

Персональные особенности синтагматического членения речи телеведущего Ю. Сенкевича

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
lobanov@newman.bas-net.by

В докладе описываются результаты экспериментального изучения персональных особенностей синтагматического членения речи телеведущего Ю.Сенкевича, проведенные в рамках продолжающихся исследований по клонированию голоса и дикции человека. Исследования проведены на базе фонограмм ТВ-передач «Клуба кинопутешественников». Используемые фонограммы были предварительно проаудированы опытным фонетистом, расставлены маркеры сильных и слабых ударений, указаны границы синтагм и соответствующие им интонационные типы.

Приводятся результаты сравнительного статистического анализа просодических структур речи Ю. Сенкевича и двух других профессиональных дикторов телевидения. Получены сравнительные статистические оценки количества акцентных единиц в синтагмах, длительности межсинтагменных пауз, вероятности сочетаемости синтагм и др. Результаты исследования использованы в системе клонирования персонального голоса при синтезе речи по тексту.

Введение

На конференции Диалог-2001 [1] было предложено рассматривать синтезатор речи по тексту как компьютерное средство “клонирования” персонального голоса при условии максимально полной имитации акустических, фонетических и просодических средств индивидуальности голоса и речи диктора. На конференции Диалог-2003 [2] обсуждались вопросы автоматизации процедуры клонирования персонального голоса и дикции. В настоящей работе в рамках продолжающихся исследований по клонированию голоса и дикции описываются результаты экспериментального изучения персональных особенностей синтагматического членения речи телеведущего Ю.Сенкевича, проведенного на базе фонограмм ТВ-передач «Клуба кинопутешественников». В докладе приводятся результаты сравнительного статистического анализа просодических структур речи Ю. Сенкевича и двух других профессиональных дикторов телевидения. Объём использованных звуковых файлов для каждого диктора составлял порядка 25 МБ, что соответствует примерно 1000 словам орфографического текста стенограммы. Основная цель исследования заключалась в создании алгоритмических основ клонирования персональных просодических характеристик речи и, в частности, алгоритмов членения на синтагмы. Другая цель заключалась в сравнительной оценке существенности различий изучаемых просодических характеристик речи для нескольких дикторов.

1. Методика эксперимента

Предварительная обработка фонограмм и текста

Прежде всего расшифровывается и записывается дословный текст фонограммы. По результатам расшифровки производится коррекция текста и фонограммы. Из них убираются ошибочно произнесённые слова и звуки, участки с разного рода помехами (вдохи, шум, музыка, слова с малым уровнем звука и др.). Проводится при необходимости корректировка акустических характеристик фонограммы (выравнивания звуковых уровней, корректировка АЧХ звуков)

Разбиение фонограмм и текста на фонетические синтагмы

Путём последовательного прослушивания фонограммы 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 структуры (контура силы звука) и определённой ритмической структуры (контура длительности звуков). При членении фонограммы на синтагмы во внимание принимается также присутствие знаков препинания в соответствующем ей тексте, а также некоторых других формальных признаков текста.

Просодическая маркировка синтагм.

После членения фонограммы на синтагмы каждая из них последовательно прослушивается и маркируется следующим образом. Каждое слово синтагмы маркируется ударением, т.е. указывается место ударения и его тип: сильное (+), слабое (-) или ударение отсутствует. Слова, в которых ударение отсутствует, объединяются в тексте с одним из соседних слов в единое фонетическое слово. Затем слова со слабым ударением объединяются в единую акцентную группу (АГ) с одним из слов с сильным ударением и в тексте указываются границы АГ. После этого на основе фонетического аудирования каждой синтагме присваивается определённый интонационный подтип с указанием количества АГ в синтагме (например, 2-С, что означает: 2-х акцентная синтагма с интонацией незавершённости), а также длительности синтагматической паузы. Для обозначения интонационного типа используются знаки: С (comma) - незавершённость, Р (point) - завершённость, Q (question) - вопрос, Е (exclamation) – восклицание.

В таблице 1 приведен пример разметки текста на синтагмы и маркировки словесных ударений для части файла 5 фонограммы Ю. Сенкевича после её аудитивного анализа. Знаком “ / “ отмечены границы АГ.

Таблица 1. Пример синтагматической разметки текста

Номер синтагмы	Текст, маркированный ударениями	Интонационный тип	Длительность паузы, мс
1.	/Здра+вствуйте/,	1-С	200
2.	/дороги+е/ /люби+тели/ /путеше+ствий/.	3-Р	1100
3.	/Сего+дня/	1-С	75
4.	/мы+/ /отпра+вимся/ /сва+ми/	3-С	50
5.	/вФинля+ндию/,	1-С	750
6.	/и+/,	1-С	150
7.	/мне+ /ка+жется/,	2-С	0
8.	/что-э+то/ /путеше+ствие/ /бу+дет/ /для+вас+ /	4-С	400
9.	/интере+сным/,	1-С	0
10.	/поско+льку/	1-С	400
11.	/путеше+ствие/ /э+то/ /нето+лько/ /впростра+нстве/,	4-С	50
12.	/но+ /	1-С	120

13.	/и- вове+мени/.	1-Р	650
14.	/Мы+ /расска+жем/ /в+ам/	1-С	120
15.	/о+ /стари+нных/	2-С	900
16.	/фи+нских/ /крепостя+х/.	2-Р	400
17.	/Крепостя+х/ /за+мках/.	2-Р	2400
18.	/Впроста+нстве/,	1-С	0
19.	/потому+ что-/ /они+ /располо+жены/ /вра+зных/ /частя+х/	5-С	460
20.	/Финля+ндии/,	1-С	170
21.	/авове+мени/,	1-С	0
22.	/потому+ что-/	1-С	330
23.	/мы+ /коснее+мся/ /исто+рии/,	3-С	150
24.	/есте+ственно/.	1-Р	850
25.	/И- де+ло/ /всё+ /вто+м/,	3-С	0
26.	/что- э+ти/ /кре+пости/	2-С	300
27.	/жи+вы/ /до+сих/ /по+р/,	3-С	360
28.	/но-/ /живу+т/	2-С	100
29.	/они+ /, /коне+чно/,	2-С	380
30.	/соверше+нно/ /друго+й/	2-С	90
31.	/жи+знью/ /не+жели/	2-С	400
32.	/вте+ /времена+ /,	2-С	0
33.	/когда+ /	1-С	90
34.	/они+ / /бы+ли/	2-С	150
35.	/постро+ены/.	1-Р	900

2. Статистические характеристики синтагматического членения

Статистическая обработка результатов экспериментальных исследований фонограмм речи проводилась с целью получения некоторых количественных характеристик, полезных с точки зрения персонализации синтезируемой речи. К таким характеристикам относятся: сравнительные частоты встречаемости синтагм с различным количеством АГ, частоты встречаемости пауз различной длительности, частоты встречаемости пар синтагм с различным количеством АГ. Основные результаты статистической обработки материалов, полученные в результате экспериментальных исследований, приведены на рисунках 1 – 3.

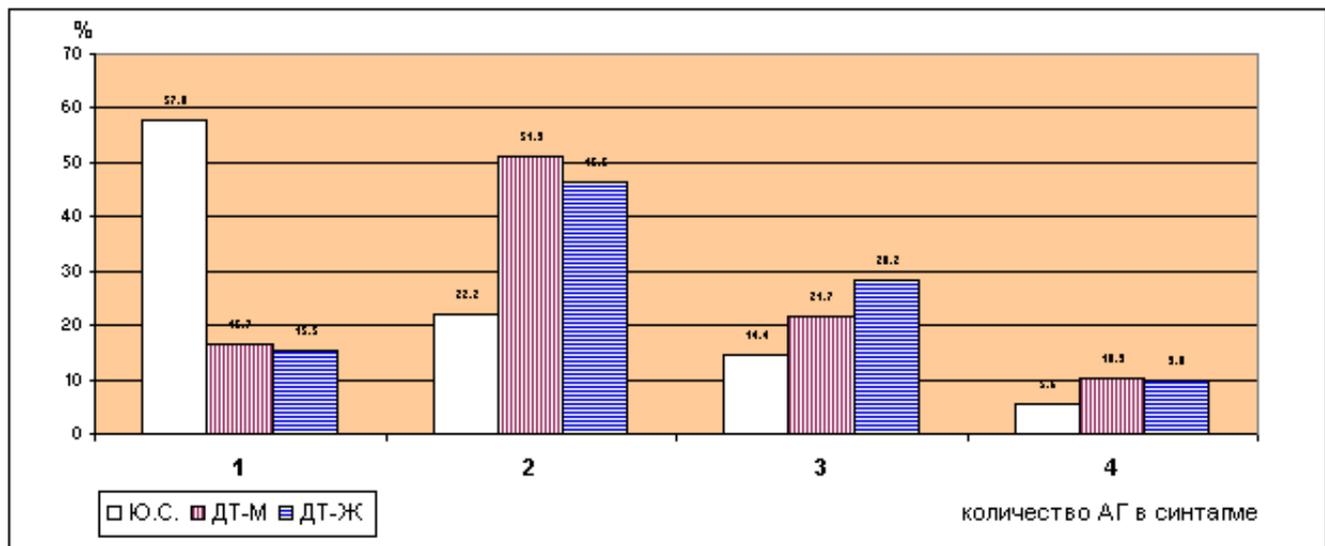


Рис.

1. Сравнительная частота встречаемости синтагм с различным количеством АГ у диктора Ю.Сенкевича и двух других дикторов телевидения (ДТ-М и ДТ-Ж).

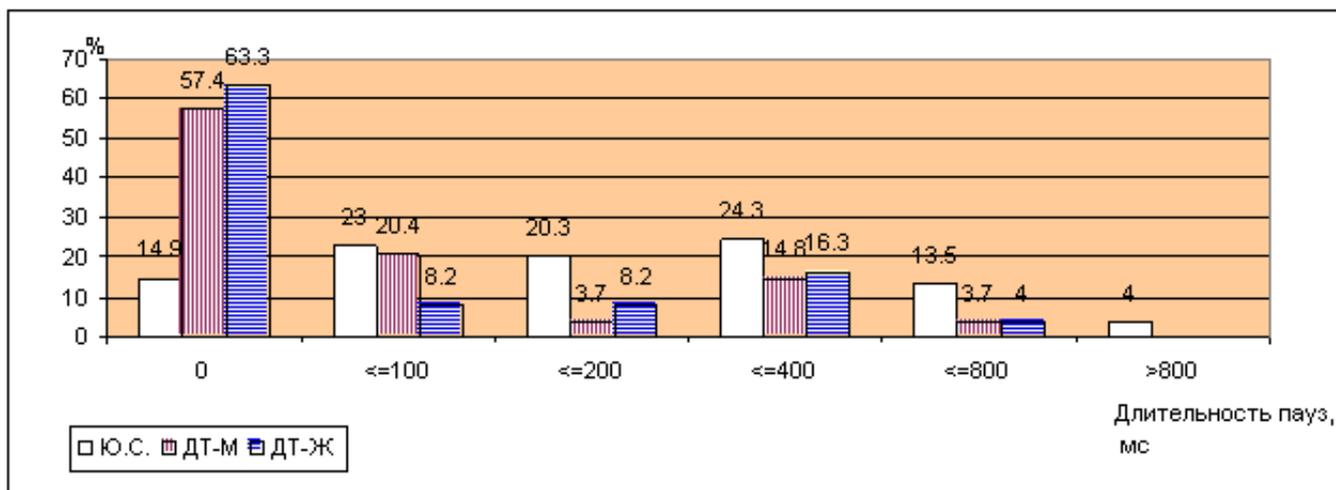
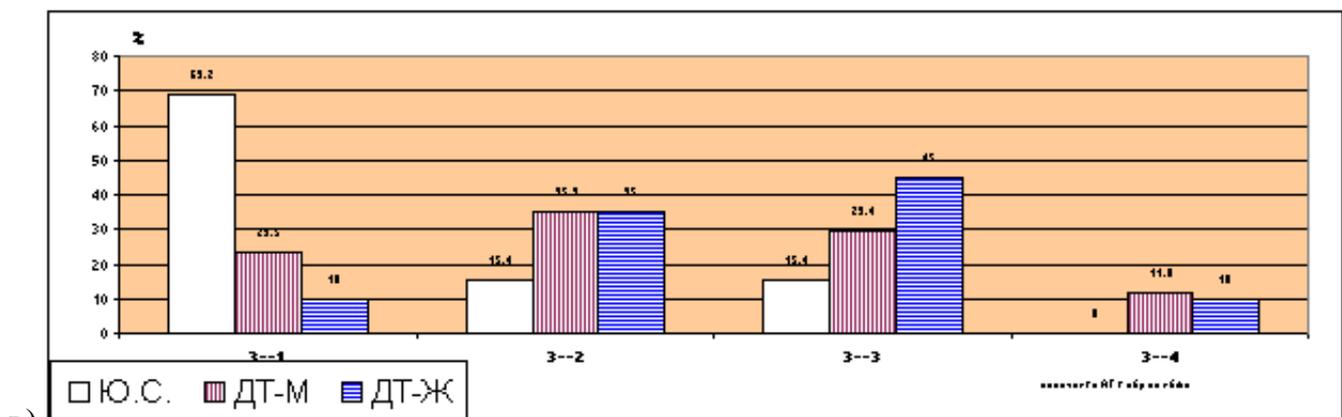
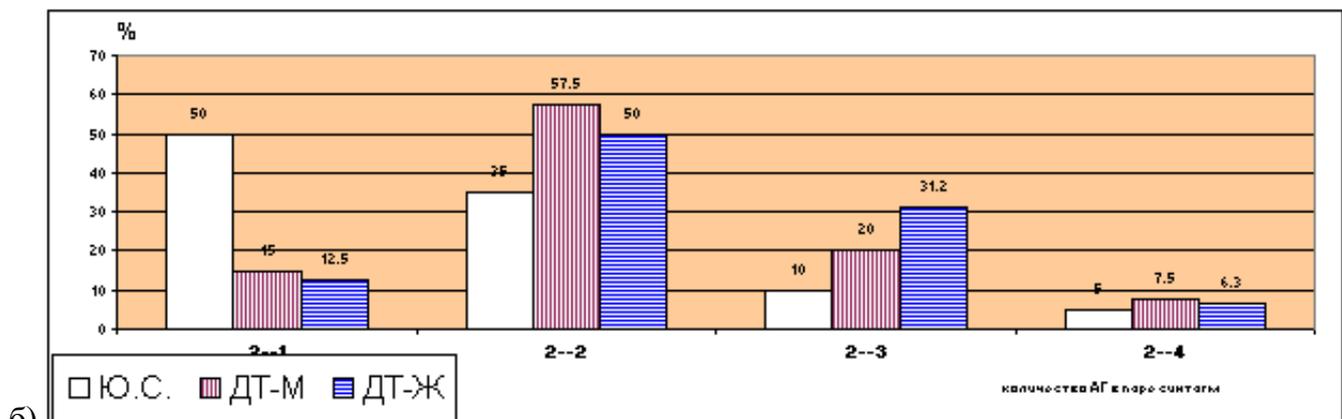
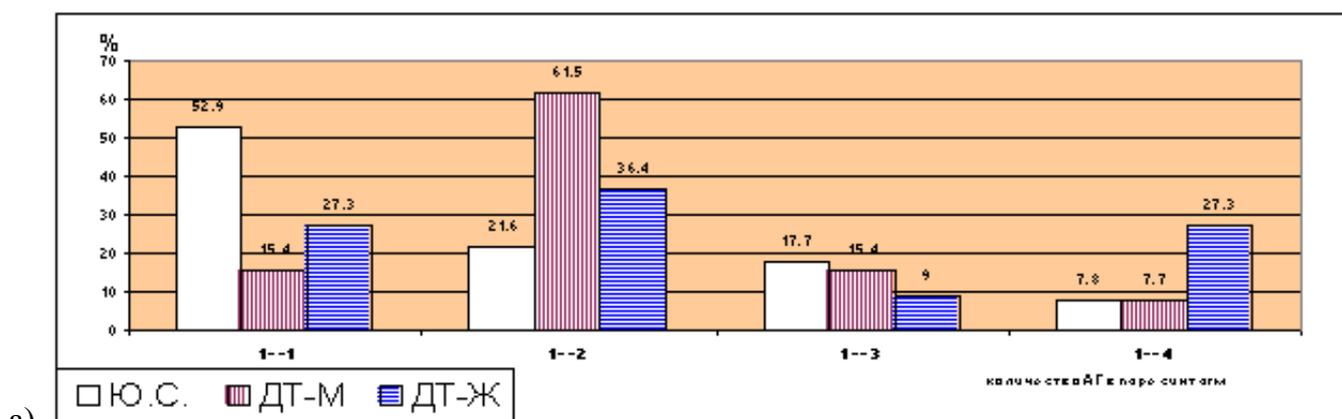


Рис. 2. Сравнительная частота встречаемости пауз различной длительности (между синтагмами с интонацией незавершённости) у диктора Ю.Сенкевича и двух других дикторов телевидения (ДТ-М и ДТ-Ж).



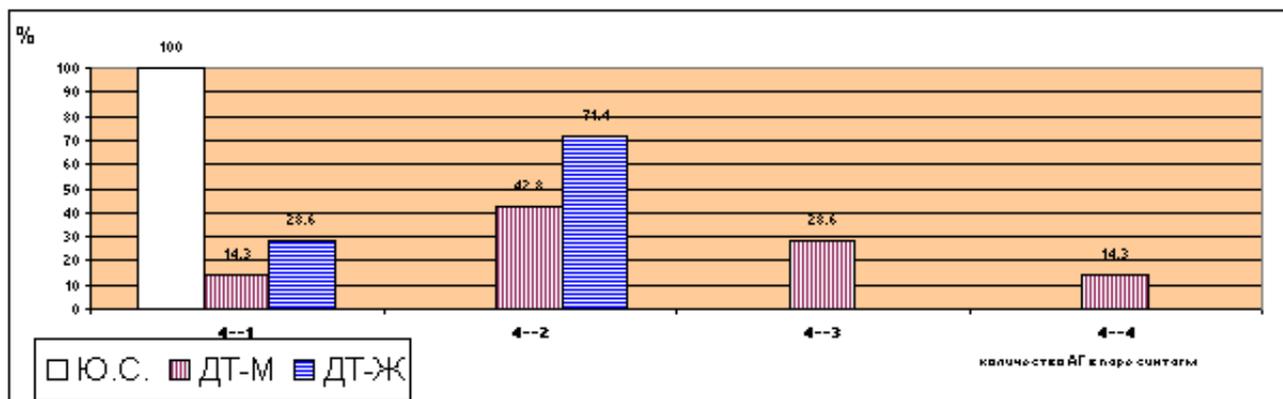


Рис. 3. Сравнительная частота встречаемости пар синтагм с различным количеством АГ у диктора Ю.Сенкевича и двух других дикторов телевидения (ДТ-М и ДТ-Ж), в которых пара начинается с: а) одноакцентной синтагмы; б) двухакцентной синтагмы; в) трехакцентной синтагмы; г) четырехакцентной синтагмы.

Обсуждение результатов

Приведенные на рис. 1-3 статистические характеристики особенностей синтагматического членения устной речи для 3-х дикторов показывают, что интересующая нас речь диктора Ю. Сенкевича обладает ярко выраженными отличиями от речи 2-х других дикторов. Из рис. 1 видно, что только у него наблюдается значительное преобладание количества одноакцентных синтагм и сравнительно равномерное распределение 2-х и 3-х акцентных. С другой стороны, из рис. 2 видим, что только у Ю. Сенкевича распределение длительностей межсинтагменных пауз относительно равномерно, в то время как у двух других дикторов наблюдается ярко выраженный максимум при нулевой длительности синтагмы. Подобные выводы можно сделать также из анализа распределений на рис. 3а,б,в,г.

Полученные экспериментальные характеристики распределений планируется использовать при клонировании речи диктора в рамках разрабатываемой стохастической модели синтагматического членения в синтезаторах речи по тексту. Проведенные статистические исследования явлений синтагматического членения являются далеко не исчерпывающими. Вне рассмотрения остались пока закономерности членения, связанные с наличием некоторых морфологических или синтаксических указателей в тексте, таких как определённые части речи или знаки препинания, перед которыми в какой-то степени допускается или запрещается наличие синтагматической границы. Опыт, приобретённый в процессе аудитивного анализа фонограмм, будет активно использован также при создании автоматизированной системы синтагматического членения аудио файлов речи дикторов.

Проведенные исследования касались лишь одной, и, пожалуй, не самой значительной части просодических явлений, в которых отражены персональные особенности речи диктора, а именно: особенностей синтагматического членения устной речи. Вне рассмотрения остались вопросы автоматического анализа, отображения и количественной оценки основного комплекса просодических характеристик речи: F0 – частоты основного тона (мелодический контур), A – амплитуды звуков (динамический контур) и T – длительности звуков (ритмический контур). Изучению комплекса этих характеристик с точки зрения проявления в них персональных особенностей речи диктора будут посвящены дальнейшие исследования авторов.

Литература

1. Б.М.Лобанов, Е.Б. Карневская, Т.В. Левковская. Синтезатор речи по тексту как компьютерное средство "клонирования" персонального голоса. Тр. Международной конференции Диалог-2001, Москва, 2001, сс. 265-272

2. Б.М. Лобанов, В.В. Киселёв. Автоматизация клонирования персонального голоса и дикции для систем синтеза речи по тексту. Тр. Международной конференции Диалог-2003, Москва, 2003, сс 417-424.