

ИНФОКОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ INFOCOMMUNICATION TECHNOLOGIES

УДК 004.934.2

ПАРАМЕТРИЧЕСКОЕ ОПИСАНИЕ ЗВУКОВ РЕЧИ В ЗАДАЧЕ РАСПОЗНАВАНИЯ

PARAMETRIC DESCRIPTION OF THE SOUND OF SPEECH IN THE TASK OF RECOGNITION

А.В. Болдышев¹, А.А. Медведева², Е.И. Прохоренко²
A.V. Boldyshev¹, A.A. Medvedeva², E.I. Prokhorenko²

¹⁾ ПАО «Ростелеком», просп. Богдана Хмельницкого, 81, Белгород, Россия

²⁾ Белгородский государственный национальный исследовательский университет,
Россия, 308015, г. Белгород, ул. Победы, 85

PJSC «Rostelecom», Ave. Bogdan Khmel'nitsky, 81, Belgorod, Russia

Belgorod State University, 85 Pobeda St, Belgorod, 308015, Russia

E-mail: boldyshev@bsu.edu.ru, medvedeva_aa@bsu.edu.ru, prokhorenko@bsu.edu.ru

Аннотация

Рассматривается задача распознавания речевых сигналов, анализируются временные, мел-кепстральные и частотные характеристики речевых сигналов, порожденные различными звуками русской речи. Предлагается комплексный подход использования различных характеристик звуков речи, который заключается в формировании групп и подгрупп звуков, в зависимости от значения исследуемых параметров, на основании которых проводится сравнительный анализ при попытке распознавания отдельного звука. Данный подход может быть использован, как элемент системы распознавания, с помощью которого можно ограничить количество вероятных типов звука для дальнейшего исследования.

Abstract

The problem of recognition of speech signals, temporal, mel-cepstral and frequency characteristics of speech signals generated by various sounds of russian speech are analyzed. The complex approach of using various characteristics of speech sounds is proposed, which consists in the formation of groups and subgroups of sounds, depending on the value of the investigated parameters, on the basis of which a comparative analysis is carried out in an attempt to recognize a separate sound. The given approach can be used as an element of the recognition system, by means of which it is possible to limit the number of possible types of sound for further investigation.

Ключевые слова: речь, распознавание речи, характеристика звука

Keywords: speech, speech recognition, sound characteristics

Введение

В настоящее время научные сообщества вкладывают огромные средства в научно-исследовательские разработки для разрешения проблемы распознавания речи. Основанием для этого служат требования постоянно развивающейся информационно-телекоммуникационной среды, в которую вовлекаются как все



более широкие слои населения, так и разнообразные системы коммерческого назначения. Основные проблемы, возникающие при создании систем распознавания речи, определяют такие факторы, как индивидуальные особенности речевой системы человека. Даже для четко произнесенных и аккуратно сформулированных слов обученным диктором без выраженных дефектов речи вероятность правильного распознавания зависит от множества сопутствующих обстоятельств от качества записывающей аппаратуры до эмоционального состояния диктора. Спонтанная же речь в естественных условиях сопровождается множеством грамматических ошибок, так называемым речевым «беспорядком» и речевыми дефектами. Влияние также могут оказывать наличие изменяющихся фоновых искажений и помех. Именно поэтому актуальная задача – распознавание естественного языка в локальных системах, остается пока недостижимой. Существующие технологии распознавания нашли применение в тех случаях, когда используемый словарь сокращен до 10 знаков или при наличии возможности создания огромной сравнительной базы [Гребнов, 2009, Аграновский, Леднов, 2004, Винцюк, 1987].

Таким образом, создание систем распознавания речи является комплексной задачей, в которой необходимо предусмотреть использование различных подходов к анализу разнообразных параметров и факторов.

В данной работе представлены анализ различных характеристик речевых сигналов и возможности применения полученных параметров в системах распознавания. Объектом исследования являются характеристики речевых сигналов, а предметом исследования – речевые сигналы, соответствующие звукам русской речи. В процессе работы над научно-исследовательской работой были использованы методы субполосного анализа/синтеза цифровой обработки речевых данных.

Параметры речевых сигналов

Речевой сигнал имеет двойственную природу. С одной стороны – это обычный акустический сигнал, который представляет собой процесс распространения энергии акустических колебаний в упругой среде и отличие различных звуков речи друг от друга определяется различием в их формировании речевым аппаратом человека [Алдошина, 2002, Fant, 1964]. С другой стороны речевой сигнал – это информационный код, «зашифрованный» посредством аддитивного смешения гармонических составляющих различной частоты и амплитуды, причем различные звуки имеют различные наборы частотных составляющих, т.е. частотный спектр, что собственно и позволяет слуховой системе человека отличить один звук от другого. Таким образом, речевой сигнал может быть представлен в виде определенного набора объективных временных и частотных характеристик [Higgins, 1985].

К исследуемым в данной работе параметрам временных характеристик относятся:

1) длительность сигнала – продолжительность звука со свойственным ему количеством колебаний в единицу времени;

2) количество переходов сигнала через ноль, которая может служить простейшей характеристикой спектральных свойств сигнала, хотя сама обработка производится во временной области [Rabiner, Schafer, 1979]. Кратковременная функция среднего числа переходов через ноль или нулевых пересечений основана на сравнении знаков соседних отсчетов и определяется, как

$$Z_n = \frac{1}{N} \sum_{m=1}^N \frac{|\operatorname{sgn}(S(m)) - \operatorname{sgn}(S(m-1))|}{2} w(n-m)$$

где N – количество отсчетов речевого сигнала $x(n)$ при $n=1, \dots, N$, $S(x)$ – знаковая функция, которая принимает значение 1, если $x(n) \geq 0$ и -1, если $x(n) < 0$, w – оконная функция;

3) значение энергии – позволяет выявить зависимость относительного значения энергии от типа звука речи и определяется из выражения [Kaiser, 1990]

$$E = \|\vec{X}\|^2.$$



Частотные характеристики представлены в виде двух групп параметров – мел-кепстральных и полученных на основе субполосного частотного анализа.

Кепстр – энергетический спектр функции $\ln|S(\omega)|^2$, определяемый выражением

$$C_s(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln|S(\omega)|^2 e^{i\omega q} d\omega$$

где $S(\omega)$ – амплитудный спектр сигнала. Поскольку $S^2(\omega)$ имеет смысл спектральной плотности энергии сигнала $x(n)$, то C_s истолковывается как энергетический спектр функции $\ln|S(\omega)|^2$. При этом, что аргумент этого «спектра» имеет размерность времени, а не частоты. Логарифм спектра мощности колебания, содержащего отраженный сигнал, имеет аддитивную периодическую компоненту, созданную этим сигналом, и поэтому преобразование Фурье от логарифма спектра мощности имеет пик (локальный максимум) на месте, соответствующем задержке отраженного сигнала [Жилияков и др., 2012].

Таким образом, получение мел-кепстральных коэффициентов позволяет вычислить следующие параметры звука речи:

- 1) количество локальных максимумов в кепстре;
- 2) взаимное расположение локальных максимумов кепстра;
- 3) доля энергии максимальных значений в кепстре.

Другая группа частотных параметров вычисляется на основе метода субполосного анализа. В некотором смысле, субполосный анализ – это задача описания свойств объекта с позиций разбиения области пространственных частот на подобласти. При этом основное внимание уделяется задаче вычисления точных значений долей энергии в заданных подобластях пространственных частот и оптимальному выделению субполосных компонент [Жилияков и др., 2007, Жилияков, Прохоренко, 2006, Баронин, 1969].

В работе используется метод, позволяющий вычислять точное распределение энергии, причем при любом желаемом разбиении на частотные интервалы подробно рассмотренный в [Жилияков и др., 2011, Жилияков, 2007, Жилияков и др., 2006].

Если отсчеты речевого сигнала представлены в виде компонентов вектора

$$\vec{x} = (x(1), \dots, x(N))^T$$

где $x(n)=x(n\Delta t)$, $n=1, \dots, N$, где Δt – интервал дискретизации по времени, то выражение

$$P_V(\vec{x}) = \frac{1}{2\pi} \int_{V} |X(\nu)|^2 d\nu \tag{1}$$

представляет собой долю энергии отрезка сигнала (евклидовой нормы вектора), соответствующую частотному интервалу

$$V = [-\nu_2, -\nu_1] \cup [\nu_1, \nu_2].$$

Здесь $X(\nu)$ представляет собой трансформанту Фурье (амплитудный частотный спектр) отрезка отсчетов сигнала (вектора), в качестве области определения которой естественно рассматривать $-\pi \leq \nu \leq \pi$ (нормированная частота).

Используя матричное обозначение, выражение (1) можно представить в виде

$$P_V(\vec{x}) = \vec{x}^T A_V \vec{x}, \tag{2}$$

где $A_V = \{a_{nk}\}$ – симметричная матрица [Гантмахер Ф.Р., 2004], элементы которой определяются как

$$a_{nk} = \begin{cases} \frac{\sin[\nu_2(n-k)] - \sin[\nu_1(n-k)]}{\pi(n-k)}, & n \neq k \\ \frac{\nu_2 - \nu_1}{\pi}, & n = k \end{cases}$$

Таким образом, долю энергий отрезка сигнала в любом частотном интервале можно вычислить на основе представления (2), не вычисляя при этом соответствующую трансформанту Фурье.



Если для речевого сигнала, соответствующего звуку речи, оценить распределение энергии в частотной области, используя субполосный анализ при разбиении частотного диапазона на некоторое количество равновеликих частотных интервалов (субполос), то можно вычислить следующие параметры:

1) номер частотного интервала с максимальной энергией;

2) количество формантных областей и их расположение. Форманта – термин фонетики, обозначающий акустическую характеристику звуков речи (прежде всего гласных), связанную с уровнем частоты голосового тона и образующую тембр звука. Тембр голоса человека – это ограниченная по своему спектру звуковая волна, которая состоит из нескольких групп частот, усиленных резонаторной системой вокального тракта, т.е. формант. В обычной разговорной речи для создания всего спектра гласных звуков достаточно трех-пяти формант, которые отличаются как по частоте, так и по энергетической составляющей. Теоретически, если описать количество, расположение и энергию формант для каждого звука, то можно сравнением однозначно определить тип любого звука, т.е. распознать его. Но индивидуальная изменчивость речевого аппарата человека и сопутствующие факторы, о которых было сказано выше, на практике препятствуют подобному способу распознавания. Тем не менее, описание формант по неким усредненным значениям, может служить одним из параметров, определяющих частотные характеристики речевого сигнала. В данном случае под формантной областью понимается субполоса, в которой сосредоточена значительная доля энергии [Белов, Белов, 2008].

3) частотная концентрация $f_{NR}^m = \min d_{NR}^m$ – минимальное количество частотных интервалов, в которых сосредоточена заданная доля энергии (например 95%) звукового отрезка:

$$W_{NR} = f_{NR}^m / R,$$

где R – общее количество равновеликих интервалов, на которые разбивается ось частот, d_{NR}^m – количество частотных интервалов, в которых сосредоточена заданная доля энергии. Исследования данного параметра были показаны в [Фирсова и др., 2011]

Итак, мы имеем три группы характеристик речевого сигнала, в пределах которых можно получить некоторое количество параметров. Кроме того, известно, что звуки речи имеют фазы нарастания, квазистационарную и затухания, таким образом, если выполнить разбиение звука на сегменты «начало», «середина» и «конец», определить для каждого из них значение энергии, то можно вычислять некоторые дополнительные параметры. Например, отношения энергий «начала» к «концу», «середины» к «концу», «середины» к «началу». Или же доли значений энергий «начала», «середины» и «конца» относительно значения энергии всего звука.

Вычислительные эксперименты

В работе были получены значения параметров всех групп, на основе речевой базы, составляющей порядка пятисот звуков русской речи, в которую вошли звуки, соответствующие всем буквам алфавита. Речевые сигналы были записаны с частотой дискретизации 16 кГц, разрядностью 16 бит в одних и тех же акустических условиях при использовании одного и того же оборудования. Отдельные звуки, соответствующие буквам алфавита, были выделены «на слух» из слитно записанных фраз.

Для каждого типа звука были определены средние значения каждого параметра каждой группы характеристик. Данные проведенных экспериментов позволили разбить все звуки на определенное количество групп и подгрупп для каждого параметра каждой группы характеристик. Принадлежность звуков той или иной группе определялась близостью значений рассматриваемого параметра. В таблицах 1 и 2 приведены примеры подобного рода группировки для параметров «длительность сигнала» и «частотная концентрация» соответственно.



Таблица 1
Table 1

Группировка звуков речи в зависимости от параметра «длительность сигнала»
Grouping of speech sounds depending on the parameter «signal duration»

№ группы	Тип звука речи	Минимальное значение параметра	Максимальное значение параметра	Среднее значение параметра
1	п	12,56	79,75	35,62
	й	24,38	60,50	39,08
2	к	9,00	164,19	46,77
	р	18,31	129,50	50,91
	т	20,25	104,06	53,28
	и	22,69	132,38	57,24
	г	20,31	99,13	57,98
	э	22,56	128,56	58,45
	в	23,63	102,75	60,55
	л	13,06	226,00	65,25
	у	22,25	117,19	68,86
3	а	26,50	157,06	71,13
	ж	31,44	139,31	73,14
	н	21,94	169,81	73,18
	ы	21,38	143,38	73,80
	д	27,31	159,56	74,00
	е	33,63	142,94	74,04
	м	34,19	149,31	77,00
	б	24,75	145,00	79,19
4	ю	74,31	95,69	83,23
	ф	42,75	122,06	86,53
	з	54,56	132,88	89,08
	х	65,56	110,56	90,01
	о	39,69	193,31	91,61
	ё	74,75	134,75	95,29
	ч	56,19	133,06	96,09
5	ц	49,75	234,44	106,69
	с	52,69	222,88	110,80
	я	46,44	190,88	112,33
	ш	90,50	162,56	122,94
	щ	72,56	176,81	137,96

Анализ значений параметра «длительность сигнала» позволил разбить все звуки русской речи на 5 групп.

Из таблицы 1 видно, что звуки речи расположились в группах без очевидной привязки к традиционным фонетическим классификациям, например приведенным в [Белашапкина и др., 1989, Буланин, 1970]. Например, самая однородная пятая группа, хотя и содержит в основном согласные звуки, которые классифицируются как шумные глухие, «ц», «с», «ш», «щ», содержит гласный составной звук «я». В остальных группах состав звуков еще более неоднороден.



Таблица 2

Table 2

Группировка звуков речи в зависимости от параметра «частотная концентрация»
 Grouping of speech sounds depending on the parameter "frequency concentration"

№ группы	Тип звука речи	Минимальное значение параметра	Максимальное значение параметра	Среднее значение параметра
1	б	3,00	10,00	3,95
	н	3,00	10,00	4,26
	в	3,00	10,00	4,34
	м	3,00	10,00	4,81
	у	3,00	9,00	4,96
	ю	3,00	15,00	5,06
	д	3,00	15,00	5,10
2	г	3,00	29,00	6,07
	ё	3,00	14,00	6,94
	е	3,00	15,00	6,95
	л	4,00	27,00	7,06
	э	3,00	17,00	7,16
	ы	3,00	14,00	7,21
	и	3,00	17,00	7,22
	о	5,00	13,00	7,52
	й	3,00	16,00	8,26
	а	5,00	19,00	8,85
	я	4,00	18,00	9,18
	х	3,00	26,00	9,63
	р	4,00	20,00	9,75
3	ж	3,00	37,00	10,31
	п	3,00	30,00	11,09
	з	3,00	40,00	12,62
	к	4,00	33,00	13,67
4	т	4,00	30,00	18,99
	щ	6,00	33,00	21,35
	ц	4,00	39,00	21,41
	ф	4,00	44,00	21,45
	ш	3,00	42,00	21,80
	ч	6,00	36,00	21,87
	с	3,00	37,00	24,23

В свою очередь, в результате анализа значений параметра «частотная концентрация» было получено 4 группы звуков речи. В данном случае состав каждой группы так же не однороден.

Кроме того, что для двух параметров было получено различное количество групп звуков речи, но так же очевидно, что состав групп различается. Аналогичные результаты наблюдаются и при анализе значений других параметров звуков.

То есть звуки речи группируются по разному при применении шкалы определенного параметра. Этот эффект так же можно визуализировать с помощью гистограмм параметров временных и частотных характеристик. На рисунке представлены подобные гистограммы некоторых параметров на примере гласного звука «о» (слева) и согласного «л» (справа).

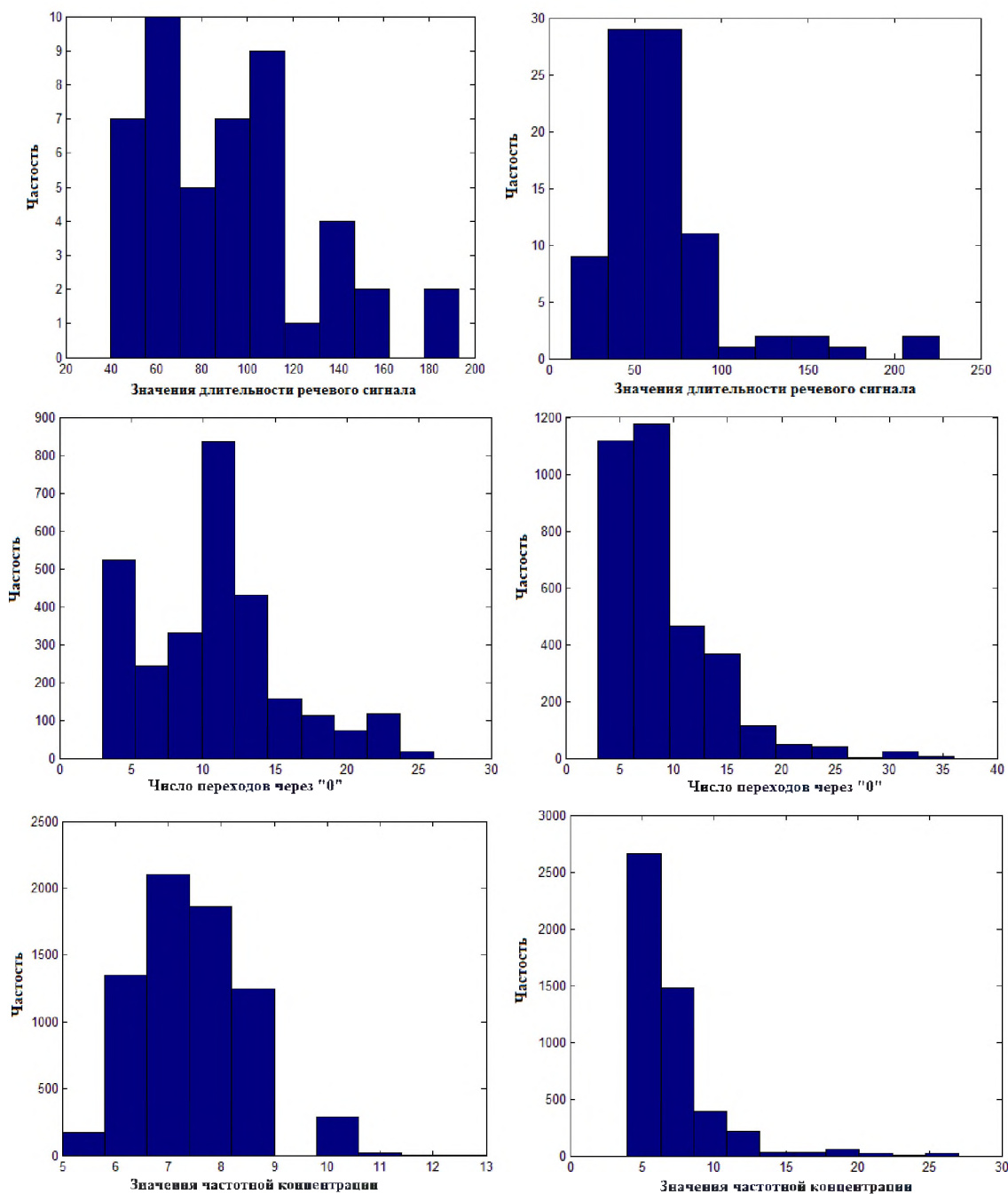


Рис. 1. Статистическое распределение параметров на примере гласного звука «о» (слева) и согласного «л» (справа), где по горизонтальной оси представлены значения соответствующих параметров, а по вертикали – количество анализируемых отрезков звуков речи

Fig. 1. The statistical distribution of parameters by the example of the vowel sound "o" (left) and the consonant "l" (right), where the horizontal axis represents the values of the corresponding parameters, and vertically - the number of analyzed segments of speech sounds

Гистограммы показывают, что различные звуки могут иметь сходные значения некоторых параметров, но большинство параметров для двух различных звуков чаще всего будут иметь разные значения. Таким образом, сравнение звуков по одному параметру не имеет смысла, а сравнение по всему набору может быть использовано как один из подходов в системах распознавания. Данное утверждение подтверждается исследованием набора параметров для отдельных звуков по сравнению с наборами параметров других звуков. В таблице 3, в качестве примера, представлен сравнительный анализ схожести значений всех параметров звука «к».



Таблица 3

Table 3

Оценка схожести значений параметров звука «к» по сравнению с параметрами других звуков
 Estimation of the similarity of the values of the sound parameters "k" in comparison with the parameters of other sounds

Звуки речи	Параметры временных характеристик	Параметры кепстральных характеристик	Параметры субполосных характеристик	Всего
	Количество параметров соответствующих звуку «к»			
К	9	3	4	16
	Количество параметров соответствующих звуку «к» схожих с параметрами другого указанного звука			
р	6	2	2	10
п	3	3	3	9
т	4	2	2	8
ж	2	3	2	7
з	2	3	2	7
х	2	3	2	7
а	3	1	2	6
и	4	2	0	6
л	4	2	0	6
о	3	2	1	6
ц	5	1	0	6
й	1	3	1	5
м	5	0	0	5
с	3	2	0	5
ч	3	1	1	5
я	4	0	1	5
б	2	2	0	4
в	3	1	0	4
г	3	1	0	4
у	3	1	0	4
ш	3	0	1	4
д	2	1	0	3
е	2	1	0	3
ф	2	1	0	3
э	3	0	0	3
н	2	0	0	2
щ	1	1	0	2
ы	2	0	0	2
ё	1	0	0	1
ю	0	0	0	0

В таблице 3 показано, что из 16 параметров, рассчитанных для звука «к», десять совпадают с параметрами звука «р», в девяти случаях совпадают с параметрами звука «п», в восьми случаях совпадают с параметрами звука «т», и так далее. При этом, параметры временных характеристик звука «к» совпадают с параметрами временных характеристик звука «р» в 6 случаях, параметры кепстральных характеристик звука «к» совпадают с параметрами кепстральных характеристик звука «р» в 2 случаях, параметры субполосных характеристик звука «к» совпадают с параметрами субполосных характеристик звука «р» в 2 случаях.

Аналогичные сравнения были выполнены и для других звуков. Результаты перекрестного сравнения показывают, что вероятность распознавания отдельного звука с использованием подобного набора параметров колеблется в интервале от 20 до 60 процентов в зависимости от звука и индивидуальной особенности голоса диктора.

Заключение

Анализ существующих методов распознавания речи позволяет сказать, что основной целью обработки речевых сигналов является получение наиболее удобного и компактного представления содержащейся в них информации. Точность представления определяется той информацией, которую необходимо сохранить или выделить

Временные характеристики речевых сигналов могут существенно меняться в зависимости от скорости воспроизведения, а также зависят непосредственно от типа звука. При анализе переходов сигнала через ноль характерно проявление квазипериодичности. Особенно ярко это проявляется для звуков, которые принято относить к классу вокализованных. При исследовании частотных характеристик интерес представляет ширина субполосных интервалов, а так же доля энергии, которую можно считать подавляющей при оценивании частотной концентрации, поскольку варьирование значений этих переменных влияет на результаты группирования звуков речи.

Результаты перекрестного сравнения параметров различных звуков речи показывают, что построить самостоятельную систему распознавания, основанную лишь на параметрическом подходе весьма затруднительно. Такой подход может быть полезен как элемент системы распознавания, с помощью которого можно ограничить количество вероятных типов звука для дальнейшего исследования.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 1507-01463.

Список литературы

References

1. Аграновский А.В., Леднов Д.А. 2004. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов. М., Радио и связь, 164.
Agranovsky A.V., Lednov D.A. 2004. Theoretical aspects of algorithms for processing and classifying speech signals. Moscow, Radio and Communication, 164.
2. Алдошина И.А. 2002. Основы психоакустики. Слух и речь. Часть 2. Информационно-технический журнал «Звукорежиссер», 3: 54–58.
Aldoshina I.A. 2002. Fundamentals of psychoacoustics. Rumor and speech. Part 2. Information and Technology magazine "Sound producer", 3: 54-58.
3. Баронин С.П. 1969. Спектральный анализ и проблема сокращенного описания речевых сигналов. В кн.: Спектральный анализ звуков речи и интонация. М., Издательство МГУ: 13–30.
Baronin S.P. 1969. Spectral analysis and the problem of reduced description of speech signals. In: Spectral analysis of speech sounds and intonation. Moscow, Moscow State University Publishing House: 13-30.
4. Белов С.П., Белов А.С. 2008. О различиях частотных свойствах информационных и неинформационных звуковых сигналов речевого диапазона. Научные ведомости БелГУ. Серия: Информатика, 7(38): 214–221.
Belov S.P., Belov A.S. 2008. On the differences in the frequency properties of information and non-information sound signals of the speech range. Nauchnye vedomosti BelGU. Informatika. [Belgorod State University Scientific Bulletin. Information technologies], 7 (38): 214-221. (in Russian).
5. Белошапкина В.А., Брызгунова Е.А., Земская Е.А. 1989. Современный русский язык: Учебник для филологических специальностей университетов. М., Высшая школа, 800.
Beloshapkova V.A., Bryzgunova E.A., Zemskaya E.A. 1989. Modern Russian language: Textbook for philological specialties of universities. Moscow, Higher School, 800.
6. Буланин Л.Л. 1970. Фонетика современного русского языка. М., Высшая школа, 206.
Bulanin L.L. 1970. Phonetics of the modern Russian language. Moscow, Higher School, 206.
7. Винцюк Т.К. 1987 Анализ, распознавание и интерпретация речевых сигналов. Киев, Наукова думка, 264.
Vinciuk T.K. 1987 Analysis, recognition and interpretation of speech signals. Kiev, Naukova Dumka, 264. (in Ukrainian)
8. Гантмахер Ф.Р. 2004. Теория матриц. М., Физматлит, 560.
Gantmakher F.R. 2004. The theory of matrices. Moscow, Fizmatlit, 560. (in Russian)



9. Гребнов С.В. 2009. Аналитический обзор методов распознавания речи в системах голосового управления. Вестник ИГЭУ, Вып.3.: 83–85.

Grebnov S.V. 2009. Analytical review of methods of speech recognition in voice control systems. Vestnik of ISEU, Issue 3: 83-85. (in Russian).

10. Жиляков Е.Г., Фирсова А.А., Чеканов Н.А. 2012. Алгоритмы обнаружения основного тона речевых сигналов. Научные ведомости БелГУ. Серия История. Политология. Экономика. Информатика, 1(120), выпуск 21/1: 135–143.

Zhilyakov E.G., Firsova A.A., Chekanov N.A. 2012. Algorithms for detecting the pitch of speech signals. Nauchnye vedomosti BelGU. Istoriya. Politologiya. Ekonomika. Informatika. [Belgorod State University Scientific Bulletin. History. Political science. Economy. Information technologies], 1 (120), issue 21/1: 135-143. (in Russian)

11. Жиляков Е.Г., Прохоренко Е.И., Болдышев А.В., Фирсова А.А., Фатова М.В. 2011. Сегментация речевых сигналов на основе анализа распределения энергии по частотным интервалам. Научные ведомости Белгоро БелГУ. Сер. История. Политология. Экономика. Информатика, 7 (102): 187–196.

Zhilyakov E.G., Prohorenko E.I., Boldyshev A.V., Firsova A.A., Fatova M.V. 2011. The segmentation of speech signals based on the analysis of the energy distribution of frequency intervals. Nauchnye vedomosti BelGU. Istoriya. Politologiya. Ekonomika. Informatika. [Belgorod State University Scientific Bulletin. History. Political science. Economy. Information technologies], 7 (102): 187-196. (in Russian)

12. Жиляков Е.Г. 2007. Вариационные методы анализа и построения функций по эмпирическим данным. Белгород, Издательство БелГУ, 160.

Zhilyakov E.G. 2007. Variational methods of analysis and construction of functions by empirical data. Belgorod, Belgorod State University Publishing House, 160

13. Жиляков Е.Г., Белов С.П., Прохоренко Е.И. 2007. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений. Белгород, Издательство БелГУ, 136.

Zhilyakov E.G., Belov S.P., Prokhorenko E.I. 2007. Methods of processing voice data in information and telecommunication systems on the basis of frequency representations. Belgorod, Belgorod State University Publishing House, 136.

14. Жиляков Е.Г., Белов С.П., Прохоренко Е.И. 2006. О субполосном преобразовании звуковых сигналов. Труды Российского научно-технического общества радиотехники, электроники и связи имени А.С.Попова. Выпуск VIII-1:167–169.

Zhilyakov E.G., Belov S.P., Prokhorenko E.I. 2006. About subband conversion of sound signals. Proceedings of the Russian Scientific and Technical Society of Radio Engineering, Electronics and Communications named after A.S. Popov. Issue VIII-1: 167-169.

15. Жиляков Е.Г., Прохоренко Е.И. 2006. Частотный анализ речевых сигналов. Научные ведомости БелГУ. Серия Информатика и прикладная математика, 2(31), выпуск 3: 201–208.

Zhilyakov E.G., Prokhorenko E.I. 2006. Frequency analysis of speech signals. Nauchnye vedomosti BelGU. Informatika i prikladnaya matematika. [Belgorod State University Scientific Bulletin. Information technologies and Applied Mathematics], 2 (31), Issue 3: 201-208. (in Russian)

16. Фирсова А.А., Белов А.С., Курлов А.В. 2011. О различии концентрации энергии по частотным диапазонам на отрезках сигналов, соответствующих шипящим звукам русской речи и шумам. Научные ведомости БелГУ. Сер. История. Политология. Экономика. Информатика, 3(108), выпуск 19/1: 186–191.

Firsova A.A., Belov A.S., Kurlov A.V. 2011. On the difference in energy concentration over frequency ranges on signal segments corresponding to the hissing sounds of Russian speech and noise. Nauchnye vedomosti BelGU. Istoriya. Politologiya. Ekonomika. Informatika. [Belgorod State University Scientific Bulletin. History. Political science. Economy. Information technologies], 3 (108), issue 19/1: 186-191. (in Russian)

17. Fant G. 1964. Acoustic theory of speech production. Mouton & Co, 304.

18. Higgins A. 1985. Keyword recognition using template concatenation. Acoustics, Speech, and Signal Processing. IEEE International Conference on ICASSP.

19. Kaiser J. 1990. On a simple algorithm to calculate the 'energy' of a signal. IEEE ICASSP: 381-384

20. Rabiner Lawrence R., Schafer Ronald W. 1979. Digital Processing of Speech Signals. Prentice-Hall. Inc, Englewood Cliffs, 509