

2. Ландэ Д. В. Поиск знаний в Интернет. Профессиональная работа: пер. с англ. / Д. В. Ландэ. – М.: Диалектика, 2005. – 272 с.

3. Поискковые системы. Алгоритмы машинного обучения [Электронный ресурс]. – Режим доступа: <http://www.scu.edu.au/programme/fullpapers/1938/com1938.htm>.

УДК 004.934.2

**Т.В. Шарий**, канд. техн. наук

Донецкий национальный университет, г. Донецк, Украина

## МОДЕЛЬ ПОСТОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ FCAS

*В статье рассматривается актуальный вопрос обработки параметров речевых сигналов в задаче распознавания речи. Указаны недостатки современных статистических моделей, с учетом которых предложена многоуровневая нечеткая когнитивная модель FCAS. Ядро FCAS представляет собой сеть фонетических процессоров, учитывающих вес речевого сегмента и работающих на признаковом, фонемном и словесном уровнях. Рассмотрена динамика и алгоритмы функционирования FCAS. Приведены результаты экспериментов с программной реализацией FCAS на тестовом словаре.*

**Ключевые слова:** речевой сигнал, когнитивное моделирование, фонема, элементарный фонетический процессор, FCAS.

*У статті розглядається актуальне питання обробки параметрів мовних сигналів у завданні розпізнавання мови. Вказано недоліки сучасних статистичних моделей, з урахуванням яких запропоновано багаторівневу нечітку когнітивну модель FCAS. Ядро FCAS являє собою мережу фонетичних процесорів, що враховують вагу мовного сегмента і діють на ознаковому, фонемному і словесному рівнях. Розглянуто динаміку й алгоритми функціонування FCAS. Наведено результати експериментів із програмною реалізацією FCAS на тестовому словнику.*

**Ключові слова:** мовний сигнал, когнітивне моделювання, фонема, елементарний фонетичний процесор, FCAS.

*The paper deals with an actual issue of speech signal parameters processing in context of speech recognition task. Starting with the shortcomings of state-of-the-art statistical models, pointed out in the paper, multilevel fuzzy cognitive model FCAS is proposed. The FCAS kernel is a network of elementary phonetic processors, processing the weights of speech segments and functioning at feature, phoneme and word levels. Model dynamics and algorithms are given. The results of experiments with FCAS program implementation on a test vocabulary are presented.*

**Key words:** speech signal, cognitive modeling, phoneme, elementary phonetic processor, FCAS.

**Постановка проблемы.** Проблема построения эффективных командных систем голосового управления не перестает быть актуальной в сфере информационных технологий на протяжении нескольких десятилетий. На сегодняшний день разработано множество подходов, моделей и методов автоматического распознавания речи, но на практике они не отличаются необходимой точностью. Значительный прогресс был достигнут за последние два года благодаря технологиям GoogleVoice и Apple Siri [1], но успехи этих решений обусловлены применением облачных вычислений, а не разработкой принципиально новых и эффективных моделей и алгоритмов. Рынок русскоязычных программ представлен единичными разработками, демонстрирующими посредственные результаты даже в условиях отсутствия шума.

**Анализ последних исследований и публикаций.** Архитектура современных систем включает два основных модуля – модуль параметризации сигнала (front-end), производящий цифровую обработку речевого сигнала (PC) и формирующий последовательность векторов признаков (ВП) (некоторое компактное описание сигнала), и модуль постобработки сигнала (back-end) [2], выполняющий распознавание слов на основе полученных ВП и закона условных вероятностей Байеса. Такие системы сначала обучаются на многочасовых коллекциях речевых данных, и затем, на этапе распознавания, производят сопоставление входных образов с ранее введёнными по обученным моделям.

В компьютерном распознавании речи часто делается акцент на учет особенностей слуховой и голосовой систем человека [2-8]. На этапе параметризации стандартом является применение метода MFCC (мел-частотных кепстральных коэффициентов) [2]. Данный метод позволяет получить компактное описание спектра сигнала с учетом свойственной человеку логарифмической частотной шкалы мел. Тем не менее, статис-

тический подход (Скрытые Марковские модели), используемый на *этапе постобработки*, не отражает явно ни специфики РС, ни особенностей восприятия речи человеком. Он демонстрирует хорошие результаты только при большом размере обучающих данных, собранных при большом числе различных условий. Поэтому модель постобработки РС по-прежнему открыта для исследований и модификаций.

**Целью статьи** является разработка и экспериментальное исследование новой модели постобработки РС, в меньшей степени зависящей от речевой статистики, в которой адекватно отображены особенности восприятия речи человеком.

**Изложение основного материала статьи.** В статье предлагается новая нечеткая когнитивная модель постобработки речевых сигналов FCAS (Fuzzy Cognitive Accented Speech). Структура модели FCAS отображена на рисунке 1.

Основным блоком модели является ядро FCAS, представляющее собой когнитивную сеть элементарных фонетических процессоров. Эти процессоры обрабатывают информацию, получаемую от трех моделей, – модели вычисления весов фонетических сегментов, признаковой и фонемной моделей. Также учитывается фонетическая статистика русского языка некоторой предметной области (словаря).

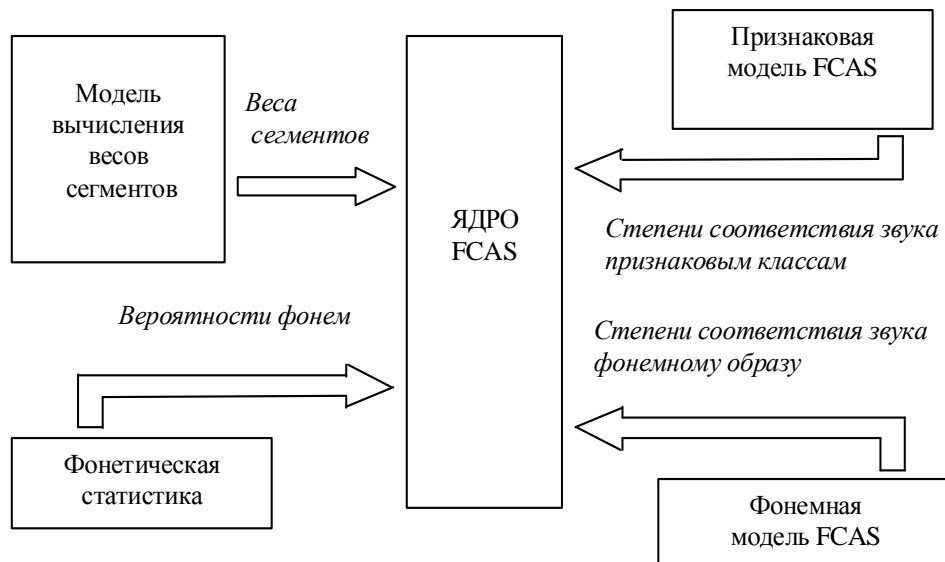


Рис. 1. Структура модели FCAS

Основным отличием предлагаемой модели от Скрытых Марковских моделей является многоуровневый учет акустико-фонетической информации на этапе постобработки речевого сигнала. В модели FCAS не только «заложена» статистика звуков, слогов и слов русской речи и их спектральные прототипы, но и предусмотрены различные параметры алгоритма обработки информации в зависимости от веса сегмента (модель вычисления весов [4]), типа сегмента (признаковая модель [5]) и последовательности сменяющихся сегментов звуков речи (ядро FCAS).

Далее в статье сосредотачивается внимание на ядре FCAS, в предположении, что степени соответствия текущего звука речи признаковым классам (в признаковой модели) и фонеме (в фонемной модели) уже вычислены, и имеется фонетическая статистика предметной области.

Формально ядро FCAS (сеть FCAS) определим как шестерку:

$$\langle w, \{\mu_{D_i}\}_{i=1}^{ND}, \{\mu_{G_j}\}_{j=1}^{NG}, \{\Phi_k\}_{k=1}^N, \{u_{ij}\}_{i,j=1}^N, y \rangle, \quad (1)$$

где  $w$  – вес речевого сегмента;  $\mu_{D_i}$  – степень соответствия речевого сегмента  $i$ -ому признаковому классу (например, классы «звонкий», «гласный», «компактный» и т. д.);  $N_D$  – число признаков классов, обрабатываемых моделью;  $\mu_{G_j}$  – степень соответствия речевого сегмента  $j$ -ой фонеме (например, фонемы [а], [к], [о] и т. д.);  $N_G$  – число фонем, обрабатываемых моделью;  $\{\Phi_k\}_{k=1}^N$  – набор элементарных фонетических процессоров (ЭФП);  $N$  – количество ЭФП в сети;  $u_{ij}$  – связь между  $i$ -м и  $j$ -м ЭФП;  $y$  – выход сети.

Первые три объекта являются входом модели ядра FCAS, поступающим от соответствующих моделей. ЭФП является узлом сети FCAS. Выход  $y$  представляет собой последовательность  $M$  распознанных текстовых символов  $\{y_1, y_2, \dots, y_M\}$ .

На рисунке 2 приведена иллюстрация сети FCAS.

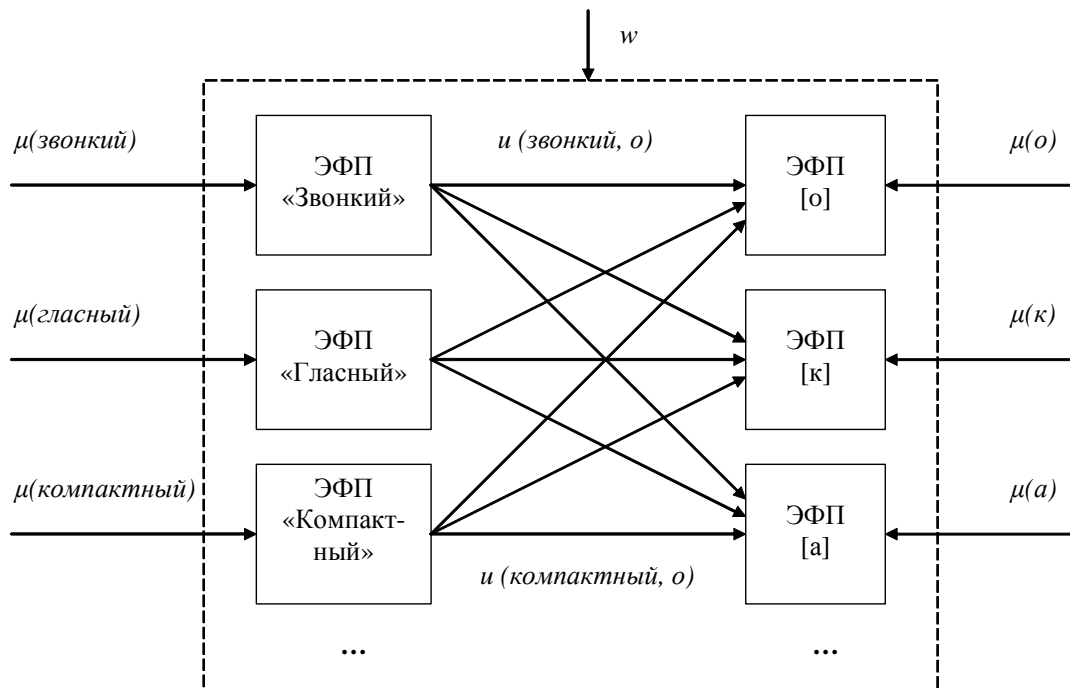


Рис. 2. Иллюстрация фрагмента сети FCAS

На рисунке 2 в левом столбце отображены три из  $N_D = 22$  элементарных фонетических процессоров *признакового* уровня, реагирующих на присутствие определенного фонологического признака в текущем звуке речи; в правом столбце – три из  $N_G = 26$  ЭФП *фонемного* уровня, реагирующих на сходство текущего звука речи с определенной фонемой. В таблице 1 указаны признаки и фонемы, обрабатываемые моделью FCAS. В общем случае в сети FCAS могут также присутствовать ЭФП словесного уровня, соответствующие отдельным словам из лексикона FCAS.

Речевой сигнал обрабатывается с помощью модели FCAS во времени. Сигнал поступает на обработку порциями (квантуется на перекрывающиеся фреймы длительностью 40 мс со степенью перекрытия 10 мс). Таким образом, шаг по времени  $\Delta t$  модели FCAS равняется 10 мс. Вычисления модели (пересчет внутренних состояний узлов сети) происходят с указанной периодичностью. Кроме того, речевой сигнал подвергается автоматической фонемной сегментации [4], в связи с чем в нем присутствуют маркеры фонетических сегментов (границы звуков речи). Моменты времени, соответствующие данным границам, будем называть *моментами принятия фонемного решения*. В эти временные срезы модель выдает максимальные степени соответствия звучащего речевого сегмента конкретным фонемам. Отметим, что для любого алгоритма сегментации

характерен некоторый процент лишних (вставляемых) границ сегментов. С учетом механизмов накопления / убавления состояний ЭФП эта ситуация может быть «сглажена», т. к. соседние сегменты будут считаться одной фонемой (с различными лишь, возможно, степенями соответствия). Кроме этого, как будет показано далее, модель FCAS может выдать большую степень соответствия некоторой фонеме не только в момент принятия фонемного решения, но и в любой другой момент времени  $t$  при выполнении определенных условий. Таким образом частично учитывается вторая проблема алгоритмов автоматической сегментации речевых сигналов – пропуск маркера сегментации.

Таблица 1

## Перечень признаков и фонем FCAS

Признаки	Звонкий	Глухой	Нешумовой	Шумовой	Гласный	Сонорный												
	Низкий	Плавный	Носовой	Звонкий взрывной	Звонкий шипящий	Высокий диффузный												
	Низкий диффузный	Низкий компактный	Признак С	Признак Ш	Признак Ф, Х	Глухой взрывной												
	Шипящий взрывной	Высокий	Высокий компактный		Глухой шипящий													
Фонемы	и	ы	э	у	а	о	л	р	м	н	б	д	г	з	ж	в	п	т
	ш	ф	х	ц	й	к	ч	с										

Всей сети узлов присвоено название «когнитивной», т. к. активация элементов, отвечающих за восприятие, производимая входным речевым сигналом, оставляет за собой «след» анализа входа на каждом из уровней обработки (фонетическом, признаковом, словесном). Этот след особый в том смысле, что он динамический, т. к. он складывается из активаций элементарных фонетических процессоров, а эти элементы продолжают взаимодействовать в течение времени. Различие между восприятием и первичной памятью полностью размыто, т.к. объект восприятия развертывается в тех же структурах, которые служат и в роли памяти, и перцептивная обработка ранних порций данных продолжается, даже когда в систему поступают новые порции. Такие продолжающиеся взаимодействия теоретически позволяют модели учитывать эффекты с «правым контекстом», а также конкретные аспекты кратковременной памяти, такие как факт, что в короткие промежутки времени можно сохранить больше информации, если она представлена единым связным целым. С этой точки зрения данная модель концептуально напоминает модель TRACE, предложенной Макклелландом и Элманом в [6]. Однако непосредственно алгоритмы обработки акустической и фонетической информации, рассматриваемые далее, принципиально отличаются. Кроме того, в отличие от модели FCAS, в модели TRACE отсутствует учет весов речевых сегментов, и в ней фиксированы фонемные интервалы, что накладывает ограничения на ее практическое использование.

ЭФП признакового уровня (рис. 3) определим как пятерку:

$$D = \langle w, \mu_D, s_D, y_D, A_D \rangle, \quad (2)$$

где  $w$  – вес текущего речевого сегмента;  $\mu_D$  – степень соответствия текущего речевого сегмента признаку  $D$ ;  $s_D$  – состояние признакового ЭФП;  $y_D$  – выход признакового ЭФП;  $A_D$  – алгоритм, вычисляющий значение выхода признакового ЭФП:

$$A_D : w, \mu_D, s_D \rightarrow y_D. \quad (3)$$

Алгоритм  $A_D$  включает три шага:

1. Накопление состояния. Состояние  $s_D$  ЭФП изменяется по формуле:

$$s_D(t) = s_D(t-1) \oplus \mu_D, \quad (4)$$

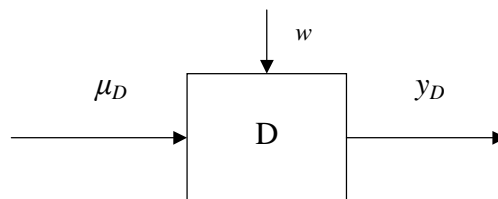


Рис. 3. Схематическое изображение ЭФП признакового уровня

где запись  $s_D(t)$  означает состояние ЭФП в момент времени  $t$ . Операция  $\oplus$  – это операция накопления. В когнитивной психологии предложена и исследована модель обработки информации человеком, известная как модель логогена Мортон [7], отражающая процессы, в результате которых человек приходит к принятию решения с учётом многих факторов: входной сенсорной информации, поступающей порциями во времени, контекста и накопленной ранее уверенности. Логоген – это абстрактная конструкция, которая определенным образом суммирует информацию, пока не будет накоплена уверенность, достаточная для реакции (ответа на вопрос или принятия решения). Такая модель обработки информации положена в основу формализованной модели логогена и хорошо зарекомендовала себя при решении ряда прикладных задач. В теории логогена предлагается следующий вид операции накопления уверенности:

$$a \oplus b = a + b - a \cdot b. \quad (5)$$

2. Расчет выхода. Выход ЭФП вычисляется по формуле:

$$y_D = s_D \cdot w. \quad (6)$$

3. «Забывание» некоторой порции информации. В качестве операции «забывания» выбрана операция вычитания:

$$s_D = \begin{cases} s_D - k_D & , \quad s_D > k_D \\ 0 & , \quad s_D \leq k_D \end{cases}. \quad (7)$$

В формуле (7)  $k_D$  – это коэффициент забывания (для ЭФП признакового уровня), вещественное число из диапазона  $[0,1]$ . Все признаковые ЭФП должны регулярно уменьшать свои состояния на некоторые небольшие значения во избежание насыщения нескольких ЭФП к моменту принятия фонемного решения. В работе применяется значение  $k_D = 0,15$ .

ЭФП фонемного уровня (рис. 4) определим как шестерку:

$$G = \langle w, \mu_G, \{v_{iG}\}_{i=1}^{N_D}, s_G, y_G, A_G \rangle, \quad (8)$$

где  $w$  – вес текущего речевого сегмента;  $\mu_G$  – степень соответствия текущего речевого сегмента фонеме  $G$ ;  $v_{iG}$  – вход с  $i$ -го ЭФП признакового уровня (зависящий от связи между выходом  $i$ -го признакового ЭФП и данным фонемным ЭФП);  $s_G$  – состояние фонемного ЭФП;  $y_G$  – выход фонемного ЭФП;  $A_G$  – алгоритм, вычисляющий значение выхода фонемного ЭФП:

$$A_G : w, \mu_G, \{v_{iG}\}_{i=1}^{N_D}, s_G \rightarrow y_G. \quad (9)$$

Связи между ЭФП являются по сути активационными:

$$u_{ij} = u(D_i, G_j). \quad (10)$$

Связь (10) отражает активацию фонемного ЭФП  $G_j$  от признакового ЭФП  $D_i$ . Каждая связь имеет свой вес. Если фонема  $G$  не обладает фонологическим признаком  $D$  (например, фонема [к] не является звонкой), то значение веса связи равно 0,0. В противном случае вес активационной связи равен вероятности  $P$  корректного выделения фонологического признака (данная статистика берется из признаковой модели FCAS).

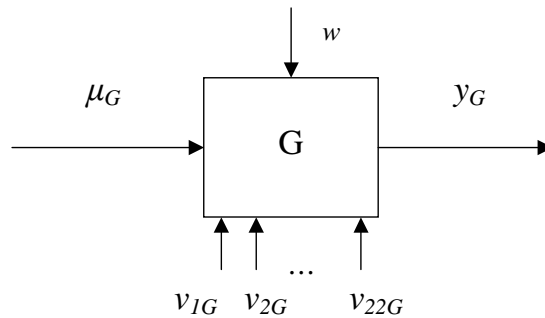


Рис. 4. Схематическое изображение ЭФП фонемного уровня

Значения признаковых входов фонемного ЭФП  $G$  рассчитываются по формуле:

$$v_{iG} = u_{iG} \cdot y_{Di} \cdot k_D, \quad (11)$$

где  $u_{iG}$  – вес активационной связи между выходом  $i$ -го признакового ЭФП и фонемным ЭФП  $G$ ;  $y_{Di}$  – выход  $i$ -го признакового ЭФП;  $k_D$  – коэффициент, характеризующий степень влияния признакового уровня вообще на принятие фонетического решения (свободный параметр модели FCAS).

Алгоритм  $A_G$  включает четыре шага:

1. Вычисление суммарной степени активации от признакового уровня.

$$v_{total} = \frac{1}{N} \sum_{i=1}^N v_{iG}, \quad (12)$$

2. Накопление состояния. Состояние  $s_G$  ЭФП изменяется по формуле:

$$s_G(t) = s_G(t-1) \oplus (\mu_G \cdot v_{total}), \quad (13)$$

где запись  $s_G(t)$  означает состояние фонемного ЭФП в момент времени  $t$ . Операция  $\oplus$  – это операция накопления (5).

3. Расчет выхода. Выход ЭФП вычисляется по формуле:

$$y_G = s_G \cdot w. \quad (14)$$

4. «Забывание» некоторой порции информации. В качестве операции «забывания», как и на признаковом уровне, используется операция вычитания:

$$s_G = \begin{cases} s_G - k_G & , \quad s_G > k_G \\ 0 & , \quad s_G \leq k_G \end{cases}. \quad (15)$$

В формуле (15)  $k_G$  – это коэффициент забывания (для ЭФП фонемного уровня), вещественное число из диапазона  $[0,1]$ . Все фонемные ЭФП должны регулярно уменьшать свои состояния на некоторые небольшие значения во избежание насыщения нескольких ЭФП к моменту принятия фонемного решения. В работе применяется значение  $k_G = 0,15$ .

Выходом модели FCAS является последовательность текстовых символов. С выходом каждого фонемного ЭФП связан «генерируемый» им текстовый символ из алфавита фонем русского языка (табл. 1). В момент принятия фонемного решения происходит выполнение двух действий:

- 1) вычисление величины «скачка состояний ЭФП»:

$$tv = \frac{1}{N_G} \sum_{i=1}^{N_G} \Delta s_{G_i}. \quad (16)$$

Если значение  $tv$  не превышает некоторого порога  $\epsilon_{tv}$ , то выход FCAS не изменяется (добавление нового символа к выходной последовательности не происходит). В противном случае, к выходной последовательности добавляется символ, генерируемый фонетическим процессором с максимальным накопленным состоянием;

2) сброс состояний всех ЭФП (уменьшение значений в  $K_{reset}$  раз):

$$s_{G_i} = s_{G_i} / K_{reset}, \quad i = 1..N_G. \quad (17)$$

Таким образом, состояния ЭФП не обнуляются при переходе к обработке следующего речевого сегмента, что отражает свойства восприятия человеком речи и взаимодействия соседних фонем [2; 6].

Моменты принятия фонемных решений дополняются также теми моментами времени, когда выполняется условие большого скачка состояний ЭФП при большом весе сегмента:

$$(w > \delta_w) \& (tv > \delta_{tv}), \quad (18)$$

где  $\delta_w$  – порог веса сегмента;  $\delta_{tv}$  – порог скачка состояний. При выполнении условия (18) к выходной последовательности символов также добавляется текстовый символ, генерируемый фонемным ЭФП с максимальным накопленным состоянием.

Особенно полезным в модели FCAS является учет фонетической статистики русского языка той или иной предметной области. Как известно, фонемы встречаются в тексте и, в особенности, следуют друг за другом, с определенными вероятностями [7; 8]. Например, следование гласного звука после звонкого шипящего гораздо более вероятно, чем следование после него глухого взрывного и т. д. В модель FCAS включен учет моментов такого рода: в момент принятия фонемного решения все признаковые и фонемные ЭФП, которые будут работать на следующем этапе (речевом сегменте), получают дополнительную активацию, в зависимости от текущего звука речи.

Эта дополнительная активация пропорциональна условным вероятностям появления конкретного признака после текущего признака:

$$s_{D_i} = s_{D_i} \oplus k_{PD} \cdot P(D_i \setminus D_{cur}), \quad (19)$$

где  $P(D_i \setminus D_{cur})$  – условная вероятность появления признака  $D_i$  после текущего признака  $D_{cur}$ ;  $k_{PD}$  – коэффициент пропорциональности.

Аналогично, дополнительная активация фонемных ЭФП пропорциональна условным вероятностям появления конкретной фонемы после текущей фонемы:

$$s_{G_i} = s_{G_i} \oplus k_{PG} \cdot P(G_i \setminus G_{cur}), \quad (20)$$

где  $P(G_i \setminus G_{cur})$  – условная вероятность появления фонемы  $G_i$  после текущей фонемы  $G_{cur}$ ;  $k_{PG}$  – коэффициент пропорциональности.

ЭФП словесного уровня определим как пятерку:

$$W = \langle \{w_t\}_{t=1}^n, \{d_t\}_{t=1}^n, \{g_t\}_{t=1}^n, y_W, A_W \rangle, \quad (21)$$

где  $\{w_t\}_{t=1}^n$  – последовательность весов речевых сегментов в интервале между двумя моментами принятия фонемного решения на словесном уровне (иначе говоря, последовательность весов предполагаемых фонем в слове);  $\{d_t\}_{t=1}^n$  – последовательность выходных значений признаков ЭФП, соответствующих фонетическим признакам, присутствующим в данном слове;  $\{g_t\}_{t=1}^n$  – последовательность выходных значений фонемных ЭФП, генерирующих текстовые символы, входящие в слово, представленное данным ЭФП словесного уровня;  $n$  – число предполагаемых фонем в слове, представ-

ленном данным ЭФП словесного уровня;  $y_W$  – выход словесного ЭФП;  $A_W$  – алгоритм, вычисляющий значение выхода словесного ЭФП:

$$A_W : \{w_t\}_{t=1}^n, \{d_t\}_{t=1}^n, \{g_t\}_{t=1}^n \rightarrow y_W. \quad (22)$$

На рисунке 5 проиллюстрирована модель ЭФП словесного уровня, генерирующего слово «ОДИН». Momentами принятия фонемного решения на уровне словесных ЭФП, в простейшем случае, можно считать речевые сегменты пауз.

Алгоритм  $A_W$  включает три шага:

1. Вычисление суммарного состояния ЭФП по признакам:

$$s_{WD} = \frac{1}{n} \sum_{t=1}^n w_t \cdot d_t. \quad (23)$$

2. Вычисление суммарного состояния ЭФП по фонемам:

$$s_{WG} = \frac{1}{n} \sum_{t=1}^n w_t \cdot g_t. \quad (24)$$

3. Расчет выхода:

$$y_W = \frac{s_{WD} + s_{WG}}{2}. \quad (25)$$

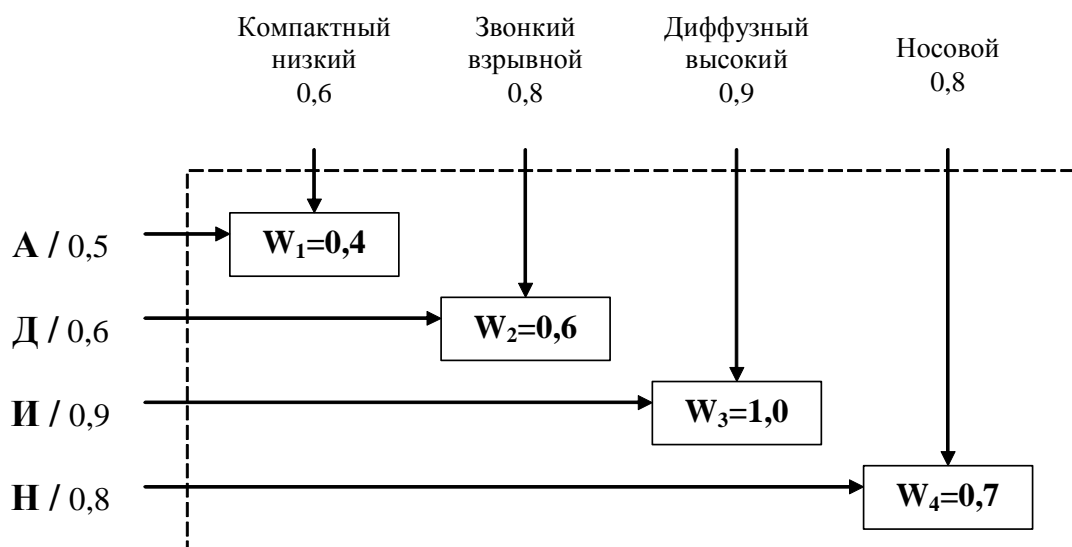


Рис. 5. Схематическое изображение словесного ЭФП

Если максимальный выход среди всех ЭФП словесного уровня превышает порог  $k_W$ , к выходной последовательности FCAS добавляется последовательность символов, представляющая слово, генерируемое данным ЭФП. В противном случае (слово отсутствует в лексиконе FCAS) к выходной последовательности добавляются текстовые символы, сгенерированные на фонемном уровне FCAS.

**Эксперименты с FCAS.** Модель была реализована программно и проверена на большом словаре (табл. 2), с использованием бюджетного микрофона и звуковой карты. В таблице 2 D – число удалений слов (число случаев, когда система не выдала никакого слова на выходе); S – число неверно распознанных слов; WER (Word Error Rate) – процент ошибок распознавания слова. В экспериментах участвовали 3 диктора (мужские голоса), произносившие все слова из таблицы по 50 раз.



Как видно из таблицы 2, показатель WER различается для разных слов, но не превышает 10 % (WER короткого слова «ДВА»); среднее значение WER составило 5 %. Это можно считать хорошим результатом, несмотря на ограниченность эксперимента (малые размер словаря и число дикторов), т. к. система не требует ни обучения на многочасовых коллекциях речевых данных, ни подстройки под диктора. Кроме того, предложенная в статье конфигурация модели – лишь одна из многих возможных (блоки на рисунке 1 являются взаимозаменяемыми).

Таблица 2

*Результаты экспериментов*

Слово	D	S	WER	Слово	D	S	WER
Ноль	1	10	7,33	Шесть	0	2	1,33
Один	0	7	4,67	Семь	0	4	2,67
Два	2	13	10	Восемь	0	6	4
Три	0	10	6,67	Девять	0	8	5,33
Четыре	0	3	2	Отмена	1	10	7,33
Пять	0	4	2,67	Вызов	0	9	6

**Выводы и предложения.** В статье рассмотрена модель постобработки РС, которая, в отличие от современных аналогов, не базируется исключительно на статистическом подходе к извлечению и представлению речевой информации. В модели FCAS отражены данные мировых исследований нескольких десятилетий, касающихся просодических, фонологических признаков речи (наряду с традиционно используемыми мел-частотными кепстральными), а фонетическая статистика используется как вспомогательный инструмент. Любой блок модели может быть заменен или удален для конкретной практической задачи. Достоинством модели является также ее относительно малая вычислительная сложность (отсутствие времязатратных операций), что позволяет применять ее в критичном к ресурсам аппаратном обеспечении.

Предложенную модель можно рассматривать как перспективный вариант для построения более точных систем, а также для интеграции с современными статистическими моделями постобработки речи. Одним из направлений усовершенствования модели FCAS является адаптивная настройка коэффициентов модели и применение робастных методов извлечения признаков.

**Список использованных источников**

1. Apple Siri [Электронный ресурс]. – Режим доступа: <http://www.apple.com/iphone/features/siri.html> / 25.05.2012.
2. Huang X. Spoken Language Processing: A guide to theory, algorithm, and system development / X. Huang, A. Acero, H. Hon. – Prentice Hall. – 2001. – 980 p.
3. Шарий Т. В. О проблеме параметризации речевого сигнала в современных системах распознавания речи / Т. В. Шарий // Вісник Донецького національного університету. Серія А. Природничі науки. – Вип. 2 (2). – 2008. – С. 536-541.
4. Каргин А. А. Анализ речевых сигналов с учетом просодических характеристик / А. А. Каргин, Т. В. Шарий // Сборник трудов X международной конференции «Интеллектуальный анализ информации ИАИ-2010». – К., 2010. – С. 339-344.
5. Каргин А. А. Применение нечеткой логики в системах фонологической классификации звуков речи / А. А. Каргин, Т. В. Шарий // Искусственный интеллект. – 2010. – № 3. – С. 210-219.
6. McClelland, J. L. The TRACE Model of Speech Perception / J. L. McClelland, J. L. Elman // Cognitive Psychology. – Vol. 18. – 1986. – P. 1-86.
7. Morton, J. Word recognition / J. Morton, J. C. Marshall // Psycholinguistics 2: Structures and Processes. – 1979. – P. 107-156.
8. Klatt, D. Speech perception: A model of acoustic-phonetic analysis and lexical access / D. Klatt // Journal of Phonetics. – Vol. 7. – 1979. – P. 279-312.