АЛГОРИТМ ПЕРСОНАЛИЗИРОВАННОГО СИНТАГМАТИЧЕСКОГО ЧЛЕНЕНИЯ ТЕКСТА ДЛЯ TTS-СИСТЕМ КЛОНИРОВАНИЯ РЕЧИ

Б.М. Лобанов, Л.И. Цирульник

Объединённый институт проблем информатики НАН РБ
Беларусь 220012, Минск, ул. Сурганова, 6
Тел. (017)284-2773, факс (017)231-8403
Эл.почта: lobanov@newman.bas-net.by

Введение.

В работах [1,2] было предложено рассматривать синтезатор речи по тексту как компьютерное средство клонирования речи при условии максимально полного воспроизведения индивидуальных акустических, фонетических и просодических характеристик голоса и речи диктора. В настоящей работе в рамках продолжающихся исследований по клонированию описываются результаты экспериментального изучения персональных особенностей синтагматического членения речи телеведущего Ю.Сенкевича в сравнении с особенностями синтагматического членения речи одного из профессиональных дикторов (диктор М). Для исследования использовались фонограммы ТВ-передач “Клуб кинопутешественников”, а также фонограммы 2-х рассказов в исполнении диктора М. Объём использованных звуковых файлов для каждого диктора составлял порядка 25 МБ, что соответствует примерно 1000 словам орфографического текста стенограмм. Основная цель исследования заключалась в создании алгоритмических основ клонирования персональных просодических характеристик речи и, в частности, алгоритма персонализированного синтагматического членения текста для синтеза речи.

Предварительная обработка фонограмм и текста.

Прежде всего составляется и записывается дословный текст фонограммы. Затем производится коррекция текста и фонограммы. Из них убираются ошибочно произнесённые слова и звуки, участки с разного рода помехами (шум, музыка, слова с малым уровнем звука и др.). При необходимости проводится корректировка акустических характеристик фонограммы (выравнивания звуковых уровней, корректировка АЧХ записи).

Разбиение фонограмм и текста на фонетические синтагмы.

Путём прослушивания на фонограмме и в тексте отмечаются границы синтагм (под синтагмой понимается самостоятельная в интонационном смысле часть фразы или вся фраза). Решение о наличии конца синтагмы принимается на основе ряда признаков, таких как: присутствие дыхательной паузы, комплексная реализация одного из возможных интонационных типов синтагмы, наличие определённой динамической и ритмической структуры. При членении фонограммы на синтагмы во внимание принимается также присутствие знаков препинания в соответствующем ей тексте и некоторых других формальных признаков текста.

Просодическая маркировка синтагм.

Каждая синтагма в фонограмме прослушивается опытным аудитором и маркируется следующим образом: для каждого слова синтагмы указывается место ударения и его тип: сильное (+), слабое (-) или ударение отсутствует. Затем слова со

слабым ударением объединяются в единую акцентную группу (АГ) с одним из слов с сильным ударением и в тексте указываются границы АГ.

Статистические характеристики синтагматического членения.

Статистическая обработка результатов экспериментальных исследований фонограмм речи проводилась с целью получения количественных характеристик, полезных с точки зрения персонализации синтезированной речи. К таким характеристикам относятся сравнительные частоты встречаемости пар синтагм с различным количеством АГ, а также пар слов с различными грамматическими категориями на стыке синтагм, не отмеченных знаком препинания.

Для получения последней характеристики для каждой пары слов текста, не разделенной знаком препинания, определяются соответствующие части речи и подсчитывается общее количество пар различных частей речи в тексте. Затем определяется количество различных пар частей речи, разделенных границей синтагмы. По результатам обработки составляется двумерный массив частот появления конца синтагмы между указанными частями речи. В процессе анализа результатов было принято решение выделять только 7 наиболее "важных" частей речи: глагол, существительное, местоимение, наречие, прилагательное, союз, предлог, а все остальные объединить как "другая часть речи".

Результаты статистического анализа.

В таблице 1 приведены результаты анализа статистической вероятности встречаемости пар синтагм с различным количеством АГ, начиная с пар (0-1), (0-2), … (0-4), т.е. для синтагм в начале фразы, а затем: (1-1)…(1-4), (2-1)…(2-4), …, …, (4-1)…(4-4). Синтагмы с числом АГ >4 (редко встречающиеся) не рассматривались.

Таблица 1. Вероятность встречаемости пар синтагм с различным количеством АГ

	Диктор Ю.С. / Диктор М.			
	1	2	3	4
0	**0,45 / 0,15**	**0,30 / 0,50**	**0,20 / 0,20**	**0,05 / 0,15**
1	**0,50 / 0,15**	**0,20 / 0,60**	**0,20 / 0,15**	**0,10 / 0,10**
2	**0,50 / 0,15**	**0,35 / 0,50**	**0,10 / 0,20**	**0,05 / 0,15**
3	**0,70 / 0,20**	**0,15 / 0,35**	**0,15 / 0,30**	**0,00 / 0,15**
4	**1,00 / 0,15**	**0,00 / 0,40**	**0,00 / 0,30**	**0,00 / 0,15**

Приведенные статистические характеристики особенностей синтагматического членения устной речи для 2-х дикторов показывают, что интересующая нас речь диктора Ю. Сенкевича обладает ярко выраженными отличиями от речи диктора М. Из таблицы 1 видно, что у него наблюдается значительное преобладание количества одноакцентных синтагм и сравнительно равномерное распределение 2-х и 3-х акцентных при небольшом числе 4-х акцентных.

В таблице 2 приведены результаты анализа частоты встречаемости пар слов с различными грамматическими категориями на стыке синтагм. Частота появления конца синтагмы указывается в процентах, причем 0 означает, что такая пара либо вообще не встретилась в анализируемых текстах (в пределах 10-ти процентной статистической

достоверности результатов анализа), либо ни разу не была разделена границей синтагмы в речи данного диктора.

Таблица 2. Частота встречаемости (в %) границ синтагм между различными частями речи

	Диктор Ю.С. / Диктор М.						
	Глаг	Нареч	Прилг	Сущ	Мест	Союз	Пред
Глагол	**25 / 15**	**55 / 0**	**75 / 0**	**55 / 5**	**35 / 40**	**0 / 0**	**40 / 0**
Наречие	**35 / 30**	**0 / 0**	**20 / 35**	**0 / 10**	**15 / 0**	**0 / 0**	**45 / 25**
Прилагательное	**0 / 0**	**0 / 0**	**75 / 50**	**40 / 0**	**0 / 0**	**0 / 0**	**0 / 0**
Существительное	**70 / 30**	**80 / 70**	**80 / 50**	**55 / 0**	**20/ 0**	**90 / 0**	**80 / 30**
Местоимение	**35 / 5**	**0 / 0**	**55 / 0**	**15 /0**	**0 / 0**	**0 / 0**	**45 / 25**
Союз	**45 / 0**	**10 / 20**	**75 / 0**	**55 / 0**	**45 / 0**	**0 / 0**	**50 / 0**
Предлог	**0 / 0**	**0 / 0**	**30 / 0**	**30 / 0**	**20 / 0**	**0 / 0**	**0 / 0**

Приведенные статистические характеристики особенностей встречаемости пар слов с различными грамматическими категориями на стыке синтагм для 2-х дикторов показывают, что интересующая нас речь диктора Ю. Сенкевича обладает определённо выраженными отличиями от речи диктора М. Из таблицы 2 видно, что у него есть характерная особенность наличия границы синтагмы после предлогов и союзов, а также наблюдается значительное преобладание количества границ синтагм перед существительными, после глаголов и после местоимений.

***Алгоритм разбиения текста на синтагмы*.**

Входные данные:

- *Грамматический словарь русского языка*.

Словарь включает максимально возможное количество словоформ, каждая из которых сопровождается пометами названия части речи.

- *Произвольный орфографический текст.*

Предварительно текст должен быть приведен к каноническому орфографическому виду, т.е. в нём должны быть преобразованы к текстовому виду цифры, формулы, аббревиатуры и сокращения.

- *Двумерный массив P(m,n) частот встречаемости пар синтагм с различным количеством АГ*

Индекс m принимает значения 0,1,...4, индекс n - значения 1,2...4. Значения элементов массива P(m,n) находятся в диапазоне [0 - 1].

- *Двумерный массив G(p,k) частот появления конца синтагмы после p-й и перед k-й частью речи.*

Индексы i и j могут принимать следующие значения: глагол, местоимение, наречие, предлог, прилагательное, союз, существительное, другая часть речи. Значения элементов массива G[i,j] находятся в диапазоне [0 - 100].

Выходные данные:

Текст, разбитый на синтагмы, без указания длительности паузы после каждой синтагмы.

Структура алгоритма:

1. *Входной текст разбивается на слова.*

2. *Для каждого слова входного текста определяется (из грамматического словаря) часть речи. Формируется последовательность $\{S_i\}$ частей речи, соответствующих словам текста.*

3. Для каждой пары S_i, S_{i+1}, не разделенной знаком препинания, по данным из таблицы 2 формируется последовательность $\{G_i (p,k)\}$.

4. Из последовательности $\{G_i (p,k)\}$. *начиная с* $G_1(p,k)$ *выбирается первая подпоследовательность* $\{G_i (p,k)/1\}$, *для которой выполняется одно из двух условий:*

- количество знаменательных частей речи в подпоследовательности достигло 5 и при этом не встретилось знака препинания;

- подпоследовательность заканчивается знаком препинания при i<5.

5. Элементы полученной подпоследовательности $\{G_i(p,k)/1\}$ *умножаются на соответствующие данные из таблицы 1 и формируется подпоследовательность* $\{G_i(p,k)*P_i(m,n)/1\}$ *и осуществляется поиск ее r-го максимального элемента, т.е.* $r=ArgMax_{/i}\{G1_i(p,k)*P_i(m,n)/1\}$. *После r-го слова исходного текста отмечается рекомендуемое положение конца синтагмы.*

6. Если значения всех элементов подпоследовательности равны 0, то добавляется ещё одно слово и повторяется процедура по п.5.

7. После того, как выбрана первая подпоследовательность $\{G_i (p,k)*P_i(m,n)/1\}$, *начиная с (r+1)-го элемента, выделяется следующая подпоследовательность* $\{G_i (p,k)/2\}$, *для которой выполняется хотя бы одно из указанных выше условий. Затем повторяется шаг 5 алгоритма.*

8. Действия п.5-7 алгоритма повторяются до конца входного текста.

Заключение

Статистические данные, методика получения которых описана в данной статье, будут использоваться для дальнейших исследований по клонированию просодических характеристик речи. Основные результаты основаны на сравнительно небольшом объёме экспериментальных данных. В дальнейшей работе предполагается использовать появившуюся недавно базу данных Национального фонда русского языка [3].

Графическая иллюстрация результатов статистической обработки фонограмм и текстов, примеры автоматического разбиения текста на синтагмы и образцы клонов синтезированной речи будет приведены во время доклада.

Литература

1. Лобанов Б.М. и др. *Синтезатор персонализированной речи по тексту "ЛобаноФон-2000"* Тр. Международной конференции, посвящённой 100-летию российской экспериментальной фонетики. Ст.-Петербург, 2001, сс.101-104.
2. Lobanov B. and Karnevskaya H**.** *TTS-Synthesizer as a Computer Means for Personal Voice Cloning (On the example of Russian)***.** In the Book: Phonetics and its Applications. Stuttgart: Steiner. 2002, pp. 445-452.
3. http://ruscorpora.ru