)$, $U_{g2} = abs(ff_2)$ – огибающих сигналов в указанных диапазонах. Другие обозначения раскрываются в описании работы устройства выделения признаков.

Устройство для выделения слоговых сегментов в речевом сигнале работает таким образом (рис. 3).

Речевой сигнал воспринимается акустическим датчиком, превращается в электрический и поступает на вход усилителя. Электрический сигнал, усиленный до значения достаточного для работы последующих каскадов, поступает на входы двух полосовых фильтров с полосами $\Delta_1 = [p_1; p_2] = 800 – 2500$ Гц и $\Delta_2 = [p_3; p_4] = 250 – 540$ Гц. Амплитудный детектор, вход которого подключен к выходу первого полосового фильтра, выделяет огибающую U_{g2} речевого сигнала в диапазоне 250 – 540 Гц. Амплитудный детектор, вход которого подключен к выходу второго полосового фильтра, выделяет огибающую U_{g1} речевого сигнала в диапазоне 800 – 2500 Гц. Напряжения U_{g1} и U_{g2} , поступающие на входы умножителя сигналов, перемножаются, в результате чего на его выходе появляется напряжение U_n , равное:

$$U_n = U_{g1} \cdot U_{g2}.$$

Диапазоны частот первого и второго полосовых фильтров подобраны таким образом, что в результате выполнения операции умножения получают корреляцию энергии только гласных звуков. Это приводит к устранению в огибающей U_n максимумов, соответствующих участкам высокоэнергетических фрикативных звуков. Это напряжение поступает на вход дифференциатора 9, формирователя признаков 8 (рис. 3), на выходе которого образуется напряжение, пропорциональное производной от напряжения U_n . Напряжение поступает на первые входы пороговых схем 10 и 11. На второй вход пороговой схемы 10 поступает положительное пороговое напряжение U_{n1} , а на второй вход пороговой схемы 11 поступает отрицательное пороговое напряжение U_{n2} , а U_{n1} и U_{n2} выбраны таким

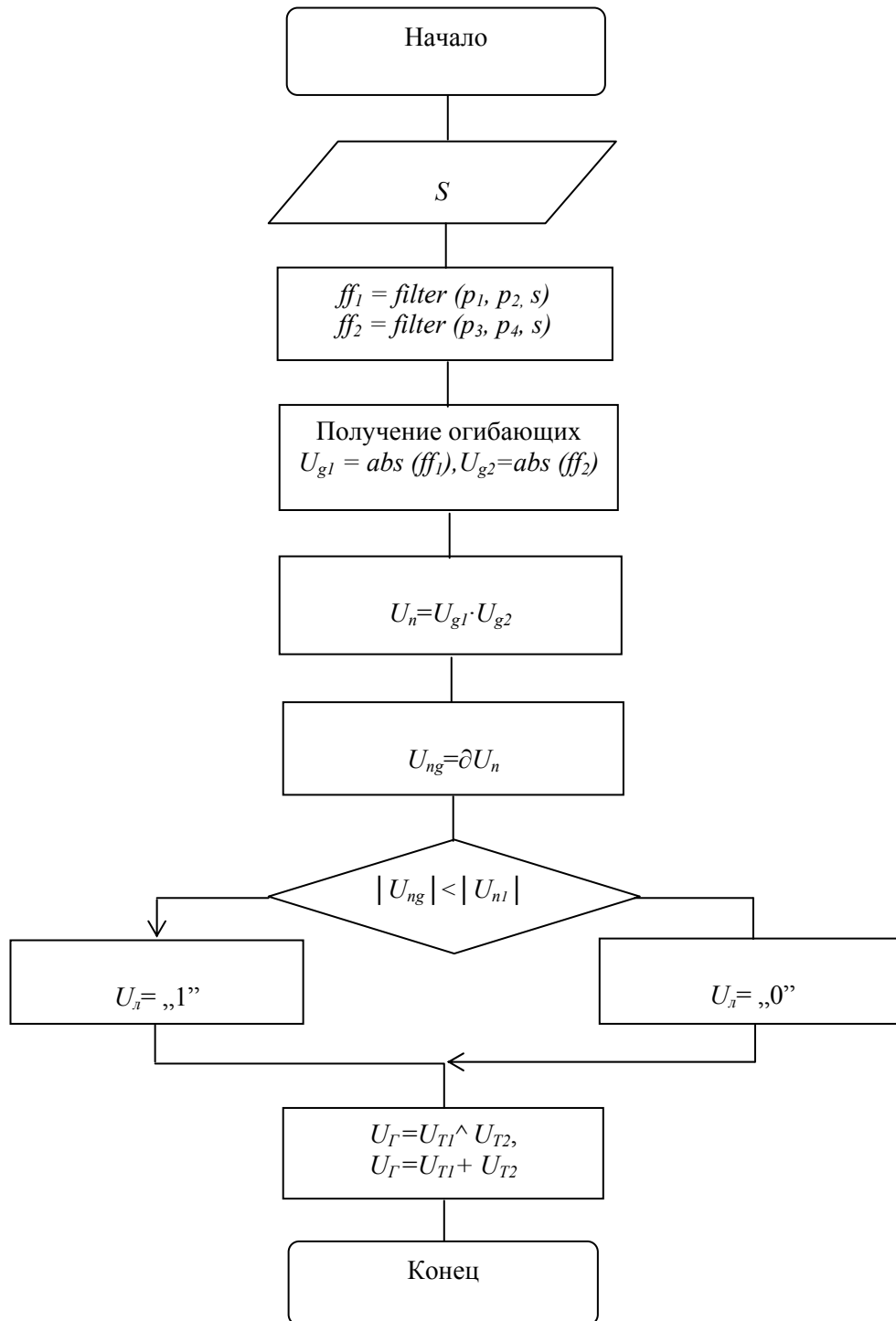


Рис. 2. Алгоритм выделения признаков слогов

образом, что $|U_{n1}| = |U_{n2}| \approx 50$ мВ, при этом устраняются ошибочные срабатывания пороговых схем из-за фоновых шумов в моменты перехода сигнала через нулевые значения. С выходов пороговых схем 10 и 11 получают клипированные сигналы, приведенные к значениям стандартных уровней цифровых сигналов. Эти сигналы поступают на входы схемы ИЛИ 12, на выходе которой образуется цифровой сигнал U_d . Фронты сигнала, поступающего на вход счетного триггера с прямым динамическим управлением 13, переключают его каждый раз в противоположное состояние, в результате чего на его выходе образуется цифровой

Наукові праці ВНТУ, 2007, № 1

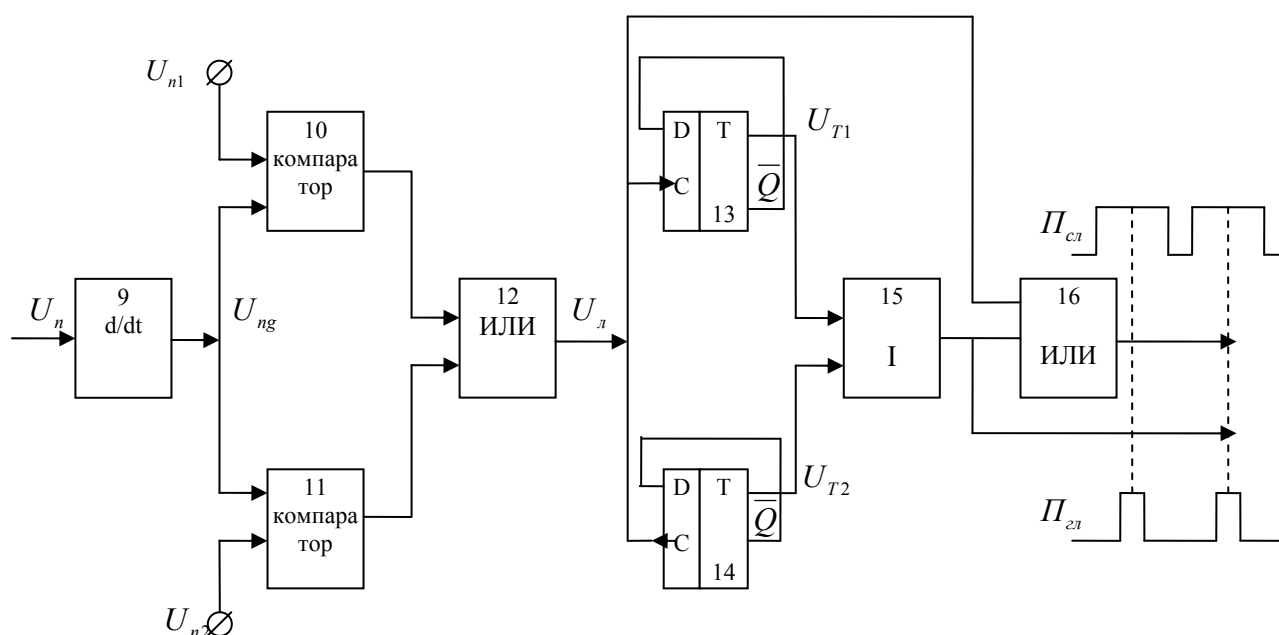


Рис. 3. Функциональная схема формирователя сегментных границ слогов

сигнал U_{T1} . Счетный триггер с обратным динамическим управлением 14 переключается спадами входного сигнала $U_л$ и выдает на выходе сигнал U_{T2} . Сигналы U_{T1} и U_{T2} поступают на входы схемы совпадения 15, на выходе которой образуются короткие импульсы $U_э$, которые локализуют центры слоговых ядер в слове. Эти импульсы поступают на первый вход схемы ИЛИ 16, а также на выход устройства для выделения слоговых сегментов в слове, являясь признаком $\Pi_{гл}$ определения местоположения центра гласных звуков в слове. На второй вход схемы ИЛИ 16 поступает цифровой сигнал $U_л$, который схемой ИЛИ объединяется с сигналом $U_э$, в результате чего на выходе схемы 16 вырабатывается сигнал $\Pi_{сл}$, который выделяет в слове сегменты, соответствующие положению слогов в слове. Длительность сигнала $\Pi_{сл}$ и количество сигналов $\Pi_{гл}$ используются системой распознавания для классификации словаря на подмножества, сформированные по данным признакам.

Для тестирования предложенного метода был проведен эксперимент, который состоял в сегментации 650 слогов с использованием предложенного метода. Статистическая обработка экспериментальных данных позволила рассчитать надежность этого метода, которая составила 96,4%. Надежность других методов на эквивалентной тестовой выборке не превышала 76%.

Выводы

Предложенные в работе метод и алгоритм выделения признаков слогов, основанные на уточненной модели слуховой системы человека, позволяют повысить надежность сегментации речевого сигнала на слоговые сегменты и определить такие их признаки, как длительность, местоположение и количество. Использование этих признаков позволяет увеличить скорость распознавания путем сокращения альтернатив для поиска в 2 – 4 раза, а также повысить надежность этого устройства. Разработанное в работе устройство, которое реализует этот метод, имеет большую надежность и меньшую сложность по сравнению с

известными устройствами, а также может быть применено при разработке автономных систем распознавания речи.

СПИСОК ЛИТЕРАТУРЫ

1. Быков Н.М. Методы и средства измерения и преобразования информации в системах машинного распознавания речи. – Дис. на соискание уч. ст. канд. техн. наук. – Винница, ВПИ, 1985. – 243 с.
2. N.M. Bykov, I.V. Kuzmin, A.I. Yakovenko. Development of effective strategy of pattern recognition. – Proceedings of SPIE, 2001, Vol. 4225, pp.76 – 83.
3. Джелинек Ф. Разработка экспериментального устройства, распознающего раздельно произносимые слова. // Тр. ин-та инженеров по электронике и радиоэлектронике.: Пер. с англ. 1985. – Т. 73. – № 11. – С. 91 – 100.
4. Биков М.М., Гришук Т.В. Методи підвищення дикторонезалежності опису і розпізнавання мовної інформації в мережі INTERNET // “Інтернет – Освіта – Наука – 2002”, третя міжнародна конференція ІОН – 2002, 8 – 12 жовтня 2002 р. Збірник матеріалів конференції. – Вінниця: УНІВЕРСУМ – Вінниця, 2002. – Том 2. – С. 329 – 332.
5. Методы автоматического распознавания речи / Под ред. У. Ли.– М.: Мир, 1983. – Т.1. – 200 с.
6. Ruske C., Schotola F. An approach to speech recognition using syllabic decision units. – Proc. 1978, IEEE ICASSP, Tulsa, 1978. – N.Y., 1978, pp. 772 – 725.
7. Ковтун В.В. Вибір інформативних ознак в задачі ідентифікації диктора // МКІМ – 2002. Міжнародна конференція з індуктивного моделювання. Львів, 20 – 25 травня 2002: Праці в 4-х томах. – Львів, ДНДІ, 2002. – Т.1, ч. 2 – С. 280 – 287
8. Биков М.М., Гришук Т.В. Розпізнавання мовних образів з використанням нейромережевого підходу // МКІМ – 2002. Міжнародна конференція з індуктивного моделювання, Львів, 20 – 25 травня 2002: Праці в 4-х томах. – Львів: Державний НДІ інформаційної інфраструктури, 2002. - Том 1., Ч. 2. – С. 203 – 207.

Быков Николай Максимович – профессор кафедры;

Ковтун Вячеслав Васильевич – ассистент кафедры;

Савинова Наталия Геннадиевна – студентка.

Кафедра компьютерных систем управления, Винницкий национальный технический университет