

СИНТЕЗ РЕЧЕПОДОБНЫХ СИГНАЛОВ С ИСПОЛЬЗОВАНИЕМ АЛЛОФОНОВ

^{1,2} Белорусский государственный университет информатики и радиоэлектроники
Беларусь, 220660, Минск, ул. П.Бровки,6
Тел.(017) 239-8939, факс (017) 231-0914
e-mail: nil53@bsuir.edu.by

³Объединенный институт проблем информатики НАН Беларуси
Беларусь 220012, Минск, ул. Сурганова, 6
Тел. (017)284-2773, факс (017)231-8403
e-mail: lobanov@newman.bas-net.by

Предложен и проанализирован алгоритм формирования акустических речеподобных сигналов (РПС), основанный на использовании статистики элементов речи русского языка (фоноабзацев, синтагм, слов, слогов и фонем) и предварительно созданной базы аллофонов. Приведены структурная схема указанного алгоритма и результаты испытания реализующей его компьютерной модели. Рассмотрены вопросы реализации автономного синтезатора РПС на базе микроконтроллера AT89S8252 фирмы Atmel для 8-битных аллофонов и AT91M40800 - для 16-битных. Достоинством предлагаемого устройства является возможность оперативной замены аллофонной базы и используемых статистических характеристик элементов речи с целью подстройки РПС под речь конкретного человека.

Речеподобные сигналы (РПС) находят применение в телефонетрии, психологических исследованиях, при защите информации [1,2].

При любом способе синтеза РПС отправным пунктом является использование характеристик естественной речи на выбранном языке. В [2] обсуждались вопросы придания РПС повышенного сходства с естественными речевыми сигналами путем использования известной технологии Text-to-Speech (TTS) [3]. При этом за пределами рассмотрения осталась задача озвучивания формируемых текстов. Для решения этой задачи оказалось весьма плодотворным использование следующих соображений.

Устная речь - это звуковой поток, представленный ее элементами в иерархии: Фоноабзац - Фраза - Синтагма - Слово - Слог - Фонема (аллофон). Каждый из элементов речи высшего ранга составляется из элементов более низкого ранга и подчиняется определенному закону распределения вероятностей [4,5]. Некоторое представление о статистических характеристиках элементов речи русского языка можно получить из данных, приведенных в табл.1.

Таблица 1. Статистические характеристики различных элементов речи

Число элементов речи		1	2	3	4	5	6	7
Частота появления [%]	Фраз в фоноабзаце	5	10	15	30	25	10	5
	Синтагм во фразе	5	10	30	25	15	10	5
	Слов в синтагме	5	10	30	25	15	10	5
	Слогов в слове	2	13	30	28	16	7	4

После синтагмы, фразы и фоноабзаца обязательно должна присутствовать пауза, длительность которой выбирается следующим образом.

Таблица 2 — Формирование пауз

	Синтагма	Фраза	Абзац
Длительность	0.3 сек.	0.8 сек.	1.5 сек.

Номер ударного слога в слове зависит от длины последнего и подчиняется следующему закону.

Таблица 3 — Правило расположения ударного слога в слове

Число слогов в слове	1	2	3	4	5	6	7
№ ударного слога	1	2	2	3	3	4	5

Подобно синтезу речи по технологии TTS синтез РПС целесообразно осуществлять [2] путем преобразования формируемого с учетом упомянутых статистических закономерностей орфографического текста в акустические колебания звукового диапазона частот.

При различных требованиях к качеству имитации естественной речи число аллофонов в русском языке может быть выбрано от 500 до 1500.

Рассматриваемый ниже метод формирования РПС основан на использовании базы аллофонов, подготовленной в научной лаборатории Лобанова Б.М., и состоит в следующем (см.Рис.1).

- Из фраз формируется фоноабзац, при обязательном учете частоты встречаемости фоноабзацев, состоящих из определенного числа фраз.

- Фразы из синтагм составляются по тому же принципу, что и в предыдущем пункте.

- Формируется синтагма из слов.

- Слово составляется из слогов, определяется положение ударного и первого предупредительного слогов. Для создания очередного слога используется модель, позволяющая осуществлять учет предыдущего слога.

- Слова орфографического текста преобразуются в слова фонемного текста.

- Последовательности букв фонемного текста преобразуются в последовательность аллофонов и озвучиваются. При озвучивании учитываются паузы, расставляемые в конце синтагмы, фразы и фоноабзаца, соответствующей длительности, а так же просодическое оформление сформированных синтагм и фраз.

Для формирования слогов с учетом предыдущего выполняются следующие действия.

Статистически представительный текст разбивается на слоги в соответствии с правилами:

- в слове всегда одна гласная буква;

- все согласные до первой гласной в слове принадлежат первому слогу;

- все согласные после последней гласной принадлежат последнему слогу;

- если между двумя соседними гласными находится более одной согласной, то самая первая согласная относится к начальному слогу, а остальные согласные к следующему;

- если между двумя соседними гласными находится одна согласная, то она относится к следующему слогу.

При разбиении текста на слоги учитывается число одинаковых слогов, встречающихся в тексте. После того, как процедура разбиения текста на слоги закончена, формируется словарь наиболее часто встречающихся слогов, покрывающих 90% текста.

По сформированному в соответствии с приведенными правилами словарю слогов текст заново разбивается на слоги. Но теперь задача состоит не в выделении слогов, а в получении статистических данных относительно условной встречаемости слогов словаря (с учетом

предыдущего слога или отсутствием такового). Данные заносятся в таблицу, вид которой поясняется Рис.2.

Статистические данные о слогах

	Столбец первого слога	Столбец второго слога	...	Столбец последнего слога
Нет предыдущего слога	***	***		***
Предыдущий слог — первый слог словаря	***	***		***
Предыдущий слог — второй слог словаря	***	***		***
...				
Предыдущий слог — последний слог словаря	***	***		***

Рис. 2.

Для получения условных вероятностей появления слогов в тексте сформированная таблица нормируется построчно.

Алгоритм формирования последовательности слогов состоит в генерировании случайного числа, равномерно распределенного от 0 до 1, и поиска слога, указываемого этим числом. Распределение, используемое при поиске слога, определяется предыдущим сформированным слогом, указывающим на соответствующую ему строку таблицы статистических данных.

Сформированные слова орфографического текста преобразуются в слова фонемного текста с использованием правила перевода “Буква-фонема”. Слова, составленные из фонем, преобразуются в последовательность аллофонов с помощью алгоритма “Фонема-Аллофон”. Просодическое (интонационное) оформление сформированных синтагм и фраз осуществляется в соответствии с рядом правил. Более подробно приведенные процедуры описаны в [3].

При решении задач эффективной защиты речи для составления таблицы вероятностей следует использовать тексты статей, речей, докладов защищаемого лица. Наряду с обработкой текстов, настоятельно рекомендуется применение базы аллофонов этого лица.

В результате проведенной работы создана компьютерная модель, реализующая рассмотренный алгоритм синтеза РПС. Предусмотрена демонстрация работы модели с озвучиванием образцов формируемых сигналов.

Ведется разработка автономного синтезатора РПС на базе микроконтроллера AT89S8252 фирмы Atmel для 8-битных аллофонов и AT91M40800 для 16-битных аллофонов. Его достоинством является возможность оперативной замены аллофонной базы и статистических данных для защиты речи конкретного человека.

ЛИТЕРАТУРА

1. Recommendation P.501 ITU-T (08/96).
2. Воробьев В.И., Давыдов А.Г. Синтез речеподобных сигналов // Труды XI Сессии РАО, т. 3, Акустика речи, медицинская и биологическая акустика. – М.: ГЕОС, 2001. – С. 83-87.
3. Лобанов Б.М. Синтез речи по тексту. Сб. Научных трудов 4-й Международной школы-семинара по искусственному интеллекту. -Мн.: БГУИР, 2000. – С. 57-76.
4. Обратный словарь русского языка. Около 125 000 слов. - М.: Сов. Энциклопедия, 1974. -944с.
5. Русская грамматика. Том 1. – АН СССР, М.: Наука, 1982. –783с

5.

Структурная схема алгоритма формирования речеподобных сигналов

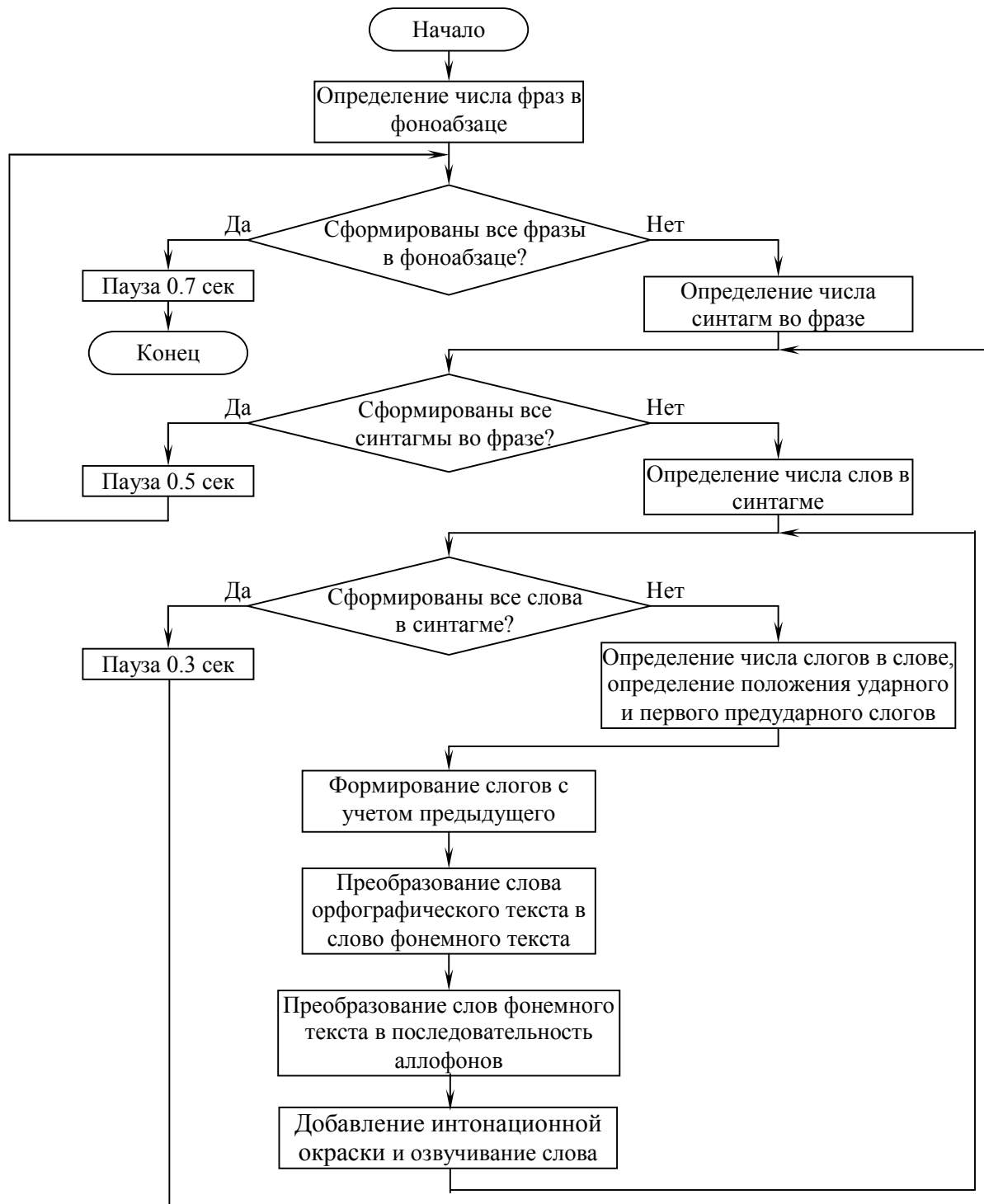


Рис. 1.