



УДК 004.822:514

### О СИНТЕЗАТОРЕ ТАДЖИКСКОЙ РЕЧИ ПО ТЕКСТУ

Худойбердиев Х.А.\*

\* *Худжандский политехнический институт Таджикского технического Университета,  
г. Худжанд, Республика Таджикистан*

**tajlingvo@gmail.com**

В настоящей работе излагается основная идея реализации синтезатора таджикской речи по тексту.

**Ключевые слова:** синтез речи; обработка текста; структура текстовой информации; синтезатор.

1. Напомним ряд понятий, используемых в дальнейшем.

*Текст* – это последовательность предложений, построенных согласно правилам данного языка и данной знаковой системы и образующая сообщение, [1].

В свою очередь, *предложение* будем рассматривать как совокупность упорядоченных элементов 7 типов, называемых *словом, числом, символом, пробелом, внутренним знаком препинания* (запятая, двоеточие, точка с запятой, тире), *внешним знаком препинания* (точка, многоточие, знак вопросительный, знак восклицательный) и, наконец, *служебным символом окончания абзаца* (в письменном тексте его нет, но он присутствует в электронном тексте как непечатаемый знак ¶).

Отметим, что смысл, который мы заключаем в названия элементов, следует понимать в общепринятых значениях. Подчеркнем также, что в конкретном предложении некоторые элементы могут отсутствовать (например, числа, символы, внутренние знаки препинания и т.д.) в то время как присутствие других – обязательно (например, внешний знак препинания).

Нам понадобятся 5 типов пауз, используемых в речи:

$P_s$  – пауза между слогами при произношении слова;

$P_w$  – пауза между словами при чтении предложения (соответствует пробелу между словами);

$P_i$  – пауза, отмечающая внутренний знак препинания;

$P_e$  – пауза, отмечающая внешний знак препинания;

$P_a$  – пауза, отмечающая конец абзаца.

2. Теперь опишем идею синтезирования речи по тексту в виде принципиальной блок-схемы, (см. рисунок 1).

Работа синтезатора происходит следующим образом. После ввода очередного предложения оно анализируется по составу своих элементов. Если очередной элемент – слово, то в блоке 1 оно разделяется на слоги с указанием ударного слога и затем осуществляется его озвучивание с использованием базы “слог-звук”.

Если очередной элемент – число, то оно в блоке 2 преобразуется в текст и затем его озвучивание происходит через блок 1.

Если очередной элемент – символ, то его озвучивание происходит в блоке 3 путем извлечения соответствующего звучания из базы “символ-звук”.

Если очередной элемент – пробел, внутренний или внешний знак препинания или же знак окончания абзаца, то для них из соответствующего блока извлекается соответствующая пауза.

Синтезатор речи, представленный в виде блок-схемы, подсказывает, что в его основу закладывается принцип конкатенации озвученных слогов.

Поскольку слог выступает в качестве основной звуковой единицы речи, то для реализации синтезатора *требуется описать многообразие все слогов соответствующего естественного языка*, что и сделано в [6].

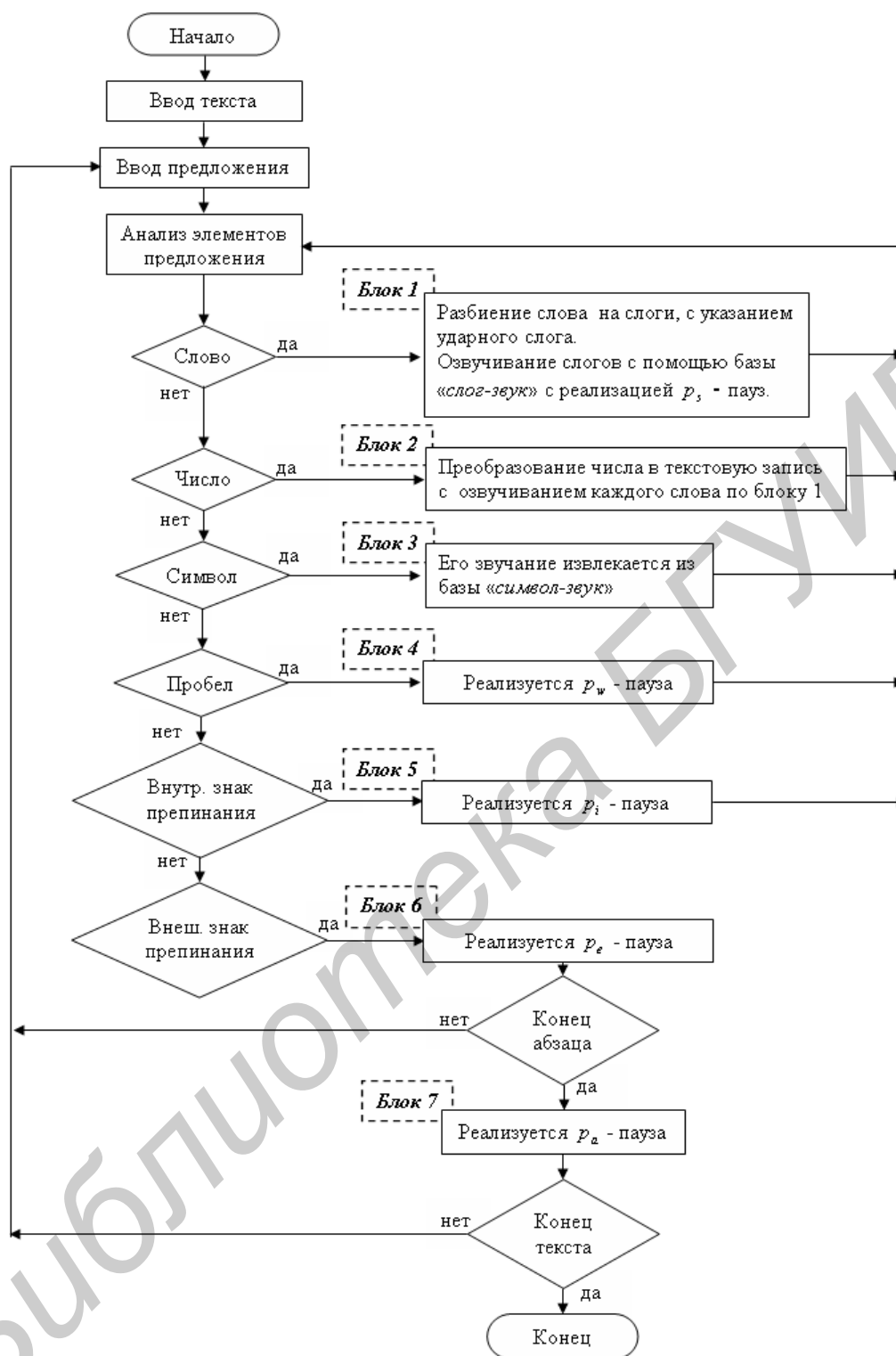


Рисунок 1 - Принципиальная блок-схема синтеза речи по тексту

Поскольку каждый слог, представленный в виде цепочки букв, нуждается в его звуковом образе, то *требуется создание базы "слог-звук"* что и сделано в [3].

Поскольку синтезатор предусматривает озвучивание чисел и символов, то для реализации синтезатора разработаны соответствующие алгоритмы и программы для трансформирования

числа в текст и создания базы "символ-звук".

И, наконец, требуется настроить длительности пауз  $p_s$ ,  $p_w$ ,  $p_i$ ,  $p_e$  и  $p_a$  таким образом, чтобы получить, по-возможности, естественную и разборчивую синтетическую речь.

3. На основе детальных проработок

концептуальной схемы был создан синтезатор Tajik Text-to-Speech, т.е. комплекс программ для синтеза таджикской речи по тексту. Настройка параметров синтезатора проводилась путем вычислительных экспериментов. Установлено удовлетворительные значения длительностей пауз:

- для границ абзацев  $P_a = 900$  мс;
- для границ предложений  $P_e = 600$  мс;
- для запятых  $P_i = 400$  мс;
- для межсловной и межслоговой пауз,

соответственно,  $p_w = 200$  мс и  $p_s = 20$  мс.

Структурная схема программного комплекса представлена на рисунке 2.

**Блок 1.** Подсистема «Интерфейс пользователя» состоит из двух компонентов - «Ввод текста» и «Озвученная речь», которые имеют одностороннюю связь, т.е. пользователь имеет возможность ввести текстовую информацию и в результате получить речевой вариант вводимого текста. Для получения результатов блок 1 связывается с блоком 2 по двум направлениям - по предоставлению информации для лингвистического анализа и получению результатов озвучивания. Блок 1 взаимодействует также и с блоком 3 непосредственно для использования необходимых данных о настройках системы (выбор мужского или женского голоса, громкость и скорость озвучивания).

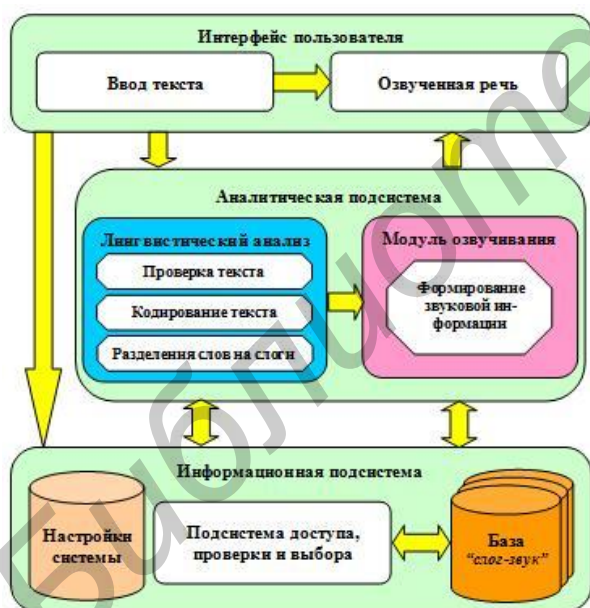


Рисунок 2 - Структурная схема Tajik Text-to-Speech

**Блок 2.** Аналитическая подсистема состоит из двух частей - «Лингвистического анализа» и «Модуля озвучивания». Первый из них состоит из подмодулей «Проверка текста», «Кодирование текста» и «Разделение слов на слоги». «Проверка текста» используется для проверки вводимой информации, которая включает в себя такие текстовые элементы, как слова, целые числа, символы и знаки препинания. Данный подмодуль

проверяет текстовые элементы, преобразует в тестовый вариант целые числа и символы и затем передает их для кодирования.

Процесс кодирования осуществляет одноименный подмодуль, который каждое слово  $W$  входного текста преобразует в упорядоченную совокупность  $W_{0,1}^*$  нулей и единиц (напомним, что цифрой 1 замещаются гласные буквы, а цифрой 0 – согласные), т.е. все слова представляются своими слоговыми структурами.

Закодированный текст передается подмодулю «Разделение слов на слоги», который работает на основе собственного алгоритма. Разделенные на слоги слова лингвистически анализируются и передаются «Модулю озвучивания».

В указанном модуле происходит формирование звуковой информации с использованием базы «слог-звук» информационной подсистемы, ударных слогов, межслоговых и межсловных пауз, а также пауз, отмечающих такие знаки препинания как запятая и точка.

Модуль озвучивания является заключительной стадией аналитической подсистемы, и звуковой вариант текстовой информации отправляется в интерфейс пользователя.

**Блок 3, «Информационная подсистема»** содержит базы данных, названные «Настройкой системы» и «Базой слог-звук». Первая из них используется для хранения временных данных настройки системы, вторая база «слог-звук» - для хранения статистических данных о звуковых файлах 3259 слогов таджикского языка. Для работы с этой базой данных используется модуль обеспечения доступа, проверки и выборки необходимых данных.

Для оценки эффективности работы синтезатора были организованы эксперименты по озвучиванию разнообразной текстовой информации (фрагменты из повестей, романов, научных статей, учебников, газет, журналов, Интернет - сайтов). Оценка полноты множества слогов, использованных для формирования синтетической речи, связывалась с процентом озвученных слов по отношению к общему количеству слов в пределах выбранных фрагментов текста. Результаты эксперимента показаны в таблице.

В приведенной таблице во 2-м столбце приводятся названия файлов, в 3-м столбце – источник, из которого взята информация и в 6-ом – доля озвученных слов в процентах по отношению к общему числу слов.

Таблица 1. Коэффициент озвучивания текста для фрагментов текстовой информации

№	Название файла	Источник	%
1	Гимн.rtf	Учебник	100,00
2	С_Айни1.rtf	«Ятим», С.Айни	99,47
3	Вестник32.rtf	Газета «Вароруд»	98,63
4	Сино.rtf	Реферат	98,88
5	Ду дилдода.rtf	Интернет сайт www.gazeta.tj	98,10
6	Президент.rtf	Газета «Чархи гардун»	96,69
Итого			98,45

Полученные результаты показали вполне удовлетворительное качество работы комплекса программ Tajik Text-to-Speech по озвучиванию таджикского текста.

Таким образом, комплекс программ Tajik Text-to-Speech, хотя и не решает полностью поставленной проблемы, *все же в настоящее время оказывается первым программным продуктом, хорошо осуществляющим компьютерное озвучивание таджикских текстов.* На данном уровне разработки комплекс уже сейчас может быть использован людьми с ослабленным зрением.

Другие эксперименты проводились на научных семинарах Института математики АН РТ. Его участники по собственному усмотрению вводили в компьютер таджикские тексты и затем оценивали естественность и разборчивость звучания синтетической речи. Общее мнение семинара - *компьютерный синтезатор, построенный по принципу конкатенации 3259 таджикских слогов, вполне успешно выполняет функцию озвучивания таджикских текстов.*

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- Советский энциклопедический словарь. -М.: Советская энциклопедия, 1980. 1600 с.
- Л.В.Златоустова, С.В.Кодзасов, О.Ф.Кривнова, И.Г.Фролова. Алгоритмы преобразования русских орфографических текстов в фонетическую запись. -М.: МГУ, 1970.
- Лобанов Б.М. Цирульник Л.И. Жадинец Д.В. Сизонов О.Г. Алгоритмы синтеза просодических характеристик речи по тексту в системе «Мультифон». Объединённый институт проблем информатики НАН Беларуси, Минск. 2007.
- Лобанов Б.М., Карневская Е.Б., Левковская Т.В. Синтезатор речи по тексту как компьютерное средство «клонирования» персонального голоса // Тр. Международной конференции Диалог-2001 / -М.:, 2001. С. 265-272.
- Лобанов Б.М., Цирульник М.И. Компьютерный синтез и клонирование речи, Минск, Белорусская наука, 2008, - 316 с.
- Тим Кинтцель. Программирование звука на ПК. Пер. с англ. -М.: ДМК Пресс, 2005, 432 с.
- Баландюк Ю. В. Просодическое устройство слова в корейском языке (на материале двуслога). Автореферат диссертации на соискание ученой степени кандидата филологических наук, Институт стран Азии и Африки при МГУ им. М.В.Ломоносова, -М.: – 2003, 23 с.
- Худойбердиев Х.А., О многообразии слогов таджикского языка // Известия АН РТ, № 2 (127), 2007. - с. 31-34.
- Усманов З.Д., Худойбердиев Х.А. Алгоритм безударного озвучивания таджикского текста // ДАН РТ, Т.50, № 4, 2007. - с. 302-305.
- Усманов З.Д., Худойбердиев Х.А. Компьютерное озвучивание таджикского текста // Патент (интеллектуальный продукт) зарегистрирован 041ТJ 04.09.2007 НПИ центром.

### ON A TAJIK SPEECH SYNTHESIZER UNDER A TEXT

Kh. A. Khudoyberdiev\*

Khujand Polytechnic Institute  
of Tajik Technical University  
tajlingvo@gmail.com

In the article a block-scheme of the program complex for realization of a synthesis of Tajik speech under a text is described. Values of parameters providing a satisfactory work of a synthesizer are specified.