

Система сегментно-слогового распознавания изолированных слов из больших словарей

Савенкова О.А.

В статье рассмотрены основные модули системы сегментно-слогового распознавания изолированных слов из больших словарей, математические модели и алгоритмы сегментно-слогового синтеза эталонной траектории, положенные в их основу.

• *система распознавания речи* • *слог* • *сегментно-слоговое распознавание* • *траектория параметров* • *алгоритм эвристического поиска*.

The basic modules of the large vocabulary segment-syllabic recognition system of isolated words, the mathematical models and the segment-syllabic synthesis algorithms of a reference parameters trajectory are considered.

• *speech recognition system* • *syllable* • *segment-syllabic recognition* • *parameters trajectory* • *heuristic search*.

Введение

Основные направления разработки систем распознавания речи (далее СРР): распознавание отдельно произносимых, не связанных по смыслу, слов, а также распознавание слитной речи как решение задачи декодирования языковой структуры речевого сообщения с использованием различных источников лингвистических знаний [1–10].

Решение задачи обработки и распознавания речи — задача структурной аппроксимации речевого сигнала (далее РС) как совокупности компонентов, так и иерархии категорий информации (информация о физической природе РС, о лингвистической структуре языка и речи). Существует актуальная проблема обеспечения соответствия между параметрическим, фонетическим и символьным представлениями речи [1, 4, 9]. Необходим комплексный подход к построению СРР, учитывающий взаимосвязи между иерархией представления информации о РС и решающий проблему акустико-фонетического декодирования [1, 2, 5, 7, 9–11]. Данная проблема не имеет точного решения. Поэтому в основу её решения положим стратегию эвристического поиска с использованием специфических знаний о речевой информации на разных уровнях иерархии. Для этого согласно [12], исходную задачу сведём, во-первых, к выбору объектов, необходимых для решения задачи (алфавит объектов распознавания), а во-вторых, к разработке стратегий эффективного поиска потенциальных решений, которые могут быть сгенерированы с помощью выбранных объектов.

Однозначного ответа на вопрос о том, какие элементарные образы (фонема, слог, слово) и в каком количестве используются в процессе распознавания, пока не получено. Однако в ряде исследований показано, что обработка речи требует перехода от распознавания слов как целостных звуковых

образов к распознаванию звуковых единиц, которые меньше слова [3]. Сравнение моделей языка по количеству используемых разных языковых единиц в словаре и процентом непокрытых слов в текстовом материале показало, что чем меньше размер элементов, тем более полно они покрывают пространство слов языка [13].

Результаты исследований в области психоакустики и экспериментальной фонетики показывают, что любую речевую реализацию можно рассматривать как последовательность открытых слогов, которые представляют собой единство слогаобразующего гласного (Г) и одного или более согласных (С). Фрагмент речи между паузами является цепочкой связанных между собой слогов. Причем единого деления на слоги внутри такого фрагмента не происходит, и в этом смысле слог не отличается от отдельного звука речи [10, 11, 14–18]. Таким образом, с одной стороны, слог — основа образования разнообразных звуков, а с другой — большого числа разных ритмических структур слов [11, 18–20]. Слоги являются оптимальными с точки зрения артикуляционной организации и надёжности распознавания. Их основное преимущество — учёт просодических свойств речи. О чём свидетельствуют существующие примеры систем анализа/синтеза речи, приведённые в [19, 21–24]

Элементы алфавита объектов распознавания для СРР из больших словарей должны иметь такую длину и быть подобраны в таком количестве, чтобы из них можно было построить любые слова или фразы и предложения [4, 25, 26]. Этим требованиям удовлетворяют слова-слоги длиной 2 и 3 символа-фонемы, причём в [25, 26] выявлено, что наиболее употребляемыми являются слоги, которые содержат 2 (С + Г) и 3 (С + С + Г) фонемы.

С учётом вышеизложенных фактов в данной работе рассмотрим модели и алгоритмы для решения задачи фонемно-слового распознавания речи, в общем виде представленной в [4].

Пусть задан алфавит слогов $SL = \{SL_1, \dots, SL_k, \dots, SL_Z\}$, $k = 1, \dots, Z$. Каждый слог SL_k содержит K символов-фонем $SL_k = (P_1, P_2, \dots, P_K)$, $\forall P_i \in P$ (P — множество всех символов-фонем). Для каждого слога SL_k задана последовательность параметров $Y_k = (y_{k1}, y_{k2}, \dots, y_{ki}, \dots, y_{kN_k})$, $i = 1, \dots, N_k$ (траектория параметров (далее ТП) в пространстве признаков), определены границы сегментов-фонем $S_k = (s_{k1}, s_{k2}, \dots, s_{kj}, \dots, s_{kL_k})$, $j = 1, \dots, L_k$. Пусть также задана последовательность параметров $X = (x_1, x_2, \dots, x_N)$ реализации РС, предъявленного для распознавания. Для X определены границы сегментов-фонем $S = (s_1, s_2, \dots, s_L)$. Сегменты-фонемы SG_i ($SG_i = (x_{s_{beg}}, \dots, x_{s_{end}})$, $s_{beg} = s_i$, $s_{end} = s_{i+1}$, $i = 1, \dots, L$), составляющие последовательность параметров $X = (SG_1, SG_2, \dots, SG_i, \dots, SG_L)$, могут быть объединены некоторым образом в M групп-слогов XSL_p по m_p сегментов-фонем каждая, причём $\sum_p m_p = L$, $p = 1, \dots, M$. Символьная последовательность

$W^* = (P_1^*, P_2^*, \dots)$, которая соответствует последовательности параметров X предъявленной реализации РС, неизвестна. Необходимо траектории параметров X наилучшим образом сопоставить траектории параметров слогов $\{Y_k\}$, вычисляя

$$dist = \sum_p \min_k (XSL_p \# Y_k), \quad p = 1, \dots, M, \quad k = 1, \dots, Z, \quad (1)$$

где $\#$ — операция сопоставления. Таким образом, необходимо построить такую последовательность параметров X^* , которая будет близкой по своим параметрам к параметрам предъявленной реализации речевого сигнала X . Такую траекторию параметров X^* будем называть *эталонной траекторией параметров* (далее ЭТП). Некоторую траекторию

параметров X_M^* , построенную из X элементов алфавита SL , которая по количеству сегментов соответствует предъявленной траектории параметров X , будем называть *решением-кандидатом*.

Для построения решений задачи (1) рассмотрим подход, основанный на применении методов поиска в пространстве состояний с учётом особенностей исследуемой задачи [12]: выбран алгоритм эвристического поиска в пространстве состояний, который в отличие от базовых стратегий поиска, использует определенного вида оценочную функцию (далее ОФ) $f(n)$, сокращающую объем перебора. В общем случае ОФ $f(n)$ для узла n имеет вид

$$f(n) = g(n) + h(n), \quad (2)$$

где $g(n)$ — длина пути от начального узла к узлу n , $h(n)$ — эвристическая оценка расстояния из узла n к целевому узлу, для определения которой используют любую эвристическую информацию о решаемой задаче [12].

Структура системы распознавания речи «SPeach»

Структура системы «SPeach» (рис. 1) состоит из модулей, которые могут работать в режимах обучения и распознавания: (а) модуль обработки РС; (б) модуль формирования алфавита слогов; (в) модуль распознавания. В режиме обучения работают (а), (б), а в режиме распознавания — (а), (в).

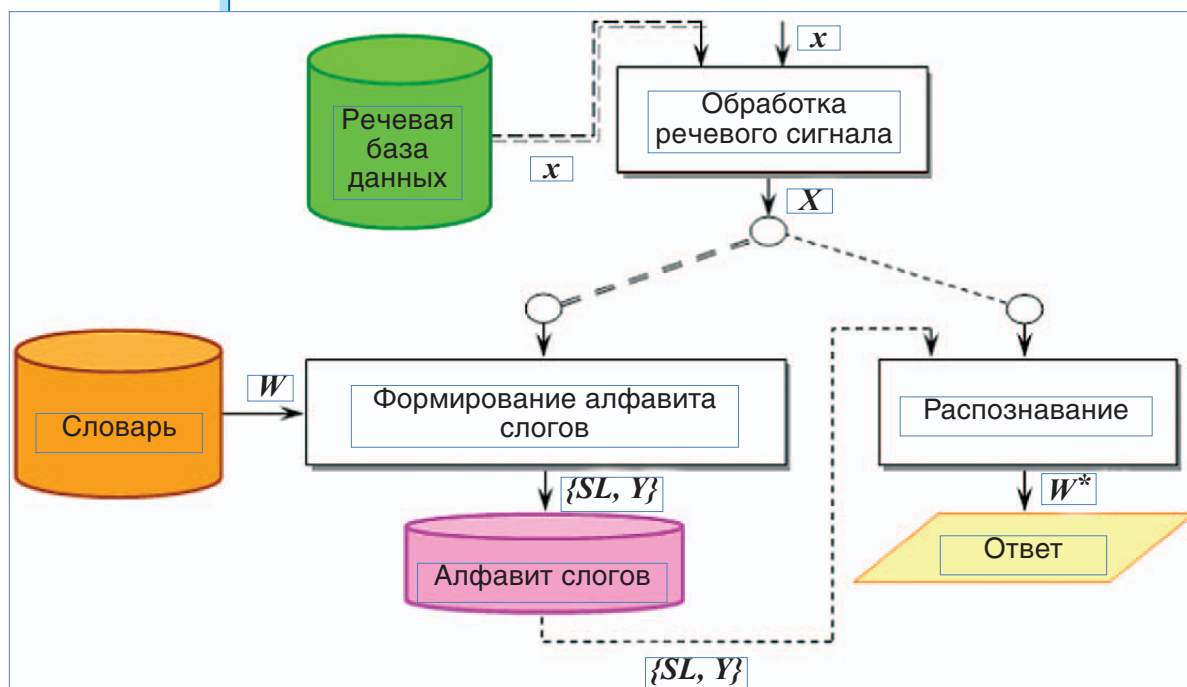


Рис. 1. Структура системы распознавания речи

Модель обучения

Последовательность этапов обработки информации о РС, которые составляют модель обучения, представлена на рис. 2.

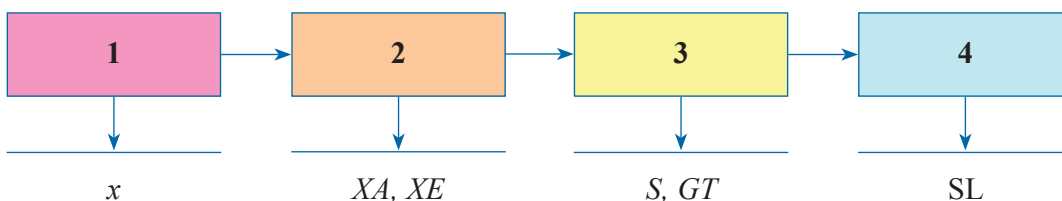


Рис. 2. Модель обучения

Шаг 1. Ввод речевого сигнала с микрофона или открытие wav-файла из речевой базы данных (далее РБД).

Частота дискретизации речевых сигналов составляет $F_s = 22050$ Гц.

Шаг 2. Первичная обработка РС, формирование траектории параметров.

Интервал анализа РС, выбранный для исследований, $\Delta T = 11,6$ мс. В данной работе траектории параметров $X = (x_1, x_2, \dots, x_i, \dots, x_T)$, $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,j}, \dots, x_{i,m})$, $i = 1, \dots, T$ (T — количество интервалов анализа РС), $j = 1, \dots, m$: спектрально-временное представление (далее СВП) $XA(\omega, t)$; спектрально-полосное представление (далее СПП) $XE(l, t)$, $l = 1, \dots, 9$ [4, 27].

Для СВП XA и СПП XE разработаны модели аналитического описания в классе функций $f \in C(\Omega)$, которые являются гладкими непрерывными функциями в соответствующих частотно-временных областях определения $\Omega: [\omega_0, \omega_1] \times [t_0, t_N]$. Согласно им можно восстановить исходную траекторию параметров РС с минимальной погрешностью.

1. Построение аналитического описания СВП слогов

СВП XA РС содержит произвольное количество всплесков/пиков спектральной энергии, которые произвольно расположены на определенной частотно-временной области $D_A: [\omega_0, \omega_M] \times [t_0, t_N]$. Естественно найти набор функций в виде пиков, с помощью которых можно описать частотные и временные свойства РС [28].

Для построения аналитического описания ТП слогов, используется колоколообразная функция (далее КФ) модифицированный локон Аньези [4, 29 — 32].

Пусть в частотно-временной области D_A таблично задана спектрально-временная функция $YA(\omega_k, t_l)$ для некоторого слога SL_j , где ω_k, t_l — дискретно заданные частота и время соответственно, $k = 0, \dots, M$, $l = 0, \dots, N$, $j = 1, \dots, Z$. Для исходной траектории параметров YA_j слога SL_j построим описание YLA_j в виде суперпозиции произведений КФ:

$$Zt_{(i)}(t_l) = \frac{a_{(i)}^3}{c_{(i)}^2 + (t_l - T_{(i)})^2}; Z\omega_{(i)}(\omega_k) = \frac{b_{(i)}^3}{d_{(i)}^2 + (\omega_k - \Omega_{(i)})^2}, \quad (3)$$

где функции $Zt_{(i)}(t_l)$, $Zt \in [t_0, t_N]$, $i = 1, \dots, L$ описывают временные свойства РС; функ-

кции $Z\omega_{(i)}(\omega_k)$, $Z\omega \in [\omega_0, \omega_M]$, $i = 1, \dots, L$ — частотные свойства РС.

Произведение $Zt_{(i)}(T) \cdot Z\omega_{(i)}(\Omega)$ описывает всплеск $i = 1, \dots, L$ спектрально-временной

функции YA , который находится на частоте Ω , в момент времени T . Областью определения произведений функций $Zt_{(i)}(T) \cdot Z\omega_{(i)}(\Omega)$ является область $D_A: [\omega_0, \omega_M] \times [t_0, t_N]$. Тогда описание $YLA(\omega_k, t_l)$ в виде суперпозиции L произведений КФ (3) в некоторой точке (ω_k, t_l) СВП имеет вид

$$YLA(\omega_k, t_l) = \sum_{i=1}^L Z\omega_{(i)}(\omega_k) \cdot Zt_{(i)}(t_l), \quad k=0, \dots, M, \quad l=0, \dots, N. \quad (4)$$

Определение неизвестных параметров КФ (3) $a_{(i)}, b_{(i)}, c_{(i)}, d_{(i)}, T_{(i)}, \Omega_{(i)}, i=1, \dots, L$ выполняется по алгоритму, предложенному в [29, 30].

Пример аналитического описания СВП для некоторых слогов приведён на рис. 3–6.

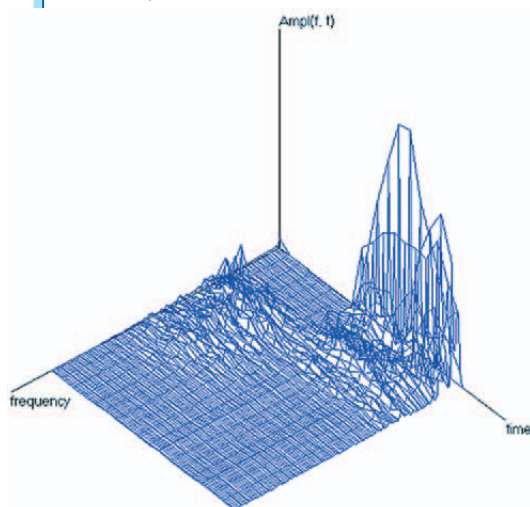


Рис. 3. Исходное СВП слога «че»

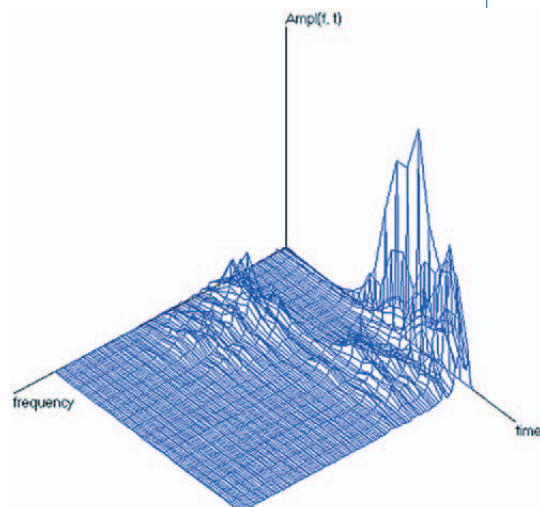


Рис. 4. Описание СВП слога «че»

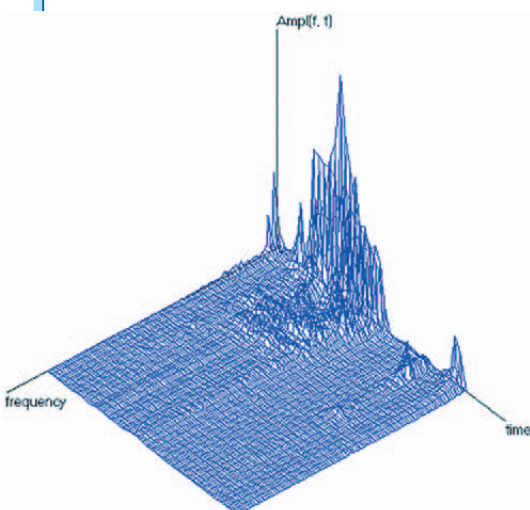


Рис. 5. Исходное СВП слога «век»

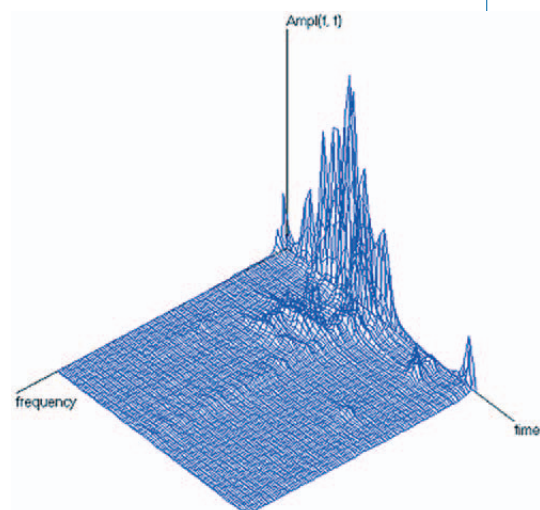


Рис. 6. Описание СВП слога «век»

2. Построение сплайн-описания СПП слогов

Анализируя вид СПП на интервалах продолжительностью сегмента-фонемы для разных звуков речи, можно сделать вывод: на каждом временном отрезке, который соответствует сегменту-фонеме, последовательности параметров СПП в каждой частотной полосе $l = 1, \dots, 9$ имеют вид, который можно описать полиномами низких порядков $P_n(x)$, $n \leq 3$. Таким образом, для аналитического описания СПП элементов алфавита можно применить математический аппарат сплайн-функций [4, 33].

Пусть для слога SL_k сформирована траектория параметров YE_k (СПП), найдены границы сегментов-фонем SG_{kn} $S = (s_1, \dots, s_n, \dots, s_{L_k})$, $k = 1, \dots, Z$, $n = 1, \dots, L_k$, $L_k = 2; 3; 4$. Тогда модель сплайн-описания траектории параметров слога SL_k , которая аппроксимирует исходную YE_k в частотной полосе, имеет вид

$$YSE_k = \begin{cases} YSE_{k,1,i}, & s_1 \leq i \leq s_2, \\ \dots \\ YSE_{k,n,i}, & s_{n-1} \leq i \leq s_n, \\ \dots \\ YSE_{k,L_k,i}, & s_{L_k-1} \leq i \leq s_{L_k}, \end{cases} \quad (5)$$

$$YSE_{k,n,i} = a_{k,n} + b_{k,n} \cdot (t_{k,i} - s_{n-1}) + c_{k,n} \cdot (t_{k,i} - s_{n-1})^2 + d_{k,n} \cdot (t_{k,i} - s_{n-1})^3, \quad (6)$$

где $a_{k,n}$, $b_{k,n}$, $c_{k,n}$, $d_{k,n}$ — коэффициенты $P_3(x)$, который описывает n -й сегмент-фонему k -го слога $n = 1, \dots, L_k$.

Для нахождения параметров модели сплайн-описания траектории параметров YE_k в каждой частотной полосе l решается задача минимизации среднеквадратического приближения с условиями в точках сегментации (на основе эмпирических данных в качестве узлов сплайн-функции выбраны точки сегментации РС), которые обеспечивают требуемую гладкость и непрерывность:

$$\sigma_l^2 = \sum_{j=1}^{L_k} \sum_{i=s_{j-1}}^{s_j} (YSE_{l,k,i} - YE_{l,k,i})^2 \rightarrow \min \quad [32].$$

Шаг 3. Сегментация речевого сигнала методом верификации [4].

Шаг 4. Формирование алфавита слогов.

1. Выбор множества слов $W = \{W_j\}$, $j = 1, \dots, NW$ и соответствующей РБД

В качестве словаря, который используется для формирования алфавита слогов, выберем список частотных слов, построенный на основе представительного корпуса современного языка. Статистические данные о зависимости процента покрытия текстового корпуса (около 16 млн слов) частотными словами ($Perc$) от количества частотных слов NW представлены в [34].

2. Выбор структуры данных для хранения и эффективного использования информации о слогах в системе распознавания речи

Для элемента $SL_{M,k}$ $k = 1, \dots, Z$, $M = 2; 3$ из множества M -символьных слогов предложена структура данных, которая является совокупностью следующих категорий информации:

а) лингвистическая информация о k -том слоге (символьное представление $SL_{M,k}$ транскрипция $SLT_{M,k}$ идентификатор слова, в состав которого входит и др.);

- б) вспомогательная информация об SL_k (идентификатор слога, количество временных отсчетов $NT_{M,k}$, количество сегментов L_k , границы сегментов $S_k = (s_1, \dots, s_{L_k})$, ТШП-транскрипция GT_k);
- в) параметрическое представление (СВП, СПП, параметры модели описания СВП в классе колоколообразных функций, параметры модели сплайн-описания СПП).

3. Декомпозиция слов W_j на слоги SL

В модуле формирования алфавита слогов реализованы алгоритмы для следующей задачи [35].

Пусть задано некоторое множество слов $W = \{W_j\}$, $j = 1, \dots, NW$, которому соответствует множество реализаций РС. Каждое слово W_j содержит LW_j символов, т.е. $W_j = (w_{j,1}, w_{j,2}, \dots, w_{j,LW_j})$. Для каждого слова W_j найдена транскрипция $WT_j = (wt_{j,1}, wt_{j,2}, \dots, wt_{j,LWT_j})$ длиной LWT_j символов, которая соответствует сегментам-фонамам траектории параметров РС.

Процесс деления слов $W = \{W_j\}$, $j = 1, \dots, NW$, с учетом транскрипции WT_j , в цепочку M -символьных слогов будем называть декомпозицией

$$Decomp^M : W_j \xrightarrow{WT_j} Decomp^M(W_j) = SL_1^M, \dots, SL_{L_{M,j}}^M, \quad (7)$$

где $L_{M,j}$ — количество M -символьных слогов в слове W_j ($M = 2, 3$).

Таким образом, в результате декомпозиции (7) для множества слов W сформирован алфавит SL , состоящий из M -символьных слогов $SL_{M,k}$ ($SL_{2,k} = \{w_{j,m}, w_{j,m+1}\}$, $SL_{3,k} = \{w_{j,m}, w_{j,m+1}, w_{j,m+2}\}$, $m = 1, \dots, LW_j - M$), общее количество которых составляет:

$$F_{SL}(W) = \sum_{j=1}^{NW} \sum_{M=2}^3 L_{M,j} \quad (M = 2, 3). \quad \text{Очевидно, что такой набор объектов}$$

распознавания является избыточным, поэтому для покрытия множества слов W формируется такой алфавит SL^* , что

$$F_{SL^*}(W) \rightarrow \min. \quad (8)$$

Проанализируем списки из 1000 ($W1000$), 5000 ($W5000$) и 9000 ($W9000$) наиболее частотных слов русского языка, для которых значения *Perc*: 64,07%, 82,06% и 87,82% соответственно [34].

Для выбранных списков частотных слов выполнено сравнение количества 2- и 3-символьных сочетаний, покрывающих указанные множества слов (результаты приведены на рис. 7). Выяснено, что множество таких 2- и 3-символьных сочетаний, полученное с помощью (7) для $W5000$, покрывает почти 90% сочетаний, полученных для $W9000$.

Множество слов $W5000$ покрывает 82% текстового корпуса, который составляет 13 млн слов, что достаточно для создания СРП с большим словарём. Кроме того, начиная с $NW = 5000$ частотных слов, значение прироста процента покрытия текстового корпуса этими словами увеличивается с достаточно малым шагом.

Поэтому считаем, что наиболее оптимальным для создания алфавита слогов является использование множества из 5000 наиболее частотных слов языка.

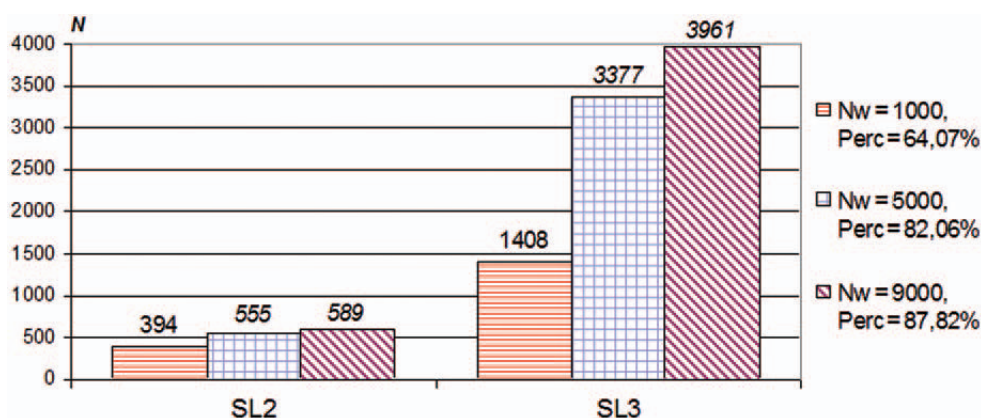


Рис. 7. Количественный анализ 1000, 5000, 9000 наиболее частотных слов

При формировании алфавита слогов необходимо также учитывать информацию о фонетическом представлении слов (обобщённая звуковая транскрипция, основанная на действующих в данном языке стандартных правилах чтения), которую можно получить с помощью правил транскрибирования (например, для украинского языка [22], для русского языка [19]) [1]. Так, например, для множества слов $W5000$ определено, что 92,8% 2- и 3-символьных сочетаний соответствуют 2-, 3- и 4-символьным сочетаниям в транскрипции, а 7,2% — соответствуют более длинным сочетаниям, которые можно получить объединением 2-, 3-, 4-символьных.

Известно, что ГС-сочетания сегментов С + Г, Г + С и С + С (Г — гласный, С — согласный) составляют три типа интеграции артикуляторных работ и составляют структуру артикуляторного жеста в таких произносимых единицах, как слог и фонетическое слово [26]. Для множества слов $W5000$ был выполнен анализ количественного состава 2- и 3-символьных ГС-сочетаний разных типов и определен порог $N_g \geq 5\%$ для включения их в алфавит (рис. 8).

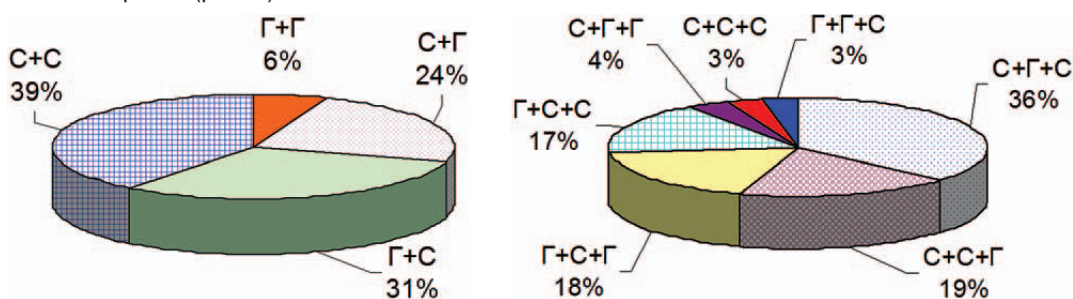


Рис. 8. Структура множеств 2- и 3-символьных сочетаний для 5000 наиболее частотных слов

Сформулируем правило выбора элементов в алфавит на основе декомпозиции наиболее частотных слов языка: выбирать 2- и 3-фонемные слоги, составляющие все типы интеграции артикуляторных работ, и которым отвечают 2, 3 или 4 символа в транскрипции и соответствующему количеству сегментов в последовательности параметров РС.

4. Группирование слогов

Организация алфавита образов, которые используются в алгоритме распознавания, должна обеспечить максимально высокую скорость доступа к каждому элементу в процессе поиска. Для этого предложено применение следующих уровней группирования элементов алфавита (рис. 9).

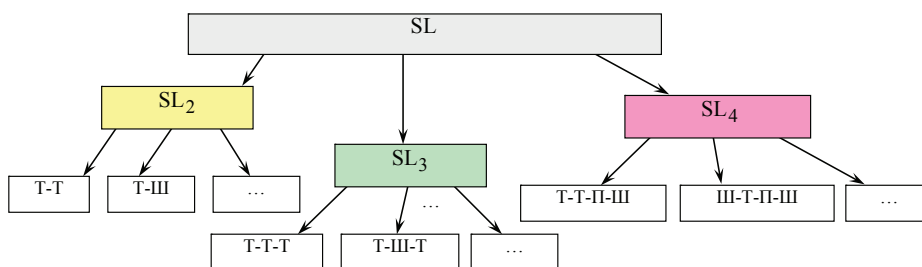


Рис. 9. Схема группирования слогов

Группирование по количеству символов $\alpha = M$ и количеству сегментов $\beta = L$:

$$SL_{2,2}, SL_{3,2}, SL_{2,3}, SL_{3,3}, SL_{2,4}, SL_{3,4} \subset SL$$

$$SL_2 \supset (SL_{2,2} \cup SL_{3,2}), SL_3 \supset (SL_{2,3} \cup SL_{3,3}),$$

$$SL_4 \supset (SL_{2,4} \cup SL_{3,4}), \text{ где } SL_{\alpha,\beta}, \alpha = 2, 3, \beta = 2, 3, 4, \text{ Коэффициент сокращения рассматриваемых элементов алфавита в каждом } \alpha - \beta\text{-подмножестве: } K_{\alpha,\beta} = Z / NSL_{\alpha,\beta}, \text{ где } NSL_{\alpha,\beta} \text{ — количество слогов в } \alpha - \beta\text{-подмножестве, } Z \text{ — общее количество слогов.}$$

Пусть для каждого сегмента траектории параметров РС по некоторым правилам можно определить тип сегмента (признаки (Ш) шумного, (Т) тонального или (П) паузы). Тогда для слога можно сформировать ТШП-транскрипцию в виде символьной последовательности из обозначений типов сегментов и выполнить группирование элементов в $\alpha - \beta$ -подмножествах по типу ТШП-транскрипции ($\alpha - \beta - \gamma$ -подмножества).

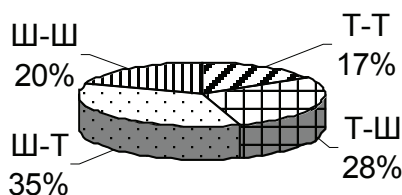
Информация о количественных данных алфавита слогов приведена в таблице 1 и рис. 10 (а, б, в).

Таблица 1

Количественные данные об алфавите слогов

Группа $SL_{\alpha,\beta}$	$\beta = 2$	$\beta = 3$	$\beta = 4$
$\alpha = 2$	438	121	10
$\alpha = 3$	0	1499	1579
Количество $NSL_{\alpha,\beta}$	438	1620	1589
Количество Z	3647		
Коэффициент $K_{\alpha,\beta}$	8,33	2,25	2,3

Схема группирования слогов на первом уровне позволяет сократить множество рассматриваемых элементов алфавита в наилучшем случае в 8,33 раза (12,01% от общего количества слогов в алфавите), а в наиболее худшем — в 2,25 раза (44,42% от общего количества слогов в алфавите).



а) 2-сегментные слоги

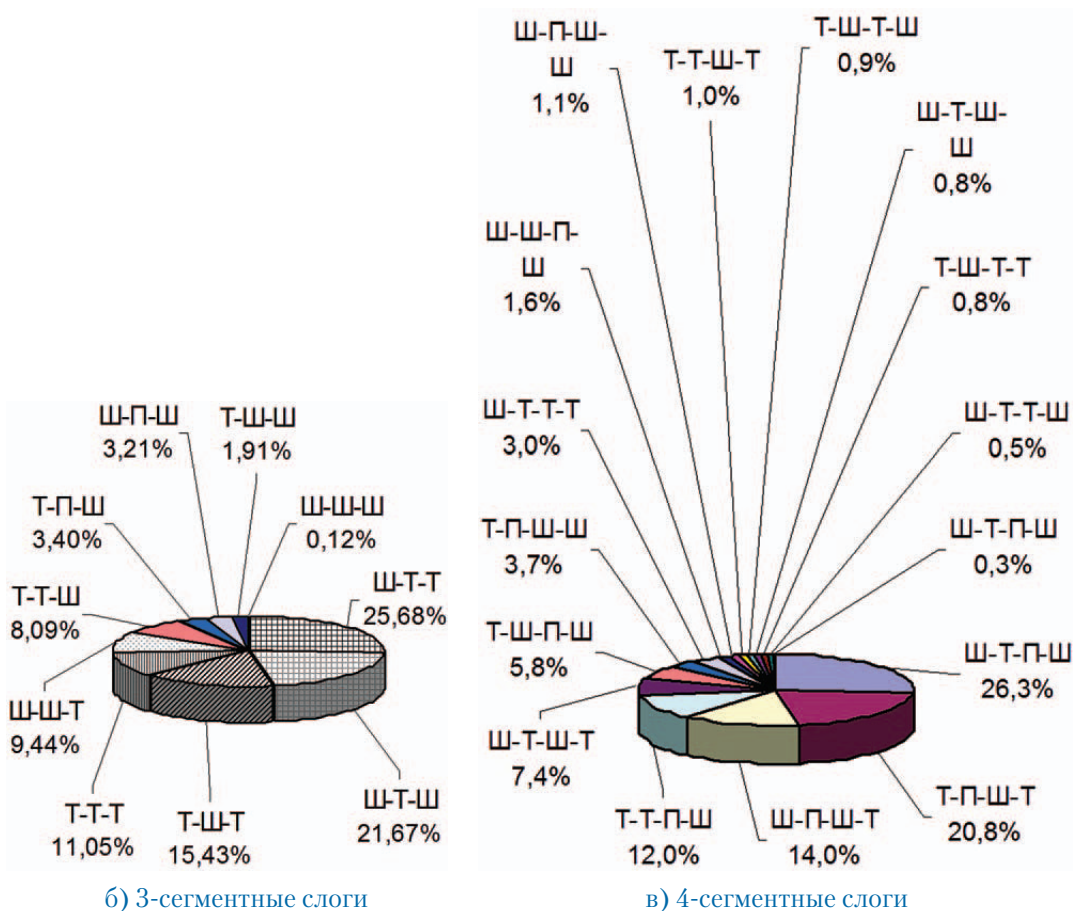


Рис. 10. Анализ алфавита слогов

5. Вычисление функции эвристической оценки состояний

Для выбора вида функции эвристической оценки состояний $h(n)$ в оценочной функции $f(n)$ (2) рассмотрены следующие характеристики сегментно-слового представления информации о РС: вложенность слогов $SL_{2,k} \in SL_{3,k}$ на уровне слогов и траекторий параметров; тип сегментов; значение среднего расстояния между ТП слогов, которые принадлежат к разным способам сочетания признаков сегментов. Перечисленные характеристики можно использовать независимо для вычисления $h(n)$. С учётом этого, выбраны следующие слагаемые оценки $h(n)$ в $f(n)$ (2):

- 1) для узла n вычисляется значение оценки $h_v(n)$, на основе анализа вложенности слогов $SL_{2,k} \in SL_{3,k}$;
- 2) оценка $h_g(n)$ вычисляется на основе анализа вариантов сочетания признаков сегментов (Т-Т, Ш-Т и т.п.);
- 3) оценка расстояния $h_d(n)$ для всех возможных переходов из узла n для разных вариантов сочетания признаков сегментов;
- 4) оценка количества нераскрытых узлов $h_o(n)$, которые остаются нерассмотренными в процессе поиска в направлении целевого узла.

Для каждого слагаемого 1)-4) экспертом задан соответствующий коэффициент (v, g, d, o), который имеет смысл веса соответствующей эвристической оценки, и правило нормирования весовых коэффициентов: $1/v + 1/g + 1/d + 1/o = 1$.

Таким образом, выражение для вычисления значения эвристической оценки $h(n)$ из узла n к целевому узлу имеет вид

$$h(n) = \hat{h}_v(n) + \hat{h}_g(n) + \hat{h}_d(n) + \hat{h}_o(n) = \\ = \frac{1}{v} h_v(n) + \frac{1}{g} h_g(n) + \frac{1}{d} h_d(n) + \frac{1}{o} h_o(n). \quad (9)$$

Для вычисления функции эвристической оценки состояний (9) был проанализирован словарь из 5000 наиболее частотных слов русского языка и соответствующие им записи из РБД [34]. Алгоритмы вычисления слагаемых функции эвристической оценки состояний и полученные результаты изложены в [36].

Модель распознавания

Последовательность этапов обработки информации о РС, которые составляют модель распознавания, представлена на рис. 11.

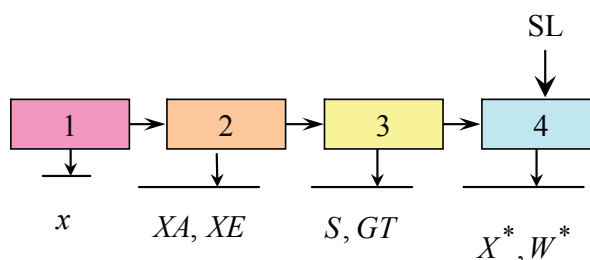


Рис. 11. Модель распознавания

Шаги 1–3 — аналогичные шагам 1–3 модели обучения, представленной на рис. 2.

В результате выполнения шагов 1–3 для предъявленного РС, сформирована траектория параметров в виде СВП XA , СПП XE , определены границы $S = \{s_1, \dots, s_L\}$ и тип GT сегментов.

Шаг 4. Поиск эвристических решений задачи сегментно-слового распознавания.

На основе экспериментальных исследований сделан следующий вывод относительно использования представления ТП (СВП, СПП) при распознавании: СПП для предварительного распознавания и выделения списка из N -list решений-кандидатов для ЭТП; СВП — для выбора одного или нескольких решений-кандидатов из списка N -list по критерию максимальной близости в смысле (1) к ТП предъявленного РС и принятие решения о распознавании (или установление факта отказа от распознавания).

1. Поиск эвристических решений первого уровня

Для нахождения решений-кандидатов задачи сегментно-слового распознавания (1) с помощью методов поиска в пространстве состояний (ПС) выполним формализацию задачи в терминах ПС:

1. ПС представим в виде корневого графа синтеза ЭТП X^* (рис. 12): количество узлов пространства состояний определяется в зависимости от количества сегментов L ТП XS предъявленного РС; нумерация узлов на графе отвечает номерам сегментов SG_i , $i = 1, \dots, L$ в ТП предъявленного РС; коэффици-

ент разветвления $Br = 3$; максимальная длина пути в ПС от начального узла к целевому $Len = \left\lceil \frac{L}{2} \right\rceil$; глубина самого поверхностного целевого узла $Dep = \left\lceil \frac{L}{4} \right\rceil$.

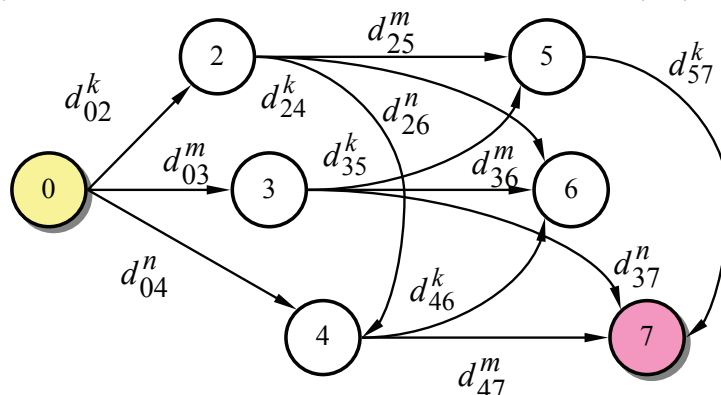


Рис. 12. Граф синтеза эталонной траектории параметров для $L = 7$

- Состояние B_i : ТП $X^*_M = (Y_1, Y_2, \dots, Y_i, \dots, Y_M)$ на основе конкатенации ТП Y_i слогов SL_i , $i = 1, \dots, M$; ТП Y_i слогов SL_i содержат $L_i = 2, 3, 4$ сегмента-фонемы.
- Начальное состояние S_0 : ЭТП X^* не содержит ТП слогов, т.е. $X^* = \emptyset$.
- Допустимые переходы на графе \mathbf{F} : дуги $l_i \rightarrow l_{i+N}$, $N = 2, 3, 4$, которые соединяют узлы на графе синтеза. Если не определены критерии для выбора дуги перехода из текущего состояния, то существует $Br!$ вариантов для перехода. Для выбранного перехода вычисляется значение стоимости дуги (евклидово расстояние) $d_{i, i+N}$, которое характеризует близость группы сегментов YSL_i предъявленного РС с траекторией параметров Y_k некоторого слога SL_k .
- Сумма стоимостей дуг, принадлежащих некоторому пути из начального в конечный узел, определяет интегральное сходство X с X^*_M

$$D = \sum_{i=1}^M d_{i, i+N}. \quad (10)$$

- Целевое состояние S_g : комбинация ТП Y_i с минимальным значением (10) и соответствующая комбинация слогов SL_i , на пути от начального к конечному сегменту предъявленного РС.

В результате анализа элементарных шагов сопоставления N -сегментных ($N = 2, 3, 4$) траекторий параметров слогов с ТП предъявленного РС, которая сегментирована на L сегментов-фонем, определена математическая модель генерирования решений-кандидатов задачи (1) без учета порядка следования слогов внутри ЭТП:

$$f(\eta, \lambda, \mu) = 2 \cdot \eta + 3 \cdot \lambda + 4 \cdot \mu, \quad (11)$$

где η, λ, μ — количество слогов, которые имеют 2-, 3- и 4-сегментные ТП соответственно. С помощью (11) можно оценить количество N var возможных решений-кандидатов для ЭТП: N var вычисляется исходя из количества наборов целочисленных значений $\{\eta, \lambda, \mu\}$, при которых выполняются условия $2 \cdot \eta + 3 \cdot \lambda + 4 \cdot \mu = L$, $\eta, \lambda, \mu < L$, $\eta, \lambda, \mu \geq 0$, т.е.

$$N \text{ var} = \sum_{\eta, \lambda, \mu} f(\eta, \lambda, \mu) \cdot \frac{(\eta + \lambda + \mu)!}{\eta! \cdot \lambda! \cdot \mu!}$$

$$\text{где } f(\eta, \lambda, \mu) = \begin{cases} 1, & 2 \cdot \eta + 3 \cdot \lambda + 4 \cdot \mu = L, \quad \eta, \lambda, \mu \geq 0, \\ 0, & \text{в другом случае.} \end{cases}$$

Обозначим некоторое решение-кандидат таким образом $Sol_j = \{path_j, X_j^*, dist_j\}, j = 1, \dots, N \text{ var}$, (путь решения $path_j = 0 \rightarrow \dots \rightarrow L$; в $S_0: path = \{\emptyset\}$), тогда множество решений $Sol = \{Sol_1, Sol_2, \dots, Sol_{N \text{ var}}\}$. Множество решений для примера графа синтеза (рис. 12) содержит $N \text{ var} = 5$ возможных путей (рис. 13): $path_1: 0 \rightarrow 2 \rightarrow 4 \rightarrow 7$; $path_2: 0 \rightarrow 2 \rightarrow 5 \rightarrow 7$; $path_3: 0 \rightarrow 3 \rightarrow 5 \rightarrow 7$; $path_4: 0 \rightarrow 3 \rightarrow 7$; $path_5: 0 \rightarrow 4 \rightarrow 5$.

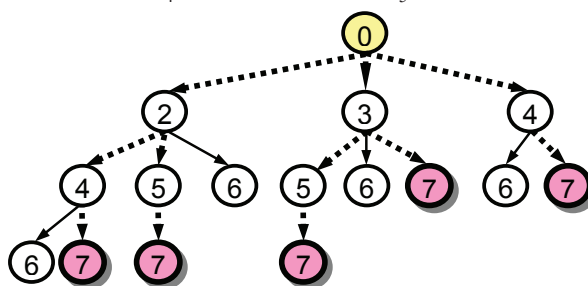


Рис. 13. Множество решений для случая $L = 7$

Зависимость $N \text{ var}$ от количества сегментов-фонем L в ТП предъявленного РС имеет экспоненциальный вид. Стратегия полного перебора вариантов практически не применима для поиска решений данной задачи вследствие большого объема вычислений при сопоставлении.

Один из способов поиска решений — вычисление стоимостей $d_{i, i+N}$ для всех допустимых переходов $l_i \rightarrow l_{i+N}$ на графе синтеза и нахождение оптимального пути из начальной вершины в конечную с помощью известных алгоритмов поиска кратчайшего пути, например, Дейкстры. Очевидно, что в зависимости от L такой поиск требует вычисления значений стоимостей в количестве

$$Cntedist(L) = \begin{cases} 1, & L = 2; \\ 2, & L = 3; \\ 4, & L = 4; \\ 3 \times (L - 3), & L \geq 5. \end{cases}$$

Если количество слогов в алфавите Z , то в общем случае для нахождения оптимального пути на графе синтеза необходимо осуществить, по меньшей мере, $N_1 = Z \times Cntedist(L)$ операций сравнения, что в свою очередь увеличивает время распознавания за счет лишних вычислений.

Другой способ использует предположение, что значения стоимостей всех дуг на графе синтеза неизвестны, и для нахождения ЭТП применяют стратегии направленного поиска, позволяющие сократить общее количество вычислений. Очевидно, что в наихудшем случае, используя второй способ поиска решений, необходимо осуществить $N_2 = Z \times Len$ операций сравнения ($Len < Cntedist(L)$ для $L \geq 5$).

Эвристический алгоритм поиска решений задачи сегментно-слового распознавания на основе алгоритма эвристического поиска в пространстве состояний, который использует оценочную функцию вида (2) с учетом (9)

и схему группирования (рис. 9), состоит из последовательности шагов А.1-А.7. Введены следующие обозначения: n_s, n_g — начальный и целевой узлы соответственно; $nSpisok, pSpisok$ — списки неразвёрнутых и развёрнутых узлов; $CntPall$ — счётчик общего количества развёрнутых узлов; $CntPg$ — счётчик количества узлов на пути от начального до целевого узла; $WsortSpisok$ — упорядоченный список построенных решений-кандидатов, для которых выполняется условие $D_{gr} \min \leq D \leq D_{gr} \max$ ($D_{gr} \min, D_{gr} \max$, определённые экспериментально границы оптимальности решения); $CntW$ — счётчик в списке $WsortSpisok$.

- А.1.** Внести начальный узел $n_s = 0$ в $pSpisok$, для каждого дочернего узла которого $n = n_s + N$ ($N = 2, 3, 4$) вычислить значение ОФ $f(n)$ (2) и внести эти узлы в $nSpisok$, $cntPall = 1$; $cntPg = 0$, $cntW = 0$. Особенность вычисления ОФ $f(n)$ на этом шаге — для эвристической оценки состояний $h(n)$ (9) не используется оценка h_v ($h_v(n_s) = 0$, поскольку для начального узла не существует родительских узлов, что делает невозможным её вычисление).
- А.2.** Если $nSpisok$ пустой — окончание алгоритма, иначе — перейти к шагу А.3.
- А.3.** Выбрать из $nSpisok$ узел $n = n_{f \min}$ с минимальным значением ОФ $f(n)$:

$$f_{\min} = \min([f(n_s + 2), f(n_s + 3), f(n_s + 4), \dots], \dots, f(n + 2), f(n + 3), f(n + 4)).$$
 Изменить решение-кандидат для X^* , добавив ТП слога из соответствующего α -, β -, γ -подмножества алфавита SL в i -тую позицию (i определяется номером текущего сегмента ТП предъявленного РС X). Вычислить значение стоимостей $D_{i, i+N} \cdot D$ (10).
- А.4.** Если узел $n_{f \min}$ целевой, т.е. $n_{f \min} = n_g$, то выполняется операция композиция символьной последовательности-решения задачи распознавания W^* , которая отвечает найденной X^* .
- А.5.** Если $D < D_g \min$, то завершается работа алгоритма поиска и на выход поступает найденный ответ W^* , в противоположном случае выполняются следующие операции: узел $n = n_{f \min}$ вносится в $pSpisok$ ($cntPall = cntPall + 1$) и удаляется из $nSpisok$; если $D_g \min \leq D \leq D_g \max$, то найденное решение вносится в $WsortSpisok$ для следующего этапа принятия решения ($cntW = cntW + 1$); выполняется переход к шагу А.3.
- А.6.** Если $n_{f \min} \neq n_g$, развернуть узел $n_{f \min}$, построив все его дочерние узлы $n = n_{f \min} + N$ ($N = 2, 3, 4$). Внести узел $n_{f \min}$ в $pSpisok$ ($cntPall = cntPall + 1$), удалив из $nSpisok$. Если для узла $n_{f \min}$ отсутствуют дочерние узлы, то перейти к шагу А.2, в противоположном случае — к шагу А.7.
- А.7.** Для каждого дочернего узла $n = n_{f \min} + N$ ($N = 2, 3, 4$) вычислить значение ОФ $f(n)$ и внести все узлы в $nSpisok$. Перейти к шагу А.2.

После завершения эвристического поиска на следующем этапе принятия решения выполняется анализ списка $WsortSpisok$ и формирование списка «потенциальных» $Nlist$ ответов распознавания W^* или устанавливается факт отказа от распознавания. Нахождение списка из $Nlist$ решений-кандидатов для ЭТП X^* из условия наилучшей близости к траектории параметров XE предъявленного РС в смысле (1) осуществляется методом динамического программирования [1].

Построение траекторий параметров решений-кандидатов для ЭТП выполняется для случая СПП XE^* с помощью сегментно-слового сплайн-синтеза на основе следующей модели.

Пусть известно, что ЭТП для предъявленного РС состоит из ТП R слогов. Модель сегментно-слового сплайн-синтеза ЭТП в частотной полосе (для СПП) имеет вид:

$$Y(t; \bar{\theta}) = \begin{cases} Y_{1,1}(t; \bar{\theta}_{1,1}), & K_0 \leq t \leq K_1, \\ \dots \\ Y_{R,1}(t; \bar{\theta}_{R,1}), & K_{S-1} \leq t \leq K_S, \end{cases}$$

где

$$Y_{j,k}(t; \bar{\theta}_{j,k}) = a_{j,k} + b_{j,k} \cdot (t - K_s) + c_{j,k} \cdot (t - K_s)^2 + d_{j,k} \cdot (t - K_s)^3,$$

$$\bar{\theta}_{j,k} = \{a_{j,k}, b_{j,k}, c_{j,k}, d_{j,k}, K_s\}, \quad j = 1, \dots, R, \quad k = 1, \dots, L_j,$$

$$L_j = 2; 3; 4, \quad s = 1, \dots, S.$$

В каждой частотной полосе l параметры сплайн-моделей

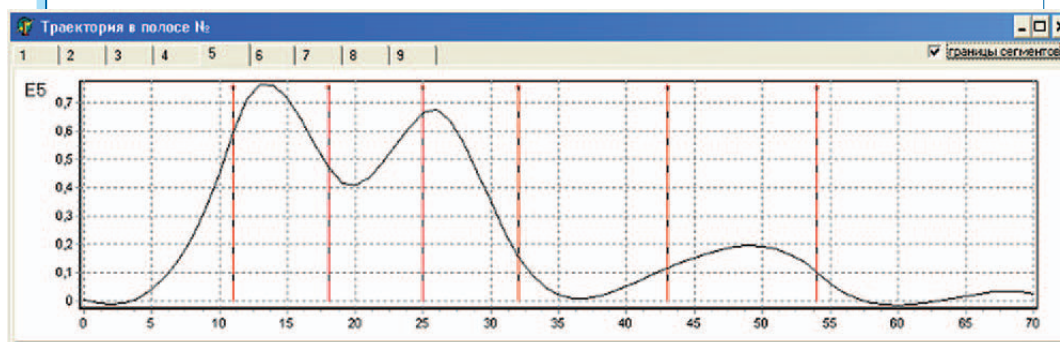
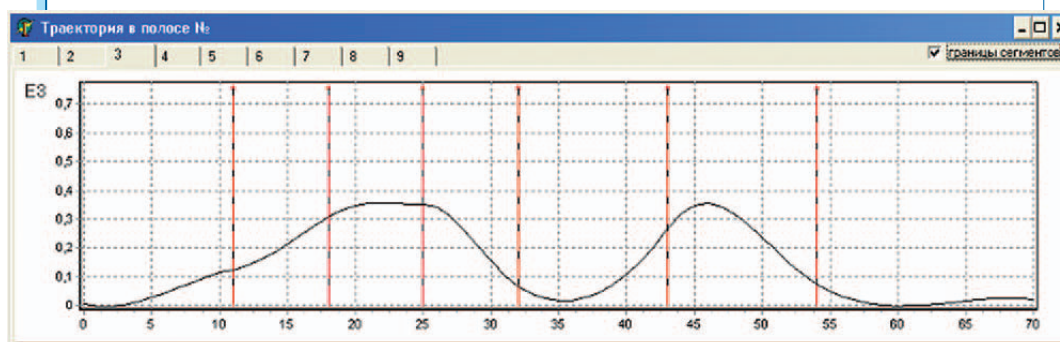
$$\bar{\theta}_{j,k} = \{a_{j,k}, b_{j,k}, c_{j,k}, d_{j,k}, K_s\}, \quad j = 1, \dots, R, \quad k = 1, \dots, L_j,$$

$L_j = 2; 3; 4$ для траекторий параметров слогов алфавита, составляющих ЭТП, вычисляются из условия минимизации ошибки аппроксимации

$$\sigma_l^2 = \sum_j \sum_k (X_{l,j,k} - Y_{l,j,k})^2 \rightarrow \min \text{ и условий, обеспечивающих гладкость}$$

и непрерывность в узлах сплайн-функций $K = (K_1, K_2, \dots, K_s, \dots, K_S)$, в качестве которых выбраны точки сегментации ТП слогов и точки конкатенации ТП слогов внутри ЭТП.

На рис. 14 приведён пример (для 3, 5, 9 полос СПП) сегментно-слового сплайн-синтеза ЭТП для предъявленного к распознаванию слова «человек».



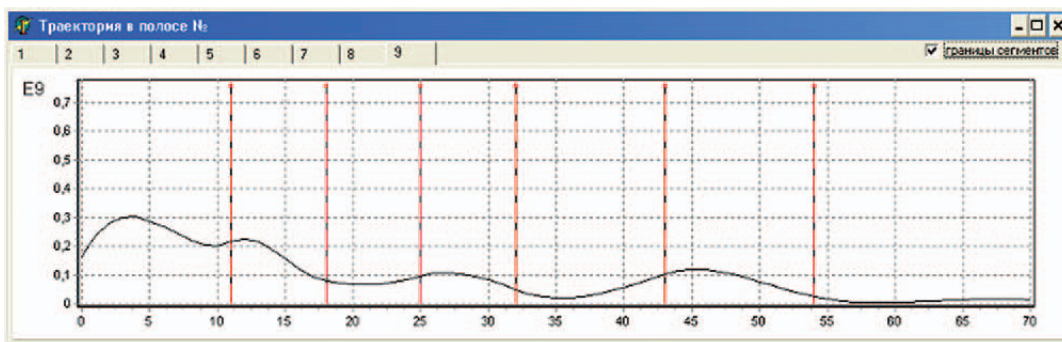


Рис. 14. Пример сплайн-синтеза ЭТП в полосе

2. Поиск решений второго уровня

Для выбранных $Nlist$ решений-кандидатов выполняется сегментно-слоговой синтез ТП XLA^* в классе КФ из ТП YLA слогов алфавита для случая СВП и сопоставление с ТП XA предъявленного РС методом динамического программирования [1]. В результате сопоставления находится такая XLA^* , которая наилучшим образом соответствует XA в смысле (1).

Модель сегментно-слоговой синтеза ЭТП, состоящая из R слогов, в классе КФ имеет вид:

$$XLA^* = YLA_{1i} + YLA_{2i} + \dots + YLA_{mi} + \dots + YLA_{Ri}, \quad N'_0 \leq i \leq N'_R, \quad (12)$$

где $m = 1, \dots, R$; N_m — количество временных отсчетов ТП YLA_{mi} (4), $t_1 \in [1, N_1], \dots, t_m \in [1, N_m], \dots, t_R \in [1, N_R]$; $N'_0 = 1, N'_R = N_1 + N_2 + \dots + N_R$.

Алгоритм сегментно-слоговой синтеза ЭТП для СВП состоит из шагов Б.1-Б.3 [31].

Б.1. Нахождение области определения ЭТП $D: [\omega_0, \omega_M] \times [t_0, t_N]$.

Временной диапазон ЭТП $[t_0, t_N]$, где $t_N = t_{N_1} + t_{N_2} + \dots + t_{N_R}$. Диапазон частот ЭТП $[\omega_0, \omega_M]$ совпадает с диапазоном частот YLA_m .

Б.2. Восстановление ТП YLA_m выбранных R слогов из алфавита SL согласно (4) на всей области определения ЭТП $XLA^* D: [\omega_0, \omega_M] \times [t_0, t_N]$.

ТП j -го слога определена в прямоугольной области $D_A^j[\omega_0, \omega_M] \times [t_0, t_{N_j}]$ и представлена параметрами $a_{(i)}, b_{(i)}, c_{(i)}, d_{(i)}, T_{(i)}, \Omega_{(i)}$ функций $Zt_{(i)}(t_l), Z\omega_{(i)}, (\omega_k)$ (3), $i = 1, \dots, L_j, k = 1, \dots, M, l = 1, \dots, N_j$.

Б.3. Синтез ЭТП XLA^* согласно модели (12).

Описание СВП XLA^* в некоторой точке (ω_k, t_l) представлено в виде суперпозиции R функций YLA_m ($m = 1, \dots, R$), которые являются аналитическим описанием СВП соответствующих m слогов из алфавита SL , таким образом

$$XLA^*(\omega_k, t_l) = \sum_{m=1}^R YLA_m(\omega_k, t_l) = \sum_{m=1}^R \left(\sum_{i=1}^{L_m} Z\omega_{(i)}(\omega_k) \cdot Zt_{(i)}(t_l) \right), \quad (13)$$

где L_m — количество параметров КФ $Zt_i(t_l), Z\omega_{(i)}(\omega_k)$ (3), $i = 1, \dots, L_m$ для соответствующего m -го слога ЭТП.

На рис. 15 — 18 приведен пример реализации сегментно-слового синтеза СВП для РС «человек».

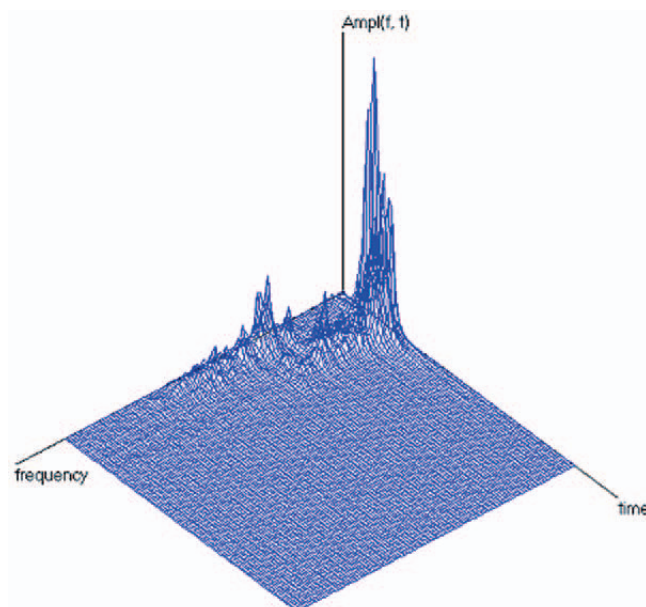


Рис. 15. ТП слога «че» из алфавита, восстановленная на области определения ЭТП

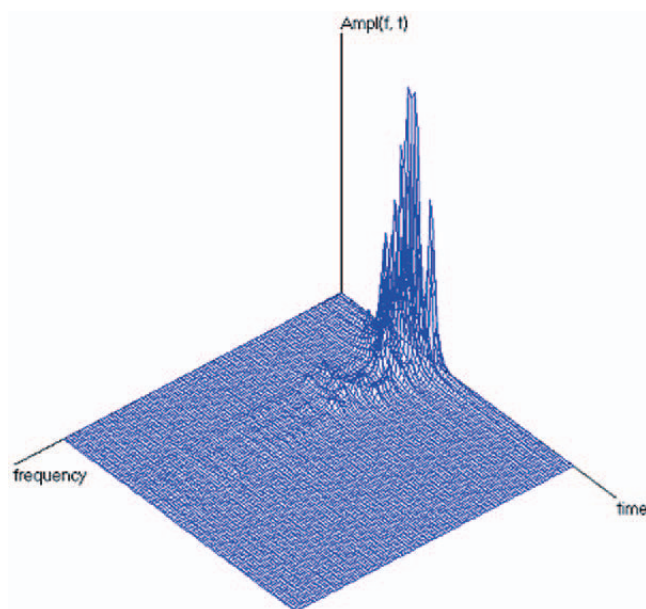


Рис. 16. ТП слога «ло» из алфавита, восстановленная на области определения ЭТП

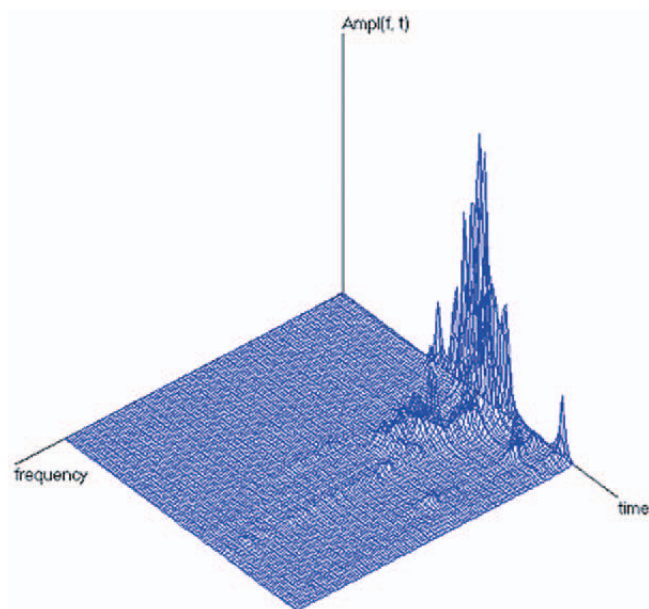


Рис. 17. ТП слога «век» из алфавита, восстановленная на области определения ЭТП

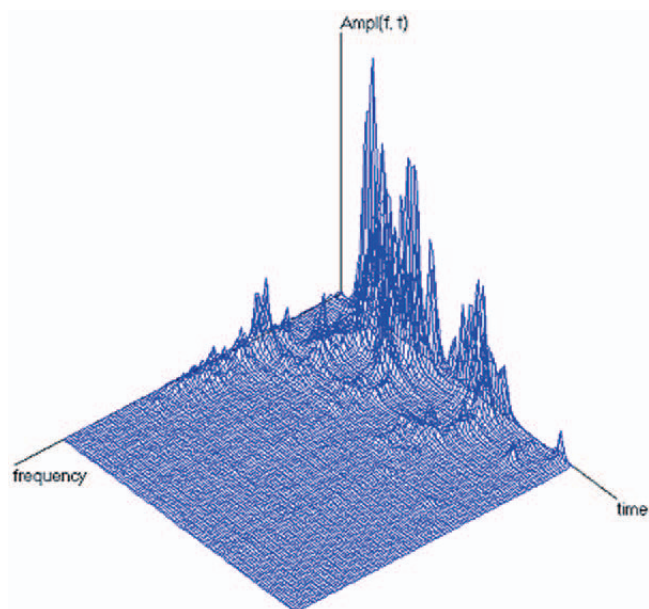


Рис. 18. ЭТП для РС «человек»

3. Композиция символического ответа распознавания W^*

Для предъявленного РС выполняется композиция символического ответа распознавания W^* на основе конкатенации (& — операция конкатенации) символической информации слогов алфавита, которые входят в ЭТП XLA^* : $W^* = SL_1 \& SL_2 \& \dots \& SL_R$, или в фонемном виде — $W^* = (P_{1,1} \& P_{2,1} [\& P_{3,1}]) \& (P_{1,2} \& P_{2,2} [\& P_{3,2}]) \& \dots \& (P_{1,R} \& P_{2,R} [\& P_{3,R}])$.

Эксперимент

Для исследований в «SPeacH» использована РБД для списка $W5000$, которая содержит записи РС 10 мужчин и 10 женщин. Записи РБД сгруппированы в следующие выборки: SW_1 (выборка РС для множества слов W_1 , из которых сформировано обучающее множество РЕ); SW_2 (выборка с другими реализациями РС для множества слов W_1); SW_3 (выборка РС для множества слов W_2 , причём W_1, W_2 такие, что $W_1 \cap W_2 = \emptyset$). Средняя длина реализации РС в сегментах без учета продолжительности сегментов: $\bar{L} = 7$. Средняя длительность реализации РС в РБД составляет $\bar{t} = 1,20$ с. Примем это значение сопоставимым с реальным временем.

Проведены исследования адекватности моделей сегментно-слогового синтеза ЭТП, надёжности и быстродействия распознавания РС из РБД. Оценка адекватности моделей сегментно-слогового синтеза ЭТП выполнена по критерию надёжности распознавания для каждой выборки РС. Также вычислена оценка разборчивости синтезированного ответа системы распознавания, которая характеризует качество синтезированной ЭТП (оценка разборчивости измеряется процентом правильно распознанных слов аудиторами [37]). В экспериментах принимали участие 20 человек, которым было предложено записать услышанные слова (объём тестового словаря составил 100 слов). Количество верно распознанных слов аудиторами — 94%, что является приемлемым. Это свидетельствует о том, что эвристический алгоритм поиска решений задачи сегментно-слогового распознавания адекватно выбирает слоги из алфавита, а предложенная модель сегментно-слогового синтеза СВП в классе КФ позволяет с достаточной точностью восстановить ЭТП для распознаваемого РС из ТП слогов алфавита. Результаты исследований каждого этапа поиска эвристических решений приведены в таблице 2 (Tcp — среднее время распознавания одной реализации РС; Err — ошибка распознавания).

Таблица 2

Анализ алгоритмов сегментно-слогового синтеза

Критерии	СВП			СПП		
	SW_1	SW_2	SW_3	SW_1	SW_2	SW_3
Tcp , сек	0,5	0,55	0,45	0,5	0,35	0,40
Err , %	2,70	3,00	5,00	7,00	9,00	15,00
$Nlist$, %	7,00	10,00	10,00	10,00	12,00	15,00

Для предъявленной реализации РС время поиска на каждом этапе в среднем составляет 0,50 сек. Согласно результатам экспериментов, задержка между окончанием ввода РС с микрофона и ответом системы распознавания составляет в среднем 1,00 сек., что является допустимым для современных диалоговых систем. Использование эвристической оценочной функции позволяет уменьшить ошибку распознавания до 5%. Проведен анализ ошибок распознавания РС, в ходе которого выявлено, что большая часть ошибок распознавания связана с ошибками при определении границ сегментов, например, между безударной гласной и сонорными согласными. Поэтому одним из дальнейших направлений исследований является усовершенствование алгоритмов сегментации.

Литература

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. К.: Наукова думка, 1987.
2. Винцюк Т.К. Образный компьютер: Концепции, методология, подходы. // Системы технического зрения и искусственного интеллекта с обработкой и распознаванием изображений. К., 2001. С. 125–138.
3. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. М.: РГГУ, 2001.
4. Карпов О.Н. Технология построения устройств распознавания речи. Д.: Изд-во Днепрпетр. ун-та, 2001.
5. Волошин В.Г. Компьютерная лингвистика. С.: ВТД, 2004.
6. Жожикашвили В.А., Петухова Н.В., Фархадов М.П. Компьютерные системы массового обслуживания и речевые технологии. // Проблемы управления. 2006. № 2. С. 3–7.
7. Ronzhin A.L., Yusupov R.M., Li I.V., Leontieva A.B. Survey of Russian speech recognition systems // SPECOM'2006. St. Peterburg, 2006. P. 54–60.
8. Pylypenko V. Information retrieval based algorithm for extra large vocabulary speech recognition // SPECOM'2006. St. Peterburg, 2006. P. 67–69.
9. Кушнир Д.А., Харламов А.А. Распознавание речи в базе многомерного сигнального пространства // Информационные технологии. 2005. № 5. С. 30–36.
10. Загоруйко Н.Г. Методы распознавания и их применение. М.: Сов. Радио, 1972.
11. Распознавание слуховых образов / Под ред. Н.Г. Загоруйко, Г.Я. Волошина. Н.: Наука, 1970.
12. Рассел С., Норвиг П. Искусственный интеллект. М.: Вильямс, 2006.
13. Карпов А.А. Модели и программная реализация распознавания русской речи на основе морфемного анализа: Автореф. дисс. канд. техн. наук: 05.13.11. Санкт-Петербург, 2007.
14. Егоров А.И., Дубровский В.В. Об анализе слуховых образов речевого сигнала. http://www.kcn.ru/tat_en/science/fccl/ar1.htm
15. Теория слога. www.erudition.ru
16. Коресёк I. Speech Recognition and Syllable Segments. // Workshop on Text, Speech and Dialogue TSD'99. Lectures Notes in Artificial Intelligence. Springer-Verlag, 1999. P. 203–208.
17. Shastri L., Chang S., Greenberg S. Syllable detection and segmentation using temporal flow neural networks // Int. Congress of Phonetic sciences. San Francisco, 1999. P. 138–146.
18. Белявский В.М., Светозарова Н.Д. Слоговая фонетика и три фонетики Л. В. Щербы. <http://www.auditech.ru/doc/cherba.htm>
19. Орлов И.А. Слоговой компиляционный синтез русской речи // Речевая информатика. М.: Наука, 1989. С. 119–139.
20. Лингвистический энциклопедический словарь. М.: Сов. энциклопедия, 1990.
21. Vasylyeva N., Sazhok M. Text selection for training procedures under phoneme units variety // SPECOM'2005, St. Peterburg, June 25 — 29, 2005. St. Peterburg, 2005. P. 629–631.
22. Крак Ю.В., Горбань В.В. Один из подходов к разработке системы автоматического озвучивания текстов на украинском языке // Искусственный интеллект. 2004. № 1. С. 196–203.
23. Togawa F., Hakaridani M., Iwahashi H. Voice activated word processor with automatic learning for dynamic optimization of syllable templates // ICASSP'86. Int. Conf. Acoust., Speech and Signal Process. New York, 1986. Vol. 2. P. 1121–1124.
24. Tsuboi T., Tomihisa A., Sugamura N. Japanese linguistic processing for continuous speech recognition // ICASSP'87. Int. Conf. Acoust., Speech and Signal Process. New York, 1987. Vol. 2. P. 805–808.

25. Бондарко Л.В., Зиндер Л.Р., Штерн А.С. Некоторые статистические характеристики русской речи // Слух и речь в норме и патологии. Л., 1977. Вып. 2. С. 3–16.
26. Дудник З.В., Затайдух О.В., Майдигов П.В. Использование Perl и Excel для создания базы данных и статистической оценки текстовых массивов в русском и украинском языках. <http://www.philol.msu.ru/~rlc2004/ru/participants/psearch.php?pid=19229>
27. Савенкова О.А., Карпов О.Н. Некоторые эксперименты по повышению надежности распознавания слов заданного словаря // Системные технологии. 2004. Вып. 35. С. 60–66.
28. Романенко С.В., Стромберг А.Г. Классификация математических моделей аналитических сигналов в форме пиков // Журнал аналитической химии. 2000. Т. 55. № 11. С. 1144–1148.
29. Карпов О.Н. Вычислительные схемы представления функций многих переменных в классах функций меньшего числа переменных. Д.: Изд-во Днепрпетр. ун-та, 2003.
30. Карпов О.Н., Габович А.Г., Марченко Б.Г. Компьютерные технологии распознавания речевых сигналов. К.: Полиграф-Консалтинг, 2005.
31. Савенкова О.О., Карпов О.Н. Применение колоколообразных функций в алгоритме сегментно-слогового синтеза // Математическое моделирование. 2008. Т. 1(18). С. 5–9.
32. Савенкова О.А., Карпов О.Н. Технология построения интеллектуальной системы распознавания речи // Искусственный интеллект. 2008. № 4. С. 785–795.
33. Де Бор К. Практическое руководство по сплайнам. М.: Радио и связь, 1985.
34. Частотный словарь. <http://www.artint.ru/projects/frqlist.asp>
35. Савенкова О.А. Разработка нейросетевого алгоритма поиска решений задачи распознавания речи // Вестник академии таможенной службы Украины. 2010. № 43. С. 137–144.
36. Савенкова О.А. Вычисление слагаемых функции эвристической оценки состояний в пространстве состояний задачи сегментно-слогового распознавания речи // Труды конф. УкрОбраз'2010. Киев, 2010. С. 69–72.
37. Людовик Т.В., Сажок Н.Н. Использование речевых баз данных большого объема при синтезе речи в системах искусственного интеллекта. // Проблемы управления и информатики. 2003. № 6. С. 82–87.

Сведения об авторе

Савенкова Ольга Александровна —

Область интересов: обработка и распознавание речевых сигналов, нейронные сети. E-mail: 2sol@ukr.net