УДК 004.934.8'1

***Имамвердиев Я.Н.*[1]*, Сухостат Л.В.*[2]**
Институт Информационных Технологий НАНА, Азербайджан, Баку
[1]yadigar@lan.ab.az, [2]lsuhostat@hotmail.com

AZ-SRDAT – РЕЧЕВАЯ БАЗА ДАННЫХ ДЛЯ АЗЕРБАЙДЖАНСКОГО ЯЗЫКА

В этой статье описывается собранная речевая база данных для распознавания диктора AZ-SRDat (AZerbaijani Language Speaker Recognition DATa). База данных содержит речевые высказывания, полученные от 86 дикторов, говорящих на азербайджанском языке, за две сессии. Основной целью AZ-SRDat является предоставление данных для оценки различных методов распознавания диктора.

Ключевые слова: *речевая база данных, речевая база данных для азербайджанского языка, распознавание диктора, международный фонетический алфавит.*

Введение

В настоящее время задача создания больших, разнообразных и информационно богатых речевых корпусов становится все более актуальной как для компьютерных приложений, так и для фундаментальных научных исследований [1, 2]. Современные системы распознавания речи базируются преимущественно на методах статистического моделирования речевых и языковых явлений и требуют обучения на больших массивах аннотированной звучащей речи.

При разработке и оценке системы верификации диктора стандартная база данных играет важную роль [2, 3]. Она облегчает сравнение результатов различных методов, что позволяет нам узнать, какой из них является наиболее перспективным для исследования. Она также помогает в определении производительности систем при определенных экспериментальных условиях и основных вопросов, которые требуют дальнейшего изучения.

До сих пор в основном английский язык был центром исследований распознавания диктора. Гораздо меньше работ было сделано по другим национальным языкам, в том числе по азербайджанскому языку [4, 5].

Мы намерены восполнить этот пробел путем создания речевой базы данных. Она может быть использована для проведения различных исследований, касающихся речи. Весь корпус поможет в исследованиях автоматического распознавания диктора в направлении построения таких систем.

В Институте Информационных Технологий Национальной Академии Наук Азербайджана на протяжении нескольких лет проводятся исследования в области распознавания личности по голосу [6–8]. В результате накопленного опыта был разработан прототип системы распознавания диктора [9] и собрана речевая база данных для азербайджанского языка AZ-SRDat (AZerbaijani language Speaker Recognition DATa). Эта база данных создана с целью поддержки и оценки автоматических систем распознавания диктора. Она включает набор данных, полученных от 86 дикторов за две сессии в офисных условиях.

AZ-SRDat рассматривается как дополнение к стандартным базам данных. Этот дополнительный корпус может помочь исследователям оценить технологии распознавания речи и диктора.

Обзор речевых баз данных

Многие речевые корпуса для английского и других национальных языков были разработаны и стали доступны благодаря Консорциуму лингвистических данных LDC

[10], Европейской ассоциации языковых ресурсов ELRA [11] и Агентству по оценке и распространению языковых ресурсов ELDA [12].

В настоящее время имеются следующие речевые базы данных: Британский национальный корпус [13], корпус для шведского языка [14], корпус для распознавания диктора для болгарского языка [15], SIVA, Polyvar, POLYCOST (для английского и 13 других европейских языков) [16], KING-92 [1], Switchboard I-II, AHUMADA [17] и др.

Все упомянутые выше речевые базы данных собирались в различных акустических средах(в основном в офисе). Запись производилась с разных типов устройств.

Все корпуса включают одновременно мужские и женские голоса. Тип речевого материала довольно разнообразен – от числовых комбинаций и прочитанных предложений до спонтанной речи.

Описание речевой базы данных для азербайджанского языка

Производительность автоматической системы верификации очень сильно зависит от речевой базы данных. Есть много факторов, которые влияют на производительность системы автоматического распознавания. К ним относятся условия записи, окружающая среда, устройства записи, длительность, пол диктора, возрастная группа и т.д. Не зная условий записи, бессмысленно ожидать хорошего результата системы автоматической верификации.

В нашем описании AZ-SRDat мы остановимся на следующих факторах:

Общая информация: корпус содержит речевые данные и был разработан для проведения экспериментов в области распознавания диктора. Он также подходит для распознавания речи, а также идентификации языка и акцента.

Субъекты записи: субъектами записи являются сотрудники Института Информационных Технологий НАНА. Все дикторы являются носителями языка. Содержит записи 86 дикторов (21 мужчина и 65 женщин), которым присвоены идентификаторы от 1001 до 1086. Корпус, в основном, включает людей среднего возраста. Таким образом, группа дикторов имеет сравнительно небольшие изменения в возрасте, профессии и образовании. Большинство дикторов было записано в течение месяца.

Речевой материал: все данные включают изолированные цифры, изолированные слова, комбинации цифр и текстовый фрагмент. Они неизменны от сессии к сессии. Вторая сессия была записана после перерыва, длившегося около двух месяцев.

Каждое высказывание сохраняется в отдельном файле в формате WAV. Записи имеют частоту дискретизации 11025 Гц при разрешении 16 бит. Никакой дополнительной цифровой обработки к речевым файлам не применяется.

Общий объем данных для каждого диктора составляет приблизительно 2500 Кб. Средняя продолжительность речи для каждого говорящего составляет около 100 сек.

Устройства записи: запись производилась с помощью программного обеспечения Cool Edit Pro в офисных условиях (рис. 1). При этом окна и двери были закрыты, чтобы избежать любого внешнего шума. Для записи были использованы наушники с шумоподавляющим микрофоном A4Tech HS-800 с частотным диапазоном 20 Гц – 20 КГц.

Для большей эффективности мы выбрали фонетически богатые слова, в которых согласные доминируют над гласными. База включает прочитанный текст, состоящий из 124 слов в 6 предложениях. Его средняя продолжительность составляет около 60 секунд.

В приведенном ниже фрагменте текста мы попытались показать произношение текста на азербайджанском языке, используя символы из Международного фонетического алфавита (МФА) [18]. А также указали точное произношение фонем азербайджанского языка.

Рис.1. Снимок процесса записи

[Xoʤɑ'lɯ] [sojgɯrɯ'mɯ] [min] [dok'guz] [jyz] [dox'sɑn] [ikin'ʤi] [il] [fevrɑ'lɯn] [ijir'mi] [bɛʃin'næn] [ijir'mi] [ɑltɯsɯ'nɑ] [ke'ʧæn] [gʲɛ'ʤæ] [Ɛrmænis'tɑn] [Silɑh'lɯ] [Gʲyvvælæ'ri] [tæræfin'næn] [Rusijɑ'nɯn] [yʧ] [jyz] [ɑlt'mɯʃ] [ɑltmɯn'ʤɯ] [motoɑtɯ'ʤɯ] [ɑlɑjɯ'nɯn] [iʃtirɑ'kɯ] [ilæ] [Xoʤɑ'lɯ] [ʃæhæri'nin] [sɑkinlæri'næ] – [ɛt'nik] [ɑ:zærbɑjʤɑnlɯlɑ'rɑ] [gɑr'ʃɯ] [tørædil'miʃdir]. [Xoʤɑ'lɯ] [fɑʤiæ'si] [ijirmin'ʤi] [æs'rin] [Bɑbi'jar], [Xɑ'tɯn], ['Liditsɛ], ['Sonmɯ] [ki'mi] [æn] [dæhʃæt'li] [væ] [gʲæd'dɑr] [fɑʤiælærin'dæn] [bi'ridir]...

[Xoʤɑ'lɯ] [sojgɯrɯmɯ'nɯ] [u'nutmɑjɯn]!

Структура речевых файлов

Всем речевым файлам были присвоены имена с уникальным идентификационным кодом, которые содержат поля, соответствующие идентификатору диктора, сессии и полу диктора.

Детали идентификационного кода даны ниже:

1) Имя файла содержит 9 символов, где первые 4 цифры представляют уникальный идентификатор диктора, а оставшиеся 4 символа, следующие за символом подчеркивания (_), представляют переменные, характеризующие каждое высказывание.
2) Последовательность присваивания имен:
   <ID диктора><символ подчеркивания><пол диктора><ID высказывания><номер сессии>.
   a. ID диктора – это уникальный идентификатор, представляющий субъекта.
   b. Пол диктора. Для мужчин начинается с «m», а для женщин – с «f».
   c. ID высказывания описывают каждый файл с помощью нескольких

параметров.

d. Номер сессии. Это поле показывает сессию записи. Цифра «1» представляет первую сессию, а «2» – вторую.

Например: 1001_faa2.wav означает, что диктор – женщина с ID 1001, вторая сессия, было произнесено слово «Azərbaycan» (Азербайджан), запись сделана в офисных условиях.

Структура файла на диске «ID_диктора/номер_сессии/имя_высказывания.wav». Например, для указанного выше файла имя его пути следующее: «1001/2/1001_faa2.wav».

Мы не делим наши речевые данные на две группы – для регистрации и распознавания, как это делается в других базах данных [19].

Мы оставляем право исследователю решать, какая сессия будет использоваться для регистрации, и какая с целью распознавания. Для того чтобы помочь в поиске речевых данных имеются так называемые Таблицы описания высказываний. Каждый диктор имеет собственную таблицу для всех его высказываний.

Если конкретный диктор имеет две сессии, то для него создаются две таблицы.

Пример такой таблицы для диктора с ID = 1001 показан в таблице 1.

Таблица 1

Таблицы описания высказываний для диктора с ID=1001

ID диктора=1001			
№	Имя файла	№ сессии	Тип высказывания
1	1001_faa1.wav	1/2	число
…	…	…	…
12	1001_fal1.wav	1/2	комбинация чисел
…	…	…	…
20	1001_fap1.wav	1/2	текстовый фрагмент

На рис. 2 показаны форма волны и узкополосная спектрограмма для одного WAV файла из первой сессии для слова «Azərbaycan».

Заключение

AZ-SRDat – речевой корпус для азербайджанского языка, записанный в офисных условиях. Он может быть использован в основном для верификации и идентификации фиксированного текста, а также текстонезависимой идентификации дикторов. Релиз этого корпуса включает только речевые файлы. Для его улучшения нужно увеличить количество дикторов и дополнить речевой материал. Кроме того, транскрипции речевых файлов должны быть включены в следующую версию базы данных. Учитывая последние достижения в области верификации диктора [20, 21], требуется добавить данные, записанные в различных условиях.

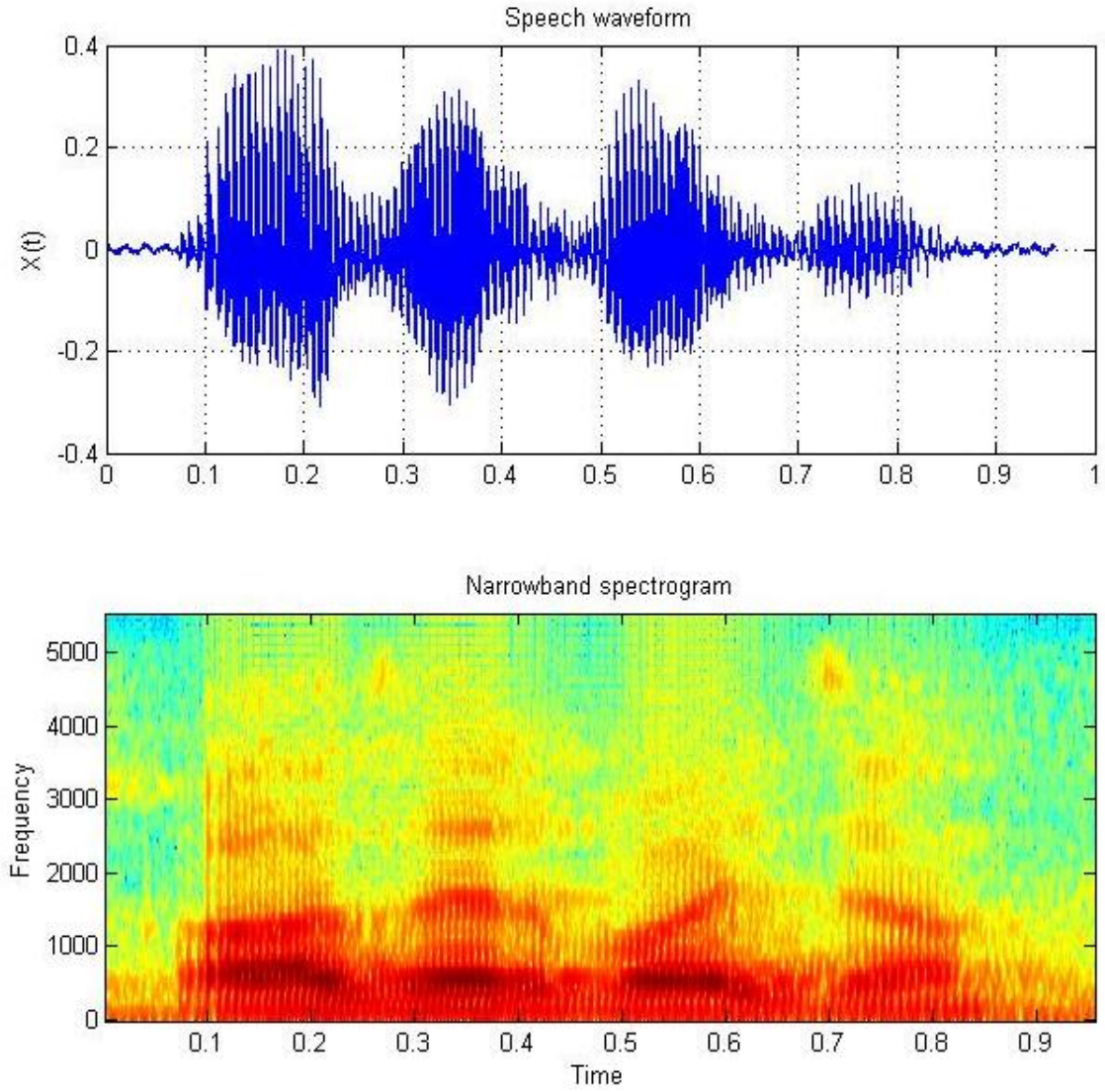

Рис.2. Звуковой файл из первой сессии

Благодарность

Авторы хотели бы выразить благодарность всем сотрудникам Института Информационных Технологий НАНА за участие в сборе речевых образцов для азербайджанского языка. А также особую благодарность выражаем А.Ганбарову за большую работу, проделанную при создании речевой базы данных.

Литература

1. Doddington G. Speaker recognition-identifying people by their voices // Proc. IEEE, 1985, vol. 73, no. 11, pp. 1651–166.
2. Reynolds D.A. An overview of automatic speaker recognition technology// Proc. IEEEInternational Conference on Acoustics, Speech, and Signal Processing, 2002, vol. 4, pp. 4072–4075.
3. Campbell J.P., Reynolds D.A. Corpora for the evaluation of speaker recognition systems // Proc. IEEEInternational Conference on Acoustics, Speech, and Signal Processing, 1999.
4. Abbasov A., Fatullayev R., Fatullayev A. HMM-based large vocablary continuous speech recognition system for Azerbaijani// Proc. of PCI-2010, 2010, vol. 1, pp. 23–26.

5. Imamverdiyev Y.N., Sukhostat L.V. SVM based recognition of Azerbaijani vowels // 5th Int-l Conf. on Application of Information and Communication Technologies (AICT), 12–14 Oct. 2011, Baku.
6. Имамвердиев Я.Н., Сухостат Л.В. Речевые базы данных для систем распознавания диктора // Вопросы защиты информации, 2011, № 4, с. 27–32.
7. Сухостат Л.В. Разработка методов и алгоритмов для синтеза систем биометрической идентификации личности по голосу, Науч. семинар, 30 ноября 2012, Баку, с. 29–30.
8. Имамвердиев Я.Н., Сухостат Л.В. Об одном методе извлечения признаков для систем распознавания диктора // İnformasiya texnologiyaları problemləri, 2012, №2, pp. 14–19.
9. Сухостат Л.В. Разработка прототипа системы распознавания личности по голосу //Azərbaycan xalqının ümummilli lideri Heydər Əliyevin 90 illik yubileyinə həsr olunmuş "İnformasiya təhlükəsizliyi problemləri üzrə I respublika elmi-praktiki konfransı, 2013, c. 151–154.
10. LDC, Lingustic Data Consortium. Сайт: http://www.ldc.upenn.edu/
11. ELRA, European Language Resource Association. Сайт: http://www.elra.info/
12. ELDA, Evaluations and Language resources Distribution Agency. Сайт: http://www.elda.org/
13. British National Corpus, http://www.natcorp.ox.ac.uk/
14. Allwood J., Bjornberg M., Gronqvist L., Ahlsen E. and Ottesjo C. Spoken Language Corpus at the Department of Linguistics // Forum: Qualitative Social Research, Goteborg University, 2000, vol. 1, no. 3.
15. Ouzounov A. BG-SRDat: A Corpus in Bulgarian Language for Speaker Recognition over Telephone Channels // Cybernetics and Information Technologies, 2003, vol.3, no.2, pp.101–108.
16. Melin H. Databases for Speaker Recognition: Activities in COST250 Working Group 2, COST 250 - Speaker Recognition in Telephony, Final Report 1999, European Commission DG-XIII, Brussels, August 2000.
17. Ortega-Garsia J., Gonzalez-Rodriguez J., Marrero-Aguiar V. AHUMADA: A large speech corpus in Spanish for speaker characterization and identification // Speech Communication, 2000, vol. 31, pp. 255–264.
18. Handbook, IPA: Handbook of the International Phonetic Association, Cambridge University Press.1999, 214 p.
19. Barlow M., Booth L. and Parr A. The Collection of Two Speaker Recognition Targeted Speech Databases // Proc. 4th Aust. Int. Conf. Speech Science and Technology, 1992, pp. 706–711.
20. Yin S.-C., Rose R., Kenny P. A joint factor analysis approach to progressive model adaptation in text-independent speaker verification // IEEE Transactions on Audio, Speech, and Language Processing, 2007, vol. 15, no. 7, pp. 1999–2010.
21. Dehak N., Kenny P., Dehak R., Dumouchel P., Ouellet P. Front-end factor analysis for speaker verification // IEEE Transactions on Audio, Speech, and Language Processing, 2011, vol. 19, no. 4, pp. 788–798.

UOT 004.934.8'1

İmamverdiyev Yadigar N.[1], Suxostat Lyudmila V.[2]
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
yadigar@lan.ab.az[1], lsuhostat@hotmail.com[2]

AZ-SRDat – Azərbaycan dili səs nümunələrin bazasıdır

Bu məqalədə səsə görə şəxsin tanınması üçün toplanmış AZ-SRDat (AZerbaijani language Speaker Recognition DATa) səs nümunələri bazası təsvir edilir. Bu bazada Azərbaycan dilində danışan 86 şəxsdən iki sessiya ərzində alınmış səs nümunələri toplanmışdır. AZ-SRDat-ın əsas məqsədi səsə görə şəxsin tanınması metodlarının qiymətləndirilməsi üçün zəruri olan verilənlərin təqdim edilməsidir.

Açar sözlər: *səs nümunələri bazası, Azərbaycan dili üçün səs nümunələri bazası, səsə görə şəxsin tanınması, Beynəlxalq Fonetik Əlifba.*

Yadigar N.Imamverdiyev[1], Lyudmila V.Sukhostat[2]
Institute of Information Technology of ANAS, Baku, Azerbaijan
yadigar@lan.ab.az[1], lsuhostat@hotmail.com[2]

AZ-SRDat – a speech database for Azerbaijani language

This paper describes AZ-SRDat (AZerbaijani language Speaker Recognition DATa), a speech database for speaker recognition. The database contains speech utterances produced by 86 speakers in Azerbaijani language in two sessions. The main purpose of AZ-SRDat is to provide data for evaluation of different methods for speaker recognition.

Keywords: *speech database, speech database for Azerbaijani language, speaker recognition, International Phonetic Alphabet.*