

УДК 004.622+004.934+534.86=411.21

ФОРМИРОВАНИЕ БАЗ ДАННЫХ НА РУССКОМ ЯЗЫКЕ ДЛЯ ВЕРИФИКАЦИИ АРАБСКОЯЗЫЧНЫХ ДИКТОРОВ

М.О. АЛЬ-ХАТМИ, М.Ш. МАХМУД, Л.М. ЛЫНЬКОВ, А.Г. ДАВЫДОВ, Д.А. БОРИСЕВИЧ

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь*

Поступила в редакцию 5 октября 2009

Рассмотрены принципы построения и описаны базы данных на русском языке для возможного распознавания речи дикторов арабского происхождения, учитывающие артикуляционные отличия согласных и гласных звуков русского и арабского языков и результаты измерений статистических спектральных характеристик.

Ключевые слова: речь, диктор, спектр, артикуляция.

Введение

При оценке принадлежности диктора к какой-либо этнической группе (национальности) используются экспертные методики определения разборчивости речевых сигналов, в частности по артикуляционным таблицам [1], которые представляют собой измерения относительного числа правильно переданных слов, слогов и звуков через испытываемый (тестируемый) канал. Разрабатываемая методика формирования баз данных на русском языке, используемая диктором — носителем арабской речи, в первую очередь предназначена для систем распознавания русскоязычной речи для экспертных систем, систем оценки и улучшения качества связи, анализа речи возможных террористов и злоумышленников [2].

Целью данной работы является разработка методологических основ и составление артикуляционных таблиц на русском языке путем сравнения систем согласных и гласных звуков арабского и русского языков [3], различий в статистических характеристиках спектра речи для оценки ее разборчивости и идентификации.

Органы речи и классификация речи

Звуки речи образуются с помощью речевого аппарата. Органы, составляющие речевой аппарат человека, делятся на: 1) дыхательные органы (легкие, бронхи, дыхательное горло); 2) гортань с голосовыми связками; 3) надгортанные полости. Дыхательные органы подают струю воздуха, необходимого для образования звука, которая попадает в гортань, где находятся голосовые связки. Последние представляют собой две мускульные складки. Когда голосовые связки сближены и напряжены, струя воздуха, проходящая между ними, с силой раздвигает их и приводит в колебания, в результате чего образуется музыкальный тон (голос).

Надгортанные полости состоят из полости глотки, полости рта и полости носа. Работа органов речи при образовании звуков речи называется артикуляцией. В зависимости от того, какое участие органы речи принимают в образовании звуков, они делятся на активные (подвижные) и пассивные (неподвижные).

Активные органы речи, к которым относятся губы, язык, язычок, голосовые связки и мышцы гортани, определяют звуки по способу их артикуляции.

Пассивные органы речи, к которым относятся зубы, нёбо и гортань, определяют звуки по месту их артикуляции.

Кроме классификации звуков речи с точки зрения физиологической, их можно также классифицировать с точки зрения акустической, т.е. по тому, какое впечатление на слух производит тот или иной звук.

Все звуки речи делятся, прежде всего, на согласные и гласные.

Согласные звуки

В основе артикуляции согласных звуков лежит шум, образуемый в полости рта или гортани. По способу артикуляции согласные звуки делятся на: смычные (мгновенные), например, русские *б, т* ; щелевые (фрикативные), например, русские *с, з*; и смычно-щелевые (аффрикаты). Аффрикаты – это сложные звуки, в которых первый образующий их звук смычный, а второй – фрикативный. В русском языке, например, аффрикатой является звук *ч*, состоящий из звука *т*, произносимого с последующим звуком *ш* слитно.

В табл. 1 представлена классификация согласных звуков арабского языка [4].

Таблица 1. Классификация согласных звуков арабского языка

По месту артикуляции По способу артикуляции			По месту артикуляции		По способу артикуляции					По месту артикуляции	
			Губные		Нёбные					Гортанные	
			Губно-губные	Губно-зубные	Переднёнбные			Среднёнбные	Глубокозаднёнбные	Связочные	Зевные
Смычные	Шумные взрывные	Звонкие	б			д	д				'айн
		Глухие				т	т		к	к	'хамза
	Сонанты носовые		м					н			
Щелевые (фрикативные)	Шумные	Звонкие			з	з	з		г		
		Глухие		ф	с	с	с	ш	х	х	х
	Сонанты		й					л	й		
Смычно-щелевые (аффрикаты)		Шумные звонкие						дж			
Дрожащие (вибранты)		Сонанты						р			

По месту артикуляции согласные делятся на губные (губно-губные, губно-зубные), нёбные (переднёнбные, среднёнбные, глубоко-заднёнбные), гортанные (связочные и зевные).

Артикуляция согласных характеризуется наличием:

- шума;
- музыкального тона (голоса). В зависимости от наличия голоса согласные делятся на глухие и звонкие;
- палатализации, т.е. смягчения. Согласные, подвергающиеся смягчению, называются палатализованными;
- эмфатичности (напряженности звука). Звуки, произносимые с напряжением, называются эмфатическими (или зычными).

И, наконец, в зависимости от того, что преобладает при артикуляции согласного звука – шум или голос, согласные делятся на шумные (например, русские *б, с, з*) и сонорные (*м, р, н, л*).

Сравнение системы согласных звуков арабского и русского языков показывает, что между ними имеются существенные различия: 1) в русском языке 35 согласных звуков, а в

арабском – 28; 2) в большинстве своем русские согласные различаются по признаку мягкости и твердости, чего в арабском языке не наблюдается; 3) в арабском языке есть такие согласные звуки, которых нет в русском языке (например, межзубные, эмфатические, зевные и др.); с другой стороны в русском языке есть такие согласные звуки, которых нет в арабском языке (например: *п, в, ц, ч* и др.)

Гласные звуки

Основой артикуляции гласных звуков является музыкальный тон (голос). Гласные звуки классифицируются в зависимости от положения основных органов речи, участвующих при их образовании: языка, губ и мягкого нёба. В зависимости от положения языка в полости рта гласные звуки различаются по ряду и степени подъема языка. По ряду гласные классифицируются на гласные заднего, среднего и переднего ряда. По степени подъема языка различаются гласные нижнего, среднего и верхнего подъема (например, русская *а* – нижнего подъема, переднего ряда). Губы при артикуляции гласных либо округлены (лабиализованы), либо находятся в нейтральном положении. В зависимости от положения губ гласные делятся на лабиализованные (например, русские *о, у*) и нелабиализованные (например, русские *а, и*).

В зависимости от степени раскрытия рта гласные подразделяются на гласные открытого типа (например, русская *а* в слове «база») и гласные закрытого типа (например, русская *и* в слове «милый»).

Помимо приведенной выше классификации гласных звуков в качественном отношении, в арабском языке гласные различаются также количественно, т.е. по признаку долготы и краткости, причем этот признак имеет смысловозначительное значение.

Другой фонетической особенностью арабского языка является то, что в нем не допускается стечение двух гласных, стоящих рядом (например, как в слове «сообщение»).

Разработка принципов построения и описания баз данных на русском языке для сегментации речевых сигналов арабскоязычных дикторов

Для возможного автоматизированного распознавания дикторов арабского происхождения, которые произносят речь на русском языке, предлагается следующая схема построения (рис. 1) на основе расчета и измерения ее разборчивости.



Рис 1. Схема построения методики анализа речи на русском языке, произносимой арабскоязычным диктором.

Первый этап построения такой методики заключается в определении русских согласных и гласных звуков, имеющих отличие по месту артикуляции. Второй этап характеризуется

определением русских согласных и гласных звуков, которые не употребляются в арабской речи. На третьем этапе проводится анализ русскоязычных текстовых таблиц [3] и их направленный выбор для оценки разборчивости речи.

Основной особенностью данной методики является установление и учет распределения звукового давления такой речи по октавным полосам.

Статистические характеристики спектра речи

Анализ статистических характеристик спектра речи по литературным источникам показывает, что имеются различия спектров в зависимости от языка. В связи с этим была проведена работа по изучению различий в спектре речи разных языков, так как эти данные могут быть использованы при оценке разборчивости речи в каналах связи и построении систем защиты речевой информации для каждого конкретного языка.

При изучении статистических характеристик речи использовались аудиозаписи с речью дикторов, которые являются носителями данных языков. Запись аудиозаписей выполнялась на персональном компьютере с помощью аппаратуры RFT для акустических измерений с линейной частотной характеристикой в диапазоне 20 Гц – 10 кГц и программы Sound Forge 9.0. Предварительная обработка аудиофайлов проходила с использованием программы Sound Forge 9.0. Запись дикторов из Беларуси и арабских стран проводилась в акустически заглушенной комнате. Данные по русской и английской речи были взяты из литературных источников [2], [3].

Построение спектра речевых сигналов и дальнейшее изучение различий спектров проводилось с помощью программы MatLAB R2008. В данной программе получали значение мощности всего сигнала и характеристики звукового давления по октавам. Так как аудиозаписи разных дикторов были получены с разными интегральными уровнями, полученные данные нормировались на один уровень – 70 дБ. Статистическая обработка проводилась на основании полученных данных и посчитанных средних значений [5,6].

Для подсчета средних значений данные уровней звукового давления, выраженного в децибелах, переводились в значение звукового давления, выраженного в паскалях, по формуле:

$$P = 10^{(\bar{P}/20)} \cdot P_0, \quad (1)$$

где P – полученное значение уровня звукового давления, дБ; P_0 – пороговый уровень слуха, равный $2 \cdot 10^{-5}$ Па.

Далее для всех значений P одной октавы находилось среднее значение звукового давления P_{cp} . Средние значения звукового давления P_{cp} затем переводились в звуковое давление, выраженное в децибелах.

Для оценки полученных результатов и задания доверительного интервала была посчитано среднеквадратичное отклонение σ в каждой октаве по формуле:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (P_i - P_{cp})^2}{n}}, \quad (2)$$

где P_i – значение звукового давления в октаве для i -го диктора, Па; i – для всех дикторов, представителей одного языка; n – количество дикторов.

В табл. 2 представлены полученные результаты анализа спектров речи для различных языков. В графе «белорусская» представлены спектры жителей Республики Беларусь, которые читали текстовые таблицы Д50 и Д70 на русском языке из СТБ ГОСТ Р50840-2000. Также на рис. 2–5 представлены графики с распределением звукового давления по октавам для каждого языка отдельно. На приведенных графиках видно, что статистические характеристики речи разных языков различны. При оценке разборчивости речи в каналах связи и построении систем защиты речевой информации необходимо учитывать полученные данные. При построении более точных систем можно использовать характеристики звукового давления по третьоктавным полосам, которые являются более точными.

Таблица 2. Распределение звукового давления по октавным полосам

Речь	Среднее значение звукового давления, дБ				
	1-я октава, $f_{cp}=250$ Гц	2-я октава, $f_{cp}=500$ Гц	3-я октава, $f_{cp}=1000$ Гц	4-я октава, $f_{cp}=2000$ Гц	5-я октава, $f_{cp}=4000$ Гц
белорусская	$P_{cp}=65,4$ $P_{cp}+\sigma_+=66,6$ $P_{cp}-\sigma_-=63,9$	$P_{cp}=67,1$ $P_{cp}+\sigma_+=68,0$ $P_{cp}-\sigma_-=66,0$	$P_{cp}=59,0$ $P_{cp}+\sigma_+=61,5$ $P_{cp}-\sigma_-=55,7$	$P_{cp}=55,2$ $P_{cp}+\sigma_+=57,6$ $P_{cp}-\sigma_-=52,0$	$P_{cp}=50,4$ $P_{cp}+\sigma_+=52,9$ $P_{cp}-\sigma_-=46,7$
русская [2]	66,3	66,0	60,8	56,1	53,0
арабская	64,8	64,2	63,2	57,9	49,8
английская [3]	65,1	67,0	62,6	53,7	44,0

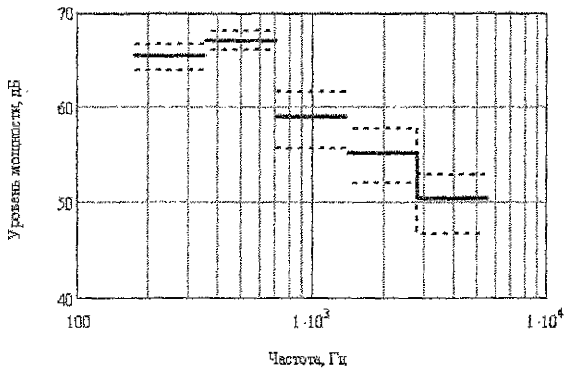


Рис. 2. Белорусская речь

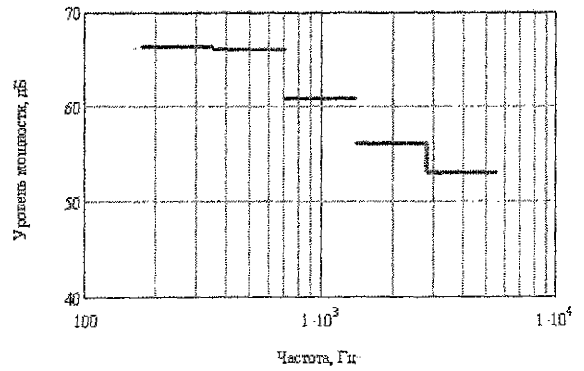


Рис. 3. Русская речь

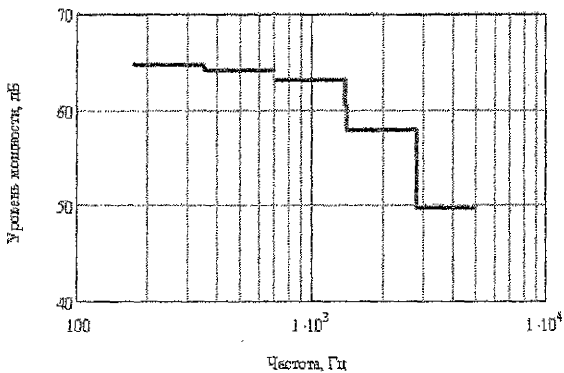


Рис. 4. Арабская речь

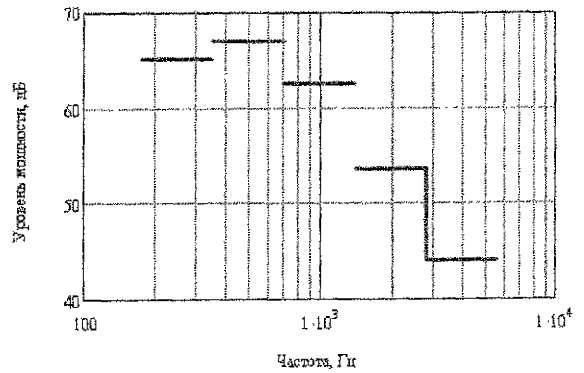


Рис. 5. Английская речь

Заключение

В результате сравнения системы согласных и гласных звуков русского и арабского языков показаны их различия по принципу мягкости и твердости. При этом в русском языке употребляются такие согласные звуки, которых нет в арабском языке, что позволило создать словесные и артикуляционные таблицы на русском языке для арабскоязычных дикторов. Экспериментально установлена значительная разница в статистических спектральных характеристиках речи по октавам для каждого языка.

FORMATION OF DATABASES IN RUSSIAN LANGUAGE FOR ARABIC ANNOUNCER SPEECH SIGNALS VERIFICATION

M.O. ALHATME, M.SH. MAHMOUD, L.M. LYNKOV, A.G. DAVIDOV, D.A. BORISEVICH

Abstract

The database in Russian for possible speech recognition of Arabic announcer is described and principles of its construction are considered. These databases take into account the articulation differences of consonants and vowels in Russian and Arabic languages. The results of statistical measurements of spectral characteristics are used.

Литература

1. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи. М., 1987.
2. Хорев А.А. Защита информации от утечки по техническим каналам. Ч.1. Технические каналы утечки информации. Учебное пособие. М., 1998.
3. Покровский Н.Б. Расчет и измерение разборчивости речи. М., 1962.
4. Ковалев А.А., Шарбатов Г.Ш. Учебник арабского языка. М., 1979.
5. Железняк В.К., Макаров Ю.К., Хорев А.А. Некоторые методические подходы к оценке эффективности защиты речевой информации.
6. ISO/TR4870 Acoustics – The construction and calibration of speech intelligibility test. P. 22.