

В. Н. ЖУРАВЛЕВ (Запорожский национальный технический университет),  
И. В. ЖУКОВИЦКИЙ (ДИИТ)

## АНАЛИЗ ДЕВИАЦИИ ЧАСТОТЫ ФОНАЦИИ ПРИ ДЕМОДУЛЯЦИОННОЙ ОБРАБОТКЕ РЕЧЕВОГО СИГНАЛА

Розглянуто алгоритм цифрової обробки ідентифікаційного фрагменту мови, що дозволяє на основі аналізу значення несучої частоти сигналу мови, який наведений у вигляді сигналу з фазовою модуляцією, підвищити точність ідентифікації параметрів фонем.

Рассмотрен алгоритм цифровой обработки идентификационного фрагмента речи, позволяющий на основе анализа девиации несущей частоты речевого сигнала, представленного в виде сигнала с фазовой модуляцией, повысить точность идентификации параметров фонем.

In clause the algorithm of digital processing of an identification fragment of speech allowing is considered, on the basis of the analysis deviation of carrying frequency of a speech signal submitted as a signal with phase modulation, to increase accuracy of identification of parameters phoneme.

Широкое распространение информационных систем на железнодорожном транспорте предопределяет необходимость разработки методов защиты информации, которая хранится и обрабатывается в этих системах, в том числе – методов идентификация абонентов системы. Одними из известных методов идентификации являются биометрические методы. При этом основная трудность применения речевых технологий в системах биометрической идентификации абонентов сетей связи – это недостаточная устойчивость методов распознавания слитной речи по отношению к входным ошибкам и ошибкам обработки. Большинство методов, которые применяются в настоящее время, основаны на моделях генерации гипотез с использованием семантико-синтаксических или стохастических ограничений. Методы базируются на применении теории цепей Маркова, теории нейронных сетей [1], выделения гипотез фонем методом скользящего анализа сигнала с последующей оценкой гипотез морфем на основе акустического правдоподобия словоформ и правдоподобия временных интервалов [2],

оценивания спектральной энергии сегментов речевого сигнала [3].

Минимальными структурно-функциональными звуковыми единицами в большинстве языков являются фонемы. Они сами по себе не обладают смысловыми значениями, но потенциально связаны со смыслом как элементы единой знаковой системы. В сочетании друг с другом, и нередко по отдельности, они образуют слова и обеспечивают опознавание (идентификацию) и различение (дифференциацию) языковых знаков как значимых единиц.

Как правило, полное описание речевого сигнала только его спектром невозможно. Наряду со спектральной информацией, необходима еще и информация о динамике речи, которую определяют параметры активного состояния фонем. Для расчета параметров и характеристик, идентифицирующих активное состояние фонем в составе слов, наиболее часто применяются классические методы спектрально-временного анализа [4].

Известен [1] алгоритм идентификации параметров активного состояния фонем, функциональная схема которого представлена на рис. 1.

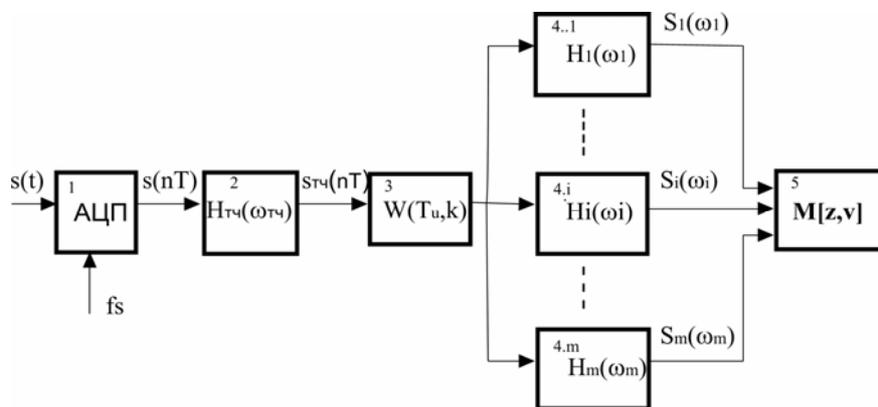


Рис. 1. Функциональная схема алгоритма идентификации параметров активного состояния фонем

Сигнал  $s(t)$  преобразовывается в блоке 1 (см. рис. 1) в цифровую форму  $s(nT)$  аналого-цифровым преобразователем АЦП на основании обобщенной теоремы отсчетов с частотой дискретизации

$$f_s = 1/T,$$

где  $T$  – интервал дискретизации, которая должна быть не менее удвоенной верхней частоты канала тональных частот (ТЧ).

Для последующей обработки сигнал  $s(nT)$  обрабатывается в блоке 2 фильтром с передаточной характеристикой  $H_{\text{ТЧ}}(\omega_{\text{ТЧ}})$ , соответствующей каналу ТЧ:

$$s_{\text{ТЧ}}(nT) = s(nT)H_{\text{ТЧ}}(\omega_{\text{ТЧ}}). \quad (1)$$

Блок 3 осуществляет оконное  $W(T_n, k)$  сегментирование сигнала  $s_{\text{ТЧ}}(nT)$  с параметрами:

$T_n$  – определяет длительность сегмента идентифицируемого сигнала на основании значения полосы частот  $\Delta\omega^{\text{ТЧ}}$  канала ТЧ и постоянной времени слуха  $T_c$  [5]

$$T_n = f(\Delta\omega^{\text{ТЧ}}, T_c); \quad (2)$$

$k$  – определяет количество сегментов идентификации и является функцией от  $T_n$  и коэффициента перекрытия сегментов идентификации  $K_{\text{пер}}$ , который обычно [3] принимается равным 0,5.

Идентификацию амплитуд  $A$  и мощности  $N$  спектральных  $S_i$  составляющих формант по  $n$  полосам (с серединными частотами  $\omega_1 \dots \omega_i \dots \omega_m$ ) осуществляют блоки фильтров 4.1...4.i...4.m, с передаточными характеристиками  $H_i(f_i)$ , на выходах которых образуются сигналы

$$S_i(\omega_i) = s_{\text{ТЧ}}(nT)H_i(\omega_i). \quad (3)$$

В блоке 5 осуществляется обработка сигналов  $S_i(\omega_i)$  с записью параметров формант ( $A, N, S_i$ ) в матрицу эталонов идентификационных признаков абонентов  $M[z, v]$ , где  $z$  – вектор абонентов,  $v$  – вектор идентификационных параметров абонента.

Рассмотренный алгоритм позволяет анализировать фрагменты речевых сигналов на временном отрезке  $T_n$ . Таким образом, дискретность определения времени активного состояния фоновы зависит от  $T_n$ , что вносит погрешность в идентификационные параметры  $v$  абонента.

В данной работе предложен алгоритм идентификации фонем, основанный на анализе изменения несущей частоты при фазовой демодуляции речевого сигнала абонента. В основе алгоритма лежит наблюдение, что при фазовой демодуляции речевого сигнала, значение несущей частоты изменяется на границах фонем, а значит, может быть представлено как идентификационный параметр.

Сигнал  $s(t)$  на интервале времени идентификации  $T_p$  является сигналом с конечной энергией, не содержит дельта-функций и разрывов. Основываясь на этом можно предположить, что параметры его изменения можно определить, исследуя девиацию несущей частоты  $\Delta\omega_n$ , при представлении речевого сигнала как сигнала с угловой модуляцией.

Представим речевой сигнал  $s(t)$  на сегменте идентификации  $T_p$  в виде сигнала несущей с частотой  $\omega_n$ , который модулирован информационным сигналом  $s_y(t)$ . Мгновенные амплитуда  $A_y(t)$  и фаза  $\varphi_y(t)$  сигнала  $s(t)$  изменяются по закону  $Y$  передаваемого сообщения

$$s(t) = A_y(t) \cos[\omega_n t + \varphi_y(t) + \varphi_0], \quad (4)$$

где  $A_y(t)$  – амплитудная огибающая сигнала  $s(t)$ ;  $\varphi_y(t)$  – мгновенная фаза сигнала  $s(t)$ ;  $\varphi_0$  – начальная фаза сигнала  $s(t)$ .

Весь аргумент функции  $\cos$  выражения (4) представляет полную фазу  $\Psi(t)$  сигнала  $s(t)$ :

$$\Psi(t) = \omega_n t + \varphi_y(t) + \varphi_0. \quad (5)$$

Для выделения мгновенной амплитуды и фазы применим метод низкочастотного эквивалента [6], который заключается в замене исходного сигнала  $s(t)$  его комплексной огибающей. Запишем сигнал  $s(t)$  в экспоненциальной форме

$$s(t) = \text{Re}\{A_y(t)e^{j\Psi(t)}\}. \quad (6)$$

Выражение в фигурных скобках является комплексной функцией действительной переменной

$$\dot{S}(t) = A_y(t)\{\cos[\Psi(t)] + j \sin[\Psi(t)]\}, \quad (7)$$

действительная часть которой совпадает с исходным сигналом  $s(t)$ , а мнимая часть

$$u(t) = A_y(t)\sin[\Psi(t)] \quad (8)$$

отличается от исходного сигнала поворотом фазы на угол  $(-\pi/2)$  и является сопряженным гармоническим сигналом. Сопряженный гармонический сигнал получаем из исходного  $s(t)$ , проведя над ним преобразование Гильберта

$$u(t) = \frac{1}{\pi} \int_0^{T_c} \frac{s(\tau)}{\tau - t} d\tau, \quad (9)$$

где  $t$  – текущее время;  $\tau$  – подынтегральное время свертки.

Данный интеграл представляет собой линейную свертку сигнала  $s(\tau)$  и функции  $1/\pi\tau$  и, таким образом, является идеальным фазовращателем, вносящим на всех частотах фазовый сдвиг, равный  $-90^\circ$ .

Можно определить модуль мгновенной огибающей  $A_y(t)$  сигнала  $s(t)$

$$A_y(t) = \sqrt{s^2(t) + u^2(t)}, \quad (10)$$

полную мгновенную фазу

$$\Psi(t) = \arctg \left[ \frac{u(t)}{s(t)} \right] \quad (11)$$

и мгновенную частоту  $\omega'(t)$

$$\begin{aligned} \omega'(t) &= \frac{d\Psi(t)}{dt} = \frac{d}{dt} \left\{ \arctg \left[ \frac{u(t)}{s(t)} \right] \right\} = \\ &= \frac{s(t) \frac{du(t)}{dt} - u(t) \frac{ds(t)}{dt}}{s^2(t) + u^2(t)}. \end{aligned} \quad (12)$$

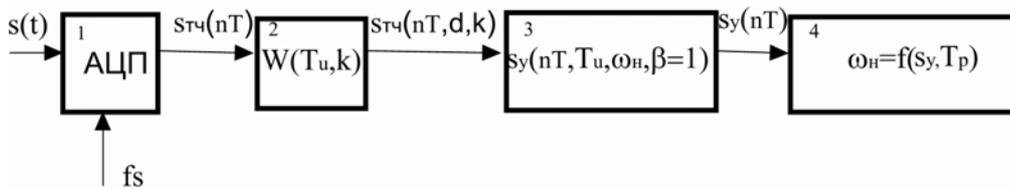


Рис. 2. Функциональная схема алгоритма идентификации параметра несущей частоты фонем

Блок 2 осуществляет оконное  $W(T_u, k)$  сегментирование сигнала  $s_{Tч}(nT)$ . Таким образом, сигнал на выходе блока 2 будет определяться

$$s(nT, T_u, k) = s(nT)W(T_u, k). \quad (15)$$

Блок 3 производит декодирование сигнала  $s(nT, T_u, k)$  в соответствии с выражением (13) на сегменте длительностью  $T_u$  с определением значения несущей частоты  $\omega_n$

$$s_y(nT, T_u, \omega_n, \beta=1) = \Psi(nT) - \omega_n nT - \varphi_0, \quad (16)$$

Выделив в мгновенной частоте постоянную несущую  $\omega_n$  и начальную фазу  $\varphi_0$ , можно вычислить мгновенную фазу  $\varphi_y(t)$

$$\varphi_y(t) = \Psi(t) - \omega_n t - \varphi_0 = s_y(t), \quad (13)$$

которая является искомой информационной составляющей  $s_y(t)$  в фазомодулированном речевом сигнале  $s(t)$ .

Под индексом фазовой модуляции  $\beta$  понимают [6] отношение девиации несущей частоты  $\Delta\omega_n$  к частоте  $\omega_y$  модулирующего сигнала  $s_y(t)$ . Потребуем неизменности индекса модуляции на интервале времени  $T_p$  и положим его равным единице  $\beta=1$

$$\beta = \frac{\Delta\omega_n}{\omega_y} \Big|_{T_p} = 1. \quad (14)$$

В этом случае можно предположить, что изменения мгновенной фазы информационной составляющей  $\varphi_y(t)$  (13) на интервале  $T_p$  будут отражены в функции девиации несущей частоты  $\Delta\omega_n$ .

С целью увеличения точности идентификации параметров фонем предлагается дополнить существующие методы следующим алгоритмом.

Сигнал  $s(t)$  преобразовывается в блоке 1 (рис. 2) в цифровую форму  $s(nT)$  аналого-цифровым преобразователем АЦП с частотой дискретизации  $f_s$ .

где  $T_u$  – время идентификации несущей частоты  $\omega_n$ .

В блоке 4 определяются значения несущей частоты  $\omega_n$  на интервале времени  $T_p$ .

По предлагаемому алгоритму был произведен расчет несущей частоты  $\omega_n$  морфем «АКУЛА», «НАУКА» и «КУКЛА», содержащих одинаковые фонемы.

В результате расчетов, которые проводились в программной среде MatLab 6.5, получены следующие графики (рис. 3–5).

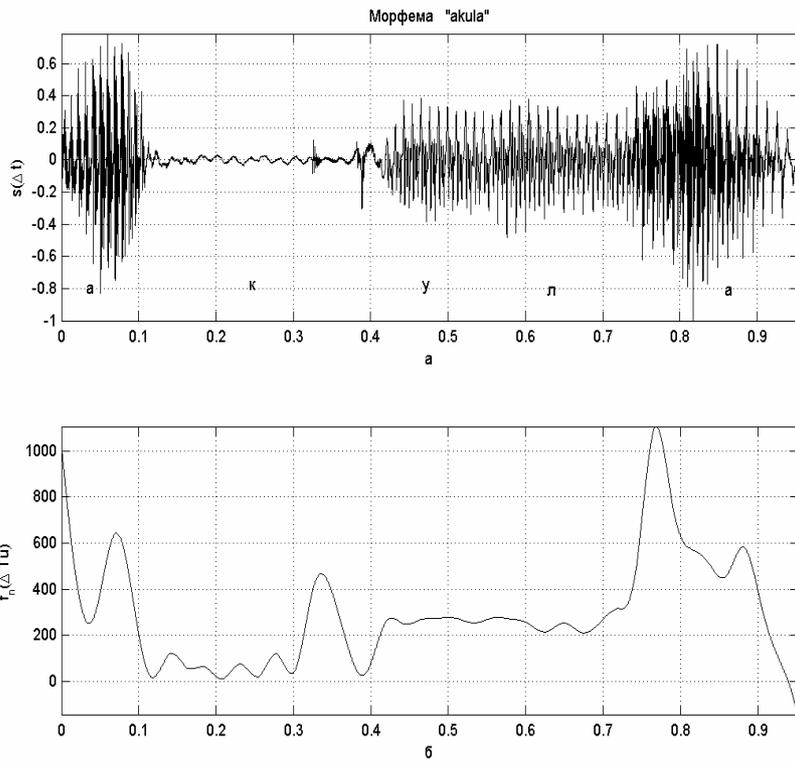


Рис. 3. Графики анализа морфемы «АКУЛА»:  
 а – график зависимости сигнала морфемы от времени  $s(nT)$ ; б – график девиации параметра несущей частоты на интервале времени идентификации  $\omega_n(T_p)$

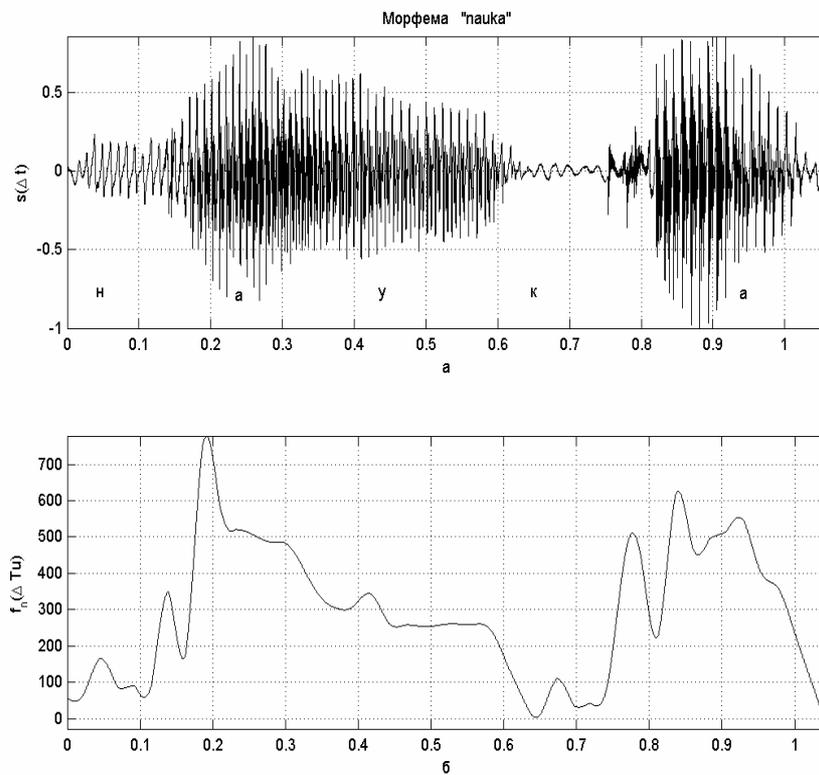


Рис. 4. Графики анализа морфемы «НАУКА»:  
 а – график зависимости сигнала морфемы от времени  $s(nT)$ ;  
 б – график девиации параметра несущей частоты на интервале времени идентификации  $\omega_n(T_u)$

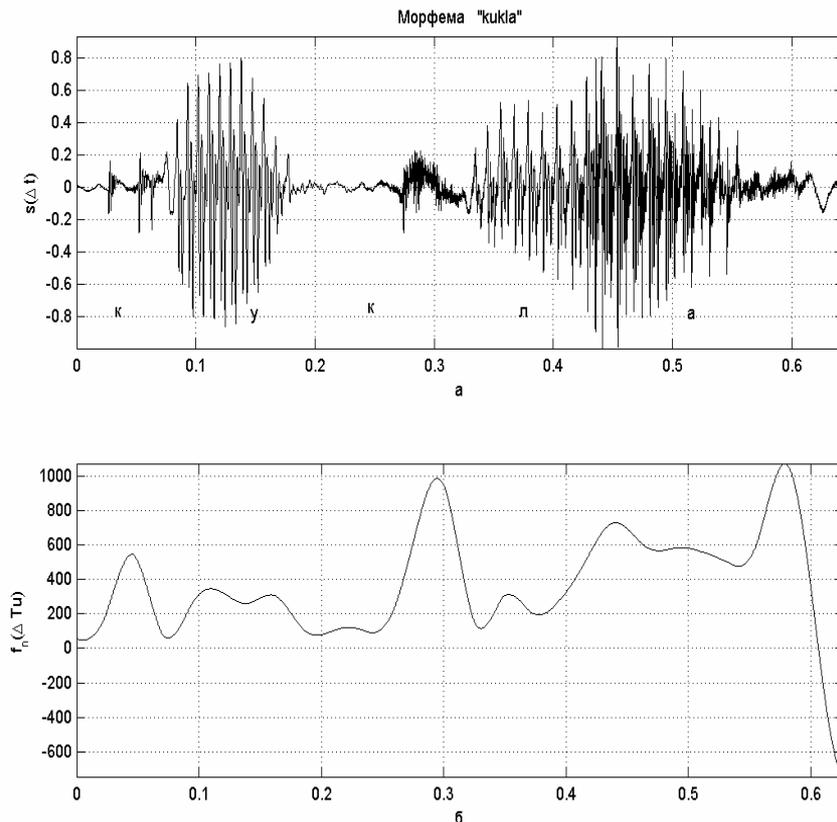


Рис. 5. Графики анализа морфемы «КУКЛА»:  
 а – график зависимости сигнала морфемы от времени  $s(nT)$ ;

б – график девиации параметра несущей частоты на интервале времени идентификации  $\omega_n(T_u)$

Речевой сигнал морфем записывался на жесткий диск ПЭВМ с частотой дискретизации 96 кГц и 16-разрядным квантованием по уровню.

Проведя анализ вариации параметра несущей частоты  $\omega_n(T_u)$  для вокализованных фонем «а», «у» и «л» на интервале  $T_p$ , можно сделать следующие выводы:

- несущая частота фонемы «а» лежит в пределах 480...650 Гц;
- несущая частота фонемы «у» лежит в пределах 220...250 Гц;
- несущая частота фонемы «л» лежит в пределах 180...270 Гц.

Однако характер девиации несущей частоты  $\omega_n(T_u)$  на интервале идентификации вокализованных фонем разный, поэтому для определения обобщенных зависимостей параметров идентификации требуется проведение дополнительных исследований.

### Выводы

Полученные результаты позволяют сделать вывод о целесообразности применения предложенного алгоритма при идентификации параметров фонем.

Повышение эффективности идентификации достигается на основании проведения, дополнительно к существующим методам, анализа девиации несущей частоты фонем речи, представленной в виде сигнала с фазовой модуляцией.

Увеличение точности идентификации возможно при анализе статистической обработки девиации значений параметра несущей частоты для фонем идентифицируемого абонента сети.

### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Лалетин П. А. Использование сетей с самостоятельной адаптацией для распознавания слов человеческой речи / П. А. Лалетин, Э. Г. Лалетина, Ю. П. Ланкин // Научная сессия МИФИ-2000. 2-я Всероссийская научно-техническая конференция «Нейроинформатика-2000»: Сборник научных трудов. В 2-х частях. Ч. 2. – М.: МИФИ, – 2000. – С. 88–95.
2. Старушко Д. Г. Новая система признаков для распознавания речевых единиц / Д. Г. Старушко, В. Ю. Шелепов // Штучный интеллект. – 2002. – № 4. – С. 48–53.

3. Филимонов В. В. Опыт создания системы автоматического распознавания русской речи: Тезисы международной конференции «Информатизация правоохранительных систем», – М.: 1998. Ч. 2, – С. 104–106.
4. Бондарко Л. В. Звуковой строй современного русского языка. – М.: Просвещение, 1997. – 175 с.
5. Цвикер Э. Ухо как приемник информации / Э. Цвикер, Р. Фельдкеллер; Пер. с нем. под ред. Б. Г. Белкина. – М.: Связь, 1971. – 225 с.
6. Гоноровский И. С. Радиотехнические цепи и сигналы: Учебник для вузов. – М.: Радио и связь, 1986. – 512 с.

Поступила в редколлегию 29.10.2005.