Technical report about King Saud University Database

"KSU-Speech Database"

A versatile component in the research in Arabic speech processing is the availability of a public speech database; hence in a project funded by King Abdulaziz City for Science and Technology (KACST), facilitated by the National Plan for Science and Technology (NPST) in King Saud University, we built a very rich database speech database and made it available to the researchers worldwide. Though the major use of the database in the project was for Arabic speaker recognition applications, but we designed it to be also beneficial for other speech applications such as speech recognition, computer aided pronunciation training and race/nationality recognition.

The database is rich in different aspects: (a) it has more than 200 speakers of both genders; (b) the speakers are from different races, for example, Saudis, Arabs, and non-Arabs; (c) utterances are both read text and semi- spontaneous (answers to question); (d) scripts are of different dimensions, such as, isolated words, digits, phonetically rich words, sentences, phonetically balanced sentences, paragraphs, etc.; (e) different sets of microphones with high and medium quality; (f) telephony and non-telephony speech; (g) three different recording environments – office, soundproof room, and cafeteria; (h) three different sessions, where the sessions are separated by at least three weeks. The database was carefully manually verified, where the manual verification was complemented with automatic verification. We named the database King Saud University speech database.

The KSU database is very rich and can be used in many researches in speech processing research as we mentioned before. Hence, we were able to use it in many research areas of speech processing such as:

- Computer aided pronunciation training
- Gender Effect in Trait Recognition
- Automatic Identification of Arabic L2 Learners Origin
- Voice based Gender Classification
- Voice and Unvoiced Classification
- Comparison of Voice Features for Arabic Speech Recognition
- Noisy channel and cross channel speaker recognition systems
- Effect of Arabic text on the speaker recognition system.

To encourage researchers to work on different applications of Arabic speech processing and advance the research in computerization of Arabic language we made the database open to the research community by putting it with resources of the Language Data Consortium (LDC), as shown in Figure 1, which licenses it for commercial and non-commercial use. We share it with LDC since Feb / 2014.

Linguistic Data Consortium

King Saud University Arabic Speech Database

Item Name:	King Saud University Arabic Speech Database
Author(s):	Mansour Alsulaiman, Ghulam Muhammad, Bencherif Mohamed Abdelkader, Awais Mahmood, Zulfiqar Ali
LDC Catalog No.:	LDC2014S02
ISBN:	1-58563-669-X
ISLRN:	789-673-729-277-5
DOI:	https://doi.org/10.35111/vpqe-bz17 (https://doi.org/10.35111/vpqe-bz17)
Release Date:	February 17, 2014
Member Year(s):	2014
DCMI Type(s):	Sound
Sample Type:	pcm
Sample Rate:	48000
Data Source(s):	microphone speech
Application(s):	speech recognition, speaker identification
Language(s):	Arabic
Language ID(s):	ara
License(s):	King Saud University Arabic Speech Database (/license/ksu-arabic-speech-database.pdf)
Online Documentation:	LDC2014S02 Documents (/docs/LDC2014S02/)
Licensing Instructions:	Subscription & Standard Members, and Non-Members (http://www.ldc.upenn.edu/language-resources/data/obtaining)
Citation:	Alsulaiman, Mansour, et al. King Saud University Arabic Speech Database LDC2014S02. Hard Drive. Philadelphia: Linguistic Data Consortium, 2014.
Related Works:	View

Introduction

King Saud University Arabic Speech Database was developed by Speech Group (SG) at King Saud University (http://ksu.edu.sa/en/) and contains 590 hours of recorded Arabic speech from 269 male and female speakers. The utterances include read and spontaneous speech. The recordings were conducted in varied environments representing quiet and noisy settings.

Data

The corpus was designed principally for speaker recognition research. However, other possible applications include first language recognition, mobile effect, multichannel effect, and use of different type of microphones. The speech sources are word lists, sentence lists, paragraphs and question and answer sessions. Read speech text includes the following:

- Sets of sentences devised to cover allophones of each phoneme, phonetic balance, and differentiation of accents.
- Word lists developed to minimize missing phonemes and to represent nasals fricatives, commonly used words, and numbers.
- Two paragraphs selected because they included all letters of the alphabet and were easy to read.

Spontaneous speech was captured through question and answer sessions where speakers answer questions displayed on screen. The questions were on general topics such as the weather and food and included the speaker name or number.

The speakers were Saudis and non-Saudis. Among the non-Saudi participants were Arabs and non-Arabs. All female speakers were either Saudis or non-Saudi Arabs. Male speakers included non-Arabs from the Indian subcontinent, Africa, South East Asia and East Europe. Non-Arab participants were required to be able to read Arabic at an acceptable level. Most of the Non-Arab speakers were from the fourth level in the Arabic Linguistics Institute (http://ali.ksu.edu.sa/en) at King Saud University. The non-Saudi participants represented 28 nationalities and were chosen from clusters of areas or countries.

Each speaker was recorded in three different environments: in a soundproof room , in an office and in a cafeteria. The recordings were collected via different microphones and a mobile phone and averaged between 16-19 minutes. The recordings were done in three sessions with a time-gap of an approximately 6 weeks.

The data was verified for missing recordings, problems with the recording system or errors in the recording process. All files are presented as two channel 48 kHz 16-bit FLAC compressed PCM wav files. Note that sizes and file names in the documentation are for the uncompressed wav files.

Samples

Please view this male sample (desc/addenda/LDC2014S02.m.wav) and female sample (desc/addenda/LDC2014S02.f.wav).

Updates

None at this time.

Copyright

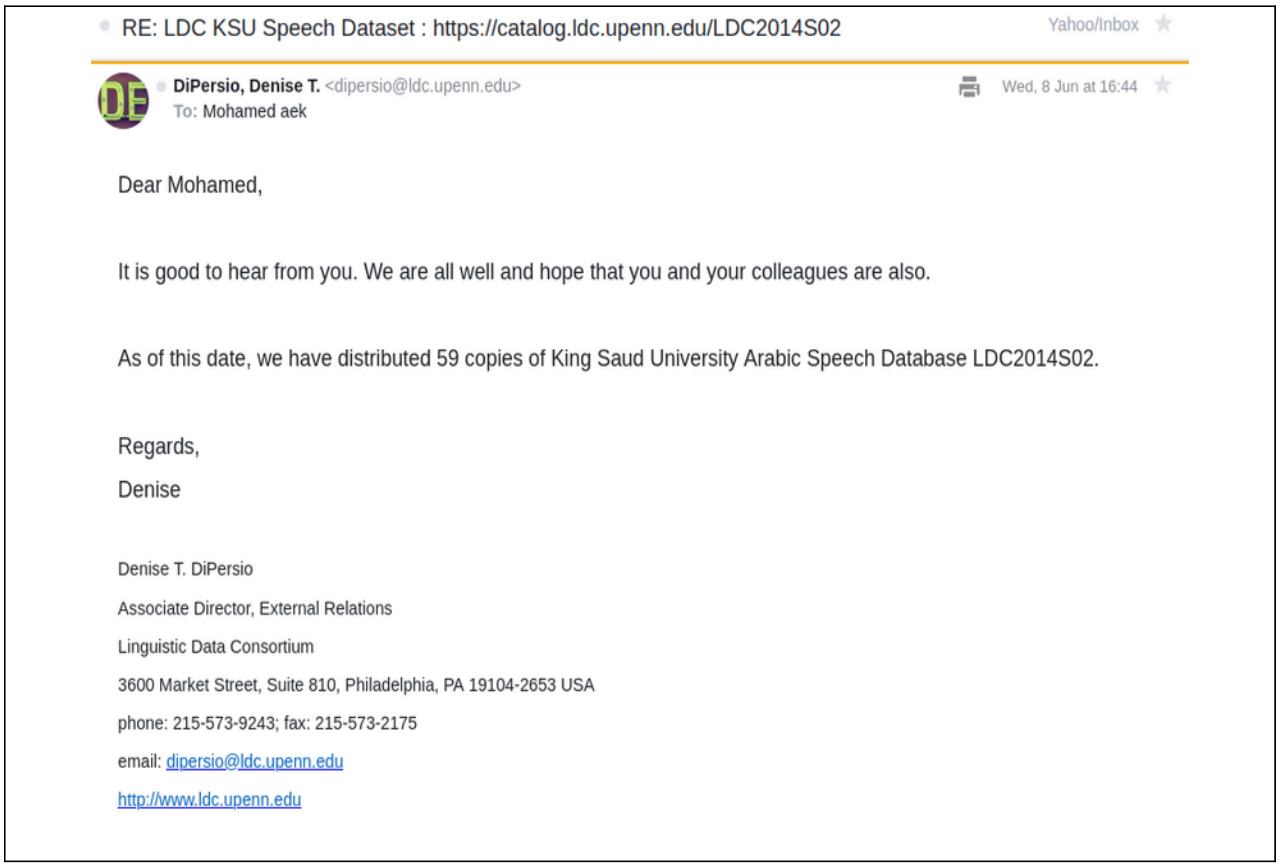
Available Media

◉ Web Download

Figure 1 : KSU Speech database @LDC

Website : https://catalog.ldc.upenn.edu/LDC2014S02

Due to the richness of the database, it was licensed to many researchers worldwide. According to LDC latest email to us, in Figure 2, LDC distributed 59 copies of King Saud University Arabic Speech Database LDC2014S02, up to the 8th June 2022.



RE: LDC KSU Speech Dataset : https://catalog.ldc.upenn.edu/LDC2014S02 Yahoo/Inbox

DiPersio, Denise T. <dipersio@ldc.upenn.edu>
To: Mohamed aek
Wed, 8 Jun at 16:44

Dear Mohamed,

It is good to hear from you. We are all well and hope that you and your colleagues are also.

As of this date, we have distributed 59 copies of King Saud University Arabic Speech Database LDC2014S02.

Regards,

Denise

Denise T. DiPersio

Associate Director, External Relations

Linguistic Data Consortium

3600 Market Street, Suite 810, Philadelphia, PA 19104-2653 USA

phone: 215-573-9243; fax: 215-573-2175

email: dipersio@ldc.upenn.edu

http://www.ldc.upenn.edu

Figure 2: Message from LDC (Wed 8 June 2022)

In addition to licensing the database by LDC, we also made the database available free of charge to many Arabic researchers who contact us directly. We only ask that they sign a license for non-commercial use.

Moreover, due to the richness and versatility of the database Samsung Jordan approached us to license it to them for commercial use. We took the permission of the project supporter and licensed it to Samsung, as shown in Figure 3.

26 تقارير

رسالة الجامعة

العدد ١٢٣٩ - الأحد ٢٤ رجب ١٤٣٧هـ الموافق ١ مايو ٢٠١٦

أنشأتها مجموعة «معالجة الكلام» وسجلتها باتحاد البيانات اللغوية

قاعدة «بيانات المتحدث العربي» تجذب شركات عالمية للاستفادة من مميزاتها

أنشئت قاعدة البيانات الصوتية ضمن مشروع مدعوم من الخطة الوطنية للعلوم والتقنية والابتكار

طورت مجموعة «معالجة الكلام» بكلية علوم الحاسب والمعلومات، بدعم من الخطة الوطنية للعلوم والتقنية، قاعدة بيانات غنية للمتحدث العربي، تحتوي على تسجيل ما يقارب 590 ساعة من الكلام لحوالي 362 من الذكور والإناث، وتتميز بغناها من حيث تنوع النصوص وأماكن وطرق التسجيل وجنسيات المشاركين، حيث شارك في القاعدة متطوعون من عدة جنسيات مختلفة، سعوديون وعرب «من 8 دول» وغير عرب «من 20 دولة»، وبعد تسجيل القاعدة ومراجعتها والتحقق منها والتأكد من سلامتها من الأخطاء، تم وضعها في اتحاد البيانات اللغوية «Linguistic Data Consortium – LDC» وإطلاق اسم «قاعدة بيانات جامعة الملك سعود» عليها، وذلك في عام 2014.

اتحاد البيانات

يعد تسجيل قاعدة بيانات المتحدث العربي الصوتية في اتحاد البيانات اللغوية «Linguistic Data Consortium – LDC» إنجازاً واعترافاً بهذه القاعدة، واتحاد البيانات اللغوية هو اتحاد مفتوح من الجامعات والمكتبات والشركات والمختبرات البحثية الحكومية تأسس عام 1992 ويضم 737 قاعدة بيانات صوتية منها 91 قاعدة بيانات صوتية عربية، وتم خلال السنوات الثلاث الأخيرة استخدامها من أكثر من 36 جهة بحثية عالمية لإجراء أبحاث متعلقة باللغة العربية.

توقيع العقد

بعد اطلاع شركة سامسونج على قاعدة البيانات الصوتية للمتحدث العربي التابعة لقاعدة بيانات جامعة الملك سعود، أعجبت بها وبغنى محتواها، واتصلت بمجموعة «معالجة الكلام» البحثية للحصول على حق استخدام هذه القاعدة في أبحاثها ومنتجاتها، ومن ثم وقعت شركة سامسونج ومجموعة معالجة الكلام عقداً يتيح للشركة استخدام قاعدة بيانات جامعة الملك سعود في إجراء أبحاث على اللغة العربية ومن ثم تطوير مُنتجات تخدم المجتمع العربي، ومن المتوقع أن يكون لهذا التعاون دور إيجابي كبير في تطوير تطبيقات تعتمد على معالجة اللغة العربية يستفيد منها المواطن العربي وتسهل إجراء كثير من العمليات وعلى الأخص في الأجهزة الذكية من خلال التعامل باللغة العربية.

مساهمة فاعلة

عميد كلية علوم الحاسب والمعلومات، د. منصور الزعير، عبر عن اعتزازه بهذا الإنجاز الذي حققته مجموعة «معالجة الكلام» في إطار المساهمة الفاعلة في إرساء مجتمع المعرفة وتحقيق الأهداف الوطنية للتنمية، وعد توقيع هذا العقد يأتي في إطار تحقيق أهداف الكلية وتعزيز دورها وتأثيرها في المجتمع والمساهمة في نقله إلى المستوى العصري للمعرفة.

وأكد وكيل كلية علوم الحاسب والمعلومات للدراسات العليا والبحث العلمي ورئيس مجموعة «معالجة الكلام» د. منصور بن محمد السليمان، أن توقيع هذه الشراكة مع شركة سامسونج يأتي في إطار المساهمة في إبتكار وإجراء البحوث الرائدة في مجال معالجة اللغات وعلى رأسها اللغة العربية مما يساهم في تطوير مُنتجات تساهم في ترقية الحياة اليومية للمجتمع العربي، وعبر عن شكره للدعم الذي توليه حكومة خادم الحرمين الشريفين - حفظه الله - في مواكبة التطورات العالمية في مجال العلوم والتقنية ممثلة ببرنامج الخطة الوطنية للعلوم والتقنية والابتكار كما عبر عن شكره للمشرف العام على الخطة د. أحمد بن عبدالله الخازم على دعمه وجهوده المتميزة.

د. السليمان

مميزات متقدمة

وأعرب الدكتور غلام محمد باحث مشارك في المشروع عن سعادته للمشاركة في هذا المشروع الذي يخدم أيضاً غير المتحدثين باللغة العربية بسبب غنى هذه القاعدة، وأوضح شريف الباحث في المشروع والمشرف على تسجيل القاعدة، أن نصوص التسجيل تنوعت بين جمل وكلمات منتقاة غنية صوتياً وفقرات كلامية، منها آيات قرآنية وكلمات شائعة وإجابات على أسئلة وأرقام عربية، وتم تسجيل قاعدة الأصوات في ثلاثة أماكن مختلفة «مكتب، غُرفة معزولة صوتياً، معلم» خلال ثلاث فترات زمنية مختلفة وكان بين كل فترة وأخرى ما لا يقل عن شهرين.

وبين أن قاعدة البيانات الصوتية تحتوي على العديد من المميزات منها أنه تم تسجيل ما يقارب من 362 متحدثاً من طلاب وطالبات جامعة الملك سعود، منهم 266 طالباً و203 عرب و 63 غير عرب»، 96 طالبة من عدة جنسيات عربية، ولضمان موثوقية هذه القاعدة تم مراجعتها من قبل عديد من الأشخاص حيث كان يتم الاستماع للتسجيل ومن ثم يتم مطابقته مع النصوص المراد تسجيلها.

د. الزعير

تسجيل الأصوات

شارك في تسجيل قاعدة البيانات العديد من طلاب وطالبات الدراسات العليا في كلية علوم الحاسب والمعلومات وكان لهم دور إيجابي وفعال في نجاح التسجيل والتحضير له، وأبدى الطلاب الذين تبرعوا بتسجيل كلامهم حماسهم وسعادتهم للمشاركة في هذا المشروع الذي يخدم لغة القرآن، حيث عبر طالب الدكتوراه محمد الجبيري عن سعادته بالمشاركة في هذا المشروع، إذ شارك في اختيار وتنظيم قوائم التسجيل ثم في عملية التسجيل والمراجعة والتدقيق.

كما أوضح طالب الدكتوراه محمد فيصل، أن عملية التسجيل تمت في ثلاثة أماكن مختلفة في درجة الضوضاء مما يتيح استخدامها في العديد من التطبيقات التي تخدم اللغة العربية.

وذكر طالب الماجستير حميد الغرباني، المسؤول الإداري عن المشروع وقوائم المتبرعين، أنه تمت دعوة الطلاب للمشاركة في المشروع عبر وضع إعلانات في كليات الجامعة والسكن الجامعي، كما تم عمل محاضرات قصيرة تعريفية في معهد اللغة العربية لغير الناطقين، وكان هنالك تجاوب كبير من الطلاب، كما تم عمل تقديم عروض تعريفية عن المشروع في العديد من المقررات الدراسية في كلية علوم الحاسب والمعلومات لتعريف الطلاب بالمشروع وأهدافه وحثهم على المشاركة في هذا المشروع الهام للغة العربية عن طريق التبرع بتسجيل كلامهم.

مرجع للباحثين

مثلت قاعدة البيانات الصوتية مرجعاً لمجموعة من الباحثين في دراساتهم وأبحاثهم، فقد استخدم د. عويس محمود، قاعدة البيانات في إجراء أبحاث رسالته للدكتوراه في قسم هندسة الحاسب، واقترح في رسالته خصائص صوتية جديدة للتعرف على المتحدث العربي بالتطبيق على قاعدة بيانات جامعة الملك سعود، وقد أعطت هذه الخصائص نتائج أفضل في التعرف على المتحدث العربي مقارنة بالخصائص الأخرى المعروفة، وبين د. عويس أن قاعدة البيانات كانت مُفيدة جداً لرسالته نظراً لغناها واحتوائها على نصوص متنوعة وبيئات وقنوات مختلفة، كما ذكر أنه نظراً لاحتواء القاعدة على تسجيل فقرات تحتوي على جميع الحروف الهجائية تمت دراسة أي الحروف الصوتية أفضل للتعرف على المتحدث وتعد هذه الدراسة أول دراسة في هذا المجال.

وكذلك استخدمت الطالبة أفنان الهندي قاعدة البيانات في رسالتها للماجستير بقسم هندسة الحاسب، واهتم البحث بالكشف التلقائي عن الأخطاء في النطق لغير المتحدثين باللغة العربية «الباكستانيين والهنود» وتم الرجوع لقاعدة بيانات جامعة الملك سعود لمعرفة أكثر الأخطاء شيوعاً لدى المتحدثين غير العرب، وبالإضافة لذلك تم إجراء تقطيع وترميز لبعض أجزاء هذه القاعدة من المتحدثين العرب وغير العرب، وقام بعملية التقطيع والترميز متخصصون في اللغة العربية وتم ذلك آلياً ومن ثم روجع يدوياً، وقد أعطت نتائج الكشف التلقائي عن الأخطاء نسب صحة عالية، وذكرت الطالبة أفنان الهندي أن محتوى قاعدة البيانات كان غنياً بالأصوات واختلاف الجنسيات مما ساعدها في إتمام البحث.

شكر وتقدير

وجه وكيل كلية علوم الحاسب والمعلومات للدراسات العليا الدكتور منصور السليمان شكره وتقديره لجميع من دعم هذا المشروع وخص بالذكر الدكتور حسن مذكور عميد كلية علوم الحاسب والمعلومات السابق، لدعمه القوي والمستمر للمشروع، وكذلك الدكتورة ناديا الغريميل عميدة أقسام العلوم والدراسات العليا، والدكتوره هند الخليفة وكيلة كلية علوم الحاسب والمعلومات لشؤون الطالبات، لدعمها للمجموعة أثناء تسجيل كلام المتبرعات.

كما شكر فريق العمل في الجانب الطلابي م. خالد خواجي، م. فايع القحطاني، محمد الجبيري، محمد فيصل، حميد الغرباني، منير الحمادي، طه الفقيه، فضل دحان، ناصر الحسامي، مصطفى قمحان، وكذلك فريق العمل في جانب الطالبات أ. مشاعل الدويس، أ. أفنان الهندي، أ. نوف الغفر، أ. صيته السلامة، أروى العيسى، اعتصام الهزاع، ابتهال المسلم، شهد القحطاني، سارة السند، ريم الأحيدب، سميرة لرضي، بشاير الحسينان.

Figure 3: KSU Speech Database citation in the KSU Newspaper-(1st May 2016)

For details of the database please refer to the papers listed below and included in this report.

List of papers about KSU DB

1) Mansour Alsulaiman, Ghulam Muhammad, Mohamed A. Bencherif, Awais Mahmood, Zulfiqar Ali, " KSU Rich Arabic Speech Database", Journal of Information, vol. 16, no. 6(B), 2013.
2) Alsulaiman M., Ali Z., Muhammad G., Bencherif M. A., Mahmood A., "KSU Speech Database: Text Selection, Recording and Verification", Proceeding of 7th European Modelling Symposium on Mathematical Modelling and Computer Simulation, 20-22 Nov., 2013.
3) Mansour M. Alsulaiman, Ghulam Muhammad, Mohamed A. Bencherif, Awais Mahmood, Zulfiqar Ali, and Mohammad Aljabri, "Building a Rich Arabic Speech Database", Proceeding of the Fifth Asia International Conference on Mathematical Modeling and Computer Simulation (AMS '11), pp. 100-106, May 23, Malaysia.
4) Awais Mahmood,Mansour M. Alsulaiman, Ghulam Muhammad,and Mohamed A. Bencherif," Verification of A Rich Arabic Speech Database" Proceeding of COCOSDA 2011, October, Taiwan.

Abstracts of the papers

1) Mansour Alsulaiman, Ghulam Muhammad, Mohamed A. Bencherif, Awais Mahmood, Zulfiqar Ali, " KSU Rich Arabic Speech Database", Journal of Information, vol. 16, no. 6(B), 2013.

Abstract

Arabic is one of the major languages in the world. Unfortunately not so much research in Arabic speaker recognition has been done. One main reason for this lack of research is the unavailability of rich Arabic speech databases. In this paper, we present a rich and comprehensive Arabic speech database that we developed for the Arabic speaker / speech recognition research and/or applications.

The database is rich in different aspects: (a) it has 752 speakers; (b) the speakers are from different ethnic groups: Saudis, Arabs, and non-Arabs; (c) utterances are both read text and spontaneous; (d) scripts are of different dimensions, such as, isolated words, digits, phonetically rich words, sentences, phonetically balanced sentences, paragraphs, etc.; (e) different sets of microphones with medium and high quality; (f) telephony and non-telephony speech; (g) three different recording environments: office, sound proof room, and cafeteria; (h) three different sessions, where the recording sessions are scheduled at least with 2 weeks interval. Because of the richness of this database, it can be used in many Arabic, and non-Arabic, speech processing researches, such as speaker / speech recognition, speech analysis, accent identification, ethnic groups / nationality recognition, etc. The richness of the database

makes it a valuable resource for research in Arabic speech processing in particular and for research in speech processing in general. The database was carefully manually verified. The manual verification was complemented with automatic verification. Validation was performed on a subset of the database where the recognition rate reached 100% for Saudi speakers and 96% for non-Saudi speakers by using a system with 12 Mel frequency Cepstral coefficients, and 32 Gaussian mixtures.

2) Alsulaiman M., Ali Z., Muhammad G., Bencherif M. A., Mahmood A., "KSU Speech Database: Text Selection, Recording and Verification", Proceeding of 7th European Modelling Symposium on Mathematical Modelling and Computer Simulation, 20-22 Nov., 2013.

Abstract

King Saud University speech database (KSU-DB) is a very rich speech database of Arabic language. Its richness is in many dimensions. It has more than three hundred speakers of both genders. The speakers are Arabs and non-Arabs belonging to twenty-nine different nationalities. The database has different types of text such as isolated words, digits, phonetically rich words and sentences, phonetically balanced sentences, paragraphs, and answers to questions. The KSU-DB was recorded in three different locations; the first is an office that represents a normal environment with low noise. The second and third locations are cafeteria and soundproof room representing noisy and quiet environments, respectively. The database has different channels of recordings; mobile, medium and high quality microphones connected to recording devices of different qualities. To track the inter-session variations of the speakers, the database was recorded in three sessions with a gap of about six weeks. Though the database main goal is for speaker recognition research, nonetheless, we made it very rich so that it can be used in many speech-processing research. A team of native Arabs verified the database manually as well as automatically.

3) Mansour M. Alsulaiman, Ghulam Muhammad, Mohamed A. Bencherif, Awais Mahmood, Zulfiqar Ali, and Mohammad Aljabri, "Building a Rich Arabic Speech Database", Proceeding of the Fifth Asia International Conference on Mathematical Modeling and Computer Simulation (AMS '11), pp. 100-106, May 23, Malaysia.

Abstract

Availability of databases is a necessity in the speech processing field. The publicly available databases in Arabic language are few. In this paper we describe a rich database for Arabic language. The database is rich in many dimensions: in text, environments, microphone type, number of recording sessions, recording system, the transmission channel, the country of origin, and the mother language. This richness makes the database an important resource for research in Arabic Language processing and very useful in many speech processing tasks, such as speaker recognition, speech recognition, and accent identification. The speakers were speaking in Modern Standard Arabic (MSA).

4) Awais Mahmood,Mansour M. Alsulaiman, Ghulam Muhammad,and Mohamed A. Bencherif," Verification of A Rich Arabic Speech Database" Proceeding of COCOSDA 2011, October, Taiwan.

KSU Rich Arabic Speech Database

Mansour Alsulaiman, Ghulam Muhammad, Mohamed A. Bencherif, Awais Mahmood and Zulfiqar Ali

Speech Processing Lab, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
Email: {msuliman, ghulam, mbencherif, awais, zuali} @ksu.edu.sa

Abstract

Arabic is one of the major languages in the world. Unfortunately not so much research in Arabic speaker recognition has been done. One main reason for this lack of research is the unavailability of rich Arabic speech databases. In this paper, we present a rich and comprehensive Arabic speech database that we developed for the Arabic speaker / speech recognition research and/or applications. The database is rich in different aspects: (a) it has 257 speakers; (b) the speakers are from different ethnic groups: Saudis, Arabs, and non-Arabs; (c) utterances are both read text and spontaneous; (d) scripts are of different dimensions, such as, isolated words, digits, phonetically rich words, sentences, phonetically balanced sentences, paragraphs, etc.; (e) different sets of microphones with medium and high quality; (f) telephony and non-telephony speech; (g) three different recording environments: office, sound proof room, and cafeteria; (h) three different sessions, where the recording sessions are scheduled at least with 2 weeks interval. Because of the richness of this database, it can be used in many Arabic, and non-Arabic, speech processing researches, such as speaker / speech recognition, speech analysis, accent identification, ethnic groups / nationality recognition, etc. The richness of the database makes it a valuable resource for research in Arabic speech processing in particular and for research in speech processing in general. The database was carefully manually verified. The manual verification was complemented with automatic verification. Validation was performed on a subset of the database where the recognition rate reached 100% for Saudi speakers and 96% for non-Saudi speakers by using a system with 12 Mel frequency Cepstral coefficients, and 32 Gaussian mixtures.

Key Words: Speaker Recognition, Speech corpus, Arabic speech database, Rich database, Phonetically, Rich Database

1. Introduction

Arabic is one of the oldest and widely spoken Semitic languages. Some of its differences from other languages are unique phonemes and phonetic features, and a complicated morphological word structure. It has been reported in the literature that major difficulties in automatic speech processing of Modern Standard Arabic (MSA) are due to distinctive characteristics of the Arabic sound system, namely, emphatic, uvular, and pharyngeal consonants, and short and long vowels [1].

A speech database is an essential component in speech processing research and in developing speech processing systems. An automatic speech/speaker recognition system can be deployed successfully in real life only if it is developed using a versatile and relevant database. Without a proper speech database, speech processing related research cannot be progressed. There are many databases in major languages, like English, Spanish, German, Japanese, Chinese, etc. These databases are rich in terms of the number of speakers, amount of speech, variability of speakers and texts, environments, and transmission channels. However, Arabic speech databases are few in numbers and most of them are private. Therefore, there is a need for a publicly available comprehensive Arabic speech database. A rich and a publicly available database is an important and essential resource for research in the Arabic speech.

While developing a speech corpus, the following consideration may be taken into account: Scope of the corpora, Content, Phonological distribution, Number of speakers, Gender, Accents and/or Regional dialects, Speaking style, Environment, Recording materials, Sessions, Partition into training and testing data sets.

Our database was designed by taking care of all these considerations. We highlight these considerations in section 1.2. In section 2, we present the different richness dimensions of the database and give justification for each dimension. We also present the recording team, the volunteers, the text verification, and the pilot recording. Section 3 gives some details of the hardware and software of the system. In section 4, the main statistics of the database are given. In section 5, we proceed with the database verification methodology and we present the results of this verification. The validation of the database is discussed in section 6, and finally in section 7 we conclude the article.

1.1. Literature review

In [2], we presented some major databases in languages other than Arabic and we also did a survey on many of Arabic speech databases. Table 1 constitutes a summary of our survey of the Arabic speech databases. We also give a short description of some non-Arabic databases for many languages including English [2]. This description will be helpful in recognizing the richness of our database, which we perceive as richer than other databases in many aspects. TIMIT is one of the mostly used English databases with large number of speakers (630) with eight different dialects of American English [19]. The speakers read ten phonetically rich sentences. The text material in the TIMIT consists of 2 dialect "shibboleth" sentences (SA), 450 phonetically-compact sentences (SX) and 1890 phonetically-diverse sentences. The SA sentences were read by all 630 speakers. Each speaker read 5 of the SX sentences and each sentence was uttered by 7 different speakers. For the SI sentences, each speaker read 3 of these sentences, with each sentence being read only by a single speaker [20]. Word and phone

level labeling are provided with the database. The database can be obtained from the Linguistic Data Consortium (LDC).

Table 1. Comparison of available Arabic speech databases

Database	speakers	Dialect	Prompts	Channel	Sampling rate	Environment
SAAVB [3]	1033	Saudi	Numbers, words, sentences, alphabets	Telephone (fixed and mobile)	8 KHz	Indoor, outdoor, car
BBL [4]	164	Levantine	Spontaneous	Microphone	16 KHz	-
QSDAS [5]	77	Quran recitation	Quranic verses	Microphone 1-channel	16 KHz	-
MSA Speech Corpus [6]	40	Levant, Gulf, Africa	Sentences	SHURE microphone, 2 channels are converted to 1 channel	44.1 KHz is converted to 16 KHz	Studio
ALGASD [7]	300	Algerian Arab	Sentences	Microphone, 1-channel	16 KHz	-
West Point [8]	110	Native, non-native	Sentences	SHURE microphone	22.05 KHz	-
NetDC Arabic BNSC [9]	-	-	News	Radio receiver	22.05 KHz	-
Global Phone Arabic [10]	78	Tunisia, Palestine, Jordan	Sentences from newspaper	Microphone	16 KHz	-
Egyptian Arabic Speecon [11]	550 (adults) 50 (child)	Egyptian	Spontaneous + Read (words, sentences)	Microphone, 4-channel	16 KHz	Office, entertainment , car, public place
A-Speech DB [12]	205	-	Continuous speech	Microphone	16 KHz	Office
OrienTel Morocco MCA [13]	772	Moroccan	Digits, words, sentence + spontaneous	Fixed & mobile phones	8 KHz	-
OrienTel Tunisia MCA [14]	792	Tunisian	Digits, words, sentence + spontaneous	Fixed & mobile phones	8 KHz	-
OrienTel Egypt MCA [15]	750	Egyptian	Digits, words, sentence + spontaneous	Fixed & mobile phones	8 KHz	-
OrienTel UAE MCA [16]	880	UAE	Digits, words, sentence + spontaneous	Fixed & mobile phones	8 KHz	-
OrienTel Jordan MCA [17]	757	Jordanian	Digits, words, sentence + spontaneous	Fixed & mobile phones	8 KHz	-
NEMLAR Broadcast News [18]	-	-	News	Radio receiver	16 KHz	

LDC also provides Switchboard 2 Phase I and II databases including NIST evaluation subsets. These databases include large number of speakers recorded in different sessions [21]. The content is spontaneous text material uttered in office and home environments.

A speech database in Castilian Spanish called AHUMADA is developed specifically to consider speaker variability and channel-dependent influences [22]. The database contains the following parameters: microphone and telephone channels; read and spontaneous speech; different speech rates while reading the texts; six different recording sessions; dialectal variations of speakers; fixed utterances and speaker specific utterances; etc. The text materials consist of (a) 24 isolated digits, (b) 10 digit strings consisting of 10 digits each, (c) 10 phonologically and syllabically balanced phrases of 8 to 12 word length, (d) One

phonologically and syllabically balanced text of about 180 words, and (e) one minute of spontaneous speech. The speakers were 104 male speakers with age between 28 to 42 years. Both microphones and telephone lines were supplied to a professional DAT device. The sampling rate is 44.1 kHz.

POLYCOST is a telephone speech database consisting of different European languages [23]. There were 134 speakers (74 male and 60 female) from different European countries speaking the following text materials: connected digits uttered in English, sentences uttered in English, and sentences in mother tongue, where one of the prompts was dedicated to free speech. The utterances were recorded in room and office environments. The database contains six sessions per speaker.

Some small databases are developed in Slovene-language at the University of Ljubljana [24]. The databases (K211d, GOPOLIS, VNTV, and VINDAT) contain isolated words, broadcast news, diaphone, etc. K211d is an isolated-word corpus designed for phonetic research studies of the Slovene spoken language. Two hundred and fifty one Slovene words were carefully selected as text prompts. Ten speakers (five female and five male) were selected to participate the recording. The recording was phonetically transcribed and labeled manually. The GOPOLIS corpus is a large speech database containing Slovene dialogues in airline timetable information services. There were 50 speakers (25 male and 25 female) speaking randomly chosen 100 sentences.

There are many other databases dedicated to English, Japanese, Chinese, German, and Spanish. These databases are publicly available either commercially or free. Publicly available databases make research in speech processing and recognition in these languages rich and diverse. Compared to these major languages, Arabic has significantly less number of publicly available speech databases, though Arabic is a major language and an official language in the United Nations.

The most widely recognizable speaker recognition evaluations (SRE) are conducted by the National Institute of Standards and Technology (NIST) [25]. Their projects contribute in finding new directions to the problem of text independent automatic speaker recognition. In the NIST SRE, the speaker recognition performance is measured by means of detection error trade-off (DET) curves and detection cost functions. The NIST releases SRE plans in different years as a part of their ongoing projects. The most recent NIST Year 2010 speaker recognition evaluation plan includes not only conversational telephone speech, but also read and conversational speech recorded in room microphone channel [26].

1.2. Guidelines for developing the database

While developing a speech corpus, the following considerations are usually taken into account by the research team that performs the recording:

Scope of the corpora: The corpora design depends on the application that will use these corpora: phonetic analysis, speech synthesis, speech recognition, or speaker recognition.

Content: In [27], it is observed that text material affects automatic speaker/speech recognition performance to a great extent. The corpus can have a variety of contents, for example, single digit, continuous digits, isolated words, phrases, sentences, paragraphs, etc.

Phonological distribution: The analysis units (words, phrases, sentences, etc.) should be carefully chosen so that the distributions of phones are balanced. Scripts should contain all possible vowels, consonants, co-articulations, etc. [28, 29, 30, 31].

Number of speakers: The total number speakers should be enough to validate the experiment under study. These speakers should speak a sufficient number of utterances. The diversity of speakers (age, education level, etc.) is an important factor to consider [32].

Gender: The corpora may contain almost equal number of male and female speakers [33].

Accent: The speakers can be chosen to cover different types of accents [34].

Speaking style: Based on the target, the corpora may contain read, spontaneous or both types of speech [32, 34].

Environment: The utterances can be recorded in different types of acoustic environments, for example, sound proof room, office room, corridor, restaurant, street, inside vehicle, etc. in order to track the effect of microphone variability on ASR [35, 36, 37, 38, 39].

Recording materials: Data can be collected with different types of microphones and transmission channels, for example, mobile phone, land phone, etc. [35, 36].

Sessions: Data may be collected in different sessions to observe the effect of intersession variability [7, 32].

Partitioning into training and testing data sets: The corpus needs to be large enough to de divided into training and testing sets to account for different types of variability [7, 32]. It is better that the experiments are closed set.

Questions and Answer: A database can contain a question and answer session to get the information of speaker such as his/her name, age, sex, profession or his spontaneous reaction to these questions [7, 27].

It is hard to cover all these points in on one database but we were able to do this. We took into consideration the points mentioned above and designed the database to be rich in many dimensions and beneficial in different applications and studies. The developed database is rich in text, text categories, environment, microphones, channels, nationality, mother language, number of recording microphones, and number of sessions. It can be used in many applications related to speech/speaker recognition and even for Arabic accent classification.

2. Characteristics of the Database

2.1. Richness of the Database

In this section, we describe the different aspects of the richness of our corpus. We also give details of the database and how we designed it.

2.1.1. Richness in Text

The corpus text consist of sentences, words, paragraphs, and answers to questions. In the following subsections, we briefly describe each.

(a) Sentences

Three different types of sentences have been used:

Rich sentences taken from SAAVB: The list given in SAAVB was designed to cover allophones of each phoneme. The list contains 934 sentences. We increased the list to 940 sentences by repeating 6 sentences to get a total number divisible by 20. To divide the 940 sentences into sub lists, each sub list has20 sentences, where each sub list should include all the phonemes with each phoneme repeated as much as possible; we did the following: We divided the 940 list randomly into 47 sub lists (47x20 = 940), each one contains 20 sentences. Each sub list was checked if it contains all the phonemes and the number of occurrence of each phoneme. The randomization was repeated again to find new sub lists, and again we count the occurrence of all phonemes in every sub list. After 20 randomization of the list into sub lists, we selected the randomization that gave optimal sub lists for the recording. Each of these sub lists contains all the phonemes.

Rich sentences from [40]: In this study, 20 lists have been suggested; each list contains 10 phonetically balanced sentences. From the 20 lists we choose 4 lists that are easier to pronounce and do not have something that may be very strange to the speakers or offending or confusing to him. We took the opinion of test speakers in selecting the easy to read lists. We fixed one list for all speakers. A second list was chosen randomly from the remaining three lists.

Accent identifying sentences: we selected two common sentences that are suggested in SAAVB due to their ability to differentiate accents.

(b) Words

Four categories of words have been used:

Rich words: These are rich words suggested in SAAVB. The SAAVB list consisted of 700 words. We divided the list into 35 sub lists randomly. Each sub list was checked for the number of missing phonemes. Randomization was repeated to find new sub lists, and again we checked for the number of missing phonemes in every sub list. The optimal sub lists,

obtained after 20 randomizations, were chosen for the recording. The optimal sub list is the one with minimum number of missing phonemes.

Phonetically distinctive words: The words were selected from SAAVB because they contain nasals fricatives, and vowels which are closely related to the speaker characteristics and can help in recognizing the speaker identity.

Common words: This list contains 20 words. We designed it to contain words that are used frequently in the everyday conservation. These words consist of almost all Arabic alphabets except two. Examples of some common Arabic words are the Arabic equivalent of [Hello], [yes], [no], [news], etc.

Numbers: This list contains Arabic digit from zero to nine. These digits contained only 17 Arabic alphabets out of 28 but we included this sub list for its importance in many applications.

(c) Paragraphs

Pronouncing paragraphs are different than pronouncing sentences or words. Therefore, two paragraphs were added in this list. The first paragraph is a verse from Quran (the Holy Book of Muslims). The second paragraph is taken from a book of a famous writer. The paragraph was selected because it included all letters, was easy to read by normal readers, and was appealing to them (it is a feel good paragraph). Each of the verse and the paragraph contains all alphabets.

(d) Question and answers

In this database, it was not easy to record an online conversation. Hence we opted for something similar, which were answers by the volunteers to questions by a team member. This is called spontaneous speech. Samples of these questions are: "What is your name?", "how is the weather today?", "what is your best food?". Of course all the questions are in the Arabic language.

(e) Richness in fixed and variable text

Some of the texts were spoken by all the speakers and some texts were distributed among the speakers.

2.1.2. Richness in Speakers

The speakers were Saudis and non-Saudis. The non-Saudis were Arabs and non-Arabs. The non-Arabs were chosen so that they could read Arabic language at an acceptable level. They were mainly from the fourth level in the Arabic language institute at King Saud University. The non-Saudis represented 28 nationalities. They were chosen to represent clusters of areas or countries.

2.1.3. Richness in Recording Sessions

We achieved three sessions of recording. Every session is verified before recording the next one.

2.1.4. Richness in the Recording Environments

Each speaker is recorded in three different environments: sound proof room (Eckel CL-11), office, and cafeteria. For a reason that will be explained later, the second and third sessions were recorded only in the office and the soundproof room.

2.1.5. Richness in the Recording System

The recording system was similar in the office and the cafeteria, and different in the soundproof room.

(a) Recording systems for office and cafeteria

The office and cafeteria system consisted of the following subsystems:

- Two professional microphones (SHURE Beta 58A) connected to a high quality mixer (Yamaha MW12CX in office, Yamaha MW8CX in cafeteria) to be recorded in stereo.
- Medium quality microphone (Sony Dynamic microphone F-V220) connected to a sound card (Sound Card Creative surrounding 5.1).
- Medium quality microphone connected directly to the computer.
- Mobile (Nokia N97) originating calls to a similar mobile connected to the sound card.

(b) Recording system for soundproof room

The soundproof room system consisted of the following subsystems:

- Professional Side-Address condenser Microphone (SHURE PG42) connected to a high quality mixer (Yamaha MW12CX) to be recorded in stereo.
- Professional quality microphone (SHURE Beta 58A) connected to sound card (Sound Card Creative surrounding 5.1).
- Mobile (Nokia N97) connected to the computer via a sound card.

2.1.6. Summary of Richness of Database

Our database is rich in many aspects, as depicted in Fig.1. This subsection summarizes the richness dimensions or aspects:

(a) Text

- Words, sentences, and paragraphs
- Read text vs. spontaneous text

- Common text vs. uncommon text
- Rich words vs. non-rich words (numbers and common words)
- Fixed text vs. variable text
- Easy to pronounce vs. hard to pronounce
- Quran vs. normal text

(b) Environment

- Very quiet (sound room), quite (office), noisy (cafeteria)

(c) Microphones

- We have many microphones vs. one microphone in some databases
- Medium quality, high quality, and high quality Phantom

(d) Different combinations of microphones and recording equipment

- e.g. low quality recording with Medium quality microphone

(e) Sessions

- 3 sessions

(f) Multi-Nationalities

- 9 Arab nationalities
- 20 non-Arab nationalities

(g) Ethnicity:

- Arabs vs. Non-Arabs
- 5 Different regions of the world.

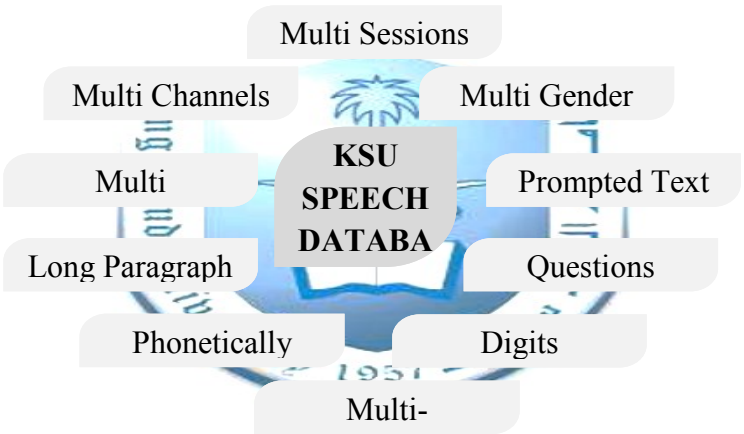


Fig.1: KSU Speech Database Diversity

2.2. Justification of the database specification

Our Corpus vs. SAAVB: SAAVB text was selected by its authors based on a scientific analysis, and since we decided to include sentences and words in our database, it was logical to use the text of SAAVB. But our database is richer because it has more dimensions than SAAVB. For example, SAAVB recorded Saudi speakers from mobile or telephone in one session. Our database recorded many nationalities in three sessions using many microphones within many environments and had more text quantity and variety.

Our Corpus vs. Ref [40]: The authors of [40] designed 20 lists; each list contains 10 sentences so that each list is rich by itself. It had richer text than SAAVB, so we chose it for the same reasons we chose SAAVB. This will make our database richer than both in texts; moreover, our database has more text and has other dimensions as we explained in the case of SAAVB above. Another reason for choosing these lists is that they are not easy to read or pronounce.

Fixed Sentences: These are two common sentences selected from SAAVB. They are designed to differentiate between dialects. We choose these sentences for the same reason.

Common Words: SAAVB sentences and words are not easy; the sentences of [40] are more difficult. Common words were selected by us and were chosen because they will be easier to pronounce, and hence will be more likely to be pronounced correctly.

Numbers: These are important for many applications and are used in daily life.

Rich Words: These words were selected from SAAVB because they have some characteristics that make them useful for speaker recognition as highlighted in section 2.2.1.2.

Paragraphs: Pronouncing paragraphs will have different features than pronouncing sentences and words; hence we included two paragraphs. The first paragraph was selected among many paragraphs because it had the following characteristics: easy to read, it is a feel good paragraph, and it has all the Arabic letters.

The second paragraph is a verse from the Holy Quran. Hence, all speakers were familiar with it. Moreover it has all the Arabic letters. Note that the speakers were asked to read it as normal text and not recite it.

Questions and Answers: The Gulf database was recorded from two speakers speaking to each other. The Babylon Arabic Levantine speech was answers to written questions; so in our database, we included a part similar to Babylon and we consider it as semi questions and answers. It is helpful because it is a different way of speaking than reading from screen or paper, and therefore, may have different characteristics and different effectiveness in recognizing the speaker.

The order of the list: The order of the list was not arbitrary. We made the order of the list such that they will be easy for the speaker.

Silent room: Recording in the silent room produces high quality recording that is very important for language analysis. It can be used to analyze the language, the speaker identity, the native language origin of the speaker.

Office: It is the normal environment in our daily life.

Cafeteria: Recording noisy speech will be helpful in studying the robustness of the speech or speaker recognition methods.

Nationalities: Allow us to investigate effect of the native origin or study characteristic of the speech of a certain nationality.

Ethnicity: Allow us to investigate the effect of ethnicity or to study characteristic of the Arabic speech with respect of ethnic groups.

2.3. The Recording Team

For the recording, and later for verification, the manpower is very critical because understanding the technicality of the computer program and the system equipment is necessary. The recording and verifying were done by a team of six researchers at the college of computer and information sciences, King Saud University. They hold a B.Sc. in computer science\engineering and are native Arabs. They were supervised by two researchers with Master degrees in computer.

2.4. The volunteers

The success of creating the database depends highly on getting volunteers and on the desired number of volunteers. The volunteers had to be Saudis, Arabs, and Non Arabs. The methods used to contact and recruit the volunteers were: Electronic and printed advertisements, presentation in the classes of college of computer, the Arabic language institute, and the personal contacts.

So, all the speakers were literate people, either students at the university, researchers or professors.

2.5. Text Verification

After selecting the text, the next step was to put the text in a displayable form in front of the speaker. The display form was originally a paper form, and then we opted for displaying the text on the screen. Arabic language is unique in some aspects. Indeed, diacritization is sometimes needed to correct pronunciation. Therefore, a great care was taken to make sure that the displayed form was 100% as in the original text.

The written text went into many stages .First of all, it was diacritized, then, rechecked to make sure that it is correct. We asked from volunteers about their opinion for the whole process. The assessment indicates that there is a need for more diacritization of the SAAVB

text. Some sentences were confusing and even some words need diacritization to clarify either it is a verb or noun.

Another suggestion was to display the text on the screen. MATLAB did not support reading from word files so we stored the sub lists in RTF format and gave to two professional editors to diacritize it and check it.

2.6. Pilot recording

Before recording the speakers, we tested the system and the whole process with some speakers. Our goals were to test the system (hardware and software), the comfort and technical soundness of the setup, the endurance of the speakers of the recording in one session and in three consecutive sessions, and to measure the time of each step and each session.
From these initial tests we found:

- Some texts needed diacritization.
- Reading from papers was not comfortable; our solution for this was to have the lists displayed on a second screen.
- Speakers were comfortable in all the other points we checked.
- The time range for speaker recording was 5-7 minutes in a location.

3. System Description

3.1. Software description

In order to fulfill the technical specifications, team developed many versions of the program. The flowchart of the main program is illustrated in Fig. 2.The main features of the last version of the program are:

- The generation of the speaker reports, per recording location, containing all the channels, aiming to detect a corrupted wave file, at the end of the recording session of the speaker.
- An automatic visual report of the recorded channels (just the mono version), this method helped the recording team a lot, as they were sure that the channels were recording in an acceptable level.
- A maximum duration of 120 sec has been allocated to each recorded sentence or paragraph, in order to control the length of the speech.

3.2. Hardware description

Table 2 gives the actual hardware configuration in the three locations. Fig. 3 presents the hardware configuration as in the sound proof room and the office. The sampling rate in all of the recordings was 48K sample/sec with 16 bits resolution.

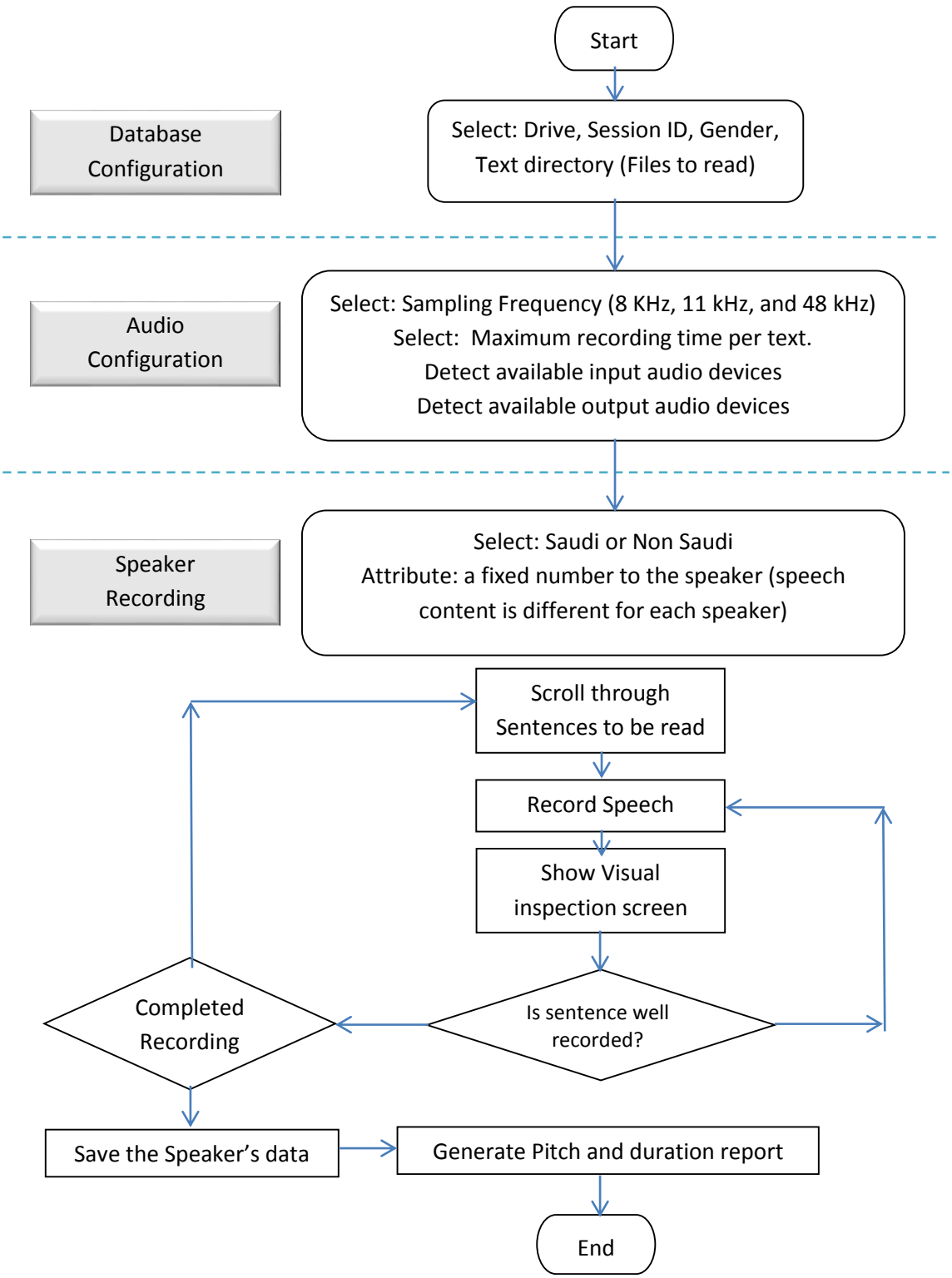
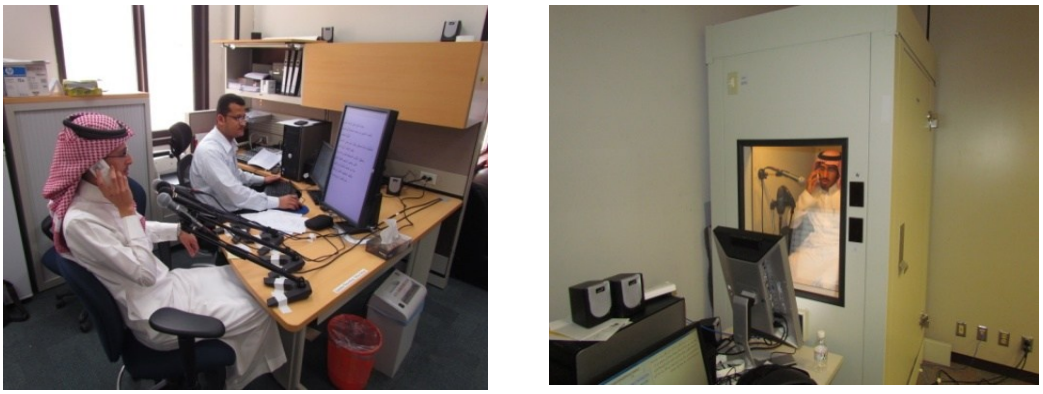


Fig. 2. Flowchart of the main program.

Table 2. Hardware configuration

Device	Brand	Office, Café	Soundproof Room
Microphone	SHUR Beta 58A	2	1
Microphone	Sony F-V220	2	---
Mobile	Nokia N97	1	1
Phantom	---		1
Mixer	Yamaha MW-12CX	1	1
Sound Cards	Creative 5.1 Surrounding	2	1



(a) Office Setup (b) Soundproof Room Setup

Fig.3. Setup for the recording

4. Main Statistics and Results of the Recording

The distribution of the male speakers who recorded in any location and in any session by nationalities is given in Table 3. The female speakers did not include non-Arabs and they were 70 Saudi, 14 Yemenis, 1 Egyptian and 2 Palestinians.

Table 3. List of the recorded male nationalities

Nationality	Number	Nationality	Number	Nationality	Number
Arabs *(Africa & Middle East)*		*Africa Non Arabs*		*Asian Non-Arabs\ Indian Subcontinent*	
Saudi	146	Nigeria	3	India	9
Yemen	15	Uganda	3	Pakistan	8
Egypt	13	Benin	2	Nepal	7
Syria	9	Kenya	2	Afghanistan	4
Tunisia	4	Mali	2	Bangladesh	3
Algeria	4	Central Africa	1	**Total**	**31**
Sudan	4	Guinea Bissau	1	*Asian Non-Arab/East Asia*	
Lebanon	1	Ivory Coast	1	Indonesia	8
Palestine	1	Liberia	1	Philippines	2
Total	**201**	Senegal	1	Thailand	1
East Europe		Togo	1	**Total**	**11**
Serbia	**1**	**Total**	**18**		
Total speakers of all nationalities			**258**		

In verifying the first session, we noticed that the effect of noise was low. So it seems that professional or quality microphones, available nowadays, have a strong noise attenuation capability. Hence the second and third sessions were recorded only in the office and the soundproof room.

The number of volunteers who recorded in the three sessions in the required locations is given in the Table 4. The nationality distribution of male and female volunteers who finished recording in all required locations for the three sessions is provided in Table 5 and Table 6, respectively.

Table 4. Number of speakers in the three sessions

session	No. of male speakers in			No. of female speakers in		
	Office	Cafeteria	Sound proof	Office	Cafeteria	Sound proof
First	253	240	240	87	87	87
Second	206	--	237	77	--	77
Third	136	--	133	64	--	64

Table 5. Male speaker distribution in the three sessions

Session	Saudi	Non-Saudi		Total
		Arab	non-Arab	
First	137	42	61	240
Second	115	41	50	206
Third	55	36	42	133

Table 6. Female speaker distribution in the three sessions

Session	Saudi	Non-Saudi		Total
		Arab	non-Arab	
First	70	17	-	87
Second	61	16	-	77
Third	48	16	-	64

From Tables 4 and 5, we notice that the number of speakers who participated in the third male session is much lower as compared to the first and second session. The reason is that when we started the third male session it was the time of final exams. The majorities of the volunteers were students and were busy in their exams.

Table 7 gives the time duration of the different lists using the mixer data for session 2. Table 7 is divided based on the nationality or race, and it also gives the average duration of the unit of the list.

5. Verification

The recording of the volunteers' speech was followed by verifying the recorded speech. Verification is as vital as the recording itself. For this reason, before starting the recording of any session we verified the previous session. A clear system was designed for the verification and was improved based on our experience. In the following, we briefly discuss the verification stage and shed light on some important findings or ideas.

Table 7. Average Time duration (in Seconds) for the different lists using the mixer data for session 2 (including silence)

Text	Saudi	Arab	Non-Arab	No. of Unit	Avg. /unit
SAAVB sentence_1	32	33	40	Sentence(10)	3.76
SAAVB sentence_2	32	34	41	Sentence(10)	3.80
Fixed sentence	7	7	8	Sentence(2)	4.00
Numbers	10	9	10	Words(10)	1.10
Common words_1	11	11	13	Words(10)	1.26
Common words_2	11	10	12	Words(10)	1.22
Rich words_1	10	10	11	Words(10)	1.13
Rich words_2	10	10	11	Words(10)	1.13
Paragraph_1	29	31	36	Paragraph	34.2
Paragraph_2	48	49	60	Paragraph	56.2
Distinctive words_1	10	9	11	Words(10)	1.11
Distinctive words_2	10	10	11	Words(10)	1.13
Phonetically balanced sent_1	19	19	23	Sentence(10)	2.2
Phonetically balanced sent_2	20	19	24	Sentence(10)	2.29
Q/A _1	20	19	26	Answers(10)	2.34
Q/A_2	17	17	21	Answer(10)	1.98

It is important to mention that the recording system was tested many times in all locations before the actual recording of the volunteers and the team members were selected and trained and supervised. But this cannot substitute the verification of the recording. Moreover the verification stage is more than just verification, it is also commenting on the quality of the recorded speech or documenting the database.

The size of session 1, session 2, and session 3 are 166 GB, 76 GB, and 42.9 GB, respectively. This will give a total database of size 284.9 GB. The number of files for session1, session2 and session 3 are 56217, 22738 and 12924, respectively. Verifying this database is a huge task.

The database is huge, and can actually be looked at as many databases depending on the text, recording system (microphone and digitizing device) the recording environment, and the session number. So human verification of the whole database was a major step by itself and needed a large number of verifiers over a long time. We performed the verification in three stages.

5.1. Stage 1 of verification

This was completed in the first session. The verifiers were given clear instructions in what to do and an excel sheet to fill for every volunteer at each location [41]. The main instructions can be summarized as follows:

- Verify that all channels (or subsystems) of the recording system worked and that there were no missing recording in any channel.

- The mobile channel may have some missing recording due to network quality. This is to be documented, but is not considered error unless it was for a whole sub list or a sizable portion of it (each user read 16 sub lists in each location).
- Not recording any part of any sub list is an error.
- If a letter is missed or substituted then this is to be counted as an error and has to be documented in the sheet. If the replacement or insertion is due to dialect then it is not counted as error but has to be documented.
- Minor stuttering is acceptable but has to be documented
- The verifier also has to comment on the pronunciation correctness

5.2. Stage 2 of verification

After verifying 20% of the first session we were confident that our recording system worked correctly except for rare instances that happened for random sub lists with random users where part of the system, e.g. sound card taking input from medium quality microphone, will not record [may be due to MATLAB having problem with reading from many opened devices]. So the verification was relaxed from verifying all the channels to only verifying the mixer and the mobile channels. This continued for the rest of the first session.

5.3. Stage 3 of verification

This stage was done for the second and third session. From the verification of the first session we were sure of the recording system and that the mobile channel was working except for missing letters or words due to network quality (the sound room is in the basement). So the verification was relaxed to verifying only the mixer channel. Moreover to enhance the verification three improvements were made.

5.3.1. Improvements in stage 3 of verification

These improvements were:

First Improvement: Concatenating the recording of the lists for each volunteer at each location. This simplified the verifier task since he did not have to close and open the recording of 16 sub lists.

Second Improvement: An automatic report was generated through a developed MATLAB program that writes results directly in an excel sheet for each volunteer. The report will flag any lists that may have not been recorded or has a problem, as shown in Fig. 4, and then the verifier has to check the corresponding speech files. The pitch was used to decide if there was a recording or not. The verifier will still do his usual verification but this is an extra help.

Third Improvement: Generation of a graphic display of the recording of all channels while recording and putting the display in a report immediately after finishing the recording of any volunteer, as shown in Fig. 5. This was of great help to catch errors while recording or immediately after recording each volunteer session, so the error can be corrected.

مشروع : التعرف على المتحدث العربي

ARABIC SPEAKER RECOGNITION PROJECT

رقم: 08-inf167-02

Automatic Verification Process

Office				Speaker	NS1		
					Office recording Errors		1
	Avg Recording (min)	Errors		(Hz)			(Hz)
Office	6.403	1	Mean Pitch	146.983	Pitch (Mobile Effect)		183.715
Cafeteria	6.034	0		106.332			195.509
Silent room	6.691	0		180.435			200.957
Session Average time	19.128						

Directory	Wave Files	Duration	Min	Max	Max-Min	Pitch	Remarks
1.SAAB sentences_1	Computer_Mic_Front	33.61	-0.29	0.3321	0.6232	147.9	
	Mic_CreativeSB	33.6	-0.5	0.5588	1.0546	147.3	
	Mobile_CreativeSB	33.63	-0.63	0.7219	1.3507	162.1	
	Yamaha Mixer	33.55	-0.6	0.4267	1.0307	149.9	

Directory	Wave Files	Duration	Min	Max	Max-Min	Pitch	Remarks
1.SAAB sentences_2	Computer_Mic_Front	38.49	-0.32	0.4341	0.7495	152.3	
	Mic_CreativeSB	38.48	-0.47	0.661	1.1312	151.6	
	Mobile_CreativeSB	38.49	-0.7	0.8932	1.598	181.7	
	Yamaha Mixer	38.44	-0.65	0.4222	1.0772	153.6	

Directory	Wave Files	Duration	Min	Max	Max-Min	Pitch	Remarks
2.Fixed Sentences	Computer_Mic_Front	8.27	-0.28	0.3222	0.6029	152.1	
	Mic_CreativeSB	8.25	-0.56	0.5963	1.155	153.5	
	Mobile_CreativeSB	8.27	-0.6	0.7252	1.3277	205.9	
	Yamaha Mixer	8.19	-0.63	0.3872	1.0167	152	

Fig. 4. A snapshot of the initial automatic report

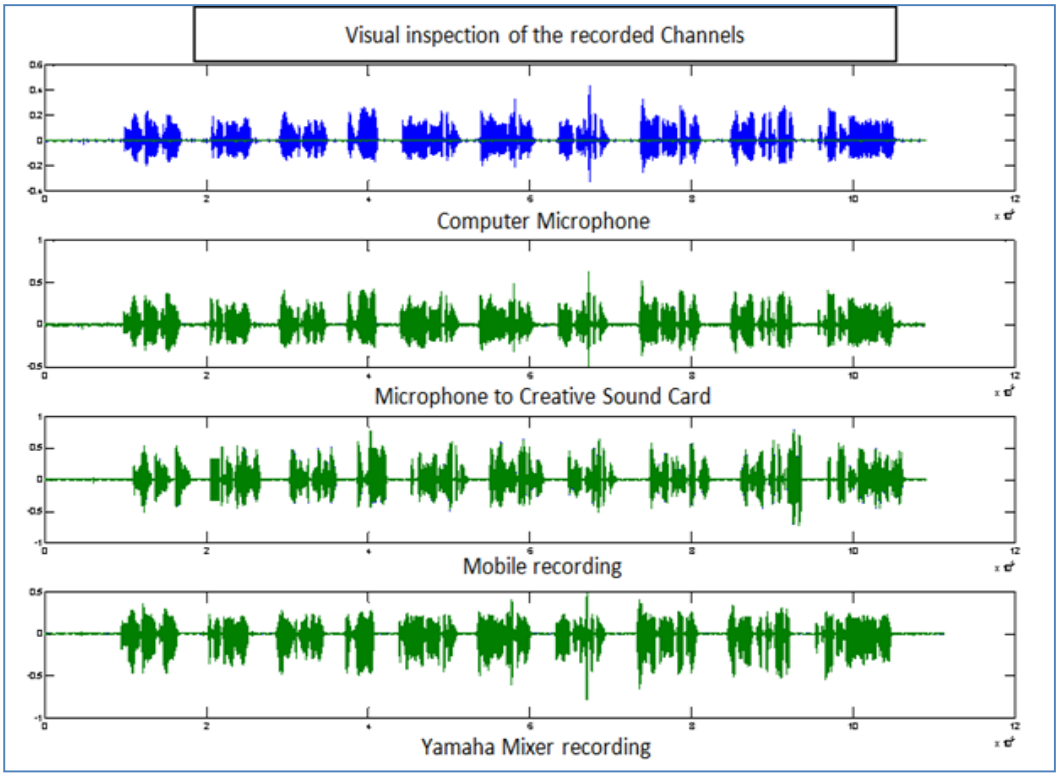


Fig. 5. A snapshot of the visual graphic display checking (office)

5.4. Verification results

Table 8 gives the number of errors detected by verification in all the three environments for the three male sessions. The verification of the female sessions is still going. A recording is considered to have an error if it contains deletion, replacement, or insertion of a character or words. Note that if this insertion or substitution is due to Arabic dialectic or non-Arab pronunciation, then it is documented but not counted as error. An important point to mention is that the severity of the errors is not as presented in the Table 8, because the database is actually 16 sub-lists and error in one of the sub lists will be counted as an error for the speaker for the whole session while in reality it is an error in one of the 16 lists.

Table 8. Verification results for the three sessions

Sessions	Place	Office	Soundproof Room	Cafeteria
First	No. of speakers	253	240	240
	Errors	42	16	24
Second	No. of speakers	206	237	-
	Errors	2	6	-
Third	No. of speakers	136	133	-
	Errors	3	2	-

By comparing the result of verification in the first and second sessions, it became clear that the improvement we made to the recording system greatly reduced the errors.

6. Validation

Validation of the database is a crucial task. Our database is a huge one. We had to select the optimum way to validate it. We are working on that question and conducting some initial experiments. Table 9 gives the results of some of the experiments on subsets of the male database. The attributes of the validation data is as in Table 10.

Table 9. Accuracy of the system with 12 MFCC and 32 Gaussians

Experiment No.	Train	Test	Number of speakers	Accuracy (%)
6 (Saudi)	Paragraph 1	Sentence 2	75	90%
7 (Saudi)	Paragraph 1	Sentence 2	138	83%
8 (Non Saudi)	Paragraph 1	Sentence 2	105	86%
9 (Saudis)	Sentence 1	Sentence 2	140	100%
10 (Non Saudis)	Sentence 1	Sentence 2	105	96%

Table 10. Attributes of the validation data.

Attribute	Value
Training set	Sentence 1
Testing set	Sentence 2
Recording room	Sound proof
Recording channel	Phantom microphone
Recording session	First
Sampling rate	16 kHz (down sampled from the original data)
Window size	20 ms
Frame rate	10 ms
Acoustic features	12 MFCC
Gaussian mixture	32

7. Conclusion

In this paper, we described a very rich and new Arabic speech database dedicated to MSA. We have presented the conditions of making a speech corpus of great quality by researchers in the field. Then we showed that our database satisfied all the conditions. The developed corpus has many dimensions of richness more than any other corpus dedicated to Arabic in the literature. We have also justified in this paper every richness aspect of our database.

Our corpus is huge in size and can be viewed as a collection of different corpora. Nonetheless, we were able to verify its content manually with documented information. We also verified it automatically by tracking the pitch value during recording sessions.

Initial validation of the database was successful and we are working on a more thorough validation.

The goal of our project is to record similar number of female Saudis utterances. We are working on that in the meantime.

The main goal of the database was to be used for speaker recognition. We went many steps ahead and made a rich and versatile database that can be used in many research areas in speech processing. For example, it can be used in the following areas: dialect/accent recognition, speaker nationality recognition, characteristics of the speech of Saudis, Arabs, and non-Arabs, effect of mobile channel in speech and/or speaker recognition, effect of low noise in speech and/or speaker recognition, the use of many channels for speech and/or speaker recognition. The list can go on and these are just examples to appreciate the richness of the database.

8. Acknowledgments

This work is supported by the National Plan for Science and Technology in King Saud University under grant number 08-INF167-02. The authors are grateful for this support. The authors also thank the project consultants Dr. Mansour Alghamdi and Prof. Sid-Ahmad Seloani for their valuable comments.

References

[1] Selouani, S. and J. Caelen, "Arabic phonetic features recognition using modular connectionist architectures. In: Proceedings of the IEEE Interactive Voice Technology for Communication, IVTTA ’98, pp. 155–160, 1998.

[2] Alsulaiman, M. and G. Muhammad, M. Bencherif, A. Mahmood and Z. Ali, “A survey on Arabic speech database”, Archives Des Sciences Journal, 2012. (submitted)

[3] Alghamdi, M., Alhargan F., Alkanhal. M., Alkhairy A., Eldesouki M. and Alenazi A., “Saudi accented Arabic voice bank,” J. King Saud University, Computer and Information Sciences, vol. 20, pp. 45-62, 2007.

[4] Makhoul, J., Zawaydeh B., Choi F., and Stallard D., “2005 BBN/AUB DARPA Babylon Levantine Arabic Speech and Transcripts,” Linguistic Data Consortium (LDC), Philadelphia, 2005. LDC Catalog Number LDC2005S08.

[5] Harrag, A. and Mohamadi T., " QSDAS: New Quranic Speech Database for Arabic Speaker Recognition", The Arabian Journal for Science and Engineering, vol. 35, no. 2C, pp. 7-13, December 2010.

[6] Abushariah, M. and Ainon R., Zainuddin R., Alqudah A., Ahmed M., and Khalifa O., "Modern standard Arabic speech corpus for implementing and evaluating automatic continuous speech recognition systems", vol. 349, no. 7, Journal of the Franklin Institute, 2011.

[7] Droua-Hamdani, G. and Selouani S. A., and Boudraa M., "Algerian Arabic speech database (ALGASD): corpus design and automatic speech recognition application", The Arabian Journal for Science and Engineering, vol. 35, no. 2C, pp. 157-166, December 2010.

[8] Stephen, Col., A. LaRocca and RajaaChouairi, "West Point Arabic Speech", LDC Catalog LDC2002S02, 2002.

[9] NetDC Arabic BNSC, ELRA Catalog ELRA-S0157.

[10] GlobalPhone Arabic, ELRA Catalog ELRA-S0192.

[11] The Egyptian Arabic Speecon Database, ELRA catalog ELRA-S0308.

[12] A-Speech DB, ELRA catalog ELRA-S0315.

[13] OrienTelMorocco MCA, ELRA catalog ELRA-B0004.

[14] OrienTelTunisia MCA, ELRA catalog ELRA-B0005.

[15] OrienTelEgypt MCA, ELRA catalog ELRA-B0006.

[16] OrienTel United ArabEmirates MCA, ELRA catalog ELRA-B0010.

[17] OrienTel Jordan MCA, ELRA catalog ELRA-B0011.

[18] NEMLAR Broadcast News Speech Corpus, ELRA catalog ELRA-S0219.

[19] Garofolo J. S., Lamel L. F., Fisher W. M., Fiscus J. G., Pallett D. S., and Dahlgren N. L., " DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM", NIST, 1993. Available at http://www.ldc.upenn.edu/Catalog/docs/LDC93S1.

[20] http://www.ldc.upenn.edu/Catalog/readme_files/timit.readme.html

[21] Graff D., and Walker K., and Canavan A., Switchboard-2 Phase I, II. Linguistic Data Consortium, Philadelphia. Available at www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC98S75; www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC99S79;

[22] Garcia J. O., Rodriguez J. G., and Aguair V. M., "AHUMADA: A large speech corpus in Spanish for speaker characterization and identification," Speech Communication, vol. 31, pp. 255-264, 2000.

[23] Petrovska D., Hennebert J., Melin H., and Genoud D., "POLYCOST: a telephone speech database for speaker recognition," RLA2C, Avignon, France, pp. 211–214, 20–23 April 1998.

[24] Mihelic F., and Gros J., Dobrisek S., Zibert J., and Pavesic N., "Spoken Language Resources at LUKS of the University of Ljubljana," International Journal of Speech Technology, vol. 6, pp. 221–232, 2003.

[25] The NIST Year 2010 Speaker Recognition Evaluation Plan, available at http://www.itl.nist.gov/iad/mig/tests/sre/2010/NIST_SRE10_evalplan.r6.pdf

[26] http://www.nist.gov/speech/tests/spk/index.htm

[27] Li K., Dammann PJ. E., and Chapman W. D., "Experimental studies in speaker verification, using an adaptive system", J. Acoust. Soc. Am., vol. 40, no. 5, pp. 966-978. 1966.

[28] Sambur M. R., "Selection of acoustic features for speaker identification," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 23, pp. 176–182, 1975.

[29] Wolf J. J., "Efficient acoustic parameters for speaker recognition," Journal of the Acoustical Society of America, vol. 51, pp. 2030– 2043, 1972.

[30] Habib S. M. M., Alam F., Rabia Sultana, Chowdhury S. A. and Mumit Khan, "Phonetically Balanced Bangla Speech Corpus", Conference on Human Language Technology for Development, Egypt, 2011.

[31] Iida A., Campbell N., "A database design for a concatenative speech synthesis system for the disabled", Fourth ISCA ITRW on Speech Synthesis (SSW-4), Interspeech 2001.

[32] Reynolds D. A., "An overview of automatic speaker recognition Technology," Proc. IEEE international conference on acoustics, speech and signal processing, ICASSP'02, vol. IV, pp. 4072–4075, 2002.

[33] Doddington G. R., and Przybocki M. A., Martin A. F., and Reynolds D. A., "The NIST speaker recognition evaluation overview: methodology systems, results, perspective," Speech Communications, 31, pp. 225–254, 2000.

[34] Kersta L. G., "Voiceprint classification for an extended population," Journal of the Acoustical Society of America (A), vol. 39, pp. 1239, 1966.

[35] Reynolds D. A., "The effects of handset variability on speaker recognition performance: Experiment on the Switchboard corpus," Proc. IEEE international conference on acoustics, speech, and signal processing, ICASSP'96 pp. 113–116, 1996.

[36] Norton R., "The evolving biometric marketplace to 2006," Biometric Technology Today, 10(9), pp. 7–8, 2002.

[37] Reynolds D. A., “Gaussian Mixture Models”, Encyclopedia of Biometric Recognition, Springer, Journal Article, February 2008.

[38] Sturim D. E., Campbell W. M., Reynolds D. A., Dunn R. B., Quatieri T. F., “Robust Speaker Recognition with Cross-Channel Data: MIT/LL Results on the 2006 NIST SRE Auxiliary Microphone Task", ICASSP 2007, Apr. 15-20, 2007.

[39] Patil H. A. and Basu T. K., “Development of speech corpora for speaker recognition research and evaluation in Indian languages”, Int. J. of Speech Tech., Springer-Verlag, vol. 11, no.1, pp.17-32, March 2008.

[40] Boudraa M., Boudraa B., and Guerin B., “Twenty Lists of Ten Arabic Sentences for Assessment”, ACUSTICA, ACTA-ACUSTICA, vol. 86, no. 5, 1998.

[41] Alsulaiman, M. and Muhammad G., Bencherif M. and Mahmood A., “Arabic speaker Recognition”, Tech. Report, Research Center, College of Computer and Information Sciences, King Saud University, Sep 2012.

*Corresponding author: Zulfiqar Ali

Speech Processing Lab, College of Computer and Information Sciences,

King Saud University,

Riyadh 11543, Saudi Arabia

Email: zuali @ksu.edu.sa

KSU Speech Database: Text Selection, Recording and Verification

Mansour Alsulaiman, Zulfiqar Ali, Ghulam Muhammed, Mohamed Bencherif, Awais Mahmood
Speech Processing Group, Department of Computer Engineering,
College of Computer and Information Sciences, King Saud University
Riyadh 11543, Saudi Arabia
{msuliman, zuali, ghulam, mbencherif, awais}@ksu.edus.sa

***Abstract*— King Saud University speech database (KSU-DB) is a very rich speech database of Arabic language. Its richness is in many dimensions. It has more than three hundred speakers of both genders. The speakers are Arabs and non-Arabs belonging to twenty-nine different nationalities. The database has different types of text such as isolated words, digits, phonetically rich words and sentences, phonetically balanced sentences, paragraphs, and answers to questions. The KSU-DB was recorded in three different locations; the first is an office that represents a normal environment with low noise. The second and third locations are cafeteria and soundproof room representing noisy and quiet environments, respectively. The database has different channels of recordings; mobile, medium and high quality microphones connected to recording devices of different qualities. To track the inter-session variations of the speakers, the database was recorded in three sessions with a gap of about six weeks. Though the database main goal is for speaker recognition research, nonetheless, we made it very rich so that it can be used in many speech-processing researches. A team of native Arabs verified the database manually as well as automatically.**

Keywords— Arabic speech database; speaker recognition; phonetically rich database; database recording, database verification.

I. INTRODUCTION

Arabic language is one of the oldest and widely spoken languages. Some of its differences from other languages are unique phonemes and phonetic features, and a complicated morphological word structure. It has been reported in the literature that the major difficulties in automatic speech processing of Modern Standard Arabic (MSA) are distinctive due to the characteristics of the Arabic sounds, namely, emphatic, uvular, and pharyngeal consonants, and short and long vowels [1].

The speech database is an essential part in building high performance speech processing systems. An automatic speech/speaker recognition system can be deployed successfully in real life only if it is built using a versatile and relevant database. Without a proper speech database, speech processing related research cannot progress. There are many databases in major languages like English, Spanish, German, Japanese, Chinese, etc. However, Arabic speech databases are few in numbers and are not rich. Therefore, there is a need for a rich publicly available comprehensive Arabic speech database. Such database can significantly increase the amount of research in Arabic speech processing.

A comparison of available Arabic databases [2]-[17] is presented in Table I. From the table, we can observe that databases on Arabic speaker recognition is not as rich as KSU-DB. Such speech databases, with less variability, cannot be utilized to investigate the effects of different variables such as text, environment, channel, session, etc. Therefore, we set a goal to record a rich Arabic speech database under the funding of National Plan for Science and Technology, Saudi Arabia. Some of the essential parameters to design and build a versatile speech database are stated in [18].

In this paper, we present the rich and comprehensive Arabic speech database that we developed for speaker recognition and other speech applications such as speech recognition, speech analysis, accent identification, ethnic groups / nationality recognition, etc. The richness of the database makes it a valuable resource for research in speech processing in general and Arabic speech in particular.

The rest of the paper is organized as follows: section II highlights the diversity and richness of KSU-DB and gives justification for the database multiple dimensions; section III describes the recording setup of the database, section IV gives the detail of the database verification, section V provides statistics about the database, and finally, section VI draws some conclusions.

II. KING SAUD UNIVERSITY SPECH DATABASE: KEY FEATURES AND THEIR JUSTIFICATIONS

A. Key Features of KSU-DB

The key features of the KSU speech database are: number of sessions, type of environments, type of recording devices, types of microphones, availability of mobile recording, number of speakers of different nationalities and different varieties of spoken text. There are more than 300 male and female speakers of different nationalities. The recording was done in three different types of environments, namely, office, cafeteria and soundproof room. Different combination of microphones (high quality and medium quality) and recording devices (mixers, built-in and high-quality sound cards). The database includes different types of text: paragraphs, sentences (rich and phonetically balanced), words (rich, digits, and common) and answers to questions. The database contains utterances of male speakers who were Saudis, non-Saudi Arabs, and non-Arabs, while the female speakers were Saudis and non-Saudi Arabs. Fig. 1 and Table II provide a summary of the richness of the database.

978-1-4799-2578-0/13 $31.00 © 2013 IEEE
DOI 10.1109/EMS.2013.41

TABLE I. COMPARISON OF AVAILABLE ARABIC SPEECH DATABASES

Database	Speakers	Dialect	Prompts	Channel + Sampling Rate in KHz	Environment
SAAVB [2]	1033	Saudi	Numbers, words, sentences, alphabets	Telephone (fixed and mobile) +8	Indoor, outdoor, car
BBL [3]	164	Levantine	Spontaneous	Microphone +16	-
QSDAS [4]	77	Quran recitation	Quranic verses	Microphone 1-channel +16	-
MSA Speech Corpus [5]	40	Levant, Gulf, Africa	Sentences	SHURE microphone, + 44.1 KHz converted to 16	Studio
ALGASD [6]	300	Algerian Arab	Sentences	Microphone, 1-channel +16	-
West Point [7]	110	Native, non-native	Sentences	SHURE microphone +22.05	-
NetDC Arabic BNSC [8]	-	-	News	Radio receiver +22.05	-
Global Phone Arabic [9]	78	Tunisia, Palestine, Jordan	Sentences from newspaper	Microphone +16	-
Egyptian Arabic Speecon [10]	550(adults) 50(child)	Egyptian	Spontaneous + Read (words, sentences)	Microphone, 4-channel +16	Office, entertainment, car, public place
A-SpeechDB [11]	205	-	Continuous speech	Microphone +16	Office
OrienTel Morocco MCA [12]	772	Moroccan	Digits, words, sentence + spontaneous	Fixed & mobile phones +8	-
OrienTel Tunisia MCA [13]	792	Tunisian	Digits, words, sentence + spontaneous	Fixed & mobile phones	-
OrienTel Egypt MCA [14]	750	Egyptian	Digits, words, sentence + spontaneous	Fixed & mobile phones +8	-
OrienTel UAE MCA [15]	880	UAE	Digits, words, sentence + spontaneous	Fixed & mobile phones +8	-
OrienTel Jordan MCA [16]	757	Jordanian	Digits, words, sentence + spontaneous	Fixed & mobile phones +8	-
NEMLAR Broadcast News [17]	-	-	News	Radio receiver +16	

Figure 1. Richness of KSU-DB.

Nine different types of text were recorded per speaker. Some of them are fixed for speakers. Some differ from a speaker to another, but the text remained the same in all sessions for a particular speaker. The text types are: i) rich sentences, ii) rich words, iii) accent identifying sentences, iv) Arabic digits, v) phonetically distinctive words,

TABLE II. SUMMARY OF THE RICHNESS OF THE KSU DATABASE

Text	Words	Sentences	Paragraphs
	Easy	Difficult	
	Common	Non common	
	Read Text	Answers to Questions (Semi-spontaneous)	
	Rich words	Digits	Common words
	Rich Sentences	Phonetically balanced sentences	
	Fixed	Variable	
	Normal	Quran	
Environment	Sound Room (Noise free)	Office (Low noise)	Cafeteria (Medium Noise)
Nationalities	Saudis	non-Saudi Arab (8 Nationalities)	non-Arab (20 Nationalities)
Ethnicity	Arabs	Non Arabs- Indian Subcontinent, Asian, African, East Europe	

(vi) common words, vii) paragraphs, viii) phonetically balanced sentences from [19], ix) answers to questions. More detail of the recorded text is given in [18]. The text

was divided into sixteen lists to make it more readable from a display screen.

B. *Justification of Key Features of KSU-DB*

In this section, we present some justifications about the specifications of the database.

Text recorded in Saudi Accented Arabic Voice Bank (SAAVB) [2] was selected based on scientific analysis; hence, we have included all of SAAVB, sentences and words, in our database. SAAVB had two common sentences that were designed to differentiate between dialects; hence, we included these sentences in our text for the same reason. Our database is richer than SAAVB because it has more dimensions whereas SAAVB recorded only Saudi speakers from mobile or telephone in just one session. The authors of [19] designed twenty lists; each list contains ten sentences so that each list is rich and balanced by itself; hence, we included four of these lists in our text. The addition of text from both [2] and [19] made our database more versatile in text.

We prepared a list of common words as they were easier to pronounce, and hence will be more likely to be pronounced correctly. Arabic digits were also included into the recording text as they are important in many applications and are used in daily life. Some words were selected from SAAVB words because they have some characteristics that make them useful for speaker recognition.

Pronouncing paragraphs may have different features than pronouncing sentences and words; hence, we included two paragraphs. The first paragraph was selected among many paragraphs because it had the following characteristics: easy to read, a feel-good paragraph, and has all the Arabic phonemes. The second paragraph is a verse from the Holy Quran; hence, all speakers were familiar with it. Moreover, it has all the Arabic phonemes. We asked the speakers to pronounce it as normal text and not recite it as a verse.

The Babylon Arabic Levantine speech database [3] is answers to written questions, so in our database, we included a part similar to Babylon, and we consider it as semi-spontaneous. It is helpful because it is a different way of speaking than reading from a screen or a paper, and therefore, may have different characteristics and different effectiveness in recognizing the speaker. It is important to mention that the order of the recorded text was not arbitrary. We made the order of the text such that the recording session will be easy for the speaker.

Recording in the silent room produces high-quality recording that is very important for speech analysis. It can be used to analyze the Arabic speech, the speaker identity, the effect of the native language of the speaker. The environment of the office was normal as in our daily life. Recording of noisy speech in the cafeteria will be helpful in studying the robustness of the speech or speaker recognition methods.

The inclusion of different nationalities allows us to investigate the effect of the nationality or study characteristic of the speech of a certain nationality. Ethnicity allows us to investigate the effect of ethnicity or to study characteristic of the Arabic speech pronounced by ethnic groups.

III. RECORDING OF KSU-DB

After the selection of the recording text, the next step was to display the text in front of the speakers from where he/she can read it with comfort and ease. Reading from a paper was a bothersome to the recording process and created problems due to movement of the speaker's head which caused variation in microphone to mouth distance. Therefore, the text was displayed on a screen positioned right behind the microphones so that the speaker will not need to move his/her head during recording as shown in Fig. 2. To facilitate correct pronunciation of the text by the non-Arab speakers, as they were an important group of the speakers, the text was diacritized by expert phoneticians where it was needed.

Figure 2. Text reading from a display screen.

The next important step was the development of the recording software according to the requirement that it should be equipped with options of selecting recording locations, sessions, Arab and non-Arab speakers, sampling frequency and be capable of detection of multi-channels. The flow chart of the developed software is depicted in Fig. 3. It contains seven GUIs including general information about the project and main menu. The other GUIs were: Database Configuration for the selection of location, session and directory to store the recording; Audio Streaming I/O Devices for detection of I/O devices, sampling frequency, and maximum time duration for recording of each list of the text; Audio Streaming I/O Devices; Checking to verify the detected I/O devices; and Record Speaker. Record speaker is the most important and vital interface of the program and it allows the selection of speaker Saudi/non-Saudi, speaker ID, text list to read, and provide text display and real time graphical view of recording of all channels. The record speaker interface is shown in Fig. 4.

The hardware used to build the recording system in office is presented in Table III. The system consists of four recording channels: two professional microphones (SHURE Beta 58A) connected to a high-quality mixer (Yamaha MW12CX) for stereo recording, medium-quality microphone (Sony Dynamic microphone F-V220) connected to a sound card (Sound Card Creative surrounding 5.1), medium-quality microphone (Sony Dynamic microphone F-V220) connected directly to the computer, and mobile (Nokia N97) connected to the sound card (Sound Card Creative Surrounding 5.1) receiving calls from a similar mobile.

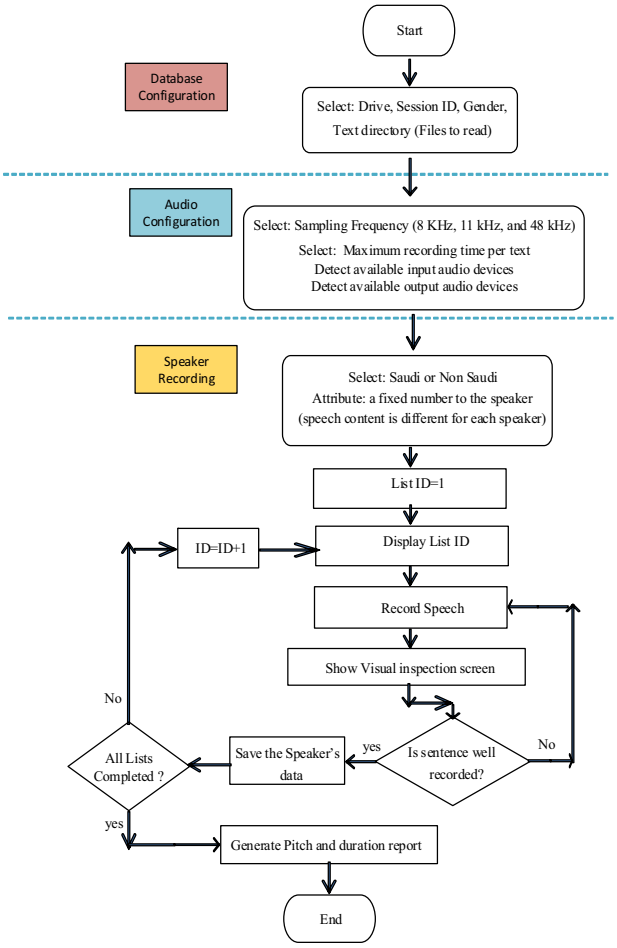


Figure 3. Flow chart of developed recording software.

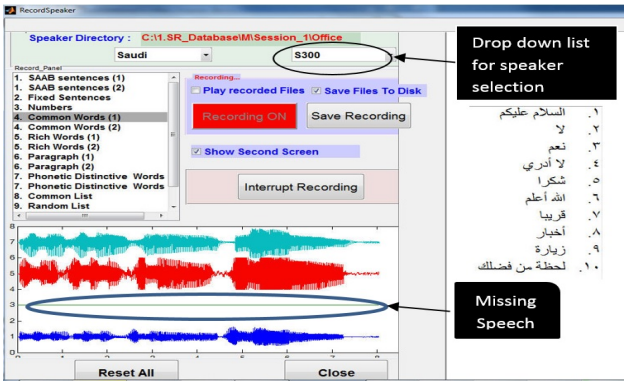


Figure 4. Speaker recording interface.

Before recording a speaker, the recording team made sure that all the devices were detected by the recording software and worked properly. The arrangement of the microphones in the office is shown in Fig. 5. The distance between speaker's mouth and microphone was adjusted to 10 to 15 cm, and it was the same for all the speakers in all sessions at each location. The recording person was listening to the speaker carefully to make sure that he read the text correctly, otherwise he would interrupt the recording and repeat the recording of that particular list. In the soundproof room, the recorder listened to the speaker through a headset because the speaker was in a soundproof room. A view of recording setup of the speakers at all locations is presented in Fig. 6.

TABLE III. LIST OF HARDWARE IN OFFICE

Device	Brand	Qty
Microphone	SHUR Beta 58A	2
Microphone	Sony F-V220	2
Mobile	Nokia N97	2
Mixer	Yamaha MW-12CX	1
Sound Cards	Creative 5.1 Surrounding	2
Computer	DELL 760	1
Recorder Screen	DELL -19"	1
Speaker Screen	DELL – 22" (Vertical position)	1

Figure 5. Arrangement of microphones in offfice.

IV. VERIFICATION OF KSU-DB

It is important to mention that the recording system was tested many times in all locations before starting the recording of the speakers. The team members were trained and supervised, but this cannot substitute the verification of the recording which was as vital as the recording itself. For this reason, before starting the recording of any session, we verified the previous session. Moreover, the verification process was more than a verification of recording only; it was also commenting on the quality of the recorded speech and documenting the spoken text. The verification team, who was also the recording team, was formed from six graduate students of computer science/engineering and were native Arabs. Two researchers holding Master degrees supervised them.

Before starting the verification, instructions were given to the team members. Some of the instructions were: (i) verify that all channels are recorded and there are no missing recording in any channel; (ii) the mobile channel may have some missing recording due to network quality, and it should not be considered an error unless it was for a whole list or a sizable portion of it (each user read 16 lists in each location); (iii) if any part of a list was missing, it should be considered an error; (iv) if a letter was deleted or substituted, then this should be counted as an error and documented in the sheet. If the replacement or insertion is due to dialect, then it should not be counted as an error but it should be documented; (v) minor stuttering is acceptable, but it should be documented; (vi) the verifier should comment on the pronunciation correctness and quality.

a) Office (b) Cafeteria (c) Sound-proof Room

Figure 6. Recording at different locations.

The manual verification was complemented by automatic verification. An automatic report was generated through a developed MATLAB program that write the results directly in an excel sheet for each speaker. The report will flag any lists that may have not been recorded or has a problem, as shown in Fig. 7, and then the verifier has to check the corresponding speech files. The pitch was used to decide if there was a recording or not. The verifier would still do his usual verification, but this was an extra help. A graphic display, showing recording of each channel, was also displayed after the recording of each list for each speaker to be sure that all channels were recorded well, as shown in Fig. 8. This is in addition to the real time display of the recording channels.

ARABIC SPEAKER RECOGNITION PROJECT

08-inf167-02 :رقم

Automatic Verification Process

Office Speaker NS1

Office recording Errors 1

	Avg Recording (min)	Errors		(Hz)		(Hz)
Office	6.403	1	Mean Pitch	146.983	Pitch (Mobile Effect)	183.715
Cafeteria	6.034	0		106.332		195.509
Silent room	6.691	0		180.435		200.957
Session Average time	19.128					

Directory	Wave Files	Duration	Min	Max	Max-Min	Pitch	Remarks
1.SAAB sentences_1	Computer_Mic_Front	33.61	-0.29	0.3321	0.6232	147.9	
	Mic_CreativeSB	33.6	-0.5	0.5588	1.0546	147.3	
	Mobile_CreativeSB	33.63	-0.63	0.7219	1.3507	162.1	
	Yamaha Mixer	33.55	-0.6	0.4267	1.0307	149.9	

Figure 7. A snapshot of the automatic report.

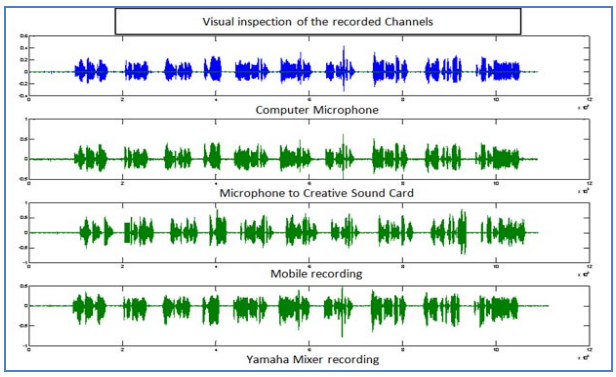


Figure 8. A snapshot of the visual graphic display checking (office).

V. STATISTIC OF KSU-DB

We accomplished the recording of three sessions recorded at three different locations. The statistics of male and female speakers recorded at all locations in each session are given in Table IV.

TABLE IV. NUMBERS OF OVERALL RECORDED SPEAKERS

Session	Saudi		Non-Saudi				Total
			Arab		non -Arab		
	M	*F*	*M*	*F*	*M*	*F*	
I	138	69	42	17	62	0	328
II	124	62	44	16	57	0	303
III	50	49	41	16	41	0	197

M stands for male speakers and F stands for female.

In the KSU-DB, various Arab and non-Arab nationalities are recorded. Table V, presents the number of male speakers for each nationality. Seventy-five percent of the Arab speakers are Saudis and the remaining twenty-five percent are from eight different Arab countries. The non-Arabs belong to four different regions of the world.

TABLE V. NATIONALITIES OF MALE ARAB AND NON-ARAB SPEAKERS RECORDED AT OFFICE IN SESSION I

Nationality	**No.**	**Nationality**	**No.**	**Nationality**	**No.**
Arabs (Africa & Middle East)		***Africa Non Arabs***		***Asian Non-Arabs (Indian Subcontinent)***	
Saudi	151	Nigeria	3	India	9
Yemen	15	Uganda	3	Pakistan	8
Egypt	13	Benin	2	Nepal	7
Syria	9	Kenya	2	Afghanistan	4
Tunisia	4	Mali	2	Bangladesh	3
Algeria	4	Central Africa	1	***Sub-total***	***31***
Sudan	4	Guinea Bissau	1	***Asian Non-Arab (East Asia)***	
Lebanon	1	Ivory Coast	1	Indonesia	9
Palestine	1	Liberia	1	Philippines	2
Sub-total	***202***	Senegal	1	Thailand	1
East Europe		Togo	1	***Sub-total***	***12***
Serbia	1	***Sub-total***	***18***		
Total number of speakers of all nationalities				**264**	

TABLE VI. RECORDING TIME (IN HOURS) OF ALL SESSIONS AT EACH LOCATION

Session	Gender	Office				Soundproof Room				Cafeteria				Total Recording Time of each gender in Each Session
		S	*NS*	*NA*	*Total*	*S*	*NS*	*NA*	*Total*	*S*	*NS*	*NA*	*Total*	
I	*Male*	11.8	4	5	20.8	11.3	3.8	4.8	20	11	3.3	4.9	19.2	60 Hours
	Female	5.66	1.4	0	7	5.5	1.5	0	7	5.8	1.4	0	7.2	21.2 Hours
II	*Male*	9.8	3.4	4.4	17.6	9.7	3.4	4.5	17.6	No Recording				35.2 Hours
	Female	4.8	1.25	0	6	4.9	1.25	0	6.1					12.1 Hours
III	*Male*	4.1	3.2	3.2	10.5	4	3.2	3.3	10.5	No Recording				21 Hours
	Female	3.8	1.2	0	5	3.8	1.2	0	5					10 Hours
Total recording time of all sessions and locations for both genders										**159.5 Hours**				

S, NS, and NA stands for Saudis, non-Saudi Arabs, and non-Arabs, respectively.

The database was recorded at 48 kHz and the size of session 1, session 2, and session 3 are 172 GB, 77 GB, and 44 GB, respectively, for male speakers. For female speakers, the sizes are 49 GB, 26 GB, and 20GB, respectively. This will give a total database of size 388 GB. Verifying this database was a huge task. Time length (in hours) for the recorded speech of the Saudi, non-Saudi Arab and non-Arabs, at every location for each session is provided in Table VI.

VI. CONCLUSION

The developed KSU-DB is a rich comprehensive database that can serve the speech research community to a great extent, and the Arabic speech community in particular. The richness of KSU-DB is due to the different varieties in the many dimensions of the database. It contains speech from a large number of male and female speakers from different nationalities and ethnic groups, different types of scripts, multiple channels, three sessions, and different environments, etc. The database was developed and used in a funded project, of two years duration, with the title *Arabic Speaker Recognition*. Due to the richness of KSU database, it can be used in other speech processing research as well. Hence, we were able to use it in a noisy channel and cross channel speaker recognition systems investigation. We also did a comprehensive study on the effect of Arabic text on the speaker recognition system. We are now using it to do research on computer-aided pronunciation training for non-Arabs learners. This database is an important contribution towards speech processing and analysis.

ACKNOWLEDGMENT

This work is supported by the National Plan for Science and Technology in KSU under grant number 08-INF167-02. The authors are grateful for this support.

REFERENCES

[1] S. Selouani and J. Caelen, "Arabic phonetic features recognition using modular connectionist architectures", In proc. of the IEEE Interactive Voice Technology for Communication, IVTTA '98, 1998, pp. 155–160.

[2] M. Alghamdi, F. Alhargan, M.Alkanhal, A. Alkhairy, A. Eldesouki and A. Alenazi, "Saudi accented Arabic voice bank," J. King Saud University-Computer and Information Sciences, vol. 20, pp. 43-58, 2008.

[3] J. Makhoul, B. Zawaydeh, F. Choi, and D. Stallard, "2005 BBN/AUB DARPA Babylon Levantine Arabic Speech and Transcripts," Linguistic Data Consortium (LDC), Philadelphia, 2005, LDC Catalog Number LDC2005S08.

[4] A. Harrag, and T. Mohamadi, "QSDAS: new Quranic speech database for arabic speaker recognition", The Arabian Journal for Science and Engineering, vol. 35, no. 2c, pp. 7-13, 2010.

[5] M. Abushariah, R. Ainon, R. Zainuddin, A. Alqudah, M. Ahmed, and O. Khalifa, "Modern standard Arabic speech corpus for implementing and evaluating automatic continuous speech recognition systems", Journal of the Franklin Institute, vol. 349, pp. 2215-2242, 2012..

[6] G. Droua-Hamdani, S. A. Selouani, and M. Boudraa, "Algerian Arabic speech database (ALGASD): corpus design and automatic speech recognition application", The Arabian Journal for Science and Engineering, vol. 35, no. 2C, pp. 157-166, 2010.

[7] Stephen Col., LaRocca A. and Chouairi R., "West Point Arabic Speech", LDC Catalog LDC2002S02, 2002.

[8] NetDC Arabic BNSC, 2011, ELRA Catalog ELRA-S0157.

[9] GlobalPhone Arabic, 2011, ELRA Catalog ELRA-S0192.

[10] The Egyptian Arabic Speecon Database, 2011, ELRA catalog ELRA-S0308.

[11] A-SpeechDB, 2011, ELRA catalog ELRA-S0315.

[12] OrienTel Morocco MCA, 2011, ELRA catalog ELRA-B0004.

[13] OrienTel Tunisia MCA, 2011, ELRA catalog ELRA-B0005.

[14] OrienTel Egypt MCA, 2011, ELRA catalog ELRA-B0006.

[15] OrienTel United Arab Emirates MCA, 2011, ELRA catalog ELRA-B0010.

[16] OrienTel Jordan MCA, 2011, ELRA catalog ELRA-B0011.

[17] NEMLAR Broadcast News Speech Corpus, 2011, ELRA catalog ELRA-S0219.

[18] M. Alsulaiman, G. Muhammad, M. A. Bencherif, A.Mahmood, Zulfiqar Ali and M. Aljabri, "Building a rich Arabic speech database", In proc. of the IEEE 5th Asia International Conference on Mathematical Modeling and Computer Simulation (AMS '11), 2011, pp. 100-105.

[19] M. Boudraa, B. Boudraa, and B. Guerin, "Twenty lists of ten Arabic sentences for assessment", ACUSTICA, ACTA-ACUSTICA, vol. 86, no. 5, 1998.

Building a Rich Arabic Speech Database

Mansour M. Alsulaiman, Ghulam Muhammad, Mohamed A. Bencherif, Awais Mahmood, Zulfiqar Ali, and Mohammad Aljabri

Speech Processing Group, College of Computer and Information Sciences,
King Saud University,
Riyadh, Saudi Arabia.
E-mail: {msuliman, ghulam, mbencherif, awais, zuali}@ksu.edu.sa

***Abstract*--Availability of databases is a necessity in the speech processing field. The publically available databases in Arabic language are few. In this paper we describe a rich database for Arabic language. The database is rich in many dimensions: in text, environments, microphone type, number of recording sessions, recording system, the transmission channel, the country of origin, and the mother language. This richness makes the database an important resource for research in Arabic Language processing and very useful in many speech processing tasks, such as speaker recognition, speech recognition, and accent identification. The speakers were speaking in Modern Standard Arabic (MSA).**

Keywords-- Arabic speech, speech database, corpus design, phonetically balanced

I. INTRODUCTION

Speech database is a core part to evaluate the performance of the system in speech processing field. The developed system can be deployed successfully in real life only if it is evaluated by a versatile and relevant database. There are many databases in major languages like English, Spanish, German, Japanese, Chinese, etc. These databases are rich in number of speakers, amount of speech, variability of speakers and texts, environments, and transmission channels. However, Arabic databases are few in numbers and most of them are private. Therefore there is a need for publicly available comprehensive Arabic speech database.

A good speech corpus is a prerequisite for research on acoustic analysis, speech and speaker recognition, etc. While developing a speech corpus, the following consideration may be taken into account.

- Whether the corpora is designed for phonetic analysis, speech synthesis, speech recognition, or speaker recognition. Depending on the scope, the corpora can be designed accordingly.
- The corpus can have a variety of contents, for example, single digit, continuous digits, isolated words, phrases, sentences, paragraphs, etc.
- Words, phrases, sentences, etc. should be carefully chosen so that the distributions of phones are balanced. Script should contain all possible vowels, consonants, co-articulations, etc. [1, 2].
- The total number speakers should be enough to validate the experiment under study. These speakers should speak sufficient number of utterances. The speakers can be of wide range of age [3].
- The corpora may contain almost equal number of male and female speakers [4].
- The speakers can be chosen to cover different types of accents [5].
- Based on the target, the corpora may contain read or spontaneous or both types of speech [3].
- The utterances can be recorded in different types of acoustic environments, for example, sound proof room, office room, corridor, restaurant, street, inside vehicle, etc.
- Data can be collected with different types of microphones and transmission channels, for example, mobile phone, land phone, etc. [6, 7].
- Data may be collected in different sessions to observe the effect of intersession variability [3].
- The corpus needs to be large enough to divide it into training and testing sets to account different types of variability [3]. It is better that the experiments are close set.

In this paper, a rich Arabic database is described. It can be used in many applications related to speech/speaker recognition and even equally good for Arabic accent classification. Moreover, it is recorded in different environments by using various

978-0-7695-4414-4/11 $26.00 © 2011 IEEE
DOI 10.1109/AMS.2011.29

mediums/equipments to analyze their effects in speech/speaker recognition tasks.

The rest of the paper is organized as follows. Section 2 provides literature survey on speech database, Section 3 describes the proposed Arabic speech database, Section 4 gives a statistic of the developed Arabic speech database, Section 5 provides discussion, and finally, Section 6 draws some conclusions.

II. LITERATURE REVIEW

The NIST Year 2010 speaker recognition evaluation plan includes not only conversational telephone speech, but also read and conversational speech recorded in room microphone channel [8]. Therefore it is desirable that the speech corpus contains both read and conversational speech recorded in mobile and microphone channels.

TIMIT is one of the mostly used English databases with large number (630) of speakers with eight different dialects of American English [9]. The speakers read ten phonetically rich sentences each in sound proof booth. Switchboard I-II including NIST evaluation subsets (LDC) include large number of speakers recorded in different sessions [10]. The content is spontaneous text materials uttered in office and home environments.

A speech database in Castilian Spanish called AHUMADA is developed specifically to consider speaker variability and channel-dependent influences [11]. The database contains the following parameters: microphone and telephone channels; read and spontaneous speech; six different recording sessions; speech of different rate; fixed utterances and speaker specific utterances; etc. The sampling rate is 44.1 kHz.

POLYCOST is a telephone speech database consisting of different European languages [12]. This allows experiments on the effect of languages on automatic speech recognition (ASR) performances. The database contains recordings in ten recording sessions which were spread over three months with a minimum spacing of three days between the sessions. The utterances were recorded in room and office environments.

Some small databases are developed in Slovene-language at the University of Ljubljana [13]. The databases contain isolated words, broadcast news, diaphone, etc. For example, K211d contains isolated words, GOPOLIS contains read speech, VNTV consists of broadcast weather forecast, and VINDAT contains read and spontaneous speech.

There are some attempts to develop Arabic speech database. A comprehensive database designed for Saudi accented speech recognition was developed at King Abdulaziz City of Science and Technology, Saudi Arabia [14]. The name of the database is Saudi accented Arabic voice bank (SAAVB), but it is not publicly available. There are 1033 native speakers and the number of utterances for each speaker is 59. The database provides telephony speech.

BBN/AUB DARPA Babylon Levantine Arabic Speech and Transcripts contains spontaneous Arabic speech from Lebanon, Jordan, Syria, and Palestine people [15]. The dataset consists of 164 speakers: 101 males and 63 females.

NIST Mixer 4 and 5 corpora contain Modern Standard Arabic (MSA), Levantine Arabic (LA), and Egyptian Arabic (EGA) speech [16]. Arabic speech database based on Algerian native speakers (ALGASD) is developed in 2009 and applied to speech analysis on Algerian accents [18]. There are some other databases, most of them are private and not concentrating to MSA.

There is no publicly available Arabic speech database that contains Saudi speakers and speakers reside in Saudi Arabia.

III. THE PROPOSED ARABIC SPEECH DATABASE

The following subsections describe the content (text) of the database, number of speakers, recording setup, and recording environments.

A. The Text Corpus

The text used for recording consist of many lists each has its own benefits. In the following we will go over these lists:

1) Rich sentences: These are rich sentences suggested in [14]. The list was designed so that it contained different Arabic phonemes in different contexts. The different allophones of each phoneme were covered. The list contained 940 sentences. The list was divided randomly into 47 sub lists, each containing 20 sentences. Each sub list was checked if it contained all the phonemes. The randomization was repeated again to find new sub lists, and again we count the occurrence of all phonemes in every sub list. The optimal sub lists, obtained after 20 randomization, were chosen for the recording. Each of these contained all the phonemes.

2) Rich words: These are rich words suggested in [14]. The list consisted of 700 words. This list was divided into 35 sub lists randomly. Each sub list was checked if it contained all the phonemes. Randomization was repeated to find new sub lists, and again we count the occurrence of all phonemes in every sub list. The optimal sub lists, obtained after 20 randomization, were chosen for the recording.

3) Accent identifying sentences: These are two sentences that are suggested due to their ability to differentiate accents in [14].

4) Numbers: This list contained Arabic digit from zero to nine. These digits contained only 17 Arabic alphabets out of 28 but are used for its importance in many applications.

5) Phonetically distinctive words: We chose these were words from SAAVB [14]. They were selected because they contain nasal (e.g. /n/, /m/), fricative (e.g. / θ /, /x/), and vowel (e.g. /a/, /i/) phonemes which are closely related to the speaker characteristics and can help in recognizing the speaker identity.

6) Common words: The list contained 20 words. This list was designed so that it contains words that are used frequently in the everyday conservation. These words consist of almost all Arabic alphabets except two. Examples of some common Arabic words are [Hello], [yes], [no], [news], etc.

7) Paragraphs: Pronouncing paragraphs are different than pronouncing sentences or words. Therefore, two paragraphs were added in this list. The first paragraph was a verse from Quran (the holy book of Muslims). This verse contains all alphabets. The second paragraph was taken from a book of a famous writer. The paragraph was chosen because it includes all letters and easy to read by normal readers and is appealing to them.

8) Two lists from [17]: M. Boudraa et. al suggested 20 lists; each list contained 10 phonetically balanced sentences. From these 20 lists we chose 4 lists that were the easiest to read and did not contain strange words or words that may not be appropriate to include in the database. We fixed one list for all speakers. A second list was chosen randomly from the remaining three lists.

9) Question and answers: In this database it was not easy to record an online conversation. Hence we opted for something similar, which were answers by the volunteers for questions by team member. Samples of these questions were (What is your name? how is the weather today? what is your best food) all in Arabic language.

10) Speaker dependent and independent task: The text is assigned to the speakers in such a way that we can perform both speaker dependent and speaker independent ASR. Some of the texts are spoken by all the speakers and some texts are distributed among the speakers.

B. Speakers

The speakers were Saudis and non Saudis: Males and females. The non Saudis were Arabs and non Arabs. The non Arabs were chosen so that they could read Arabic language in an acceptable level. They were mainly from fourth level on the Arabic language institute at King Saud University. The non Saudis represented 27 nationalities.

C. The Recording Environments

The speaker recorded the lists in three different environments: office, sound proof room (Eckel CL-11), and restaurant. The system setups in the office and the restaurant were the same while that for the soundproof room was different. Recording in these environments produces a rich database that can be used for different investigations and researches.

The office environment is the common environment for work for many people. The restaurant is a place with noise. The sound room is to build a clean high quality database.

D. Recording Sessions

We have completed recording in one session that included three different environments. Second and third sessions are to record in one month apart. The recording in three different sessions enriches the database with session variability.

E. The recording System

The recording system was similar in the office and the restaurant, and different in the soundproof room.

1) Recording systems for Office and Restaurant: The office and restaurant system consisted of the following subsystems:

- Two professional microphones (SHURE Beta 58A) were connected to a high quality mixer (Yamaha MW12CX in office, Yamaha MW8CX in cafeteria) to be recorded in stereo.
- Medium quality microphone (Sony Dynamic microphone F-V220) was connected to a sound card (Sound Card Creative surrounding 5.1).

- Another medium quality microphone (Sony Dynamic microphone F-V220) went directly to the computer.
- Mobile (Nokia N97) originating call to a similar mobile which was connected to the sound card (Sound Card Creative surrounding 5.1)

The setup of microphones was the two high quality microphones were at the edge and the two medium quality mics on the middle. The distance between the medium quality mics was 10 cm apart to make sure they were not in front of the mouth of the speaker. The mouth of the speaker was on a level above all mics. The photo in Figure 1 shows this setup. In the figure we see that the speaker's mouth is little bit up of the level of the four microphones, and the speaker is holding mobile phone in his right ear.

2) Recording system for Soundproof Room: The soundproof room system consisted of the following subsystems:

- One Professional Side-Address condenser Microphone: SHURE PG42) was connected to a high quality mixer (Yamaha MW12CX) to record in stereo.
- High quality microphone (SHURE Beta 58A) was connected to a high quality mixer (Sound Card Creative surrounding 5.1).
- Mobile (Nokia N97) to the computer via a medium quality sound card (Creative).

The sampling rate in all of the above was 48 K sample/sec with 16 bit resolution.

F. Text Reading

Instead of reading the text from paper we designed the system so that the speaker would read the text from the screen. All the required lists were stored under the speaker code. The project team member navigated from a list to another in his screen instructing the speaker to read from another large screen at the correct time. Figure 2 shows this setup.

IV. STATISTICS

The first session of recording took place in three different environments, namely sound-proof room, office room, and restaurant. Every volunteer had to start his recording from office room environment and finish speaking all the scripts he was assigned to. Then he repeats the recording in the sound-proof room, and cafeteria, in sequence. The project team members responsible for recording the speech at different locations,

Figure 1. Microphones and mobile phone setup in office environment.

Figure 2. Setup of text reading from the screen

kept logs to make sure that every speaker recorded the three environments.

There were a total of 240 speakers. The speakers consist of 137 Saudi native male and 103 non-Saudi male. Among these non-Saudis, 42 were Arab natives and the rest were non Arabs. Average recordings per speaker in office, sound-proof, and cafeteria were 19, 18, and 16 minutes, respectively. The total size of the recorded data was 166 GB. Table 1 lists all these statistics.

V. DISCUSSION

The designed database is rich in many dimensions. It is rich in text, rich in environments, rich in recording system, in the country of origin, and in the mother language. This richness will make it very beneficial in many applications. For example the sound proof system

can be used to build a high quality data base for Saudis, Non-Saudi Arabs, and Arabic speaking non Arabs. This database is very important for building a speech recognition system that can be used by Saudis, Arabs, and non-Arabs. The difference in countries of origin can be used to study the effect of this feature in speech or speaker recognition. The recording in restaurant can be used to study the effect of noise in speech or speaker recognition. Recording using mobile can be used to study the effect of noisy channels, or the effect of missing speech, on speech and speaker recognition. The database can also be used to study if the recognition rate depends on the type of text (word, sentence, or paragraph). These are only examples of the usefulness of the database.

The first session of the recording has been just finished for the male speakers. Partial verification of the recording was carried while recording. Complete verification was started after finishing the recording. Table 1 below gives snapshot statistics of the first session of the database.

Table1. Statistics of the completed part of the Arabic Speech Database.

Number of Recording Locations	3
Number of speakers	240
Saudi native	137
Non Saudi: Arabs	42
Non Saudi: Non Arabs	61
Number of Different Nationalities: Arabs	9
Number of Different Nationalities: Non Arabs	18
Average recording in Office	19 min
Average recording in Cafeteria	16 min
Average recording in sound proof Room	18 min
Database Size	166 GB

VI. CONCLUSION

We described a rich Arabic speech database. The database contains sufficiently large number of male and female speakers and a variety of text materials. The database is to be recorded in three sessions and in three different environments per session. Different types of microphones and mobile phones are used as transmission channels. Once it is fully developed, it will contribute significantly to Arabic speech processing research. The first session of recording for the male speaker was completed and we gave snapshot statistics of this recording.

ACKNOWLEDGMENT

This work is supported by the National Plan for Science and Technology in King Saud University under grant number 08-INF167-02. The authors are grateful for this support.

REFERENCES

[1] M. R. Sambur, "Selection of acoustic features for speaker identification," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 23, pp. 176–182, 1975.

[2] J. J. Wolf, "Efficient acoustic parameters for speaker recognition," Journal of the Acoustical Society of America, Vol. 51, pp. 2030– 2043, 1972.

[3] D. A. Reynolds, "An overview of automatic speaker recognition Technology," Proc. IEEE international conference on acoustics, speech and signal processing, ICASSP'02, Vol. IV, pp. 4072–4075, May 2002.

[4] G. R. Doddington, M. A. Przybocki, A. F. Martin, and D. A. Reynolds, "The NIST speaker recognition evaluation overview: methodology systems, results, perspective," Speech Communications, 31, pp. 225–254, 2000.

[5] L. G. Kersta, "Voiceprint classification for an extended population," Journal of the Acoustical Society of America (A), Vol. 39, pp. 1239, 1966.

[6] D. A. Reynolds, "The effects of handset variability on speaker recognition performance: Experiment on the Switchboard corpus," Proc. IEEE international conference on acoustics, speech, and signal processing, ICASSP'96 pp. 113–116, May 1996.

[7] R. Norton, "The evolving biometric marketplace to 2006," Biometric Technology Today, 10(9), pp. 7–8, 2002.

[8] The NIST Year 2010 Speaker Recognition Evaluation Plan, available at http://www.itl.nist.gov/iad/mig/tests/sre/2010/NIST_SRE10_evalplan.r6.pdf

[9] John S. Garofolo, et al., "TIMIT Acoustic-Phonetic Continuous Speech Corpus," Linguistic Data Consortium, Philadelphia, 1993.

[10] D. Graff, K. Walker, and A. Canavan, "Switchboard-2 Phase II," Linguistic Data Consortium, Philadelphia, 1999.

[11] J. O. Garcia, J. G. Rodriguez, and V. M. Aguair, "AHUMADA: A large speech corpus is Spanish for speaker characterization and identification," Speech Communication, Vol. 31, pp. 255-264, 2000.

[12] D. Petrovska, "POLYCOST: a telephone speech database for speaker recognition," RLA2C, Avignon, France, pp. 211–214, 20–23 April 1998.

[13] F. Mihelic, J. Gros, S. Dobrisek, J. Zibert, and N. Pavesic, "Spoken Language Resources at LUKS of the University of Ljubljana," International Journal of Speech Technology, Vol. 6, pp. 221–232, 2003.

[14] M. Alghamdi, et al., "Saudi accented Arabic voice bank," J. King Saud University, CIS, pp. 1-15, 2007.

[15] J. Makhoul, et al., "2005 BBN/AUB DARPA Babylon Levantine Arabic Speech and Transcripts," Linguistic Data Consortium, Philadelphia, 2005.

[16] http://www.nist.gov/speech/tests/spk/index.htm

[17] M. Boudraa and B. Boudraa, "Twenty Lists of Ten Arabic Sentences for Assessment," Acustica. Acta Acustica, Vol. 86, pp. 870–882, 2000.

[18] D. R. Ghania, S. A. Selouani, and M. Boudraa, "Algerian Arabic speech database (ALGASD): corpus design and automatic speech Recognition application," The Arabian Journal for Science and Engineering, Volume 35, Number 2C, pp. 157-167, December 2010.