IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 May 2023, accepted 30 May 2023, date of publication 2 June 2023, date of current version 9 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3282259

TOPICAL REVIEW

Modern Standard Arabic Speech Corpora: A Systematic Review

AMMAR MOHAMMED ALI ALQADASI[1,2], RAWAD ABDULGHAFOR[1,3], (Member, IEEE), MOHD SHAHRIZAL SUNAR[4,5], (Member, IEEE), AND MD. SAH BIN H. J. SALAM[4]

[1]National Department of Computer Science, Faculty of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia
[2]Department of Information Technology, Al-Saeed Faculty of Engineering and Information Technology, Taiz University, Taiz, Yemen
[3]Faculty of Computer Studies (FCS), Arab Open University-Oman, Muscat 130, Oman
[4]Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[5]Media and Game Innovation Centre of Excellence, Institute of Human Centered Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia

Corresponding author: Ammar Mohammed Ali Alqadasi (ammaralqadasi2@gmail.com)

ABSTRACT Speech processing applications have become integral components across various domains of modern life. The design and preparation of a reliable recognition system rely heavily on the availability of suitable speech databases. While numerous speech databases exist for English and other languages, the availability of comprehensive resources for Arabic language remains limited. In light of this, we conducted a systematic review aiming to identify, analyse, and classify existing Modern Standard Arabic speech databases. Through our review, we identified 27 publicly available databases and analysed an additional 80 subjective databases. These databases were thoroughly studied, classified based on their characteristics, and subjected to a detailed analysis of research trends in the field. This paper provides a comprehensive discussion on the diverse speech databases developed for various speech processing applications. It sheds light on the purposes and unique characteristics of Arabic speech databases, enabling researchers to easily access suitable resources for their specific applications. The findings of this review contribute to bridging the gap in available Arabic speech databases and serve as a valuable resource for researchers in the field.

INDEX TERMS Speech corpus, speech database, modern standard Arabic, MSA corpora, speech recognition, Arabic recognition.

I. INTRODUCTION

The achievement of Automatic Speech Processing (ASP), machine learning, and Deep learning led to an increase in the need for large-scale corpora (speech databases). Such database collection is often a difficult job, however, as it involves important investment involving time and money costs [1].

For this reason, speech databases have been collected for major languages like English, Spanish, German, Japanese, Chinese, Dutch, French, Indian languages, etc. () [2].

The Arabic language is a Semitic language spoken throughout the globe by more than 420 million people and as the language of the Qur'an, is the religious language of Muslims everywhere [3]. In the linguistic globe of today, when someone talks of "Arabic," they likely refer to Modern Standard Arabic, not Classical Arabic [4].

On the other extreme, in spite of the fact that Arabic is the world's most significant language, Arabic is geographically one of the most widely used world languages [5]. However, there are few Arabic speech databases and it is not rich [6], [7]. Therefore, a rich and detailed Arabic-speaking database that is publicly available is needed. Hence the databases can improve the quantity and quality of studies in Arabic language processing considerably [6], [7]. Actually, the quantity of work in the Speech domain for Arabic languages has not yet reached a critical level to be used as a true tool in speech applications, as in other developed countries' languages. This is because there are no rich databases in Arabic, unlike other languages that have a rich speech database. For example, one well-known and widely used speech database is TIMIT [8]. It consists of 630 native

The associate editor coordinating the review of this manuscript and approving it for publication was Mounim A. El Yacoubi.

speakers of American English, of whom 70% are male and 30% female. Each speaker reads 10 sentences in one session which takes approximately 30 seconds [9].

A Speech Corpus is a database of speech audio files and text transcriptions of these audio files in a format that can be used to create Acoustical Models () [2].

In this regard, some previous studies have conducted a review of English corpora and other languages, even languages less known than Arabic, such as Hindi, Amharic, and others. The studies have reported challenges related to corpora information extraction, such as inaccessibility and disorganization on personal and research group sites [10].

Determining the size of the database and the diversity of the characteristics of the speakers is also important, especially for AI and NLP projects, where a larger group is useful due to its diversity and representation of the language being studied. This necessitates database analysis by researchers to determine a database's qualities and usefulness for their research.

To the best of our knowledge, no reviews have previously looked at the PRISMA-compliant literature to Available Arabic Corpora: A Scoping Review. For example, Ahmed et al. [11] in their review surveyed the grammar of the Arabic language available for free, whether written or spoken; however, only three speech databases were discussed: ADI17 with 3033 h of Dialect Arabic, ARAB-ANDALUSIAN with 125 h of Moroccan tradition of Arab-Andalusian music and Arpod with 8.1 h of Speech data. It is noted that the three databases discussed in this review are not within the scope of our review, as they are not standard Arabic language databases.

The field of speech processing relies heavily on the availability of appropriate speech corpora, especially for Modern Standard Arabic (MSA). However, accessing and understanding the specifications of existing MSA speech databases can be challenging for researchers in the field. Therefore, the primary objective of this study is to comprehensively explore accessible MSA speech corpora, conduct an in-depth analysis, classify them based on their characteristics, and provide detailed descriptions. This effort aims to facilitate researchers in the field of speech processing in identifying and accessing the most suitable databases for their research needs.

Additionally, this systematic review encompasses an examination of previous works that utilized these databases for speech processing applications. By compiling and analyzing such studies, this review serves as a valuable resource that enables researchers to easily access not only the MSA speech databases but also the studies that have described, evaluated, or applied these databases.

Furthermore, we extend the scope of this review to include studies within the field of Arabic speech processing that have developed their own databases for training and testing purposes. The inclusion of these studies strengthens the comprehensiveness of our review and provides researchers with a broader perspective on the available resources in the field.

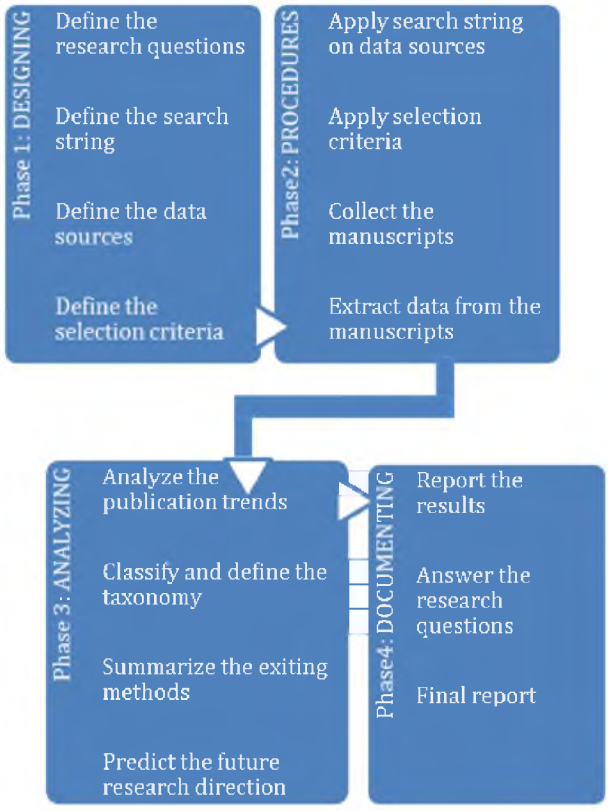

FIGURE 1. Review methodology and procedures.

Motivated by these three objectives, we conducted an extensive literature survey spanning from 2000 to 2023 to review and analyze Arabic phonetic databases, as well as studies that have utilized existing databases or created their own. The scientific contribution of this review lies in its systematic approach, comprehensive analysis, and provision of valuable insights into the landscape of MSA speech corpora. By addressing the limitations in accessing and understanding these databases, this paper aims to have a significant impact on the field of speech processing and contribute to advancements in Arabic speech recognition, synthesis, and related applications.

This paper is organized as follows: The review process methodology is presented in Section II; Section III presents the results obtained to answer the pre-determined research questions; the future research directions are discussed in Section IV; and the paper is concluded in Section V.

II. REVIEW METHODOLOGY

When developing this review, we are inspired by the methodology described by [12] and [13] which follows the guidelines recommended by [14] and [15] in which four phases are to be executed in the review process: (i) designing, (ii) conducting, (iii) analyzing, and (iv) documenting. Figure 1 depicts all of the procedures involved as well as the results of each phase.

A. PLANNING AND DESIGNING THE REVIEW

The goal of this phase is to define and outline a strategy for creating the review protocol, including research questions (RQ), search terms, data sources, and selection criteria.

This review is expressed in the following research questions in order to reach our research goal:

RQ1. What are the trends in published research studies in the field of construction and applications of spoken modern standard Arabic (MSA) databases? Objective: to categorize primary studies in assessing distributions of publication types, years, and venues based on the number of primary studies.

RQ2. What are the sources and methods of collecting Arabic data? Objective: To classify MSA speech databases based collection methods to give a general picture of data collection methods and the characteristics of each database.

RQ3. What are the most important applications and purposes for which the databases were collected? Objective: To determine the trends of the studies for the MSA speech

databases used to summarize the main characteristics of the Arabic databases based on their purpose, thus making it easier for the researcher to choose the appropriate database.

RQ4. What are the most important features of MSA speech corpora in terms of the number of speakers, the size of the database, corpus prompts, and finally the diversity in the ages, sex and nationalities of the speakers and the recording environment? Objective: To identify the most important features of the available speech Modern Standard Arabic language databases and self-datasets, and to summarize the main features of the MSA databases based on the diversity of the number, type, and ages of speakers, which makes it easier for the researcher to choose the appropriate database based on their features.

RQ5. What are the future research directions regarding the collection and construction of Arabic databases? Objective: To identify research issues and gaps related to MSA language databases.

This review aims to get useful answers for researchers who want to contribute more to the field of speech processing, as well as researchers to understand the available MSA speech databases, characteristics, design purposes, methods of collection, and how it can be used in speech processing applications, in addition to knowing the correct methods to build an Arabic speech standard database.

To define the search terms, we used a search string based on a set of important keywords. These include Arabic OR Arabic OR MSA OR Modern Standard Arabic AND Database OR Dataset OR Corpus OR Corpora.

Regarding the data sources, we identified several online digital databases and search engines that can be used to search and collect manuscripts as the primary studies in the literature review process. These include IEEE Xplore Digital Library, ScienceDirect, Scopus, Springer Link, Web of Science, Google Scholar, ResearchGate, and ACM Digital Library, in addition, a review of Arabic databases in Linguistic Data Consortium (LDC) and Open Speech and Language Resources (Open SLR).

Inclusion and exclusion criteria have been applied to manuscripts where manuscripts in the form of journal papers, conference papers, books, book chapters, technical reports, and theses in the scope of Arabic speech processing are included.

Manuscripts focused on textual or visual databases were excluded. Moreover, there were restrictions on the year of publication, as manuscripts published in 2000 and above were included.

B. REVIEW PROCEDURE

In this stage, the literature search, selection, collection, and data extraction are implemented based on several definitions in the previous stage. It is a straightforward process where the research activity begins by applying the selected search words to online digital databases and search engines to obtain candidate studies.

The inclusion and exclusion selection criteria are applied to candidate studies to filter and select relevant studies. After removing the duplication of relevant studies from different data sources, manuscripts are collected.

Based on the final selection process, we have successfully collected 229 manuscripts, 27 of which are documented databases from LDC, Open SLR, universities or scientific institutions, while the rest are distributed between manuscripts that used these databases and others that constructed their databases.

We sorted the primary studies in ascending order based on the year of publication and the name of the first author, then baseline characteristics are extracted from the primary studies and stored in tabular form. These include the year of publication, type, author name, manuscript title, problem to be addressed, database type, characteristics of speakers (number, sex, and age), database size, collection methods, collection resources, and database applications (see Appendix I, II, III).

C. REVIEW ANALYSIS

This stage is done by analyzing the data extracted from the previous stage to collect some information so that it can be used to answer our research questions, which cover Publication trends regarding the spoken modern standard Arabic language databases collected by classifying primary studies and Arabic databases.

A summary of sources and methods collection of Arabic speech databases to give a general picture of data collection methods and the characteristics of each database.

Determining the directions of MSA speech databases based on their purpose makes it easier for researchers to choose the appropriate database.

Future research directions regarding the MSA speech database collection and construction. This was done by investigating the primary studies and categorizing them based on the year of publication, their types, and their applications.

D. DOCUMENTING THE REVIEW

In the final stage, the final report was written to present the results obtained from the analysis stage, and answers to the research questions mentioned in the planning stage were provided.

III. RESULTS

This section presents the results of the review, the details of which are provided in the following subsections.

A. SEARCH RESULTS

An initial search of bibliographic databases returned a total of 1264 citations as shown in Figure 2 Among them, we found 336 duplicate citations which were therefore removed. Of the remaining 928 manuscripts, 492 were not related to the speech databases of Modern Standard Arabic according to the inclusion criteria of this review, thus 280 manuscripts were excluded because there is no discussion or application database. 59 manuscripts contained non-Arabic databases, and 33 manuscripts discussed multiple languages, the database information was not mentioned in 68 manuscripts and 52 manuscripts are not speech databases. Moreover, 76 manuscripts that are not related to the language form included in this review, which is Modern Standard Arabic, were excluded, therefore, 76 studies of classical Arabic databases and 111 studies of Arabic dialects databases were excluded. Finally, 219 studies were selected after reading the full texts. 27 publicly accessible databases were included from Database Sources.

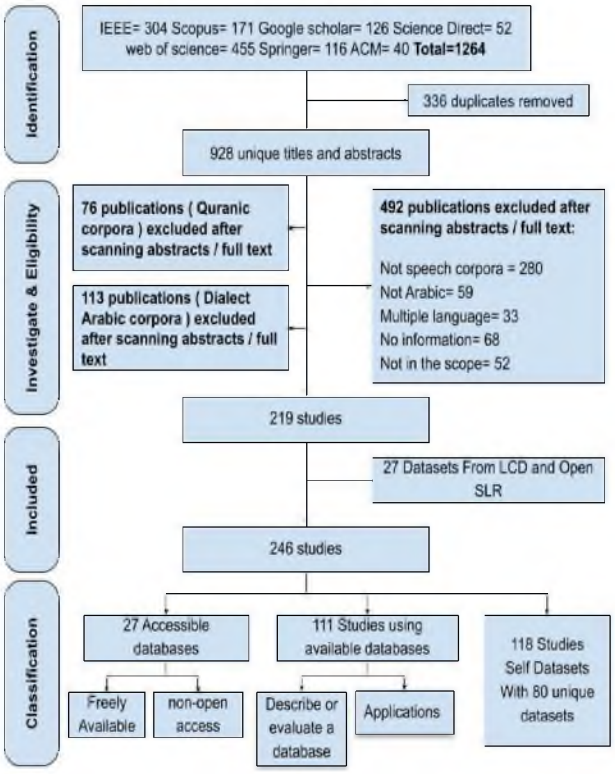

FIGURE 2. Selection of relevant publications process flow chart.

B. ANALYSIS OF THE PUBLICATION TRENDS

This section presents the analysis of publication trends in response to the first research question (RQ1), focusing on the 219 primary studies. The trends are examined based on the classification of publication type and year. Figure 3 illustrates the number of manuscripts based on different publication types. It can be observed that conference papers are the major contributors in the primary studies with 59%, followed by 37% of journal papers and 4% of other types of published manuscripts.

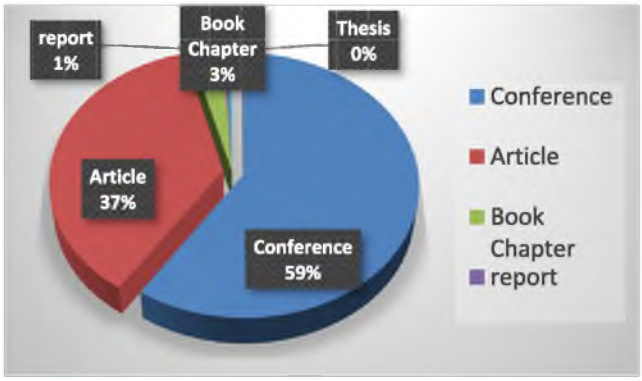

FIGURE 3. Distribution of the publication per type.

Figure 4 depicts the publishing trend for a period of 24 years, from 2000 to 2023, with relation to the distribution of publications about MSA speech databases as well as studies that described the usage and application of these databases in various fields of speech processing. As can be seen, there are 9.125 publications on average per year, with an upward trend peaking in 2013 when there were 19 publications total. Later, the trend fluctuated and the number of manuscripts published increased to 17 in 2017 and 18 in 2021. In light of this, the result supports the growing demand for speech databases for the Modern Standard Arabic language and its applications.

Figure 5 also shows the distribution of the publicly accessible Standard Arabic speech databases and the publication years. It is noted that the years 2006, 2007 and 2013 reached the highest publication number of the databases, as the number of publicly accessible published databases was 3 for each year. It is also noted that the frequency of publishing publicly accessible databases after 2017 have decreased, as only one database was published in 2018 and one in 2020, and no MSA speech databases were published after that.

C. MSA SPEECH DATABASE COLLECTION METHODS AND RESOURCES

In order to answer (Q2) and to find out the resources of Arabic databases and the methods of collecting them, in this review we have classified the publications into three categories: 1) publicly accessible databases that were often

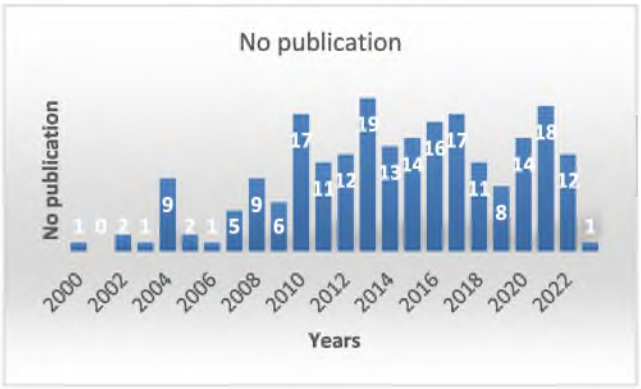

FIGURE 4. Distribution of the publication per year.

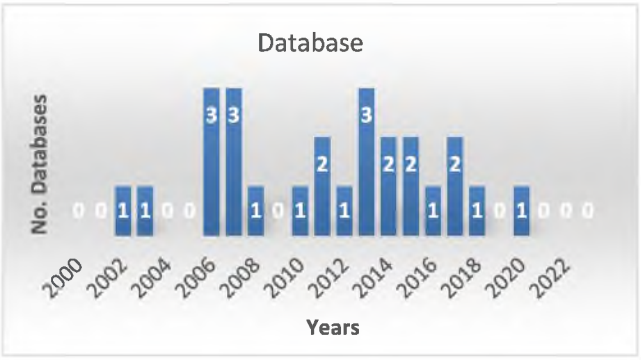

FIGURE 5. Distribution of the available MSA speech corpora per year.

collected directly from the participants (speakers) by direct recording or through other collecting method like Web-based resourcing, spontaneous telephone conversations, telephone responses of questionnaires, social media and recording broadcast, 2) studies that implement the publicly accessible MSA speech databases, 3) studies that design and construct their own MSA speech database to be used in training and testing. They also collect data from participants and are often small databases, noting that some databases can be considered Rich and standard.

Accordingly, 27 publicly accessible databases were selected, 101 manuscripts used publicly accessible databases in speech processing applications, and finally, 118 manuscripts built and designed their databases, but they are not available for researchers.

The first category was referred to as publicly accessible databases, the second category as applied studies, and the last category as self-databases.

1) PUBLICLY ACCESSIBLE MODERN STANDARD ARABIC SPEECH DATABASES

In this study, 27 Modern Standard Arabic speech databases were investigated. These databases can be accessed for free or for a charge of between $50 and $2,000, and are accessible to researchers in academic and research institutions. Table 1 shows these datasets and their publishers.

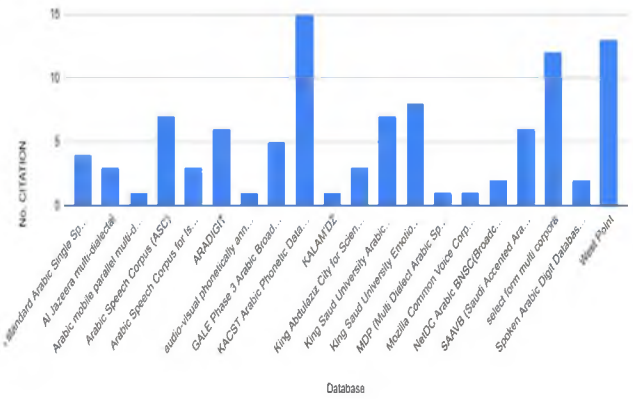

FIGURE 6. Frequency distribution of the accessible modern standard arabic speech databases.

The first set of speech corpora have taken advantage of the limited solution of telephony conversation recording. Development of the pioneer Arabic Dialects corpus began in the mid-1990s and it is CALLFRIEND Egyptian [16], [17]. OrienTel is also considered to be one of the first databases that have collected speech database for the Arabic Standard (MSA) and Arabic dialects of Palestine, Jordan, Egypt, Saudi Arabia, the UAE, Tunisia, Morocco, and United Arab Emirates countries [1], [7], [18], [19] also used the same telephone response to the questionnaire technique.

2) APPLIED STUDIES OF PUBLICLY ACCESSIBLE MODERN STANDARD ARABIC SPEECH DATABASES

In this review, we studied 101 manuscripts that used one of the available databases mentioned in Table 1. Table 2 shows the sources of the databases used in these manuscripts, and Figure 6 shows the frequency distribution of the Accessible Modern Standard Arabic Speech Databases.

3) SELF-DATASETS OF SPEECH MODERN STANDARD ARABIC

118 manuscripts were accessed during this review. The researchers of these publications built their databases, and when studying and analyzing these manuscripts, we found 80 different databases. Some of these databases are rich and balanced and can be considered standard databases, but to the best of our knowledge, they are not available to researchers, as we could not access them during our research.

Figure 7 shows that the researchers collected this data in several methods: 49 databases were collected by direct recording, 15 databases by choosing from other Arabic speech corpora, 9 databases were selected from TV broadcasting, and 2 databases for remote recording by fixed-line telephone network or Mobile and one database was collected from the web, while 5 databases lacked details on the method of data collection. 118 manuscripts were accessed during this review. The researchers of these publications built 80 unique

TABLE 1. Publicly accessible modern standard arabic speech databases * freely available for academic and research purposes, ELRA (European Language Resources Association), LDC (Linguistic Data Consortium).

Corpus Name	publisher	Ref	Available database
A standard Arabic Single Speaker Corpus (SASSC)	KACST	[20]	*
Arabic Broadcast News Speech	LDC	[21]	LDC2006S46
Arabic Learner Corpus	LDC	[22]	LDC2015S10
Arabic Speech Corpus (ASC)	ELRA	[23]	ELRA-S0384
Arabic Speech Corpus for Isolated Words	University of Stirling	[24]	ASCIW
GALE Phase 2 Arabic Broadcast News & Conversation Speech	LDC	[25]	LDC2013S02
GALE Phase 3 Arabic Broadcast News & Conversation Speech	LDC	[26]	LDC2016T17
GALE Phase 4 Arabic Broadcast News & Conversation Speech	LDC	[27]	LDC2013S02
KACST Arabic Phonetic Database (KAPD)	KACST	[28]	*
King Abdulaziz City for Science and Technology Text to Speech Database (KTD)	KACST	[29]	*
King Saud University Emotions (KSUEmotions) corpus	LDC	[30]	LDC2017S12
MSA Speech Corpus	UM and University of Jordan	[31]	MSASC
NEMLAR Broadcast News	ELRA	[32]	ELRA-S0219
NEMLAR Speech Synthesis Corpus	ELRA	[33]	ELRA-S0220
NetDC Arabic BNSC(Broadcast News Speech Corpus)	ELRA and LDC	[34]	ELRA-S0157
OrienTel Egypt MSA	ELRA	[35]	ELRA-S0222
OrienTel Jordan MSA	ELRA	[36]	ELRA-S0290
OrienTel Morocco MSA	ELRA	[37]	ELRA-S0184
OrienTel Tunisia MSA	ELRA	[38]	ELRA-S0187
Phonemes of Arabic	LDC	[39]	LDC2020S13
Spoken Arabic Digit Database (SAD)	http://www.timeseriesclassification.com	[40]	SAD
West Point	LDC	[41]	LDC2002S02
Al Jazeera multi-dialectal	Hamad Bin Khalifa University	[42]	aljazeeraSpeechCorpus
MDP (Multi Dialect Arabic Speech Parallel Corpora)	...	[43]	*
Arabic mobile parallel multi-dialect speech	Qassim University	[44]	*
King Saud University Arabic Speech Database (KSU)	LDC	[45]	LDC2014S02
Multi-Genre Broadcast (MGB-3)	Arabic speech community	[46]	mgb3

databases, which we named DB 01, DB 02,. . ., and DB 80 and listed them in Table 3 along with manuscripts that used them.

TABLE 2. Applied studies list OF PUBLICLY accessible modern standard arabic speech databases.

Database	PUBLICATION	No. CITATION
SASSC	[20], [47], [48], [49]	4
Al Jazeera multi-dialectal	[50], [42], [51]	3
Arabic mobile parallel multi-dialect speech	[52]	1
ASC	[53], [54], [55], [56], [57], [58], [59]	7
Arabic Speech Corpus for Isolated Words	[60], [61], [62]	3
ARADIGIT	[63], [64], [65], [66], [67], [68]	6
audio-visual phonetically annotated Arabic corpus	[69]	1
GALE Phase 3	[70], [71], [72], [73], [74]	5
KAPD	[75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89]	15
KALAM'DZ	[90]	1
KTD	[91], [92], [93]	3
KSU Database	[6], [94], [7], [95], [96], [97], [98]	7
KSU Emotions corpus	[99], [100], [101], [102], [103], [104], [105], [106]	8
MDP	[107]	1
Mozilla Common Voice Corpus 12.0 (MCV) corpus	[108]	1
NetDC Arabic BNSC	[109], [109]	2
SAAVB (Saudi Accented Arabic Voice Bank)	[110], [111], [112], [113], [9], [114]	6
select form multi corpora	[115], [116], [117], [118], [119], [120], [121], [122], [123], [124], [125], [126]	12
SAD	[127], [128]	2
West Point	[129], [130], [131], [132], [133], [134], [135], [136], [137], [138], [139], [140], [141]	13

It should be mentioned that some of these databases can be regarded as rich databases because they contain many speakers or are large in size, and some of these databases have been used in numerous papers in various applied fields.

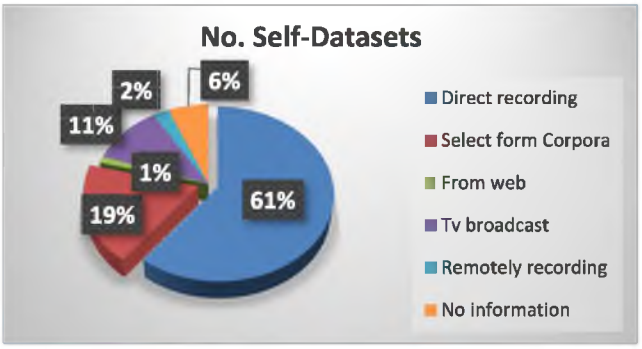

FIGURE 7. Collecting methods of datasets (self-datasets).

D. ARABIC SPEECH CORPORA ACCORDING TO PURPOSE OF COLLECTION

This section aims to answer the third question, Bougrine, et al. [1] claim that the applications using speech corpora can be classified into four main categories: speech recognition/verification, speech synthesis, speaker recognition/verification, and other spoken-language applications. Nevertheless, researchers have collected databases for a variety of purposes like general purpose, Speech synthesis / TTS, Speech recognition, speaker recognition, emotional speech recognition, specific domain, educational purposes, medical applications, language/dialect identification and speaker sex/age identification.

The developed speech databases are either for general purpose applications or for a particular application. Most speech databases that were studied were used for automatic speech recognition and Text to Speech Synthesis Systems. An example of this would be the speech Arabic Text-to-Speech (TTS) databases. The Linguistic Data Consortium (LCD) has a collection of MSA and dialectical Arabic speech and text datasets for research and development purposes.

Figure 8 shows the purposes that have been announced by the developers of the publicly Accessible Modern Standard Arabic Speech Databases. It turns out that most of the databases were created and designed for speech recognition, as 16 databases are available for speech recognition, while 4 databases were established for speech synthesis and 4 general purpose databases, and the rest of the rules are distributed to speaker recognition and those which identify feelings, educational applications and limited applications (travel and tourism). Details of the applications of each database are in Table 4.

The purposes that were declared by the researchers in their studies in which they used available databases were among these purposes consistent with the purposes declared by the developers of the database, including studies that were carried out for additional purposes. Figure 9 shows the distribution of these studies based on purposes. The distribution, in order, were as follows: speech recognition with 37 manuscripts, Corpus description/evaluation

TABLE 3. List of self-datasets of speech modern standard arabic.

Database No	Developed by	Manuscripts	Database No	Developed by	Manuscripts
DB 01	[142]		**DB 41**	[143]	
DB 02	[144]	[145]	**DB 42**	[146]	
DB 03	[147]	[148]	**DB 43**	[43]	
DB 04	[149]		**DB 44**	[44]	
DB 05	[150]		**DB 45**	[151]	
DB 06	[152]		**DB 46**	[153]	
DB 07	[154]		**DB 47**	[155]	
DB 08	[156]		**DB 48**	[157]	
DB 09	[158]		**DB 49**	[159]	
DB 10	[160]	[161] , [162]	**DB 50**	[163]	
DB 11	[164]		**DB 51**	[165]	
DB 12	[166]		**DB 52**	[167]	
DB 13	[168]		**DB 53**	[169]	
DB 14	[170]		**DB 54**	[171]	[172] , [173]
DB 15	[174]		**DB 55**	[175]	
DB 16	[176]		**DB 56**	[177]	
DB 17	[178]		**DB 57**	[179]	
DB 18	[180]		**DB 58**	[181]	[182]
DB 19	[183]		**DB 59**	[184]	
DB 20	[185]		**DB 60**	[186]	
DB 21	[187]		**DB 61**	[188]	[189] , [190] , [191] , [192]
DB 22	[193]		**DB 62**	[194]	[195] , [196] , [197] , [198] , [199] , [194] , [200] , [201] , [202]
DB 23	[203]		**DB 63**	[204]	
DB 24	[205]		**DB 64**	[206]	
DB 25	[207]		**DB 65**	[208]	
DB 26	[209]		**DB 66**	[210]	
DB 27	[211]		**DB 67**	[212]	
DB 28	[213]	[214]	**DB 68**	[215]	
DB 29	[216]	[216]	**DB 69**	[217]	
DB 30	[218]		**DB 70**	[219]	
DB 31	[220]		**DB 71**	[221]	
DB 32	[222]	[223]	**DB 72**	[224]	

TABLE 3. *(Continued.)* List of self-datasets of speech modern standard arabic.

DB 33	[225]		**DB 73**	[226]	
DB 34	[227]	[228] , [229]	**DB 74**	[230]	[231] , [232]
DB 35	[233]		**DB 75**	[234]	
DB 36	[235]		**DB 76**	[236]	
DB 37	[237]	[238] , [239] , [240], [241] , [31] , [242]	**DB 77**	[243]	
DB 38	[244]		**DB 78**	[245]	
DB 39	[246]	[247] , [248] , [249] , [250]	**DB 79**	[251]	
DB 40	[252]		**DB 80**	[253]	

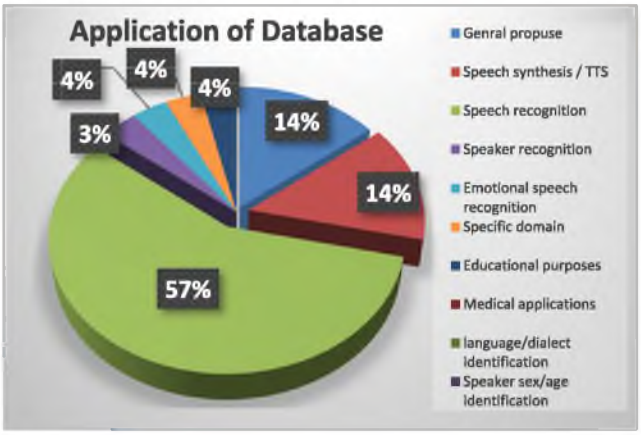

FIGURE 8. Applications of publicly accessible modern standard arabic speech databases.

TABLE 4. Lists of publicly accessible modern standard arabic speech databases based on their applications

Application of Database	Ref. Database
General purpose	[21], [22], [42], [45]
Speech Synthesis / TTS	[23], [28], [29], [33]
Speech recognition	[24], [25], [26], [27], [32], [34], [35], [36], [37], [38], [39], [20], [40], [41], [44], [46]
Speaker recognition	[31]
Emotional speech recognition	[30]
Specific domain	[43]
Educational purposes	[20]

with 17 manuscripts, General purpose with 10 manuscripts, Speaker recognition with 8 manuscripts, Emotional speech recognition with 7 manuscripts, while the rest of the

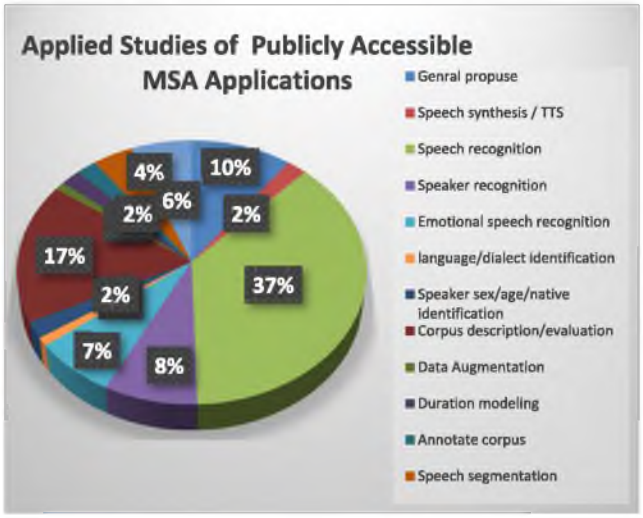

FIGURE 9. Applications of applied studies of publicly accessible modern standard arabic speech databases.

TABLE 5. Lists of applied studies based on their applications.

Application of Database	Ref. Database
General purpose	[72], [73], [74], [102], [107], [114], [126], [130], [131], [133]
Speech Synthesis / TTS	[55], [98]
Speech recognition	[48], [49], [51], [52], [58], [59], [60], [61], [62], [68], [70], [76], [77], [85], [87], [89], [92], [94], [111], [112], [113], [117], [119], [120], [121], [122], [123], [124], [125], [127], [128], [129], [132], [136], [139], [140], [141]
Speaker recognition	[63], [64], [65], [66], [67], [95], [97], [110]
Emotional speech recognition	[91], [93], [99], [101], [104], [105], [106]
language/dialect identification	[108]
Speaker sex/age/native identification	[137], [138]
Corpus description/evaluation	[20], [42], [56], [57], [69], [79], [86], [6], [7], [100], [103], [109], [109], [9], [115], [118], [134]
Data Augmentation	[47]
Duration modelling	[54], [135]
Annotate corpus	[50], [90]
Speech segmentation	[71], [75], [82], [88]
Speech Classification	[78], [80], [81], [83], [84], [96]

manuscripts are distributed among the rest of the purposes. Details of the applications of each rule are in Table 5.

As for the self-databases, it did not differ significantly with regard to a more frequent order of purposes than the databases, where the purpose of 55 manuscripts was speech recognition and then, respectively, Corpus description/evaluation with 27 manuscripts, Speech synthesis / TTS with 11 manuscripts, Speech Classification with 7 manuscripts. Figure 10 shows the frequency distribution

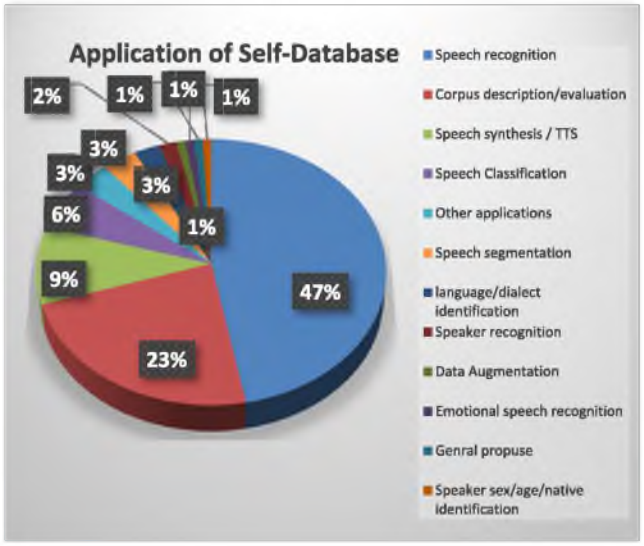

FIGURE 10. Applications of self-database.

TABLE 6. Lists of Self-DATABASES based on their applications.

Application of Database	No. Database
Speech recognition	[164], [166], [170], [176], [185], [187], [207], [209], [211], [214], [216], [220], [216], [222], [223], [227], [228], [229], [238], [241], [244], [248], [252], [249], [143], [153], [155], [157], [163], [167], [169], [173], [182], [181], [184], [189], [188], [195], [196], [201], [204], [206], [190], [191], [192], [217], [219], [221], [231], [234], [230], [232], [251], [253], [215]
Corpus description/evaluation	[142], [149], [148], [158], [160], [162], [168], [174], [183], [205], [213], [218], [235], [31], [239], [240], [237], [246], [247], [146], [43], [44], [177], [186], [194], [194], [200]
Speech Synthesis / TTS	[144], [147], [145], [152], [154], [156], [161], [178], [208], [212], [245]
Speech Classification	[150], [233], [159], [165], [197], [199], [202]
Other applications	[180], [172], [198], [224]
Speech segmentation	[225], [151], [171], [250]
language/dialect identification	[193], [226], [236]
Speaker recognition	[242], [179]
Data Augmentation	[175]
Emotional speech recognition	[243]
General purpose	[210]
Speaker sex/age/native identification	[203]

of autonomous databases based on their applications, and the details are in Table 6.

IV. DATABASE CHARACTERISTICS AND VARIETY

The fourth question in this review: What are the most significant characteristics of MSA speech corpora in terms of the number of speakers, the size of the database, the corpus prompts, and finally the diversity in the ages, sex, and nationalities of the speakers and the recording environment? This section addresses the answer to this question.

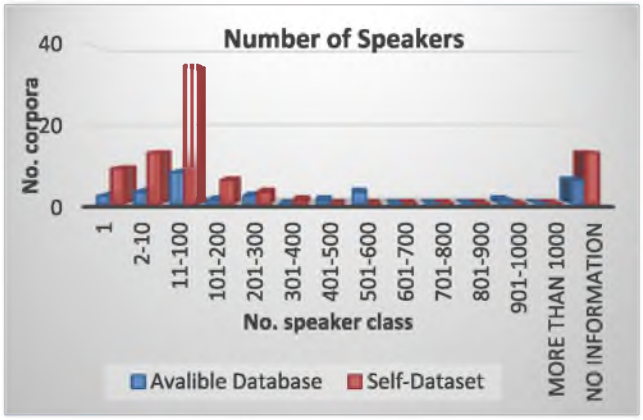

FIGURE 11. Distribution of the databases per speaker's numbers.

A. NUMBER OF SPEAKERS

The number of speakers is an important factor in many speech processing applications such as speaker recognition, as well as speaker-independent (SI) speech recognition, where the most important in these applications is the number of speakers in the database to be used in training and testing to recognize the speaker or speech.

The available databases, as well as the self-databases, varied in the number of speakers, as some of them contain one speaker and some contain a few or many speakers, and some databases did not announce the number of speakers. See Figure 11.

We observe that the majority of publicly accessible MSA databases contain fewer than 100 speakers, followed by databases that did not reveal the number of speakers. However, there are four databases with a reasonably high number of speakers—between 400 and 600—which are OrienTel-issued databases. Arabic Learner Corpus [22] announced the largest number of speakers, which is 942 speakers.

Likewise, most self-databases contain less than 100 speakers, followed by databases that did not disclose the number of speakers. Note that the largest self-database in terms of the number of speakers did not exceed 400 speakers.

B. SPEAKERS' SEX

Knowing the sex of the speakers is important in many speech processing applications such as identifying the speaker's sex applications. Some applications also require diversity of the sex of the speakers to know the ability of the model to recognize speech for both males and females. Some researchers have stated that the accuracy of speech recognition is affected by the sex of the speakers in speech recognition and showed that there is a difference in the accuracy of speech recognition based on the sex of the speaker [72].

TABLE 7. Publicly accessible MSA database information on the number of speakers.

No. of Speakers Class	Available Database Name	Ref
1	SASSC, ASC	[20], [23]
2-10	NEMLAR Speech Synthesis Corpus, KTD, KAPD	[33], [29], [28]
11-100	Phonemes of Arabic, KSU Emotions corpus, MSA Speech Corpus, Arabic Speech Corpus for Isolated Words, MDP, Arabic mobile parallel multi-dialect speech, SAD, NetDC Arabic BNSC	[39], [30], [31], [24], [43], [44], [40], [34]
101-200	West Point	[41]
201-300	NEMLAR Broadcast News, KSU Database	[32], [45]
301-400	...	...
401-500	OrienTel Egypt MSA	[35]
501-600	OrienTel Morocco MSA, OrienTel Jordan MSA, OrienTel Tunisia MSA	[37], [36], [38]
601-900	...	...
901-1000	Arabic Learner Corpus	[22]
No information	Arabic Broadcast News Speech, GALE Phase 2, GALE Phase 3, GALE Phase 4, Al Jazeera multi-dialectal, MGB-3	[21], [25], [26], [27], [42], [46]

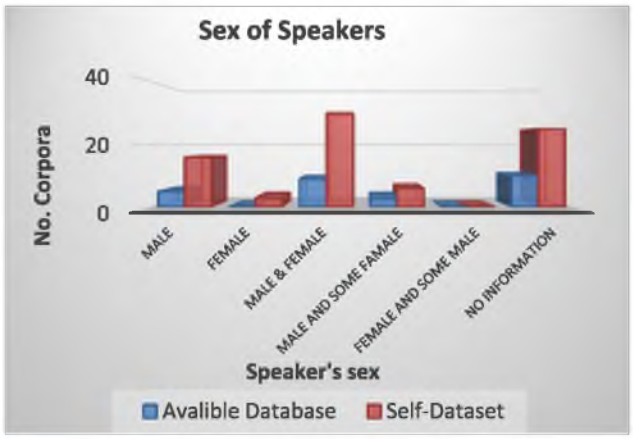

FIGURE 12. Distribution of the databases per speaker's sex.

It should be noted that male speakers are frequently more prevalent in databases than female speakers. In publicly accessible MSA Speech Databases, we discovered that 10 databases did not provide information about the speaker's sex, followed by the databases containing both sexes, then the databases containing only males, with no database for female speakers only. As for the self-databases, the majority of them contain both sexes, followed by those databases that did not mention information about the sex of the speaker, then the databases that contain only males, and three databases contain only female speakers. See Figure 12.

TABLE 8. MSA self-database information on the number of speakers.

No. of Speakers	Self-Dataset Name	Ref
1	DB 01, DB 02, DB 03, DB 04, DB 05, DB 06, DB 07, DB 08, DB 78	[142], [144], [147], [149], [150], [152], [154], [156], [245]
2-10	DB 09, DB 10, DB 11, DB 12, DB 13, DB 14, DB 15, DB 16, DB 17, DB 18, DB 19, DB 20, DB 21	[158], [160], [164], [166], [168], [170], [174], [176], [178], [180], [183], [185], [187]
11-100	DB 22, DB 23, DB 24, DB 64, DB 25, DB 26, DB 27, DB 28, DB 29, DB 30, DB 31, DB 32, DB 33, DB 34, DB 35, DB 36, DB 37, DB 38, DB 39, DB 40, DB 41, DB 42, DB 43, DB 44, DB 45, DB 46, DB 47, DB 48, DB 49, DB 50, DB 51, DB 52, DB 53, DB 54, DB 80	[193], [203], [205], [206], [207], [209], [211], [213], [216], [218], [220], [222], [225], [227], [233], [235], [237], [244], [246], [252], [143], [146], [43], [44], [151], [153], [155], [157], [159], [163], [165], [167], [169], [171], [253]
101-200	DB 55, DB 56, DB 57, DB 58, DB 59, DB 68	[175], [177], [179], [181], [184], [215]
201-300	DB 60, DB 61, DB 62	[186], [188], [194]
301-400	DB 63	[204]
No information	DB 65, DB 66, DB 67, DB 69, DB 70, DB 71, DB 72, DB 73, DB 74, DB 75, DB 76, DB 77, DB 79	[208], [210], [212], [217], [219], [221], [224], [226], [230], [234], [236], [243], [251]

Table 9 presents an overview of the characteristics of publicly available MSA Speech Databases in relation to the disclosure of speaker's sex information. It is noteworthy that the majority of these databases, accounting for 37%, did not provide any information regarding the speaker's sex. Additionally, 33% of the databases disclosed the sex information of both males and females, while 19% specifically indicated male speakers, 11% included a combination of males and some females, and 0% mentioned female speakers.

In Table 10, we classify the self-databases based on the information about the speaker's sex. Notably, a significant portion of the databases, approximately 38%, included both male and female speakers. Furthermore, 31% of the databases did not disclose any information pertaining to the speaker's sex. Among the databases that did disclose this information, 20% exclusively comprised male speakers, while 8% consisted of a combination of males and some females. Finally, a mere 4% of the databases featured only female speakers.

C. SPEAKERS' AGE

The American Medical Association have categorized age designations into Neonates or newborns (birth to 1 month),

TABLE 9. Publicly accessible MSA database information on the sex of speakers.

Sex of Speakers Class	Available Database Name	Ref
Male	SASSC, ASC, Arabic Speech Corpus for Isolated Words, KTD, Phonemes of Arabic	[20], [23], [24], [29], [39]
Female	...	...
Male & Female	MSA Speech Corpus, OrienTel Egypt MSA, OrienTel Morocco MSA, OrienTel Jordan MSA, OrienTel Tunisia MSA, West Point, KSU Emotions corpus, SAD, NEMLAR Speech Synthesis Corpus, KSU Database	[31], [35], [37], [36], [38], [41], [30], [40], [33], [45]
Male and some Female	Arabic Learner Corpus, MDP	[22], [43]
Female and some Male	...	...
No information	Arabic Broadcast News Speech, GALE Phase 2, GALE Phase 3, GALE Phase 4, KAPD, Al Jazeera multi-dialectal, Arabic mobile parallel multi-dialect speech, MGB-3, NEMLAR Broadcast News, NetDC Arabic BNSC	[21], [25], [26], [27], [28], [42], [44], [46], [32], [34]

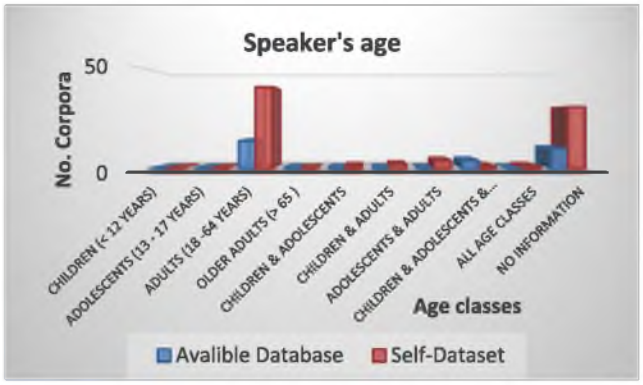

FIGURE 13. Distribution of the databases per speaker's age.

Infants (1 month to 1 year), Children (1 year through 12 years), Adolescents (13 years through 17 years, Adults (18 years or older) and Older adults (65 and older) [254].

Thus, the age of the speakers was classified based on this classification. It is noted in Figure 13 that the highest of these categories are adults, followed by the database that did not give information about the age of the speakers.

Table 11 shows the categories of databases (available) based on the ages of the speakers, and it is noted that there are no databases specifically for the age groups of speakers under 17 years old. Nevertheless, as is the case in the databases provided by OrienTel, these categories are combined with other categories in a single database. Moreover, the category over 65 years old does not exist in these databases.

TABLE 10. MSA self-database information on the sex of speakers.

Database Class	Self-Dataset Name	Ref
Male	DB 01, DB 02, DB 03, DB 04, DB 05, DB 07, DB 13, DB 25, DB 35, DB 40, DB 57, DB 60, DB 64, DB 65, DB 69, DB 80	[142], [144], [147], [149], [150], [154], [168], [207], [233], [252], [179], [186], [206], [208], [217], [253]
Female	DB 08, DB 75, DB 78	[156], [234], [245]
Male & Female	DB 38, DB 62, DB 46, DB 68, DB 09, DB 27, DB 30, DB 29, DB 61, DB 34, DB 10, DB 33, DB 37, DB 42, DB 39, DB 15, DB 14, DB 47, DB 17, DB 52, DB 16, DB 18, DB 21, DB 56, DB 22, DB 23, DB 24, DB 45, DB 51, DB 55	[244], [194], [153], [215], [158], [211], [218], [216], [188], [227], [160], [225], [237], [146], [246], [174], [170], [155], [178], [167], [176], [180], [187], [177], [193], [203], [205], [151], [165], [175]
Male and some Female	DB 32, DB 26, DB 41, DB 43, DB 49, DB 50	[222], [209], [143], [43], [159], [163]
Female and some Male	...	...
No information	DB 48, DB 63, DB 06, DB 19, DB 20, DB 31, DB 36, DB 44, DB 66, DB 67, DB 11, DB 12, DB 28, DB 53, DB 54, DB 58, DB 59, DB 70, DB 71, DB 72, DB 73, DB 74, DB 76, DB 77, DB 79	[157], [204], [152], [183], [185], [220], [235], [44], [210], [212], [164], [166], [213], [169], [171], [181], [184], [219], [221], [224], [226], [230], [236], [243], [251]

Table 12 shows the classification of self-databases based on the ages of the speakers, and we note that the categories children, adolescents, and Older adults are not found independently but are combined with other categories. For example, DB 46 contains two categories, Children and adolescents, DB 32, DB 68 combine the categories of Children and adults, DB 25, DB 43, DB 45, DB 51 combine the two categories of adolescents and adults, DB 55 includes three categories less than 65 years old. As for DB 27, it includes all groups, including the elderly, as the age of speakers at this base ranges from 4 to 75 years.

D. THE DATABASES' SIZE

Database size is an important factor in speech processing applications, as the larger the training and test data, the higher the accuracy. It is noted in Table 13 that the databases announce the size of the database through the duration in hours, the number of sentences or the number of words or the number of phonemes based on the type of database: Sentences, Words, Digits, Alphabet, News and Conversations.

It is noted that the largest database in terms of time duration is GALE Phase 4 2400 h, followed by Al Jazeera

TABLE 11. Publicly accessible MSA database information on the age of speakers.

Age class	Available Database Name	Ref
Children (< 12 years)	...	...
Adolescents (13 - 17 years)	...	...
Adults (18 -64 years)	SASSC, Arabic Learner Corpus, ASC, Arabic Speech Corpus for Isolated Words, KAPD, KTD, KSU Emotions corpus, MSA Speech Corpus, NEMLAR Speech Synthesis Corpus, Phonemes of Arabic, SAD, MDP, KSU Database	[20], [22], [23], [24], [28], [29], [30], [31], [33], [39], [20], [40], [43] , [45]
Older adults (> 65)	...	...
Children & Adolescents & Adults	OrienTel Egypt MSA, OrienTel Jordan MSA, OrienTel Morocco MSA, OrienTel Tunisia MSA	[35], [36], [37], [38]
No information	Arabic Broadcast News Speech, GALE Phase 2, GALE Phase 3, GALE Phase 4, NEMLAR Broadcast News, NetDC Arabic BNSC, West Point, Al Jazeera multi-dialectal, Arabic mobile parallel multi-dialect speech, MGB-3	[21], [25], [26], [27], [32], [34], [41], [42], [44], [46]

multi-dialectal 2000 h, MGB-3 1200 h. But it is important to note that this is the size of the audio data, for example, the base of Al-Jazeera contains only 94 h labelled while MGB-3 contained only 20 hours transcribed.

The size information of self-databases is shown in Table 14. It should be noted that these are not sizable databases, DB 43 having the biggest size at 32 hours.

E. SPEAKER'S NATIONALITY

One of the characteristics that speech databases are interested in is the nationalities of the speakers, where diversity in nationalities is an advantage of the database. Figure 14 shows the classification of databases based on the nationalities of the speakers. Table 15 shows the classifications of the publicly accessible MSA database according to the nationalities of the speakers. It is noted that all the databases stated the nationalities of the speakers except for one database [21] that did not mention this information. It should also be noted that 13 databases stated that the speakers are Native Arabs without specifying their nationality. And 6 databases contain multiple nationalities, West Point [41] database combined native and non-native Arab, and the rest of the databases were for specific nationalities such as Levantine Arabic, Egyptian, Jordanian, Moroccan and Tunisian.

Table 16 shows the classification of self-databases based on the nationalities of the speakers. Unfortunately, 32% of

TABLE 12. MSA self-database information on the age of speakers.

Age class	Self-Dataset Name	Ref
Children (< 12 years)	...	...
Adolescents (13 - 17 years)	...	...
Adults (18 - 64 years)	DB 01, DB 02, DB 03, DB 04, DB 05, DB 07, DB 08, DB 09, DB 10, DB 11, DB 13, DB 14, DB 16, DB 17, DB 18, DB 21, DB 23, DB 24, DB 26, DB 29, DB 30, DB 33, DB 37, DB 38, DB 39, DB 40, DB 42, DB 47, DB 50, DB 52, DB 53, DB 56, DB 57, DB 60, DB 62, DB 63, DB 64, DB 65, DB 78, DB 80	[142], [144], [147], [149], [150], [154], [156], [158], [160], [164], [168], [170], [176], [178], [180], [187], [203], [205], [209], [216], [218], [225], [237], [244], [246], [252], [146], [155], [163], [167], [169], [177], [179], [186], [194], [204], [206], [208], [245], [253]
Older adults (> 65)	...	...
Children & Adolescents	DB 46	[153]
Children & Adults	DB 32, DB 68	[222], [215]
Adolescents & Adults	DB 25, DB 43, DB 45, DB 51	[207], [43], [151], [165]
Children & Adolescents & Adults	DB 55	[175]
All age classes	DB 27	[211]
No information	DB 06, DB 12, DB 15, DB 19, DB 20, DB 22, DB 28, DB 31, DB 34, DB 35, DB 36, DB 41, DB 44, DB 48, DB 49, DB 54, DB 58, DB 59, DB 61, DB 66, DB 67, DB 69, DB 70, DB 71, DB 72, DB 73, DB 74, DB 75, DB 76, DB 77, DB 79	[152], [166], [174], [183], [185], [193], [213], [220], [227], [233], [235], [143], [44], [157], [159], [171], [181], [184], [188], [210], [212], [217], [219], [221], [224], [226], [230], [234], [236], [243], [251]

databases did not clarify the information of the speakers' nationalities, with a similar percentage of the databases stating that the speakers are Native Arabs without specifying the nationality, 12% of the databases contain multiple nationalities, and the rest of the databases contain Arab and non-Arab nationalities, including Egyptian, Malay, Jordanian, Moroccan, Tunisian, Algerian and Kuwaiti.

F. RECORDING ENVIRONMENT

The recording environment and the type of noise greatly affect the accuracy of speech recognition, so the researchers sought to diversify the recording environments, especially

TABLE 13. Publicly accessible MSA database information on the database's size

Corpus Name	Ref	hours	No. sentences	No. words	No. Phonemes	Unique sentences/ words
SASSC	[20]	7		51,000		627 syllables
Arabic Broadcast News Speech	[21]	10				
Arabic Learner Corpus	[22]			282,732		
ASC	[23]	3.7		1,813		
Arabic Speech Corpus for Isolated Words	[24]			10,000		20 words
GALE Phase 2	[25]	586				
GALE Phase 3	[26]	200				
GALE Phase 4	[27]	2400				
KAPD	[28]				46,000	
KTD	[29]					16 sentences
KSU Emotions corpus	[30]	05:15				
MSA Speech Corpus	[31]	23		2,110		1626 words
NEMLAR Broadcast News	[32]	40		62,000		
NEMLAR Speech Synthesis Corpus	[33]	10	2,032	42,000		
NetDC Arabic BNSC	[34]	23				
OrienTel Egypt MSA	[35]	50				
OrienTel Jordan MSA	[36]	50				
OrienTel Morocco MSA	[37]	49				
OrienTel Tunisia MSA	[38]	50				
Phonemes of Arabic	[39]			1,368	8,379	
SAD	[40]				8,800	10 digits
West Point	[41]	11.42	8,516			
Al Jazeera multi-dialectal	[42]	2000				
MDP	[43]	32	1,291			
Arabic mobile parallel multi-dialect speech	[44]				67,132	
KSU database	[45]	590				
MGB-3	[46]	1200				

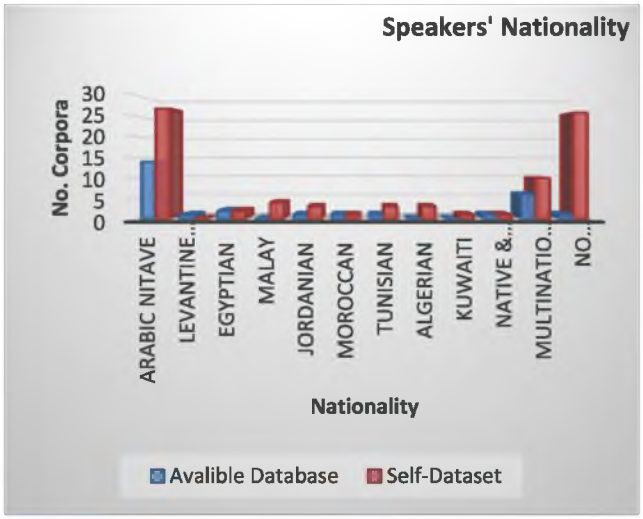

FIGURE 14. Distribution of the databases per speaker's nationality.

those applications that rely on speech recognition in public places. Figure 15 shows the classifications of databases based on the recording environment.

Table 17 shows the available databases according to the recording environment. It is noted that most of these rules are recorded in environments noise-free, in a TV studio or a soundproof room.

Table 18 shows the classification of self-databases according to the recording environment. Unfortunately, 57.5% of these databases did not give information about the recording environment.

Then comes the Databases recorded in environments noise-free, such as a sound proof room or a TV studio, while the databases with multiple environments constitute only 6%.

V. DISCUSSION AND FUTURE RESEARCH DIRECTIONS

This section answers the fifth research question (RQ5). Several research issues, current patterns of research work, and

TABLE 14. MSA self-database information on the database's size.

Database	Database	hours	No. sente	No. wor	No. Phone	Unique sentences/wor
DB 01	[142]		120	654	1190	
DB 02	[144]	4.3	1184	2639		
DB 03	[147]		202	1254	6174	
DB 04	[149]	16		8100		
DB 05	[150]		191		6179	
DB 06	[152]	1				
DB 07	[154]		148	1,48	6,787	
DB 08	[156]		434	1346	5348	
DB 09	[158]			308	1826	
DB 10	[160]		600	1000	7445	
DB 11	[164]			896		
DB 12	[166]				700	28 alphabets
DB 13	[168]	4.3		4000		
DB 14	[170]				400	20 alphabets
DB 15	[174]		791	3622	27725	
DB 16	[176]		367			
DB 17	[178]		20	53		
DB 18	[180]			500		5 words
DB 19	[183]			600		10 words
DB 20	[185]			400		10 digits
DB 21	[187]		2000			
DB 22	[193]			1200		
DB 23	[203]			6644		
DB 24	[205]	2.5	390			10 sentences
DB 25	[207]			480		8 words
DB 26	[209]				832	
DB 27	[211]			2000		
DB 28	[213]				1680	28 letters
DB 29	[216]	2.6	254	7386		
DB 30	[218]		80			16 sentences
DB 31	[220]		20			10 sentences
DB 32	[222]			920		4 words
DB 33	[225]		28			
DB 34	[227]	22.		2160	30986	
DB 35	[233]					
DB 36	[235]		300	1200		40 words
DB 37	[237]	12	415			
DB 38	[244]	1		50		
DB 39	[246]				1400	28 letters
DB 40	[252]				14500	29 alphabets
DB 41	[143]	1.2				
DB 42	[146]	2.7	415	2110		
DB 43	[43]	32	67,13	160,	778,48	
DB 44	[44]			6700		
DB 45	[151]			2100		7 words
DB 46	[153]			1650		55 words
DB 47	[155]			6600		
DB 48	[157]			1,70	14,500	10 digit + 28
DB 49	[159]			720		10 digits
DB 50	[163]			425		
DB 51	[165]			2800		7 words
DB 52	[167]			8800		10 digits
DB 53	[169]			9000		
DB 54	[171]	10		1340		
DB 55	[175]			9542		
DB 56	[177]			2704		
DB 57	[179]					
DB 58	[181]				4200	28 letters
DB 59	[184]	23	4754			
DB 60	[186]		940			700 words
DB 61	[188]	5.4	4572	39,2		
DB 62	[194]		1080			
DB 63	[204]				11200	28 letters
DB 64	[206]					
DB 65	[208]			105		
DB 66	[210]		120		517,08	
DB 67	[212]					
DB 68	[215]			1880		
DB 69	[217]		100			
DB 70	[219]	5.4		1423		
DB 71	[221]			275		11 words
DB 72	[224]			500		
DB 73	[226]	2				
DB 74	[230]	5		4000		
DB 75	[234]	7.5	400	3585		
DB 76	[236]					
DB 77	[243]		100			
DB 78	[245]		434			
DB 79	[251]			1730		
DB 80	[253]			1,70	4000	

future research directions are discussed. It will give some insights and ideas to the researchers to plan their research and development in the future.

A. FREELY ACCESS AND LIMITED AVAILABILITY OF DATABASES

There are speech databases collected by institutes or projects that are interested in collecting speech databases in several languages, including Arabic. One of these projects is the

TABLE 15. Publicly accessible MSA database information on the speaker's nationality.

Speaker's Nationality	Available Database Name	Ref
Arabic Nitave	SASSC, Arabic Speech Corpus for Isolated Words, GALE Phase 2, GALE Phase 3, GALE Phase 4, KAPD, KTD, NEMLAR Broadcast News, NetDC Arabic BNSC, Phonemes of Arabic, SAD, Arabic mobile parallel multi-dialect speech, MGB-3	[20], [24], [25], [26], [27], [28], [29], [32], [34], [39], [40], [44], [46]
Levantine Arabic	ASC	[23]
Egyptian	NEMLAR Speech Synthesis Corpus, OrienTel Egypt MSA	[33], [35]
Jordanian	OrienTel Jordan MSA	[36]
Moroccan	OrienTel Morocco MSA	[37]
Tunisian	OrienTel Tunisia MSA	[38]
Native & non-native Arab	West Point	[41]
Multinational	Arabic Learner Corpus, KSU Emotions corpus, MSA Speech Corpus, Al Jazeera multi-dialectal, MDP, KSU Database	[22], [30], [31], [42], [43], [45]
No information	Arabic Broadcast News Speech	[21]

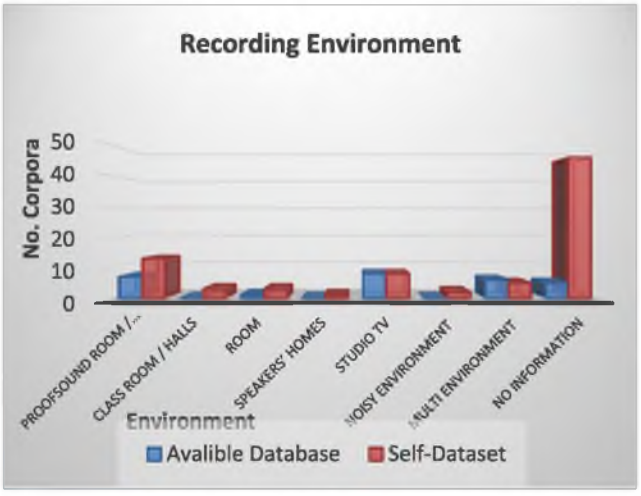

FIGURE 15. Distribution of the databases per recording environment.

OrienTel project [18], which has been dedicated to collecting speech corpora for MSA and Arabic dialects of United Arab Emirates Tunisia, Morocco, Jordan, and Egypt countries

TABLE 16. MSA self-database information on the speaker's nationality

Speaker's Nationality	Self-Dataset Name	Ref
Arabic Nitave	DB 01, DB 02, DB 04, DB 05, DB 08, DB 11, DB 18, DB 24, DB 25, DB 27, DB 32, DB 33, DB 35, DB 38, DB 40, DB 43, DB 48, DB 49, DB 54, DB 55, DB 61, DB 64, DB 65, DB 75, DB 78, DB 68	[142], [144], [149], [150], [156], [164], [180], [205], [207], [211], [222], [225], [233], [244], [252], [43], [157], [159], [171], [175], [188], [206], [208], [234], [245], [215]
Levantine Arabic	...	...
Egyptian	DB 50, DB 73	[163], [226]
Malay	DB 14, DB 28, DB 39, DB 63	[170], [213], [246], [204]
Jordanian	DB 07, DB 23, DB 69	[154], [203], [217]
Moroccan	DB 22	[193]
Tunisian	DB 10, DB 13, DB 21	[160], [168], [187]
Algerian	DB 03, DB 56, DB 62	[147], [177], [194]
Kuwaiti	DB 29	[216]
Native & non-native Arab	DB 45	[151]
Multinational	DB 16, DB 30, DB 34, DB 37, DB 41, DB 42, DB 44, DB 46, DB 60, DB 80	[176], [218], [227], [237], [143], [146], [44], [153], [186], [253]
No information	DB 06, DB 09, DB 12, DB 15, DB 17, DB 19, DB 20, DB 26, DB 31, DB 36, DB 47, DB 51, DB 52, DB 53, DB 57, DB 58, DB 59, DB 66, DB 67, DB 70, DB 71, DB 72, DB 74, DB 76, DB 77, DB 79	[152], [158], [166], [174], [178], [183], [185], [209], [220], [235], [155], [165], [167], [169], [179], [181], [184], [210], [212], [219], [221], [224], [230], [236], [243], [251]

Thus, it can be divided into two parts: The first category Modern Standard Arabic has been collected from several Arabic countries with the following details: 500 speakers from Egypt (M: 398 F:352), 757 speakers from Jordan (M: 393 F:364), 772 speakers from Morocco (M: 383, F: 389), 598 speakers from Tunisia (M: 426 F:366), and 500 speakers from United Arab Emirates (M: 254 F: 246).

West Point Arabic Speech [255] includes speech data that the project named Santiago obtained and analyzed. This corpus was originally intended to train acoustic models for automated speech recognition that could be used as an aid in teaching Arabic to West Point cadets, this corpus contains 11 hours of recording sentences spoken by 110 native and non-native speakers (native 41 males, 34 females and non-native 25 males, 10 females).

However, there are still few freely accessible speech corpora in the Arabic language, which can impede the advancement of AI-driven research. In particular, given the grow-

TABLE 17. Publicly accessible MSA database information on the recording environment.

Environment	Available Database Name	Ref
Soundproof room/studio	SASSC, ASC, KAPD, KSU Emotions corpus, MSA Speech Corpus, MDP	[20], [23], [28], [30], [31], [43]
Room	West Point	[41]
Studio TV	Arabic Broadcast News Speech, GALE Phase 2, GALE Phase 3, GALE Phase 4, NEMLAR Broadcast News, NEMLAR Speech Synthesis Corpus, NetDC Arabic BNSC, MGB-3	[21], [25], [26], [27], [32], [33], [34], [46]
Multi environment	OrienTel Egypt MSA, OrienTel Jordan MSA, OrienTel Morocco MSA, OrienTel Tunisia MSA, Al Jazeera multi-dialectal, KSU Database	[35], [36], [37], [38], [42], [45]
No information	Arabic Learner Corpus, Arabic Speech Corpus for Isolated Words, KTD, Phonemes of Arabic, SAD	[22], [24], [29], [39], [40]

TABLE 18. MSA self-database information on the recording environment

Environment	Self-Dataset Name	Ref
Soundproof room/studio	DB 02, DB 04, DB 05, DB 21, DB 24, DB 35, DB 37, DB 38, DB 39, DB 42, DB 43, DB 54, DB 59	[144], [149], [150], [187], [205], [233], [237], [244], [246], [146], [43], [171], [184]
Classroom / Halls	DB 11, DB 23, DB 26	[164], [203], [209]
Room	DB 19, DB 33, DB 69	[183], [225], [217]
Speakers' homes	DB 30	[218]
Studio TV	DB 29, DB 34, DB 41, DB 61, DB 74, DB 75, DB 76	[216], [227], [143], [188], [230], [234], [236]
Noisy environment	DB 32, DB 53	[222], [169]
Multi environment	DB 28, DB 44, DB 46, DB 49, DB 60	[213], [44], [153], [159], [186]
No information	DB 01, DB 03, DB 06, DB 07, DB 08, DB 09, DB 10, DB 12, DB 13, DB 14, DB 15, DB 16, DB 17, DB 18, DB 20, DB 22, DB 25, DB 27, DB 31, DB 36, DB 40, DB 45, DB 47, DB 48, DB 50, DB 51, DB 52, DB 55, DB 56, DB 57, DB 58, DB 62, DB 63, DB 64, DB 65, DB 66, DB 67, DB 70, DB 71, DB 72, DB 73, DB 77, DB 78, DB 79, DB 80, DB 68	[142], [147], [152], [154], [156], [158], [160], [166], [168], [170], [174], [176], [178], [180], [185], [193], [207], [211], [220], [235], [252], [151], [155], [157], [163], [165], [167], [175], [177], [179], [181], [194], [204], [206], [208], [210], [212], [219], [221], [224], [226], [243], [245], [251], [253], [215]

ing prevalence of machine learning and speech processing techniques in many industries, the high cost of corpora—up to $6000 per license—can be a significant barrier to accessibility.

B. INFORMATION OF CORPORA

The Speaker attribute is one of the most significant parameters in any speech corpus to make dealing with the corpus easier and based on the correct basis. speakers' number, sex, race, age, and distribution across the dialect's geographic region are very important examples of these characteristics [256].

The database will be significantly important and widely used whenever it has a large number of speakers and is diverse in terms of the sex of the speakers, their ages and the other attributes that the studies aim to explore.

The OrienTel databases are distinguished from other databases because it has details of all the speakers' information, as well as extensive Prompts which is digits, words, sentence and spontaneous. The same telephone response to the questionnaire method is used in these corpora, which are available via the ELRA catalogue1.

The KSU database [45] was also built and designed to be a rich database characterized by the diversity of speakers and recording environments. It was recorded in three different environments and had 300 speakers from more than 11 different countries, and included both sexes.

Nevertheless, many Arabic databases still lack detailed information for speakers. Thus, additional work could be put into tagging, transcribing, labelling and describing speakers' information in the large Arabic databases Al Jazeera multi-dialectal and GALE by transcript in order to make it more easily searchable in different fields of speech.

C. MSA DATABASE AND MULTI ARABIC FORMS CORPUS

Arabic speech Databases can be classified into Modern Standard Arabic databases and multi-dialect databases. Several studies on Arabic Automatic Speech Recognition (ASR) have concentrated on creating Modem Standard Arabic (MSA) recognizers, a standardized linguistic standard used in the Arabic-speaking world and used in the newspapers, broadcasting news, media, seminars, universities, courtrooms [257].

The databases in this category can be classified into rich databases for MSA words or sentences, MSA and Arabic dialect speech parallel corpus, and small databases for a few words such as numbers or Arabic phonemes that can be used in some speech processing applications.

The KACST Arabic Phonetic Database (KAPD) was developed by King Abdulaziz City for Science and Technology (KACST) to provide comprehensive information about Arabic sounds and their articulatory mechanisms. It consists

of over 46,000 files recorded from 7 native Arabic speakers, covering various token positions and featuring 8 repetitions of Arabic sounds. It can be used for research and development such as speech recognition, speech synthesis, speech therapy, speech perception and speech modelling [28].

Some speech databases include the three or two forms of the Arabic language. One of the most important and famous of these databases is SAAVB corpus, which is dedicated to speakers from all towns in Saudi Arabia using the questionnaire method by telephone response [258]. The primary feature of this corpus is that a preliminary selection of speakers and environment is conducted before recording. This corpus includes Modern Standard Arabic and Saudi accent.

There are not many Arabic speech corpora available for use in Arabic NLP projects such as translation. When talking about dialect resources, this dearth becomes a significant barrier for Arabic academics working in the field of NLP. As a result, Arabic NLP researchers must create some or most of their resources. Each researcher starts from nothing and takes a long time to gather enough resources. Another factor to take into account when talking about a data shortage is the problem of standardization in NLP activities.

The first parallel dialect speech corpus that has been built and collected in Arabic is Multi-Dialect Parallel Corpus (MDP), which is a free corpus of MSA and three Arabic dialects, these databases were recorded by direct recording methods from the Gulf, Egypt and Levantine. The dataset involved 1291 recordings for MSA and 1069 recordings for other dialects. In addition, 32 speech hours have been recorded by 52 participants [43].

However, this database is industry specific i.e., travel, tourism and MSA routers and contains only 1291 sentences. Arabic speech databases still need to expand the multi dialects corpora and include other dialects, to enable Arabic NLP researchers to contribute further to this important field.

D. MSA CORPORA RICHNESS AND DIVERSITY IN THE DATABASE

Some Arabic speech databases can be described as a rich database because of the number of speakers, the diversity of speakers, and different environments for recording developed by some researchers or research institutions.

King Saud University developed a KSU Rich Arabic with the largest size Arabic corpus and included 159.5 Hours, 300 speakers (males and females), from 9 Arab nationalities and 18 non-Arab nationalities, multi-dialects (MSA, Saudi dialect, and Qur'an recitation), different types of Prompts (Sentences, Words, Paragraphs Pronouncing, Question and answers), and multi-Environment (office, indoor, outdoor, car) [6], [7].

Due to the difficulty of collecting a large speech database, some researchers have constructed speech datasets with a small number of speakers (sometimes a single speaker) often asked to read a text or conversation. For example, Halabi [56] recorded a sub-database of currencies, time and numbers to use in Arabic speech synthesis. A local professional broadcaster recorded 12 hours of pure recording time of 3930 sentences. One of these corpora is a standard Arabic Single Speaker Corpus (SASSC) developed by the Research Institute King Abdulaziz City, where 7 hours and 20 minutes of the data are recorded by a professional male speaker in a studio [20].

On the other hand, some researchers collected a larger number of speakers, but the recorded text was small (some letters, numbers, . . .). It is often the repetition of that vocabulary to address the lack of speakers. For example [40] collected data from 44 male and 44 female speakers, where the digits (0-10) were recorded ten times for each speaker.

Moreover, some researchers sought to collect a larger database of Arabic speech, and some of them resorted to collecting data from radio or TV broadcasts. For example, Ali et al. [259] describe an ASR system for Modern Standard Arabic (MSA) constructed using a 200-hour broadcast news database, a combination of 76 hours of the broadcast report (BR), and 127 hours of conversational broadcast (BC). Other databases were collected with different approaches. For example, a total of 50 hours of audio transcription from the Al-Jazeera news channel [260], [261] to describe a detailed comparison of several speech recognition technologies. This type of dataset provides a larger volume of speech data, however, the process of editing, organizing and removing noise and irrelevant information is a tedious task, and the quality of the recording is often not good.

As for Modern Standard Arabic, some rich and varied databases feature a large number of speakers and a big size corpus. However, unfortunately, most of the Arabic databases do not provide explicit or implicit reports about speakers' information in the catalogue, such as age, and completely ignored this kind of significant speakers' attribute in their corpus [258]. Moreover, some databases also indicate whether speakers are Arabic native, or non-native speakers, as there are some databases for all speakers of Arabic speakers (such as SAAVB database which all speakers are Arabic native speakers). On the other hand, some databases contain both native and non-native speakers such as KSU Rich Arabic, while some Arabic speech databases have not said a thing about this very significant information that can help researchers in ASR and NLP field.

The diversity of the speakers' nations is another facet of the databases' richness. Consequently, certain databases handled it. For instance, KSU Rich Arabic corpus includes speakers from various ethnic, Arab and non-Arab (Africans and Asians) groups. They are chosen from nine Arab countries regarding Arab speakers in this corpus: Saudi, Yemen, Syria, Egypt, Algeria, Tunisia, Sudan, Palestine, and Lebanon [6], [7]. KSU includes three forms of Arabic. Qur'an verses, MSA sentences, and colloquial Arabic sentences were recorded.

Finally, some studies focused on the diversity of dialects and their multiplicity in the speech database. For example, Al Jazeera's larger-scale multi-dialect speech corpus, based

on Al Jazeera's Broadcast News [1], annotation is carried out by the technology of crowdsourcing, in addition to modern Standard Arabic. This corpus includes the four main groups of Arabic dialects.

E. DATA QUALITY

The quality of data in Arabic speech databases can vary significantly. Factors such as background noise, recording equipment, and speaker diversity can impact the accuracy and reduce the audio quality.

Most standard Arabic speech databases, especially large databases, have poor quality. Assessing the quality of audio files can be subjective and dependent on various factors. However, there are general indicators that can help evaluate whether an audio file is of poor or good quality, such as clarity and intelligibility, background noise, and frequency range.

When reviewing available speech databases for Modern Standard Arabic, it is found that most databases are recorded using various audio devices, such as those collected through mobile or telephone communications, or those collected in different environments. For example, the OrienTel project collected data from different environments such as streets, homes, offices, and others, leading to poor quality.

In addition to unwanted background noise, such as static, humming, or environmental noise, the overall sound level and volume levels should be consistent throughout the recording. Inconsistency and sudden changes in sound level or uneven volume levels can result in poor quality or inappropriate recording techniques. This often occurs in databases that lack recording device standards or a unified recording environment, such as a single studio. Despite the importance of this information, most available databases do not mention these details in their database descriptions.

Furthermore, the frequency range is an indicator of audio quality. Therefore, high-quality audio files should exhibit balanced frequency response across the audible spectrum. Severe loss of high-frequency content or low-frequency response can indicate low recording quality. The frequency range in available databases varies, such as 96 kHz, 24-bit; 96 kHz, 16-bit; 44100 Hz, 16-bit; 16 kHz, 16-bit; 8 kHz, 8-bit. The channels also vary between single-channel and multi-channel.

Based on the aforementioned indicators, some of the available databases can be considered high-quality databases due to meeting the quality standards in sound. One of these databases is the [45] database, where 590 hours of data were collected from 269 speakers through direct recording in three different sessions, which are Office, Cafeteria and one of these environments being soundproof. The recordings were made at a frequency range of 48 kHz, 16-bit, and detailed descriptions of the recording devices and environment were provided.

Similarly, [20] recorded 7 hours of speech from a professional speaker in a professional studio at a frequency range of 96 kHz, 16-bit. Also, [23] recorded 3.7 hours in a professional studio at a frequency range of 48 kHz, 16-bit for speech synthesis purposes. Additionally, [28], [29], [30] are considered high-quality databases.

F. CORPORA APPLICATIONS

Different speech processing applications require different characteristics in the databases with regard to the number of speakers, the diversity of sex, the diversity of their nationalities, the diversity of the recording environment, the type of prompts, whether phonemes, words, or sentences, in addition to the size of the database.

For example, isolated word recognition applications do not require long sentences, but require a large database in terms of the number and variety of words. Whereas, emotion recognition applications require a database of sentences that have been recorded in different emotional situations. The remaining speech processing programmes continue in the same manner; thus, each application has requirements that the researcher must provide in the database.

In fact, the majority of publicly available databases are directed towards speech recognition applications, with less focus on the rest of the applications such as speech synthesis, speaker identification/verification, and emotion recognition.

It was observed that the speech databases being developed are for the Conversion of Text to Speech, for which the database consists of phonemes or syllables. For speech synthesis applications; Al-Halabi developed an Arabic Speech Corpus (ASC) [23] containing 1813 utterances for a total of 3.7 hours recorded by a professional speaker. However, the database needs to be expanded and increased in size to include all Arabic phoneme characteristics to ensure more accurate Arabic speech syntax [262].

Due to the variety of speech recognition applications, speaker recognition databases' structure and content are also diverse. These applications need a large number of speakers to learn about the vocal features of each speaker. It should be noted that when the number of users increases, it becomes difficult to find unique features for each user, as failure to do so may lead to wrong identification, and therefore it is necessary to train a sufficient number of speakers. Alfaifi, et al. [22] constructed the Arabic Learner Corpus, which contains 942 native speakers who recorded 282,732 words.

Muhammad presents KSUEmotions corpus [30], which contains the emotional discourse of ten male speakers and thirteen female speakers from Saudi Arabia, Syria and Yemen. It includes emotions: neutral, happiness, sadness, surprise, and anger. However, emotional speech recognition in the Arabic language is insufficiently tackled in the literature compared to other languages.

G. FUTURE RESEARCH AND DIRECTIONS IN MSA SPEECH DATABASES

In order to advance the field of Arabic speech processing, it is crucial to explore potential avenues for future research and development. This section highlights several key directions

that can contribute to the expansion and enhancement of Arabic speech databases, as well as the exploration of novel research areas.

1) Develop more diverse and representative Arabic speech databases. Efforts to expand the range and diversity of Arabic speech databases could improve generalizability and research power in the areas of Arabic speech recognition, synthesis, speaker recognition, dialects, and other applications.
2) There is a need for more research into the effectiveness of different speech processing techniques, including those specifically designed for Arabic speech. This can help identify the most effective techniques and methods for different applications. For example, an exploration of the impact of dialectal differences, as Arabic speech varies greatly due to the many different dialects spoken across different regions. Therefore, one of the future research directions could be further investigation into the impact of dialect variation on speech processing techniques and the development of dialect-specific databases and tools.
3) Most databases are designed for speech recognition or speech synthesis, so they still need to build Arabic speech databases for use in some applications such as text-based or non-text-based speaker recognition, as there are not many Arabic speech databases designed for speaker recognition. Emotion recognition databases are still lacking, as well as some applications such as educational, medical and other applications.
4) Development of techniques for organizing, classifying, and segmenting MSA speech databases [263], particularly large ones such as those obtained from television broadcasts [50] or sourced from platforms like YouTube and social media [27].

Overall, there are several limitations to the study of Arabic speech databases, but there are also many exciting opportunities for further research in this area. By addressing these limitations and exploring new avenues for investigation, we can continue to improve our understanding of Arabic speech and develop more effective tools and techniques for speech processing.

VI. CONCLUSION

In this review, we have studied the Speech databases of Modern Standard Arabic. In the beginning, we describe the review methodology and research questions, then we identified online digital databases and search engines that can be used to search and collect the manuscripts including IEEE, ScienceDirect, Scopus, Springer Link, Web of Science, Google Scholar, ResearchGate, ACM. In addition, we have included collections from LDC and Open SLR, popular language corpora resources. Finally, we studied the manuscripts and classified them based on the characteristics of the databases and studied the challenges and research directions. Our review provides valuable insights into this area while highlighting the paucity of speech databases in Arabic compared to other languages. Furthermore, this study is the first to review all publicly available databases as well as those created by the researchers during their studies (self-databases). Through this review, we aim to facilitate access to the appropriate database for researchers, encourage researchers to build new Arabic datasets in the currently exposed areas, and encourage institutions to make the database freely available and easily accessible to researchers. This review is expected to provide valuable information and a quick reference for readers, especially researchers and practitioners, to understand or explore MSA databases and their characteristics, which makes it easier for the researcher to choose the appropriate database through its characteristics, thus this review will contribute to research in speech processing applications.

REFERENCES

[1] S. Bougrine, A. Chorana, A. Lakhdari, and H. Cherroun, "Toward a web-based speech corpus for Algerian dialectal Arabic varieties," in *Proc. 3rd Arabic Natural Lang. Process. Workshop*, 2017, pp. 138–146.

[2] A. Mohammed and M. S. Sunar, "Toward a rich Quranic Arabic speech corpus for Tajweed Rules," in *Proc. Universal Wellbeing (ICUW)*, p. 210.

[3] E. Saiegh-Haddad and R. Henkin-Roitfarb, "The structure of Arabic language and orthography," in *Handbook of Arabic Literacy*. Cham, Switzerland: Springer 2014, pp. 3–28.

[4] M. Aboelezz, "A history of the Arabic language and the origin of non-dominant varieties of Arabic," in *Pluricentric Languages Across Continents: Features and Usage*, K. E. F. R. Muhr, Z. Ibrahim, and C. Miller, Eds. New York, NY, USA: Peter Lang, 2018, pp. 175–187.

[5] P. Behnstedt and M. Woidich, "Arabic dialectology," in *The Oxford Handbook of Arabic Linguistics*, J. Owens, Ed. Oxford, U.K.: Oxford Univ. Press, 2013.

[6] M. Alsulaiman, Z. Ali, G. Muhammed, M. Bencherif, and A. Mahmood, "KSU speech database: Text selection, recording and verification," in *Proc. Eur. Model. Symp.*, Nov. 2013, pp. 237–242.

[7] M. M. Alsulaiman, G. Muhammd, M. A. Bencherif, A. Mahmood, and Z. Ali, "KSU rich Arabic speech database," *J. Inf.*, vol. 16, no. 6, pp. 4231–4253, 2013.

[8] Linguistic Data Consortium, "TIMIT acoustic phonetic continuous speech corpus," Philadelphia, PA, USA, Tech. Rep. LDC93S1, 1993.

[9] M. Alghamdi, F. Alhargan, M. Alkanhal, A. Alkhairy, M. Eldesouki, and A. Alenazi, "Saudi accented Arabic voice bank," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 20, pp. 45–64, Jan. 2008.

[10] A. Mohammed, M. S. Sunar, and M. S. Hj Salam, "Quranic verses verification using speech recognition techniques," *Jurnal Teknologi*, vol. 73, no. 2, pp. 99–106, Mar. 2015, doi: 10.11113/jt.v73.4200.

[11] A. Ahmed, N. Ali, M. Alzubaidi, W. Zaghouani, A. A. Abd-alrazaq, and M. Househ, "Freely available Arabic corpora: A scoping review," *Comput. Methods Programs Biomed. Update*, vol. 2, 2022, Art. no. 100049, doi: 10.1016/j.cmpbup.2022.100049.

[12] W. Algihab, N. Alawwad, A. Aldawish, and S. AlHumoud, "Arabic speech recognition with deep learning: A review," in *Proc. Int. Conf. Hum.-Comput. Interact.* Cham, Switzerland: Springer, 2019, pp. 15–31.

[13] M. N. Zamri and M. S. Sunar, "Atmospheric cloud modeling methods in computer graphics: A review, trends, taxonomy, and future directions," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3468–3488, Jun. 2022.

[14] B. Kitchenham, "Procedures for performing systematic reviews," Dept. Comput. Sci., Keele Univ., Keele, U.K., Tech. Rep. 0400011T.1, vol. 33, 2004, pp. 1–26.

[15] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *J. Bus. Res.*, vol. 104, pp. 333–339, Nov. 2019, doi: 10.1016/j.jbusres.2019.07.039.

[16] A. Canavan and G. Zipperlen, *CALLFRIEND Egyptian Arabic*. Philadelphia, PA, USA: LDC, 1996.

[17] F. Biadsy, J. Hirschberg, and N. Habash, "Spoken Arabic dialect identification using phonotactic modeling," in *Proc. EACL Workshop Comput. Approaches to Semitic Lang. Semitic*, 2009, pp. 53–61.

[18] R. Siemund, B. Heuft, K. Choukri, O. Emam, E. Maragoudakis, H. Tropf, O. Gedge, S. Shammass, A. Moreno, A. N. Rodriguez, I. Zitouni, and D. Iskra, "OrienTel—Arabic speech resources for the IT market," in *Proc. LREC (Arabic Workshop)*, 2002, pp. 1–6.
[19] S. Bougrine, H. Cherroun, D. Ziadi, A. Lakhdari, and A. Chorana, "Toward a rich Arabic speech parallel corpus for Algerian sub-dialects," in *Proc. LREC Workshop Free/Open-Source Arabic Corpora Corpora Process. Tools (OSACT)*, 2016, pp. 2–10.
[20] I. Almosallam, A. AlKhalifa, M. Alghamdi, M. Alkanhal, and A. Alkhairy, "SASSC: A standard Arabic single speaker corpus," in *Proc. 8th ISCA Workshop Speech Synthesis*, 2013, pp. 1–5.
[21] LDC, "Arabic broadcast news speech," Linguistic Data Consortium, Philadelphia, PA, USA, Tech. Rep. LDC2006S46, 2006. Accessed: 2023.
[22] A. Alfaifi and E. Atwell, "Arabic learner corpus," Linguistic Data Consortium, Philadelphia, PA, USA, Tech. Rep. LDC2015S10, 2015. Accessed: 2023. [Online]. Available: https://doi.org/10.35111/5312-x803
[23] N. Halabi, "Arabic speech corpus," Eur. Lang. Resour. Assoc. (ELRA), Paris, France, Tech. Rep. ELRA-S0384, 2016. Accessed: 2023. [Online]. Available: http://catalog.elra.info/en-us/repository/browse/ELRA-S0384/
[24] *The Arabic Speech Corpus for Isolated Words*, Comput. Sci. Math., School Natural Sci., Univ. Stirling, Scotland, U.K., 2014. Accessed: 2023. [Online]. Available: https://www.kaggle.com/datasets/mohamedanwarvic/merged-arabic-corpus-of-isolated-words.
[25] Linguistic Data Consortium, "GALE phase 2 Arabic broadcast news & conversation speech," Philadelphia, PA, USA, Tech. Rep. LDC2013S02, LDC2013S07, LDC2014S07, and LDC2015S01, 2015. Accessed: 2023, doi: 10.35111/2k7p-4w91.
[26] *GALE Phase 3 Arabic Broadcast News & Conversation Speech*, Linguistic Data Consortium, Philadelphia, PA, USA, 2017. Accessed: 2023.
[27] *GALE Phase 4 Arabic Broadcast News & Conversation Speech*, Linguistic Data Consortium, Philadelphia, PA, USA, 2018. Accessed: 2023.
[28] M. Alghmadi, "KACST Arabic phonetic database," in *Proc. 15th Int. Congr. Phonetics Sci.*, Barcelona, 2003, pp. 3109–3112.
[29] King Abdulaziz University. (2011). *King Abdulaziz City for Science and Technology Text to Speech Database (KTD)*. Accessed: 2023. [Online]. Available: http://www.ktddb.org.sa/
[30] (Jul. 18, 2017). *King Saud University Emotions (KSUEmotions) Corpus*. Linguistic Data Consortium, Philadelphia, PA, USA. Accessed: 2023. [Online]. Available: https://doi.org/10.35111/q1eh-6457
[31] M. A.-A.-M. Abushariah, R. N. Ainon, R. Zainuddin, A. A. M. Alqudah, M. Elshafei Ahmed, and O. O. Khalifa, "Modern standard Arabic speech corpus for implementing and evaluating automatic continuous speech recognition systems," *J. Franklin Inst.*, vol. 349, no. 7, pp. 2215–2242, Sep. 2012, doi: 10.1016/j.jfranklin.2011.04.011.
[32] NEMLAR_Project, "NEMLAR broadcast news speech corpus," ELRA Eur. Lang. Resour. Assoc., The Netherlands, Tech. Rep. ELRA-S0219, 2006. Accessed: 2023. [Online]. Available: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0219/
[33] NEMLAR_Project, "NEMLAR speech synthesis corpus," ELRA Eur. Language Resour. Assoc., The Netherlands, Tech. Rep. ELRA-S0220, 2006. Accessed: 2023. [Online]. Available: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0220/
[34] NEMLAR_Project, "NetDC Arabic BNSC (Broadcast News Speech Corpus)," ELRA Eur. Language Resour. Assoc., LCD Linguistic Data Consortium, The Netherlands, Tech. Rep. ELRA-S0157, 2007. Accessed: 2023. [Online]. Available: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0157/
[35] OrienTel_Project, "OrienTel Egypt MSA (Modern Standard Arabic) database," ELRA Eur. Language Resour. Assoc., Tech. Rep. ELRA-S0222, 2007. Accessed: 2023. [Online]. Available: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0222/
[36] OrienTel_Project, "OrienTel Jordan MSA (Modern Standard Arabic) database," ELRA Eur. Language Resour. Assoc., The Netherlands, Tech. Rep. ELRA-S0290, 2008. Accessed: 2023. [Online]. Available: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0290/
[37] OrienTel_Project, "OrienTel Morocco MSA (Modern Standard Arabic) database," ELRA Eur. Language Resour. Assoc., The Netherlands, Tech. Rep. ELRA-S0184, 2005. Accessed: 2023. [Online]. Available: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0184/
[38] OrienTel_Project, "OrienTel Tunisia MSA (Modern Standard Arabic) database," ELRA Eur. Language Resour. Assoc., The Netherlands, Tech. Rep. ELRA-S0187, 2005. Accessed: 2023. [Online]. Available: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0187/
[39] Linguistic Data Consortium, "Phonemes of Arabic," Philadelphia, PA, USA, Tech. Rep. LDC2020S13, Dec. 2020. Accessed: 2023, doi: 10.35111/gkk0-5980.
[40] *Spoken Arabic Digit Database (SAD)*, University of Badji-Mokhtar Annaba, Algeria, 2010. Accessed: 2023. [Online]. Available: http://www.timeseriesclassification.com/description.php?Dataset=SpokenArabicDigits
[41] R. C. Stephen and A. Larocca, "West point Arabic speech," Linguistic Data Consortium, Philadelphia, PA, USA, Tech. Rep. LDC2002S02, 2002. Accessed: 2023.
[42] S. Wray and A. Ali, "Crowdsource a little to label a lot: Labeling a speech corpus of dialectal Arabic," in *Proc. Interspeech*, Sep. 2015, pp. 1–5.
[43] K. Almeman, M. Lee, and A. A. Almiman, "Multi dialect Arabic speech parallel corpora," in *Proc. 1st Int. Conf. Commun., Signal Process., their Appl. (ICCSPA)*, Feb. 2013, pp. 1–6.
[44] K. Almeman, "The building and evaluation of a mobile parallel multi-dialect speech corpus for Arabic," *Proc. Comput. Sci.*, vol. 142, pp. 166–173, Jan. 2018.
[45] *King Saud University Arabic Speech Database*, Linguistic Data Consortium, Philadelphia, PA, USA, 2023. [Online]. Available: https://doi.org/10.35111/vpqe-bz17
[46] A. Ali, S. Vogel, and S. Renals, "Speech recognition challenge in the wild: Arabic MGB-3," in *Proc. IEEE Autom. Speech Recognit. Understand. Workshop (ASRU)*, Dec. 2017, pp. 316–322, doi: 10.1109/ASRU.2017.8268952.
[47] H. Alsayadi, A. Abdelhamid, I. Hegazy, and Z. Taha, "Data augmentation for Arabic speech recognition based on end-to-end deep learning," *Int. J. Intell. Comput. Inf. Sci.*, vol. 21, no. 2, pp. 50–64, Jul. 2021.
[48] H. A. Alsayadi, A. A. Abdelhamid, I. Hegazy, and Z. T. Fayed, "Non-diacritized Arabic speech recognition based on CNN-LSTM and attention-based models," *J. Intell. Fuzzy Syst.*, vol. 41, no. 6, pp. 1–13, 2021, doi: 10.3233/Jifs-202841.
[49] H. A. Alsayadi, A. A. Abdelhamid, I. Hegazy, and Z. T. Fayed, "Arabic speech recognition using end-to-end deep learning," *IET Signal Process.*, vol. 15, no. 8, pp. 521–534, Oct. 2021, doi: 10.1049/sil2.12057.
[50] H. Mubarak, A. Hussein, S. Absar Chowdhury, and A. Ali, "QASR: QCRI aljazeera speech resource—A large scale annotated Arabic speech corpus," 2021, *arXiv:2106.13000*.
[51] A. Ahmed, Y. Hifny, K. Shaalan, and S. Toral, "Lexicon free Arabic speech recognition recipe," in *Proc. Int. Conf. Adv. Intell. Syst. Inform.* Cham, Switzerland: Springer, 2016, pp. 147–159.
[52] B. Dendani, H. Bahi, and T. Sari, "Self-supervised speech enhancement for Arabic speech recognition in real-world environments," *Traitement du Signal*, vol. 38, no. 2, pp. 349–358, Apr. 2021.
[53] I. Zangar, Z. Mnasri, V. Colotte, and D. Jouvet, "F(0) modeling using DNN for Arabic parametric speech synthesis," in *Proc. INNS Big Data Deep Learn. Conf.* Cham, Switzerland: Springer, 2019, pp. 186–195.
[54] I. Zangar, Z. Mnasri, V. Colotte, and D. Jouvet, "Duration modelling and evaluation for Arabic statistical parametric speech synthesis," *Multimedia Tools Appl.*, vol. 80, no. 6, pp. 8331–8353, Mar. 2021.
[55] A. Houidhek, V. Colotte, Z. Mnasri, and D. Jouvet, "Evaluation of speech unit modelling for HMM-based speech synthesis for Arabic," *Int. J. Speech Technol.*, vol. 21, no. 4, pp. 895–906, Dec. 2018.
[56] N. Halabi, "Modern standard Arabic phonetics for speech synthesis," Ph.D. Philosophy, Fac. Phys. Sci. Eng., School Electron. Comput. Sci., Univ. Southampton, Southampton, U.K., 2016. [Online]. Available: https://eprints.soton.ac.uk/409695/
[57] B. Al-Diri, A. Sharieh, and T. Hudaib, "A database for Arabic speech recognition ARABIC_D," Univ. Jordan, Jordan, Tech. Rep., 2002.
[58] B. Ustubioglu, G. Tahaoglu, and G. Ulutas, "Detection of audio copy-move-forgery with novel feature matching on Mel spectrogram," *Exp. Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118963, doi: 10.1016/j.eswa.2022.118963.
[59] A. Ustubioglu, B. Ustubioglu, and G. Ulutas, "Mel spectrogram-based audio forgery detection using CNN," *Signal, Image Video Process.*, vol. 17, no. 5, pp. 2211–2219, Jul. 2023, doi: 10.1007/s11760-022-02436-4.
[60] B. Dendani, H. Bahi, and T. Sari, "Speech enhancement based on deep AutoEncoder for remote Arabic speech recognition," in *Proc. Int. Conf. Image Signal Process.* Cham, Switzerland: Springer, 2020, pp. 221–229.
[61] R. Amari, Z. Noubigh, S. Zrigui, D. Berchech, H. Nicolas, and M. Zrigui, "Deep convolutional neural network for Arabic speech recognition," in *Computational Collective Intelligence*, N. T. Nguyen, Y. Manolopoulos, R. Chbeir, A. Kozierkiewicz, and B. Trawiński, Eds. Cham, Switzerland: Springer, 2022, pp. 120–134.

[62] F. S. Al-Anzi, "Improved noise-resilient isolated words speech recognition using piecewise differentiation," *Fractals*, vol. 30, no. 8, Dec. 2022, Art. no. 2240227, doi: 10.1142/S0218348X22402277.

[63] A. I. Amrous, M. Debyeche, and A. Amrouche, "Integration of auxiliary features in hidden Markov models for Arabic speech recognition," in *Proc. 3rd Int. Conf. Signals, Circuits Syst. (SCS)*, Nov. 2009, pp. 1–5.

[64] M. Debyeche, A. Krobba, and A. Amrouche, "Effect of GSM speech coding on the performance of speaker recognition system," in *Proc. 10th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, May 2010, pp. 137–140.

[65] A. Krobba, M. Debyeche, and A. Amrouche, "Evaluation of speaker identification system using GSMEFR speech data," in *Proc. 5th Int. Conf. Design Technol. Integr. Syst. Nanosc. Era*, Mar. 2010, pp. 1–5.

[66] K. Y. Zergat, A. Amrouche, M. A. Taher, and N. Zainal, "Contribution of prosodic and cepstral features in improvment of a synthesized Arabic speaker recognition task performance," in *Proc. IEEE Student Conf. Res. Developement*, Dec. 2013, pp. 70–73.

[67] K. Y. Zergat, A. Amrouche, M. Fedila, and M. Debyeche, "Robust Arabic speaker verification system using LSF extracted from the G.729 bitstream," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2013, pp. 1–5.

[68] M. Walid, B. Souha, and C. Adnen, "Speech recognition system based on discrete wave atoms transform partial noisy environment," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 466–472, 2019.

[69] O. Abdo, S. Abdou, and M. Fashal, "Building audio-visual phonetically annotated Arabic corpus for expressive text to speech," in *Proc. Interspeech*, Aug. 2017, pp. 3767–3771.

[70] U. Nallasamy, F. Metze, and T. Schultz, "Active learning for accent adaptation in automatic speech recognition," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2012, pp. 360–365.

[71] D. Rybach, S. Hahn, C. Gollan, R. Schluter, and H. Ney, "Advances in Arabic broadcast news transcription at RWTH," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand. (ASRU)*, Jul. 2007, pp. 449–454.

[72] E. Alsharhan and A. Ramsay, "Investigating the effects of gender, dialect, and training size on the performance of Arabic speech recognition," *Lang. Resour. Eval.*, vol. 54, no. 4, pp. 975–998, Dec. 2020, doi: 10.1007/s10579-020-09505-5.

[73] H. Soltau, G. Saon, B. Kingsbury, H.-K.-J. Kuo, L. Mangu, D. Povey, and A. Emami, "Advances in Arabic speech transcription at IBM under the DARPA GALE program," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 5, pp. 884–894, Jul. 2009, doi: 10.1109/Tasl.2009.2022966.

[74] E. Alsharhan, A. Ramsay, and H. Ahmed, "Evaluating the effect of using different transcription schemes in building a speech recognition system for Arabic," *Int. J. Speech Technol.*, vol. 25, no. 1, pp. 43–56, Mar. 2022, doi: 10.1007/s10772-020-09720-z.

[75] M. Alghamdi, Y. O. M. El Hadj, and M. Alkanhal, "A manual system to segment and transcribe Arabic speech," in *Proc. IEEE Int. Conf. Signal Process. Commun.*, Jun. 2007, pp. 233–236.

[76] Y. A. Alotaibi and A. Hussain, "Comparative analysis of Arabic vowels using formants and an automatic speech recognition system," *Int. J. Signal Process., Image Process. Pattern Recognit.*, vol. 3, no. 2, pp. 11–22, 2010.

[77] I. A. Maaly and M. El-Obaid, "Speech recognition using artificial neural networks," in *Proc. 2nd Int. Conf. Inf. Commun. Technol.*, vol. 1, 2006, pp. 1246–1247.

[78] Y. M. Seddiq, Y. A. Alotaibi, and S. Selouani, "Classification of emphatic consonants and their counterparts in modern standard Arabic using neural networks," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2014, pp. 000073–000077.

[79] Y. Seddiq, A. Meftah, M. Alghamdi, and Y. Alotaibi, "Reintroducing KAPD as a dataset for machine learning and data mining applications," in *Proc. Eur. Model. Symp. (EMS)*, Nov. 2016, pp. 70–74.

[80] F. M. Aloqayli and Y. A. Alotaibi, "Analyzing vowel triangles in spoken Arabic dialects," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2018, pp. 1–6.

[81] M. A. Qamhan, Y. A. Alotaibi, Y. M. Seddiq, A. H. Meftah, and S. A. Selouani, "Sequence-to-sequence acoustic-to-phonetic conversion using spectrograms and deep learning," *IEEE Access*, vol. 9, pp. 80209–80220, 2021.

[82] A. H. Abo Absa, M. Deriche, M. Elshafei-Ahmed, Y. M. Elhadj, and B. Juang, "A hybrid unsupervised segmentation algorithm for Arabic speech using feature fusion and a genetic algorithm (July 2018)," *IEEE Access*, vol. 6, pp. 43157–43169, 2018, doi: 10.1109/Access.2018.2859631.

[83] Y. Seddiq, Y. Alotaibi, A. Meftah, S.-A. Selouani, and M. Alghamdi, "Revisiting distinctive phonetic features from applied computing perspective: Unifying views and analyzing modern Arabic speech varieties," *Int. J. Speech Technol.*, vol. 21, no. 4, pp. 907–913, Dec. 2018.

[84] K. M. Nahar, M. Elshafei, W. G. Al-Khatib, H. Al-Muhtaseb, and M. M. Alghamdi, "Statistical analysis of Arabic phonemes used in Arabic speech recognition," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2012, pp. 533–542.

[85] K. M. O. Nahar, W. G. Al-Khatib, M. Elshafei, H. Al-Muhtaseb, and M. M. Alghamdi, "Arabic phonemes transcription using learning vector quantization: 'Towards the development of fast Quranic text transcription," in *Proc. Taibah Univ. Int. Conf. Adv. Inf. Technol. Holy Quran Sci.*, Dec. 2013, pp. 407–412, doi: 10.1109/NOORIC.2013.85.

[86] M. Algabri, H. Mathkour, M. M. Alsulaiman, and M. A. Bencherif, "Deep learning-based detection of articulatory features in Arabic and English speech," *Sensors*, vol. 21, no. 4, p. 1205, Feb. 2021. [Online]. Available: https://mdpi-res.com/d_attachment/sensors/sensors-21-01205/article_deploy/sensors-21-01205-v2.pdf

[87] M. Aissiou, "A genetic model for acoustic and phonetic decoding of standard Arabic vowels in continuous speech," *Int. J. Speech Technol.*, vol. 23, no. 2, pp. 425–434, Jun. 2020, doi: 10.1007/s10772-020-09694-y.

[88] M. A. Al-Manie, M. I. Alkanhal, M. M. Al-Ghamdi, and N. Mastorakis, "Automatic speech segmentation using the Arabic phonetic database," in *Proc. WSEAS Int. Conf. Math. Comput. Sci. Eng.*, 2009, no. 10, pp. 1–9.

[89] Y. O. M. Elhadj, M. Alghamdi, and M. Alkanhal, "Approach for recognizing allophonic sounds of the classical Arabic based on Quran recitations," in *Proc. Int. Conf. Theory Pract. Natural Comput.* Cham, Switzerland: Springer, 2013, pp. 57–67.

[90] S. Bougrine, H. Cherroun, and A. Abdelali, "Altruistic crowdsourcing for Arabic speech corpus annotation," *Procedia Comput. Sci.*, vol. 117, pp. 137–144, 2017, doi: 10.1016/j.procs.2017.10.102.

[91] Y. A. Alotaibi, A. H. Meftah, and S.-A. Selouani, "Investigation of emotion classification using speech rhythm metrics," in *Proc. IEEE Digit. Signal Process. Signal Process. Educ. Meeting (DSP/SPE)*, Aug. 2013, pp. 204–209.

[92] A. B. Ibrahim, Y. M. Seddiq, A. H. Meftah, M. Alghamdi, S. Selouani, M. A. Qamhan, Y. A. Alotaibi, and S. A. Alshebeili, "Optimizing Arabic speech distinctive phonetic features and phoneme recognition using genetic algorithm," *IEEE Access*, vol. 8, pp. 200395–200411, 2020.

[93] A. H. Meftah, Y. Alotaibi, and S. Selouani, "Investigating Arabic speakers' emotions using speech rhythm metrics," in *Proc. Eur. Model. Symp. (EMS)*, Nov. 2017, pp. 73–77.

[94] M. Alsulaiman, Z. Ali, G. Muhammad, A. Al Hindi, T. Alfakih, H. Obeidat, and S. Al-Kahtani, "Pronunciation errors of non-arab learners of Arabic language," in *Proc. Int. Conf. Comput., Commun., Control Technol. (ICT)*, Sep. 2014, pp. 277–282.

[95] M. Algabri, H. Mathkour, M. A. Bencherif, M. Alsulaiman, and M. A. Mekhtiche, "Automatic speaker recognition for mobile forensic applications," *Mobile Inf. Syst.*, vol. 2017, Mar. 2017, Art. no. 6986391, doi: 10.1155/2017/6986391.

[96] M. A. Qamhan, H. Altaheri, A. H. Meftah, G. Muhammad, and Y. A. Alotaibi, "Digital audio forensics: Microphone and environment classification using deep learning," *IEEE Access*, vol. 9, pp. 62719–62733, 2021, doi: 10.1109/Access.2021.3073786.

[97] M. Alsulaiman, "Effect of spoken text on text-independent speaker recognition," in *Proc. 5th Int. Conf. Intell. Syst., Model. Simul.*, Jan. 2014, pp. 279–284.

[98] M. Algabri, H. Mathkour, M. Alsulaiman, and M. A. Bencherif, "Mispronunciation detection and diagnosis with articulatory-level feedback generation for non-native Arabic speech," *Mathematics*, vol. 10, no. 15, p. 2727, Aug. 2022, doi: 10.3390/math10152727.

[99] A. Meftah, Y. Seddiq, Y. Alotaibi, and S. Selouani, "Cross-corpus Arabic and English emotion recognition," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2017, pp. 377–381.

[100] A. H. Meftah, Y. A. Alotaibi, and S. Selouani, "Evaluation of an Arabic speech corpus of emotions: A perceptual and statistical analysis," *IEEE Access*, vol. 6, pp. 72845–72861, 2018, doi: 10.1109/Access.2018.2881096.

[101] Y. Hifny and A. Ali, "Efficient Arabic emotion recognition using deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 6710–6714.

[102] A. H. Meftah, M. Qamhan, Y. Alotaibi, and S. Selouani, "Emotional speech recognition using rhythm metrics and a new Arabic corpus," in *Proc. 16th IEEE Int. Colloq. Signal Process. Its Appl. (CSPA)*, Feb. 2020, pp. 57–62.

[103] A. H. Meftah, M. A. Qamhan, Y. Seddiq, Y. A. Alotaibi, and S. A. Selouani, "King saud university emotions corpus: Construction, analysis, evaluation, and comparison," *IEEE Access*, vol. 9, pp. 54201–54219, 2021, doi: 10.1109/Access.2021.3070751.
[104] A. Meftah, S. Selouani, and Y. A. Alotaibi, "Preliminary Arabic speech emotion classification," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2014, pp. 000179–000182.
[105] A. Mefiah, Y. A. Alotaibi, and S. Selouani, "Arabic speaker emotion classification using rhythm metrics and neural networks," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 1426–1430.
[106] A. A. Alvarez, E. Issa, and M. Alshakhori, "Computational modeling of intonation patterns in Arabic emotional speech," in *Proc. SpeechProsody*, May 2022, pp. 615–619.
[107] K. Almeman and M. Lee, "A comparison of Arabic speech recognition for multi-dialect vs. specific dialects," in *Proc. 7th Int. Conf. Speech Technol. Hum.-Comput. Dialogue (SpeD)*, Cluj-Napoca, Romania, 2013, pp. 16–19.
[108] A. A. Alashban, M. A. Qamhan, A. H. Meftah, and Y. A. Alotaibi, "Spoken language identification system using convolutional recurrent neural network," *Appl. Sci.*, vol. 12, no. 18, p. 9181, Sep. 2022, doi: 10.3390/app12189181.
[109] K. Choukri, M. Nikkhou, and N. Paulsson, "Network of data centres (NetDC): BNSC—An Arabic broadcast news speech corpus," in *Proc. LREC*, 2004, pp. 1–4.
[110] M. Alkanhal, M. Alghamdi, and Z. Muzaffar, "Speaker verification based on Saudi accented Arabic database," in *Proc. 9th Int. Symp. Signal Process. Appl.*, Feb. 2007, pp. 1–4.
[111] A. A. Altuwaim, Y. A. Alotaibi, and S. Selouani, "Investigation into the speech rhythm of two Saudi dialects using the SAAVB corpus," in *Proc. 6th Int. Symp. Commun., Control Signal Process. (ISCCSP)*, May 2014, pp. 632–635.
[112] R. F. B. Ikrish, Y. A. Alotaibi, and S. B. Sandouka, "Extended speech rhythm-based analysis of Saudi dialects using SAAVB corpus," in *Proc. 14th Int. Conf. Innov. Inf. Technol. (IIT)*, Nov. 2020, pp. 172–177.
[113] S. B. Sandouka and Y. A. Alotaibi, "Extended rhythm-based investigation of Saudi dialects using the Saudi accented Arabic voice bank corpus," in *Proc. Int. Conf. Commun., Signal Process., Appl. (ICCSPA)*, Mar. 2021, pp. 1–5.
[114] M. M. Alghamdi and Y. A. Alotaibi, "HMM automatic speech recognition system of Arabic alphadigits," *Arabian J. Sci. Eng.*, vol. 35, no. 2, p. 137, Dec. 2010.
[115] M. Elmahdy, R. Gruhn, and W. Minker, "Speech corpora," in *Novel Techniques for Dialectal Arabic Speech Recognition*, M. Elmahdy, R. Gruhn, W. Minker, Eds. Boston, MA, USA: Springer, 2012, pp. 25–32.
[116] Y. Tabet, M. Boughazi, and S. Aïfi, "Speech analysis and synthesis with a refined adaptive sinusoidal representation," *Int. J. Speech Technol.*, vol. 21, no. 3, pp. 581–588, Sep. 2018.
[117] G. Droua-Hamdani, S.-A. Selouani, Y. A. Alotaibi, and M. Boudraa, "Speech rhythm in L1 and L2 Arabic," *Arabian J. Sci. Eng.*, vol. 41, no. 3, pp. 1173–1181, Mar. 2016.
[118] A. R. Ali, "Multi-dialect Arabic speech recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7.
[119] M. Elmahdy, M. Hasegawa-Johnson, and E. Mustafawi, "Development of a tv broadcasts speech recognition system for Qatari Arabic," in *Proc. LREC*, 2014, pp. 3057–3061.
[120] M. A. Menacer, O. Mella, D. Fohr, D. Jouvet, D. Langlois, and K. Smaïli, "Development of the Arabic Loria automatic speech recognition system (ALASR) and its evaluation for Algerian dialect," *Proc. Comput. Sci.*, vol. 117, pp. 81–88, Jan. 2017, doi: 10.1016/j.procs.2017.10.096.
[121] N. Terbeh and M. Zrigui, "Vocal pathologies detection and mispronounced phonemes identification: Case of Arabic continuous speech," in *Proc. 10th Int. Conf. Lang. Resour. Eval. (LREC)*, 2016, pp. 2108–2113.
[122] M. Elmahdy, R. Gruhn, W. Minker, and S. Abdennadher, "Cross-lingual acoustic modeling for dialectal Arabic speech recognition," in *Proc. Interspeech*, Sep. 2010, pp. 1–6.
[123] N. Terbeh, M. Maraoui, and M. Zrigui, "Probabilistic approach for detection of vocal pathologies in the Arabic speech," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*. Cham, Switzerland: Springer, 2015, pp. 606–616.
[124] R. Amari, A. Mars, and M. Zrigui, "Arabic speech recognition based on a CNN-BLSTM combination," in *Proc. IEEE 9th Int. Conf. Sci. Electron., Technol. Inf. Telecommun. (SETIT)*, May 2022, pp. 259–264, doi: 10.1109/SETIT54465.2022.9875681.
[125] A. Hussein, S. Watanabe, and A. Ali, "Arabic speech recognition by end-to-end, modular systems and human," *Comput. Speech Lang.*, vol. 71, Jan. 2022, Art. no. 101272, doi: 10.1016/j.csl.2021.101272.
[126] H. A. Alsayadi, S. Al-Hagree, F. A. Alqasemi, and A. A. Abdelhamid, "Dialectal Arabic speech recognition using CNN-LSTM based on end-to-end deep learning," in *Proc. 2nd Int. Conf. Emerg. Smart Technol. Appl. (eSmarTA)*, Oct. 2022, pp. 1–8, doi: 10.1109/eSmarTA56775.2022.9935427.
[127] N. Hammami, M. Bedda, and N. Farah, "Probabilistic classification based on Gaussian copula for speech recognition: Application to spoken Arabic digits," in *Proc. Signal Process., Algorithms, Architectures, Arrangements, Appl. (SPA)*, 2013, pp. 312–317.
[128] A. H. Abo absa and M. Deriche, "On the performance of ensemble-based classifiers for Arabic speech recognition," in *Proc. 4th IEEE Int. Conf. Eng. Technol. Appl. Sci. (ICETAS)*, Nov. 2017, pp. 1–5.
[129] S.-A. Selouani and Y. A. Alotaibi, "Adaptation of foreign accented speakers in native Arabic ASR systems," *Appl. Comput. Informat.*, vol. 9, no. 1, pp. 1–10, Jan. 2011.
[130] S.-A. Selouani and Y. A. Alotaibi, "Investigating automatic recognition of non-native Arabic speech," in *Proc. Innov. Inf. Technol. (IIT)*, Nov. 2007, pp. 451–455.
[131] Y. A. Alotaibi and S.-A. Selouani, "Investigating the adaptation of Arabic speech recognition systems to foreign accented speakers," in *Proc. 10th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, May 2010, pp. 646–649.
[132] Y. A. Alotaibi, S. Selouani, M. M. Alghamdi, and A. H. Meftah, "Arabic and English speech recognition using cross-language acoustic models," in *Proc. 11th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, Jul. 2012, pp. 40–44.
[133] G. Droua-Hamdani, "ANN-MLP classifier of native and nonnative speakers using speech rhythm cues," in *Proc. Int. Conf. Stat. Lang. Speech Process.* Cham, Switzerland: Springer, 2020, pp. 149–156.
[134] Y. A. Alotaibi and S.-A. Selouani, "Evaluating the MSA west point speech corpus," *Int. J. Comput. Process. Lang.*, vol. 22, no. 4, pp. 285–304, Dec. 2009.
[135] G. Droua-Hamdani, Y. A. Alotaibi, S.-A. Selouani, and M. Boudraa, "Rhythmic features across modern standard Arabic and Arabic dialects," in *Proc. Workshop Free/Open-Source Arabic Corpora Corpora Processing Tools*, 2014, pp. 43–47.
[136] Y. A. Alotaibi, S.-A. Selouani, and W. Cichocki, "Investigating emphatic consonants in foreign accented Arabic," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 21, pp. 13–25, Jan. 2009.
[137] S. Gharsellaoui, S. A. Selouani, W. Cichocki, Y. Alotaibi, and A. O. Dahmane, "Application of the pairwise variability index of speech rhythm with particle swarm optimization to the classification of native and non-native accents," *Comput. Speech Lang.*, vol. 48, pp. 67–79, Mar. 2018, doi: 10.1016/j.csl.2017.10.006.
[138] S.-A. Selouani, Y. Alotaibi, W. Cichocki, S. Gharsellaoui, and K. Kadi, "Native and non-native class discrimination using speech rhythm- and auditory-based cues," *Comput. Speech Lang.*, vol. 31, no. 1, pp. 28–48, May 2015.
[139] Y. A. Alotaibi and G. Muhammad, "Study on pharyngeal and uvular consonants in foreign accented Arabic for ASR," *Comput. Speech Lang.*, vol. 24, no. 2, pp. 219–231, Apr. 2010, doi: 10.1016/j.csl.2009.04.005.
[140] Y. Alotaibi, S.-A. Selouani, and D. O'shaughnessy, "Experiments on automatic recognition of nonnative Arabic speech," *EURASIP J. Audio, Speech, Music Process.*, vol. 2008, Feb. 2008, Art. no. 679831, doi: Artn.67983110.1155/2008/679831.
[141] S.-A. Selouani and Y. A. Alotaibi, "Adaptation to non-native speech using evolutionary-based discriminative linear transforms," *Eng. Appl. Artif. Intell.*, vol. 26, no. 2, pp. 899–904, Feb. 2013, doi: 10.1016/j.engappai.2012.06.006.
[142] A. E. Kadhi, F. Gherri, and H. Amiri, "Building diphone database for Arabic text to speech synthesis system," in *Proc. 3rd Int. Conf. Control, Eng. Inf. Technol. (CEIT)*, May 2015, pp. 1–5.
[143] A. Alshutayri and H. Albarhamtoshy, "Arabic spoken language identification system (ASLIS): A proposed system to identifying modern standard Arabic (MSA) and Egyptian dialect," in *Proc. Int. Conf. Informat. Eng. Inf. Sci.* Cham, Switzerland: Springer, 2011, pp. 375–385.
[144] O. Zine and A. Meziane, "Novel approach for quality enhancement of Arabic text to speech synthesis," in *Proc. Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, May 2017, pp. 1–6, doi: 10.1109/ATSIP.2017.8075550.

[145] O. Zine, A. Meziane, and M. Boudchiche, "Towards a high-quality lemma-based text to speech system for the Arabic language," in *Proc. Int. Conf. Arabic Lang. Process.* Cham, Switzerland: Springers, 2017, pp. 53–66.

[146] M. A. M. Abushariah, R. N. Ainon, R. Zainuddin, O. O. Khalifa, and M. Elshafei, "Phonetically rich and balanced Arabic speech corpus: An overview," in *Proc. Int. Conf. Comput. Commun. Eng. (ICCCE)*, May 2010, pp. 1–6.

[147] A. Amrouche, A. Abed, and L. Falek, "Arabic speech synthesis system based on HMM," in *Proc. 6th Int. Conf. Electr. Electron. Eng. (ICEEE)*, Apr. 2019, pp. 73–78.

[148] A. Amrouche, A. Abed, K. Ferrat, K. N. Boubakeur, Y. Bentrcia, and L. Falek, "Balanced Arabic corpus design for speech synthesis," *Int. J. Speech Technol.*, vol. 24, no. 3, pp. 1–13, Sep. 2021, doi: 10.1007/s10772-021-09846-8.

[149] Z. Oumaima and A. Meziane, "Modern Arabic speech corpus for text to speech synthesis," in *Proc. IEEE Int. Conf. Technol. Manage., Operations Decisions (ICTMOD)*, Nov. 2020, pp. 1–6.

[150] I. Hadj Ali and Z. Mnasri, "Statistical analysis of the prosodic parameters of a spontaneous Arabic speech corpus for speech synthesis," in *Int. Conf. Stat. Lang. Speech Process.* Cham, Switzerland: Springer, 2016, pp. 57–67.

[151] S. S. AlDahri, "A study of voice onset time for modern standard Arabic and classical Arabic," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Aug. 2012, pp. 691–695.

[152] W. M. Hamza and M. A. Rashwan, "Concatenative Arabic speech synthesis using large speech database," in *Proc. 6th Int. Conf. Spoken Lang. Process. (ICSLP)*, Oct. 2000, pp. 1–5.

[153] N. Hammami, M. Bedda, N. Farah, and S. Mansouri, "/R/-letter disorder diagnosis (/r/-LDD): Arabic speech database development for automatic diagnosis of childhood speech disorders (case study)," in *Proc. Intell. Syst. Comput. Vis. (ISCV)*, Mar. 2015, pp. 1–7.

[154] F. Chouireb and M. Guerti, "Towards a high quality Arabic speech synthesis system based on neural networks and residual excited vocal tract model," *Signal, Image Video Process.*, vol. 2, no. 1, pp. 73–87, Jan. 2008.

[155] N. Merad-Boudia, A. Benyettou, and A. Rubio Ayuso, "Arabic speech recognition for connected words using HTK: Triphones expanded to gmm based Quran recognition," *Int. Rev. Comput. Softw. (IRECOS)*, vol. 11, no. 12, p. 1209, Dec. 2016.

[156] I. Rebai and Y. BenAyed, "Arabic speech synthesis and diacritic recognition," *Int. J. Speech Technol.*, vol. 19, no. 3, pp. 485–494, Sep. 2016, doi: 10.1007/s10772-016-9342-8.

[157] Y. A. Alotaibi, "Is phoneme level better than word level for HMM models in limited vocabulary ASR systems?" in *Proc. 7th Int. Conf. Inf. Technology, New Generat.*, 2010, pp. 332–337.

[158] F. Boukadida and N. Ellouze, "Arabic intonatif speech database," in *Proc. IEEE Int. Conf. Ind. Technol.*, Apr. 2004, pp. 1408–1411.

[159] G. Muhammad, M. Alsulaiman, A. Mahmood, and Z. Ali, "Automatic voice disorder classification using vowel formants," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–6.

[160] M. K. Krichi and C. Adnan, "The Arabic speech database: PADAS," *Signal Process., Int. J.*, vol. 8, no. 2, pp. 10–19, 2014.

[161] K. M. Khalil and C. Adnan, "Implementation of speech synthesis based on HMM using PADAS database," in *Proc. IEEE 12th Int. Multi-Conference Syst., Signals Devices (SSD15)*, Mar. 2015, pp. 1–6.

[162] M. K. Krichi and C. Adnan, "Phonetics Arabic database automatically segmented: PADAS," in *Proc. Conférence Internationale en Automatique Traitement de Signal (ATS)*, vol. 23. Tunisia: Proceedings of Engineering and Technology—PET, 2017, pp. 28–32.

[163] M. M. Azmi and H. Tolba, "Syllable-based recognition of Arabic & English digits in noisy environment," in *Proc. 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 583–586.

[164] M. A. Ahmad and R. M. El Awady, "Phonetic recognition of Arabic alphabet letters using neural networks," *Int. J. Electr. Comput. Sci. IJECS-IJENS*, vol. 11, no. 1, pp. 1–6, 2011.

[165] S. S. AlDahri and H. A. Alhakami, "Detection of voice onset time (VOT) for unvoiced stop sound in modern standard Arabic (MSA) based on power signal," in *Proc. IEEE 13th Int. Conf. Signal Process. (ICSP)*, Nov. 2016, pp. 551–556.

[166] M. Irfan, I. Z. Mutaqin, and R. G. Utomo, "Implementation of dynamic time warping algorithm on an Android based application to write and pronounce hijaiyah letters," in *Proc. 4th Int. Conf. Cyber IT Service Manage.*, Apr. 2016, pp. 1–6.

[167] N. Zerari, S. Abdelhamid, H. Bouzgou, and C. Raymond, "Bi-directional recurrent end-to-end neural network classifier for spoken Arab digit recognition," in *Proc. 2nd Int. Conf. Natural Lang. Speech Process. (ICNLSP)*, Apr. 2018, pp. 1–6.

[168] N. Terbeh, M. Labidi, and M. Zrigui, "Automatic speech correction: A step to speech recognition for people with disabilities," in *Proc. 4th Int. Conf. Inf. Commun. Technol. Accessibility (ICTA)*, Oct. 2013, pp. 1–6.

[169] M. C. Amara Korba, H. Bourouba, and R. Djemili, "Feature extraction algorithm using new cepstral techniques for robust speech recognition," *Malaysian J. Comput. Sci.*, vol. 33, no. 2, pp. 90–101, Apr. 2020, doi: 10.22452/mjcs.vol33no2.1.

[170] S. Ismail and A. Ahmad, "Recurrent neural network with back propagation through time algorithm for Arabic recognition," *Proc. 18th ESM*, Magdeburg, Germany, 2004, pp. 13–16.

[171] M. Nofal, E. Abdel-Raheem, H. El Henawy, and N. S. Abdel Kader, "Arabic automatic segmentation system and its application for Arabic speech recognition system," in *Proc. 46th Midwest Symp. Circuits Syst.*, 2003, pp. 697–700.

[172] M. Nofal, E. A. Reheem, H. El Henawy, and N. A. Kader, "The development of acoustic models for command and control Arabic speech recognition system," in *Proc. Int. Conf. Electr., Electron. Comput. Eng. (ICEEC)*, 2004, pp. 702–705.

[173] M. Nofal, E. Abdel-Raheem, H. E. Henawy, and N. A. Kader, "Acoustic training system for speaker independent continuous Arabic speech recognition system," in *Proc. 4th IEEE Int. Symp. Signal Process. Inf. Technol.*, Jun. 2004, pp. 200–203.

[174] B. Al-Diri, A. Sharieh, and T. Hudaib, "An Arabic speech corpus: A database for Arabic speech recognition," *Dirasat, Pure Sci.*, vol. 31, no. 2, pp. 208–219, 2004.

[175] Z. Talai, H. Bahi, and N. Kherici, "Remote spoken Arabic digits recognition using CNN," in *Proc. 11th Int. Conf. Robot., Vis., Signal Process. Power Appl.*, N. M. Mahyuddin, N. R. Mat Noor, and H. A. Mat Sakim, Eds. Singapore: Springer, 2022, pp. 639–645.

[176] M. A. M. Abushariah, R. N. Ainon, R. Zainuddin, M. Elshafei, and O. O. Khalifa, "Natural speaker-independent Arabic speech recognition system based on hidden Markov models using sphinx tools," in *Proc. Int. Conf. Comput. Commun. Eng. (ICCCE)*, May 2010, pp. 1–6.

[177] A. Touazi and M. Debyeche, "An experimental framework for Arabic digits speech recognition in noisy environments," *Int. J. Speech Technol.*, vol. 20, no. 2, pp. 205–224, Jun. 2017.

[178] T. Saidane, M. Zrigui, and M. B. Ahmed, "Arabic speech synthesis using a concatenation of polyphones: The results," in *Proc. Conf. Canadian Soc. Comput. Stud. Intell.* Cham, Switzerland: Springer, 2005, pp. 406–411.

[179] M. H. Sanchez, L. Ferrer, E. Shriberg, and A. Stolcke, "Constrained cepstral speaker recognition using matched UBM and JFA training," in *Proc. Interspeech*, Aug. 2011, pp. 1–6.

[180] I. Kamel and H. Talhami, "An efficient mechanism for searching Arabic audio libraries," in *Proc. IEEE 6th Int. Symp. Multimedia Softw. Eng.*, Dec. 2004, pp. 579–582, doi: 10.1109/MMSE.2004.24.

[181] Y. Alotaibi, M. Al-Ghamdi, and F. Alotaiby, "Speech recognition system of Arabic digits based on a telephony Arabic corpus," in *Proc. IPCV*. Princeton, NJ, USA: citeseer, 2008, pp. 35–38.

[182] Y. A. Alotaibi, M. Alghamdi, and F. Alotaiby, "Speech recognition system of Arabic alphabet based on a telephony Arabic corpus," in *Proc. Int. Conf. Image Signal Process.* Cham, Switzerland: Springer, 2010, pp. 122–129.

[183] E. M. Essa, A. S. Tolba, and S. Elmougy, "A comparison of combined classifier architectures for Arabic speech recognition," in *Proc. Int. Conf. Comput. Eng. Syst.*, Nov. 2008, pp. 149–153.

[184] S. Abed, M. Alshayeji, and S. Sultan, "Diacritics effect on Arabic speech recognition," *Arabian J. Sci. Eng.*, vol. 44, no. 11, pp. 9043–9056, Nov. 2019, doi: 10.1007/s13369-019-04024-0.

[185] F. Mouria-Beji, "Neural network optimization using genetic algorithms for speech recognition," *Int. J. Eng. Intell. Syst. Electr. Eng. Commun.*, vol. 10, no. 2, pp. 69–74, Jun. 2002.

[186] M. M. Alsulaiman, G. Muhammad, M. A. Bencherif, A. Mahmood, Z. Ali, and M. Aljabri, "Building a rich Arabic speech database," in *Proc. 5th Asia Model. Symp.*, May 2011, pp. 100–105.

[187] I. Jemaa, K. Ouni, and Y. Laprie, "Evaluation of automatic formant tracking method using Fourier ridges," *Cognit. Comput.*, vol. 2, no. 3, pp. 170–179, Sep. 2010, doi: 10.1007/s12559-010-9055-2.

[188] D. AbuZeina, W. Al-Khatib, M. Elshafei, and H. Al-Muhtaseb, "Toward enhanced Arabic speech recognition using part of speech tagging," *Int. J. Speech Technol.*, vol. 14, no. 4, pp. 419–426, Dec. 2011.
[189] D. AbuZeina, W. Al-Khatib, M. Elshafei, and H. Al-Muhtaseb, "Within-word pronunciation variation modeling for Arabic ASRs: A direct data-driven approach," *Int. J. Speech Technol.*, vol. 15, no. 2, pp. 65–75, Jun. 2012.
[190] D. AbuZeina, M. Elshafei, and W. Al-Khatib, "Part of speech tagging approach to designing compound words for Arabic continuous speech recognition systems," in *Proc. Int. Conf. Informat. Eng. Inf. Sci.* Cham, Switzerland: Springer, 2011, pp. 330–338.
[191] D. AbuZeina, W. Al-khatib, and M. Elshafei, "Small-word pronunciation modeling for Arabic speech recognition: A data-driven approach," in *Proc. Asia Inf. Retr. Symp.* Cham, Switzerland: Springer, 2011, pp. 529–537.
[192] D. Abuzeina, M. Elshafei, and J. Zhou, "Pronunciation variation modeling for Arabic ASRs: A direct data-driven approach," in *Proc. 3rd Int. Conf. Comput. Eng. Technol. (ICCET)*. New York, NY, USA: ASME Press, 2011.
[193] K. Lounnas, H. Satori, M. Hamidi, H. Teffahi, M. Abbas, and M. Lichouri, "CLIASR: A combined automatic speech recognition and language identification system," in *Proc. 1st Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET)*, Apr. 2020, pp. 1–5.
[194] S. A. Selouani and M. Boudraa, "Algerian Arabic speech database (ALGASD): corpus design and automatic speech recognition application," *Arabian J. Sci. Eng.*, vol. 35, no. 2, pp. 157–166, 2010.
[195] G. Droua-Hamdani and M. Boudraa, "Rhythm metrics in MSA spoken language of six Algerian regions," in *Proc. 15th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, Dec. 2015, pp. 325–328.
[196] G. Droua-Hamdani, S. Sellouani, and M. Boudraa, "Effect of characteristics of speakers on MSA ASR performance," in *Proc. 1st Int. Conf. Commun., Signal Process., their Appl. (ICCSPA)*, Feb. 2013, pp. 1–5.
[197] A. Ahcène, A. Aissa, D. Abdelkader, B. Khadidja, and D. Ghania, "Automatic segmentation of Arabic speech signals by HMM and ANN," in *Proc. Int. Conf. Electr. Sci. Technol. Maghreb (CISTEM)*, Oct. 2016, pp. 1–4.
[198] B. F. Zaidi, M. Boudraa, S.-A. Selouani, M. S. Yakoub, and G. Hamdani, "Control interface of an automatic continuous speech recognition system in standard Arabic language," in *Proc. SAI Intell. Syst. Conf.* Cham, Switzerland: Springer, 2020, pp. 295–303.
[199] G. Droua-Hamdani, "Classification of regional accent using speech rhythm metrics," in *Proc. Int. Conf. Speech Comput.* , Cham, Switzerland: Springer, 2019, pp. 75–81.
[200] G. Droua-Hamdani, M. Boudraa, and S. A. Selouani, "Algerian Arabic speech database project (ALGASD): Corpora's elaboration," in *Proc. 3rd Int. Conf. Arabic Lang. Process. (CITALA)*, Rabat, Morocco, 2009, pp. 4–5.
[201] G. Droua-Hamdani, S.-A. Selouani, and M. Boudraa, "Speaker-independent ASR for modern standard Arabic: Effect of regional accents," *Int. J. Speech Technol.*, vol. 15, no. 4, pp. 487–493, Dec. 2012.
[202] G. Droua-Hamdani, "Formant frequency analysis of MSA vowels in six Algerian regions," in *Proc. Int. Conf. Speech Comput.* Cham, Switzerland: Springer, 2020, pp. 128–135.
[203] A. M. Jasim, S. R. Awad, F. L. Malallah, and J. M. Abdul-Jabbar, "Efficient gender classifier for Arabic speech using CNN with dimensional reshaping," in *Proc. 7th Int. Conf. Electr., Electron. Inf. Eng. (ICEEIE)*, Oct. 2021, pp. 1–5.
[204] F. Nazir, M. N. Majeed, M. A. Ghazanfar, and M. Maqsood, "A computer-aided speech analytics approach for pronunciation feedback using deep feature clustering," *Multimedia Syst.*, vol. 29, pp. 1–17, Jul. 2021, doi: 10.1007/s00530-021-00822-5.
[205] L. Demri, L. Falek, and H. Teffahi, "Contribution to the creation of an Arabic expressive speech corpus," *Acta Acustica United Acustica*, vol. 101, no. 5, pp. 1052–1063, Sep. 2015.
[206] B. A. Q. Al-Qatab and R. N. Ainon, "Arabic speech recognition using hidden Markov model toolkit (HTK)," in *Proc. Int. Symp. Inf. Technol.*, Jun. 2010, pp. 557–562.
[207] S. S. AlDahri, "The effect of Arabic emphaticness on voice time onset (VOT)," in *Proc. Int. Conf. Audio, Lang. Image Process.*, Jul. 2012, pp. 426–431.
[208] R. Abdelmalek and Z. Mnasri, "High quality Arabic text-to-speech synthesis using unit selection," in *Proc. 13th Int. Multi-Conference Syst., Signals Devices (SSD)*, Mar. 2016, pp. 1–5.
[209] E. F. Khalaf, K. Daqrouq, and A. Morfeq, "Arabic vowels recognition by modular arithmetic and wavelets using neural network," *Life Sci. J.*, vol. 11, no. 3, pp. 33–41, 2014.
[210] M. Maamouri, A. Bies, and S. Kulick, "Expanding Arabic treebank to speech: Results from broadcast news," in *Proc. LREC*. Princeton, NJ, USA: Citeseer, 2012, pp. 1856–1861.
[211] M. Khasawneh, K. Assaleh, W. Sweidan, and M. Haddad, "The application of polynomial discriminant function classifiers to isolated Arabic speech recognition," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Dec. 2004, pp. 3077–3081.
[212] N. Halabi, "Arabic speech corpus," in *Oxford Text Archive Core Collection*. Oxford, U.K.: Univ. of Oxford, 2016. [Online]. Available: https://downloads.it.ox.ac.uk/ota-public/audio/2561.zip
[213] A. A. Almisreb, A. F. Abidin, and N. M. Tahir, "Continuous Arabic alphabets corpus uttered by Malay speakers," in *Proc. IEEE Conf. Syst., Process Control (ICSPC)*, Dec. 2014, pp. 101–104.
[214] A. AbdAlmisreb, A. F. Abidin, and N. M. Tahir, "Maxout based deep neural networks for Arabic phonemes recognition," in *Proc. IEEE 11th Int. Colloq. Signal Process. Its Appl. (CSPA)*, Mar. 2015, pp. 192–197.
[215] N. Zerari, S. Abdelhamid, H. Bouzgou, and C. Raymond, "Bidirectional deep architecture for Arabic speech recognition," *Open Comput. Sci.*, vol. 9, no. 1, pp. 92–102, Apr. 2019, doi: 10.1515/comp-2019-0004.
[216] F. S. Al-Anzi and D. AbuZeina, "Synopsis on Arabic speech recognition," *Ain Shams Eng. J.*, vol. 13, no. 2, Mar. 2022, Art. no. 101534, doi: 10.1016/j.asej.2021.06.020.
[217] M. Aissiou and M. Guerti, "Genetic algorithm application to the standard Arabic phonemes classification," *Cybern. Syst.*, vol. 39, no. 3, pp. 199–212, Apr. 2008, doi: 10.1080/01969720801944281.
[218] A. Meftah, Y. Alotaibi, and S.-A. Selouani, "Designing, building, and analyzing an Arabic speech emotional corpus," in *Proc. Workshop Free/Open-Source Arabic Corpora Corpora Process. Tools Workshop Programme*, 2014, p. 22.
[219] M. Ali, M. Elshafei, M. Al-Ghamdi, H. Al-Muhtaseb, and A. Al-Najjar, "Generation of Arabic phonetic dictionaries for speech recognition," in *Proc. Int. Conf. Innov. Inf. Technol.*, Dec. 2008, pp. 59–63, doi: 10.1109/INNOVATIONS.2008.4781716.
[220] A. Bouallègue, S. Hassairi, R. Fjbali, and M. Zaied, "Learning deep wavelet networks for recognition system of Arabic words," in *Proc. Int. Joint Conf. SOCO CISIS ICEUTE*. Cham, Switzerland: Springer, 2016, pp. 498–507.
[221] M. Ben Nasr, M. Talbi, and A. Cherif, "Arabic speech recognition by bionic wavelet transform and MFCC using a multi layer perceptron," in *Proc. 6th Int. Conf. Sci. Electron., Technol. Inf. Telecommun. (SETIT)*, Mar. 2012, pp. 803–808.
[222] E. Techini, Z. Sakka, and M. Bouhlel, "Robust front-end based on MVA and HEQ post-processing for Arabic speech recognition using hidden Markov model toolkit (HTK)," in *Proc. IEEE/ACS 14th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Oct. 2017, pp. 815–820.
[223] Z. Sakka, E. Techini, and M. Bouhlel, "Using geometric spectral subtraction approach for feature extraction for DSR front-end Arabic system," *Int. J. Speech Technol.*, vol. 20, no. 3, pp. 645–650, Sep. 2017, doi: 10.1007/s10772-017-9433-1.
[224] H. M. Mohammdi and D. M. Elbourhamy, "An intelligent system to help deaf students learn Arabic sign language," *Interact. Learn. Environ.*, pp. 1–16, Apr. 2021, doi: 10.1080/10494820.2021.1920431.
[225] H. Frihia and H. Bahi, "Embedded learning segmentation approach for Arabic speech recognition," in *Proc. Int. Conf. Text, Speech, Dialogue*. Cham, Switzerland: Springer, 2016, pp. 383–390.
[226] S. A. Chowdhury, Y. Samih, M. Eldesouki, and A. Ali, "Effects of dialectal code-switching on speech modules: A study using Egyptian Arabic broadcast speech," in *Proc. Interspeech*, Oct. 2020, pp. 2382–2386.
[227] F. S. Al-Anzi and D. AbuZeina, "The effect of diacritization on Arabic speech recogntion," in *Proc. IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. (AEECT)*, Oct. 2017, pp. 1–5.
[228] F. S. Al-Anzi and D. AbuZeina, "The impact of phonological rules on Arabic speech recognition," *Int. J. Speech Technol.*, vol. 20, no. 3, pp. 715–723, Sep. 2017.
[229] F. Al-Anzi and D. AbuZeina, "Performance evaluation of Sphinx and htk speech recognizers for spoken Arabic language," *Int. J. Innov. Comput., Inf. Control*, vol. 15, no. 3, pp. 1009–1021, 2019.
[230] K. M. O. Nahar, W. G. Al-Khatib, M. Elshafei, H. Al-Muhtaseb, and M. M. Alghamdi, "Data-driven Arabic phoneme recognition using varying number of HMM states," in *Proc. 1st Int. Conf. Commun., Signal Process., Appl. (ICCSPA)*, Feb. 2013, pp. 1–6.

[231] K. Nahar, H. Al-Muhtaseb, W. Al-Khatib, M. Elshafei, and M. Alghamdi, "Arabic phonemes transcription using data driven approach," *Int. Arab J. Inf. Technol.*, vol. 12, no. 3, pp. 237–245, May 2015.
[232] K. M. O. Nahar, M. A. Shquier, W. G. Al-Khatib, H. Al-Muhtaseb, and M. Elshafei, "Arabic phonemes recognition using hybrid LVQ/HMM model for continuous speech recognition," *Int. J. Speech Technol.*, vol. 19, no. 3, pp. 495–508, Sep. 2016.
[233] A. M. Gody, "Design of wavelet based phone pass filter (PPF) using genetic algorithm," in *Proc. 22nd Nat. Radio Sci. Conf. (NRSC)*, 2005, pp. 327–336.
[234] M. Alghamdi, M. Elshafei, and H. Al-Muhtaseb, "Arabic broadcast news transcription system," *Int. J. Speech Technol.*, vol. 10, no. 4, pp. 183–195, Dec. 2007.
[235] A. Ghandoura, F. Hjabo, and O. Al Dakkak, "Building and benchmarking an Arabic speech commands dataset for small-footprint keyword spotting," *Eng. Appl. Artif. Intell.*, vol. 102, Jun. 2021, Art. no. 104267.
[236] A. Ali, N. Dehak, P. Cardinal, S. Khurana, S. H. Yella, J. Glass, P. Bell, and S. Renals, "Automatic dialect detection in Arabic broadcast speech," 2015, *arXiv:1509.06928*.
[237] M. A. Abushariah, R. N. Ainon, R. Zainuddin, B. A. Al-Qatab, and A. A. Alqudah, "Impact of a newly developed modern standard Arabic speech corpus on implementing and evaluating automatic continuous speech recognition systems," in *Proc. Int. Workshop Spoken Dialogue Syst. Technol.* Cham, Switzerland: Springer, 2010, pp. 1–12.
[238] M. A. M. Abushariah, "TAMEEM V1.0: speakers and text independent Arabic automatic continuous speech recognizer," *Int. J. Speech Technol.*, vol. 20, no. 2, pp. 261–280, Jun. 2017.
[239] M. A. M. Abushariah, R. N. Ainon, R. Zainuddin, M. Elshafei, and O. O. Khalifa, "Phonetically rich and balanced speech corpus for Arabic speaker-independent continuous automatic speech recognition systems," in *Proc. 10th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, May 2010, pp. 65–68.
[240] M. A. M. Abushariah, R. N. Ainon, R. Zainuddin, M. Elshafei, and O. O. Khalifa, "Phonetically rich and balanced text and speech corpora for Arabic language," *Lang. Resour. Eval.*, vol. 46, no. 4, pp. 601–634, Dec. 2012.
[241] M. Sawalha and M. A. Shariah, "The effects of speakers' gender, age, and region on overall performance of Arabic automatic speech recognition systems using the phonetically rich and balanced Modern Standard Arabic speech corpus," in *Proc. 2nd Workshop of Arabic Corpus Linguistics (WACL)*, 2013, pp. 1–3.
[242] E. B. Tazi, "A robust speaker identification system based on the combination of GFCC and MFCC methods," in *Proc. 5th Int. Conf. Multimedia Comput. Syst. (ICMCS)*, Sep. 2016, pp. 54–58.
[243] D. Gamal, M. Rashwan, and S. M. Abdou, "Emotion conversion for expressive Arabic text to speech," in *Proc. IEEE/ACS 11th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2014, pp. 341–348.
[244] L. Dockendorf, D. Almubayei, and M. Benton, "Testing a large corpus of natural standard Arabic for rhythm class," in *Proc. Interspeech*, 2008, p. 771.
[245] I. Rebai and Y. BenAyed, "Arabic text to speech synthesis based on neural networks for MFCC estimation," in *Proc. World Congr. Comput. Inf. Technol. (WCCIT)*, Jun. 2013, pp. 1–5.
[246] A. A. Almisreb, A. F. Abidin, and N. M. Tahir, "Arabic letters corpus based Malay speaker-independent," in *Proc. IEEE 3rd Int. Conf. Syst. Eng. Technol.*, Aug. 2013, pp. 232–236.
[247] A. A. Almisreb, A. F. Abidin, and N. M. Tahir, "Comparison of speech features for Arabic phonemes recognition system based Malay speakers," in *Proc. IEEE Conf. Syst., Process Control (ICSPC)*, Dec. 2014, pp. 79–83.
[248] A. A. Almisreb, A. F. Abidin, and N. M. Tahir, "Investigation of dynamic time warping and neural network for Arabic phonemes recognition based Malay speakers," in *Proc. 5th IEEE Int. Conf. Syst. Eng. Technol. (ICSET)*, Aug. 2015, pp. 60–65.
[249] A. A. Almisreb, N. M. Tahir, A. F. Abidin, and N. M. Din, "Noise reduction approach for Arabic phonemes articulated by Malay speakers," in *Proc. IEEE 14th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2018, pp. 267–271.
[250] A. A. Almisreb, A. F. Abidin, and N. M. Tahir, "Segmentation of Arabic letters signal using multiscale principal component analysis and zero-crossing rate based on Malay speakers," in *Proc. IEEE Int. Conf. Control Syst., Comput. Eng.*, Nov. 2013, pp. 483–486.
[251] A. Boumehdi and A. Yousfi, "Construction of a database for speech recognition of isolated Arabic words," in *Proc. 13th Int. Conf. Intell. Syst., Theories Appl.*, Rabat, Morocco, 2020, p. 25.
[252] Y. A. Alotaibi and Y. M. Seddiq, "Mapping Arabic acoustic parameters to their articulatory features using neural networks," in *Proc. IEEE Signal Process. Signal Process. Educ. Workshop (SP/SPE)*, Aug. 2015, pp. 409–414.
[253] Y. A. Alotaibi, "Experiments on ANN based ASR systems using limited Arabic vocabulary," in *Soft Computing Models in Industrial and Environmental Applications*. Cham, Switzerland: Springer, 2011, pp. 453–463.
[254] NIH. *Age*. Accessed: 2023. [Online]. Available: https://www.nih.gov/nih-style-guide/age
[255] S. A. LaRocca and R. Chouairi, "West point Arabic speech," Linguistic Data Consortium, Philadelphia, PA, USA, Tech. Rep. LDC2002S02, 2002. Accessed: Jun. 6, 2023.
[256] Y. A. Alotaibi and A. H. Meftah, "Comparative evaluation of two Arabic speech corpora," in *Proc. 6th Int. Conf. Natural Lang. Process. Knowl. Engineering(NLPKE)*, Aug. 2010, pp. 1–5.
[257] K. Kirchhoff, J. Bilmes, S. Das, N. Duta, M. Egan, G. Ji, F. He, J. Henderson, D. Liu, M. Noamany, P. Schone, R. Schwartz, and D. Vergyri, "Novel approaches to Arabic speech recognition: Report from the 2002 Johns-Hopkins summer workshop," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2003, pp. 344–347, doi: 10.1109/ICASSP.2003.1198788.
[258] Y. A. Alotaibi and A. H. Meftah, "West point, SAAVB, and BBN/AUB Arabic speech corpora: A comparative survey," in *Proc. 4th Int. Conf. Arabic Language Process. (CITALA)*, vol. 12, 2012.
[259] A. Ali, Y. Zhang, P. Cardinal, N. Dahak, S. Vogel, and J. Glass, "A complete Kaldi recipe for building Arabic speech recognition systems," in *Proc. IEEE spoken Lang. Technol. workshop (SLT)*, Jul. 2014, pp. 525–529.
[260] H. M. M. Eljagmani, "Arabic speech recognition systems," M.S. thesis, College Eng., Florida Inst. Technol., Zawia Univ., Melbourne, FL, USA, 2017.
[261] P. Cardinal, A. Ali, N. Dehak, Y. Zhang, T. A. Hanai, Y. Zhang, J. R. Glass, and S. Vogel, "Recent advances in ASR applied to an Arabic transcription system for Al-Jazeera," in *Proc. Interspeech*, Sep. 2014, pp. 1–12.
[262] A. Mohammed, M. S. B. Sunar, and M. S. H. Salam, "Recognition of holy Quran recitation rules using phoneme duration," in *Proc. Int. Conf. Reliable Inf. Commun. Technol.* Cham, Switzerland: Springer, 2017, pp. 343–352.
[263] A. M. A. Alqadasi, M. S. Sunar, S. Turaev, R. Abdulghafor, M. S. Hj Salam, A. A. S. Alashbi, A. A. Salem, and M. A. H. Ali, "Rule-based embedded HMMs phoneme classification to improve Qur'anic recitation recognition," *Electronics*, vol. 12, no. 1, p. 176, Dec. 2022.

AMMAR MOHAMMED ALI ALQADASI received the B.S. degree in information technology from Taiz University, Taiz, Yemen, in 2008, the M.S. degree in software engineering and intelligence from UTeM, Melaka, Malaysia, in 2012, and the Ph.D. degree in computer science from UTM, Johor, Malaysia, in 2021. His research interests include speech processing, image processing, and intelligent computing.

RAWAD ABDULGHAFOR (Member, IEEE) received the bachelor's and master's degrees in computer systems engineering and informatics from Saint Petersburg Electrotechnical University "LETI," Russia, in 2004 and 2007, respectively, and the Ph.D. degree in information technology from International Islamic University Malaysia, Kuala Lumpur, Malaysia, in 2017. He was a Research Assistant with the Faculty of Information and Communication Technology, International Islamic University Malaysia, from 2014 to 2017. He has been an Assistant Professor with the Faculty of Information and Communication Technology, International Islamic University Malaysia, since 2018. He was a Postdoctoral Fellow with the Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia Kuala Lumpur, Kuala Lumpur, from 2018 to 2019. His research interests include consensus models for multi-agent systems, computation theory, wireless communications, and machine learning.

MOHD SHAHRIZAL SUNAR (Member, IEEE) received the B.Sc. degree in computer science majoring in computer graphics from Universiti Teknologi Malaysia, the M.Sc. degree in computer graphics and virtual environment from The University of Hull, U.K., and the Ph.D. degree from the National University of Malaysia, in 2008. He is currently the Director of the Institute of Human Centered Engineering (iHumEn) and the Founding Director of the Media and Game Innovation Centre of Excellence (MaGICX), Universiti Teknologi Malaysia. His research interests include real-time and interactive computer graphics and virtual environments. He received scholarships from Sultan Iskandar Johor Foundation for both his postgraduate study. He received several awards and honors, including the Frew Fellowship (Australian Academy of Science), the I. I. Rabi Prize (APS), the European Frequency and Time Forum Award, the Carl Zeiss Research Award, the William F. Meggers Award, and the Adolph Lomb Medal (OSA).

MD. SAH BIN H. J. SALAM received the B.S. degree in computer science from the University of Pittsburgh, Pittsburgh, PA, USA, and the M.S. and Ph.D. degrees in speech processing and AI from Universiti Teknologi Malaysia. He is currently the Head of the Research Vicubelab Group and a Senior Lecturer with the Faculty of Computing, Universiti Teknologi Malaysia. His research interests include speech processing and synthesizing, image processing, intelligent computing, and computer graphics.