

О СОЗДАНИИ ФОНЕТИКО-АКУСТИЧЕСКОЙ БАЗЫ В РАМКАХ СИНТЕЗА ЧЕЧЕНСКОЙ РЕЧИ

Э. С. Израилова

*Академия наук Чеченской республики,
Грозненский государственный нефтяной технический университет*

Поступила в редакцию 09.06.2017 г.

Аннотация. Статья посвящена описанию процесса создания фонетико-акустической базы чеченского языка. Рассматриваются вопросы, связанные с разработкой структуры базы данных, выбором минимального фрагмента базы, подбором текстовых материалов, вводом и разметкой речевых фрагментов. Приведен созданный предварительно чеченский фонетический алфавит, ориентированный на создание системы автоматического синтеза чеченской речи.

Ключевые слова: фонетико-акустическая база данных, система синтеза речи, чеченский язык, дифтонг, аллофон, фонетический алфавит.

Annotation. The article is devoted to the description of the process of creating a phonetic and acoustic base of the Chechen language. The issues related to the development of the database structure, selection of the minimum fragment of the database, selection of text materials, input and markup of speech fragments are considered. A previously created Chechen phonetic alphabet was created, aimed at creating a system for the automatic synthesis of Chechen speech.

Keywords: phonetic-acoustic database, speech synthesis system, Chechen language, diphon, allophone, phonetic alphabet.

В последние годы в информационно-коммуникационной сфере стремительно развиваются речевые технологии, в частности, технологии синтеза и распознавания речи. Современные разработки в области речевых технологий основаны на использовании фонетико-акустических, текстовых, речевых баз данных. Для обеспечения высокого качества синтезированной речи эти базы данных должны содержать достаточно полный набор фонетических, просодических и акустических элементов речи.

РЕЧЕВЫЕ КОРПУСА И БАЗЫ ДАННЫХ

Совокупность речевых фрагментов, и созданных на их основе баз данных, кото-

рые обеспечены программными средствами доступа к ним, представляют собой речевой корпус конкретного языка [1]. При моделировании просодических характеристик речи и индивидуальных особенностей диктора корпусный подход является определяющим для качественного синтеза. Наиболее известные речевые корпуса: корпус английского языка TIMIT (1980–1990) [2], речевой корпус русского языка ISABASE (1998) [3], речевой корпус русского языка RuSpeech (2000–2001) [4]. Таким образом, начальным этапом создания системы синтеза чеченской речи является формирование речевого корпуса чеченского языка и создание сбалансированных фонетико-акустических баз данных.

Состав элементов базы может быть различным в зависимости от типа избранной базовой единицы. Чаще всего в качестве такой

единицы выступает диффон и/или аллофон. Для диффонного синтеза требуется база данных, содержащая все возможные для заданного языка двучленные комбинации фонем, а при использовании аллофонов необходим учет всех возможных сочетаний левого и правого контекста. В работе [5] аллофоны подразделяют на позиционные и комбинаторные. Позиционные аллофоны зависимы от положения ударного гласного, комбинаторные же определяются контекстом фонемы и эффектами коартикуляции и ассимиляции.

Разработчики синтезатора татарской речи [6, 7] в качестве исходной единицы используют диффоны. Акустическая база этого синтезатора включает три типа диффонов – начальный, срединный и конечный.

Для проектируемой системы синтеза чеченской речи в качестве базового набора фонетико-акустических элементов были определены диффоны и аллофоны. Создаваемая система ставит целью формирование достаточного набора диффонов для максимального приближения характеристик синтезированной речи к персональным характеристикам естественной речи. Основная идея при построении базы диффонов – перечисление всех возможных сочетаний фонем на языке.

ФОНЕТИЧЕСКАЯ СИСТЕМА ЧЕЧЕНСКОГО ЯЗЫКА

У исследователей фонетики чеченского языка нет единого мнения относительно количества фонем в современном чеченском литературном языке. Фонетическая система чеченского языка отличается сложностью вокализма (краткие, длинные, умлаутированные, простые гласные, дифтонги, трифтонги, слабо выраженная назализация гласных) и консонантизма (простые, геминированные, абруптивные, фарингальные согласные) [9].

Так, Ю. Д. Дешериев выделяет в чеченском языке более 30 гласных и 35 согласных фонем [9]. В работе [10] Д. С. Имнайшвили приводит 30 гласных и 37 согласных фонем. Общее число диффонов языка определяется квадратом числа фонем, то есть 652 образует 4225 диффонов. Однако в потоке звучащей речи

встречаются не все парные сочетания фонем. Ориентировочно в диффонной базе чеченского языка будет 1500 диффонов. В дополнение к базовым диффонам, в качестве различных оттенков фонем, также предполагается добавить в базу некоторые варианты аллофонов. Необходимость добавления аллофонов подтверждается следующими примерами. К примеру, фонема переднего ряда **аь** [ä] в положении перед удвоенными согласными бывает более открытой, чем в других позициях (аьтта ‘крошил’, даьтта ‘масло’, но аьхна ‘пахал’, йаьхна ‘отнимал’). Мягкость согласных фонем в чеченском языке, в отличие от русского не фонематична, хотя твёрдые согласные перед гласными несколько смягчаются [i], [e], [ö], [öö]. Такая незначительная палатализация не создает особых палатализованных фонем, появляются лишь их оттенки (аллофоны).

При создании базы диффонов сохраняются без изменения переходные участки, на которых происходит фонетическое взаимодействие звуков. Речевые фрагменты из базы будут размечены на диффоны и фонемы с помощью полуавтоматической системы разметки. Для этого потребовалась разработка системы фонетической транскрипции, адаптированной под фонетические особенности чеченского языка и ориентированной на использование в системе синтеза речи.

ФОНЕТИЧЕСКИЙ АЛФАВИТ ЧЕЧЕНСКОГО ЯЗЫКА AZBAT

Ссылаясь на опыт разработчиков других фонетических систем и баз, разработан фонетический машиночитаемый алфавит чеченского языка AZBAT [11] по аналогии с алфавитом американского английского языка DARPAbet [12]. За основу была взята система чеченской фонетической транскрипции из книги «Грамматика чеченского языка» [13].

При разработке алфавита AZBAT учитывались особенности произношения и графики, правила сочетаемости и вариативности фонем, описанные филологами в работах [9, 10, 13]:

- Выражение ряда гласных звуков двухбуквенными сочетаниями (**аь**, **уь**, **оь**).

- Обозначение долготы в одних случаях двух и трехбуквенными сочетаниями (**ие**, **ий**, **уй**).

- Отсутствие графического выражения долготы (**а**, **о**, **е**).

- Обозначение одной и той же буквой нескольких гласных фонем. Например, через букву **о** передают 4 фонемы: **[o]**, **[ō]**, **[yo]**, **[yō]**.

- Обозначение нелабиализованных и лабиализованных гласных переднего образования через **яь**, **юь**.

AZBAT состоит из 63 фонем (27 гласных и 36 согласных). В нем каждая фонема представлена одной, двумя или тремя буквами (см. таблицу 1). Две буквы используются для

обозначения дифтонгов и долготы гласных звуков [**аа**, **ее**, **ii**, **uu**, **oo**, **ie**, **uo**, **oa**]. Умлаутированные гласные фонемы **аь**, **оь**, **уь** обозначены через двухбуквенные сочетания [**ia**, **io**, **iu**]. Долгие дифтонги обозначены трехбуквенными сочетаниями [**uoo**, **iee**, **iuu**, **ioo**]. Символом **@** обозначена гортанная смычка **ʔ** (**ъ**). **I** – обозначение фарингального звонкого спиранта, и **II** – это ларингальный абруптивный звонкий согласный. Цифры используются как индикаторы ударения и помещаются в конце ударного гласного (0 – безударная гласная, 1 – первичное ударение, 2 – вторичное ударение).

Для реализации программного кода необходимо задать атрибуты, описывающие

Таблица 1

Фонетический алфавит AZBAT

Фонема MFA	Фонема AZBAT	Слово	Транскрипция AZBAT	Фонема MFA	Фонема AZBAT	Слово	Транскрипция AZBAT
a	a	мах, атта	max, att-ta	v	v	ведар	vee-dar
â	ax	худар	xu-daxr	f	f	фонема	fo-ne-ma
ā	aa	ваха, са	vaa-xa, saa	n	n	нана	naa-na
i	i	диг	dig	d	d	диг	dig
ī	ii	дийца	dii-ca	s	s	сом	suom
ie	ie	беш	biesh	t	t	луьста	lius-ta
iē	iee	леха	liee-xa	tʼ	tI	тIадам	tIa-dam
u	u	дуга	du-ga	z	z	борз	borz
ū	uu	духа	duu-xa	dz	dz	зама	dzaa-ma
e	e	кеста	kes-ta	c	c	цициг	ci-cig
ē	ee	леча	lee-cha	cʼ	cI	цIазам	cIa-zam
ä	ia	даьтта	diat-ta	r	r	барам	baa-ram
ü	iu	туьха	tiu-xa	l	l	шийла	shii-la
ũ	iuu	туьйра	tiu-u-ra	dž	dj	жима	dji-ma
ö	io	доьхка	diox-ka	č	ch	чам	cham
öö	ioo	доьналла	dioo-nal-lla	čʼ	chI	чIуг	chIug
o	o	догIа	do-gIa	š	sh	шовда	shovda
ō	oo	олу	@oo-lu	ž	j	бажа	ba-ja
oa	oa	хорсам	xoar-sam	j	y	лайн	lay-na
uo/wo	uo	сом	suom	g	g	гара	gaa-ra
uō/wō	uoo	тоха	tuoo-xa	k	k	киса	ki-sa
aj	ay	вайца	vay-ca	kʼ	kI	кIурз	kIurz
ej	ey	дейтта	deyt-ta	y	gI	гIала	gIa-la
ow	ow	ловза	low-za	x	x	хаза	xa-za
öw	iow	лоьвзи	liow-zi	q	q	кхор	quor
uj	uy	буйнаца	buy-na-ca	qʼ	qq	къам	qqam
oj	oy	гойту	goy-tu	ω	I	бIов	blow
m	m	дама	da-ma	˘	II	Iа	IIa
b	b	барт	bart	h̥	hx	хьяжа	hxa-ja
p	p	полла	poll-la	ʔ	@	урам, диь	@uu-ram, di@
pʼ	pI	пIелг	pIe-lg	h	h	хIоа	huo-@a
w	w	говр	gowr				

признаки каждой фонемы. Поэтому гласные фонемы были классифицированы по следующим группам: краткие гласные, длинные гласные, краткие дифтонги, длинные дифтонги, редуцированные. Долгота гласной фонемы в чеченском языке фонематична, к примеру /бажа/ – [baaja] свояк, и /бажа/- [baja]- стадо; и таких примеров в чеченском языке очень много. Поэтому было выделено 4 группы гласных по краткости/долготе [13]:

1. долгие монофтонги [aa, ee, ii, uu, oo, iuu];
2. краткие монофтонги [a, e, i, o, u, ia, iu];
3. краткие дифтонги [ie, uo, oa, ay, ey, ow, iow, io, uy, oy];
4. долгие дифтонги [uoo, iee, ioo].

Согласные звуки по способу и месту образования собраны в следующие группы:

1. взрывные [b, d, p, g, l, pl, tl, kl, p, t, k, @];
2. спيرانты [v, z, j, gl, h, f, s, sh, x, hx, ll];
3. аффрикаты [dj, dz, c, ch, cl, ql, qq];
4. сонорные [n, l, r, y, w].

СЕГМЕНТАЦИЯ И ТРАНСКРИПЦИЯ

Для проектируемых систем автоматического распознавания и синтеза чеченской речи необходимо создать фонетические транскрипции слов, используя обозначения фонем из алфавита AZBAT. При разработке новой системы транскрипции могут быть созданы вручную, но этот процесс является трудоемким, поэтому планируется генерировать транскрипции автоматически. Для системы синтеза чеченской речи транскрипции будут создаваться по входному тексту, учитывая фонетические явления, происходящие на стыках слов.

В процессе создания и заполнения текстовой базы применяется метод записи целевых бессмысленных слов – «псевдослов», описанный в работах [14-16]. Псевдослова произносятся монотонно, с приемлемой продолжительностью и просодией, перебирая все возможные комбинации дифонов и аллофонов. Преимуществами этого метода является то, что не нужно искать природные примеры слов, в которых присутствует нужный дифон, и список элементов легче проверяется, а объём созданной речевой БД будет минималь-

ным. Данный подход имеет и недостатки, заключающиеся в трудности произношения псевдослов для диктора, и непредсказуемости просодических характеристик записанной речи. Тем не менее, выбор был сделан в пользу этого метода, в целях экономии времени и минимизации самой базы данных.

Следующим моментом подготовки акустической базы является автоматическая сегментация записанных псевдослов. Существует два основных типа автоматической сегментации речи. К первому типу относят сегментацию речи при условии, что известна последовательность фонем данной фразы. Другой тип определяет границы сегментов по степени изменения акустических характеристик сигнала. При автоматической сегментации желательно использовать только общие характеристики речевого сигнала, поскольку обычно на этом этапе нет конкретной информации о содержании речевой фразы. Для подготовки речевых фрагментов и дальнейшей сегментации речевого сигнала дифоны распределены по следующим классам:

1. гласный + согласный (ам, ар, аг, ogI, акх, уьс);
2. согласный + гласный (ма, ра, plа, glu, луь, кхо);
3. гласный + гласный (аа, ао, уо, иа);
4. согласный+ согласный (ст, хч, рп, шк).

При разметке необходимо выбирать дифон из среднего слога псевдослова, чтобы минимизировать артикуляционные эффекты в начале и в конце слов. Дифонная база содержит текстовое описание дифонов, а также их оцифрованные звуковые фрагменты, записанные женским голосом. Звуковой фрагмент каждого дифона записан в отдельном wav-файле, наименование которого совпадает с кодом дифона. При создании базы дифонов были выбраны следующие параметры квантования сигнала: частота дискретизации – 22025 Гц, разрядность выборки – 16 бит.

Таким образом, был создан фонетический алфавит AZBAT и определены основные принципы создания фонетико-акустической базы данных для организации системы синтеза чеченской речи.

СПИСОК ЛИТЕРАТУРЫ

1. Кривнова О. Ф. Области применения речевых корпусов и опыт их разработки / О. Ф. Кривнова // XVIII Сессия Российского акустического общества. Сборник трудов. – Т. 3 – Изд-во ГЕОС Москва, 2006. – С. 81–84.

2. DARPA TIMIT acoustic phonetic continuous speech corpus [Электронный ресурс]. – Режим доступа: <https://catalog.ldc.upenn.edu/ldc93s1>. – (Дата обращения: 17.04.2017).

3. База речевых фрагментов русского языка ISABASE / Д. С. Богданов, О. Ф. Кривнова, А. Я. Подрабинович В. В. Фарсобина // Интеллектуальные технологии ввода и обработки информации. – Эдиторил УРСС Москва, 1998. – С. 74–85.

4. Кривнова О. Ф. Русский речевой корпус RuSpeech / О. Ф. Кривнова // Фонетика сегодня. Материалы докладов и сообщений VII международной конференции – Институт русского языка им. В. В. Виноградова РАН Москва, 2013. — С. 54–56.

5. Цирульник Л. И. Автоматизированная система клонирования фонетико-акустических характеристик речи / Л. И. Цирульник // Информатика. № 1(9). – Мн., 2006. – С. 37–46.

6. Ибрагимов Т. И. К формированию акустической базы синтезатора татарской речи / Т. И. Ибрагимов, Р. Р. Хусаинов // Доклады международной конференции Диалог – М.: Наука. 2004. – С. 237–239.

7. Хусаинов А. Ф. Программный комплекс для анализа речи (на примере распознавания фонем татарского языка) / А. Ф. Хусаинов // Доклады ТУСУР. – 2013. – № 3(29). – С. 129–133.

8. Хусаинов А. Ф. Система автоматического распознавания речи на татарском языке / А. Ф. Хусаинов, Д. Ш. Сулейманов // Программные продукты и системы. – 2013. – № 4. – С. 301–304.

Израилова Элиса Салаудиновна – старший научный сотрудник отдела прикладной семиотики Академии наук Чеченской республики, старший преподаватель кафедры информатики и вычислительной техники Грозненского гос. нефтяного технического ун-та г. Грозный. Тел.: +7(928)-780-55-69
E-mail: uelisa@yandex.ru

9. Дешериев Ю. Д. Современный чеченский литературный язык. Часть 1. Фонетика / Ю. Д. Дешериев – Чечено-Ингуш. науч.-исслед. ин-т истории, языка и литературы. – Грозный: Чечено-Ингуш. кн. изд-во, 1960. – 122 с.

10. Имнайшвили Д. С. Историко-сравнительный анализ фонетики нахских языков / Ред. А. С. Чикобава. – АН Грузинской ССР. Институт языкознания. – Тбилиси: «Мецниереба», 1977. – 300 с.

11. Израилова Э. С. Фонетический алфавит чеченского языка – как основа речевого корпуса (в печати).

12. Fisher W. M., Doddington G. R., Goudie-Marshall K. M. The DARPA Speech Recognition Research Database: Specifications and Status // Proceedings of DARPA Workshop on Speech Recognition. Feb. 1986. – P. 93–99.

13. Халидов А. И. Грамматика чеченского языка. Том 1. Введение в грамматику. Фонетика. Морфемика. Словообразование / А. И. Халидов, А. Д. Тимаев, М. Р. Овхадов. – Академия наук ЧР; Комплексный научно-исследовательский институт им. Х. И. Ибрагимова РАН. – Грозный : ФГУП ИПК «Грозненский рабочий», 2013. – 848 с.

14. Fitt S., Stephen I. Synthesis of regional english using a keyword lexicon // Speech Communication and Technology: proceedings of the 6-th European conference EUROSPEECH'99, Budapest, Hungary, 5-9 September 1999. – Budapest, 1999. – V. 2. – P. 823–826.

15. Isard S., Miller D. Diphone synthesis techniques // Speech Input/Output:technics and applications: proceedings of IEE International Conference, London, UK, May 2-8, 1986. – London, 1986. – P. 77–82.

16. Lenzo K. A., Black A. Diphone collection and synthesis // Spoken Language Processing: proceedings of 6-th International conference ICSLP'2000, Beijing, China, 16-20 October 2000. – Beijing, 2000. – V. 3. – P. 306–309.

Izrailova E. S. – Senior researcher, Department of Applied Semiotics, Academy of Sciences of the Chechen Republic, Senior lecturer, Department of Informatics and Computer Science, Grozny State Technical Oil University. Tel.: +7(928)-780-55-69
E-mail: uelisa@yandex.ru