Arabic dialects speech corpora: A systematic review

Are you this author?

By Alqadasi, AMA (Alqadasi, Ammar Mohammed Ali) [1] , [2] ; Zeki, AM (Zeki, Akram M.) [2] ; Sunar, MS (Sunar, Mohd Shahrizal) [3] , [4] ; Hashim, SZM (Hashim, Siti Zaiton Mohd) [3] ; Salam, MSH (Salam, Md Sah Hj) [3] ; Abdulghafor, R (Abdulghafor, Rawad) [5]

View Web of Science ResearcherID and ORCID (provided by Clarivate)

Source SPEECH COMMUNICATION

View Journal Impact

Volume: 175

DOI: 10.1016/j.specom.2025.103322

Article Number 103322

Published NOV 2025

Indexed 2025-11-26

Document Type Review

Abstract

Speech processing applications are crucial in various domains, necessitating reliable speech recognition systems built upon suitable speech databases. However, the availability of comprehensive resources for the Arabic language remains limited compared to other languages like English. A systematic review was conducted to identify, analyze, and classify existing Arabic dialect speech databases. Initially, online digital databases and search engines were identified to collect a diverse range of manuscripts for thorough examination. The review encompassed 30 publicly accessible databases and an additional 39 self-databases, which were thoroughly studied, classified based on their characteristics, and subjected to a detailed analysis of research trends. This paper offers a comprehensive discussion on the diverse speech databases developed for various speech processing applications, highlighting the purposes and unique characteristics of Arabic speech databases. By providing valuable insights into their availability, characteristics, challenges, and research directions, this review aims to facilitate researchers' access to suitable resources for their specific applications, encourage the creation of new datasets in underrepresented areas, and promote open and easily accessible databases. Furthermore, the findings contribute to bridging the gap in available Arabic speech databases and serve as a valuable resource for researchers in the field.

Keywords

Author Keywords: Speech corpus; Speech database; Arabic dialects corpora; Speech recognition; Arabic dialects; Modern Standard Arabic

Keywords Plus: EMOTION RECOGNITION; PHONETICALLY RICH; TEXT; CORPUS; CLASSIFICATION; IDENTIFICATION; DISTANCE; ASR

Author Information

Corresponding Address: Alqadasi, Ammar Mohammed Ali (corresponding author)

Karabuk Univ, Fac Comp & Informat Sci, Dept Comp Engn, TR-78050 Karabuk, Turkiye

Corresponding Address: Alqadasi, Ammar Mohammed Ali (corresponding author)

Int Islamic Univ Malaysia, Fac Informat & Commun Tech, Kuala Lumpur 53100, Malaysia

E-mail Addresses :

ammarmohammed@karabuk.edu.tr
Addresses :
[1] Karabuk Univ, Fac Comp & Informat Sci, Dept Comp Engn, TR-78050 Karabuk, Turkiye
[2] Int Islamic Univ Malaysia, Fac Informat & Commun Technol, Kuala Lumpur 53100, Malaysia
[3] Univ Teknol Malaysia, Fac Comp, Johor Baharu 81310, Malaysia
[4] Univ Teknol Malaysia, Inst Human Ctr Engn, Media & Game Innovat Ctr Excellence, Johor Baharu 81310, Malaysia
[5] Arab Open Univ Oman, Fac Comp Studies FCS, POB 1596, Muscat, Oman
E-mail Addresses :
ammarmohammed@karabuk.edu.tr

Data availability statement

No data was used for the research described in the article.

Categories/ Classification

Research Areas: Acoustics; Computer Science

Web of Science Categories

Acoustics; Computer Science, Interdisciplinary Applications

Funding

View funding text

Funding agency
Malaysia International Islamic University
Universiti Teknologi Malaysia

+ See more data fields

Journal information

SPEECH COMMUNICATION

← View Journal Impact

ISSN	0167-6393
eISSN	1872-7182
Current Publisher	ELSEVIER, RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS
Table of Contents	Current Contents Connect
Research Areas	Acoustics; Computer Science
Web of Science Categories	Acoustics; Computer Science, Interdisciplinary Applications

3
Journal Impact Factor ™ (2024)

0.71
Journal Citation Indicator ™ (2024)

Citation Network

In Web of Science Core Collection

0 Citations

Create citation alert

208
Cited References

→ View Related Records

How does this document's citation performance compare to peers?

Use in Web of Science

1
Last 180 Days

1
Since 2013

Learn more →

This record is from:

Web of Science Core Collection

- Science Citation Index Expanded (SCI-EXPANDED)

Suggest a correction 32

If you would like to improve the

quality of the data in this record, please Suggest a correction

← Open comparison metrics panel

Data is from InCites Benchmarking & Analytics

208 Cited References

→ View as set of results

Showing 30 of 208 *(from Web of Science Core Collection)*