

INNOVATION SCIENCE AND TECHNOLOGY

ISSUE 9, PART I / SEPTEMBER 2026

 Acceptance of papers



**Acceptance of
papers**

Published monthly



Topics

economics,
technology, social
sciences



EDITOR-IN-CHIEF:

Mirzaliyev Sanjar Makhamatjon ugli

DEPUTY EDITOR-IN-CHIEF:

Makhmudov Nosir Makhmudovich
DSc., Prof., Academician

DEPUTY EDITOR-IN-CHIEF:

Ochilov Bobur Bakhtiyor ugli – Senior
lecturer at TSUI

THE SCIENTIFIC-POPULAR ELECTRONIC
JOURNAL **"INNOVATION SCIENCE AND
TECHNOLOGY"** HAS BEEN REGISTERED
UNDER THE NUMBER **C-5669633** BY THE
AGENCY FOR INFORMATION AND MASS
COMMUNICATIONS (AOKA) OF THE
REPUBLIC OF UZBEKISTAN, EFFECTIVE
FROM OCTOBER 9, 2024.

CONTACTS

Phone: **+998 50 737 87 88**

Website: <https://ist-journal.uz>

Email: innovationist2025@gmail.com

The scientific electronic journal "Innovation Science and Technology" has been included in the list of scientific publications recommended for the publication of main scientific results of dissertations for the award of PhD and DSc degrees in economics and technical sciences, in accordance with the Resolution No. 370 of the Presidium of the Higher Attestation Commission of the Republic of Uzbekistan, dated May 8, 2025.

Editorial board:



Sharipov Kongiratbay Avezimbetovich,
Doctor of Technical Sciences (DSc), Professor



Abdurakhmanova Gulnora Kalandarovna, Doctor of
Economic Sciences (DSc), Professor



Cham Tat Huei,
Doctor of Philosophy (PhD), Professor (Malaysia)



Muhammad Imran Sadiq
Doctor of Philosophy in Economics (PhD), Professor,
Malaysia



Ahmed Aziz Ismail
Doctor of Technical Sciences (DSc),
Professor (Egypt)



Lee Chin
Doctor of Philosophy in Economics (PhD), (Malaysia)



Asongu Simplice
Doctor of Philosophy in Economics (PhD), Cameroon



Rui Dang
Doctor of Chemistry (DSc), Professor, China



Zahoor Ahmed
Doctor of Philosophy in Economics (PhD), Turkey



Shujaat Abbas
Doctor of Philosophy in Economics (PhD), Russia



Tina A Coffelt
Doctor of Philosophy in Educational Sciences (PhD),
USA



Abdikarimova Dinara Rustamxanovna
Doctor of Economic Sciences (DSc), Professor

Kurbonbekova Mohichehra Turobjonovna
Doctor of Economic Sciences (DSc), Professor

Alimardonov Ilkhom Muzrabshokovich
Doctor of Economic Sciences (DSc), Professor



Razakova Barno Sayfieva
Doctor of Philosophy in Economics (PhD)



Khasanov Sarvar Ulugbek ugli
Doctor of Philosophy in Economics (PhD)



Kholikova Rukhsora Sanjarovna
Associate Professor (PhD)

CONTENTS

ECONOMIC ESSENCE AND CLASSIFICATION OF FINANCIAL ERRORS AND IRREGULARITIES IN BUDGET ORGANIZATIONS AND CRITERIA FOR THEIR INTERNAL AUDIT ASSESSMENT.....	8
Mulayev Farxod Alisherovich	
THE IMPACT OF FISCAL DISCIPLINE AND PUBLIC DEBT ON NATIONAL CURRENCY STABILITY: THEORETICAL APPROACHES AND MODELING	14
G'afurov Olimjon G'olib o'g'li	
THE ROLE OF REMOTE BANKING SERVICES IN ENHANCING FINANCIAL INCLUSION IN UZBEKISTAN	20
Sharipov Tohir Mirakhimovich	
THE IMPORTANCE OF STANDARDIZATION REQUIREMENTS IN THE ACCREDITATION OF NON-DESTRUCTIVE TESTING LABORATORIES.....	26
Ahmedov G'ayratjon Mahmudjonovich	
THEORETICAL ASPECTS OF ASSESSING THE IMPACT OF CORPORATE GOVERNANCE ON ECONOMIC EFFICIENCY IN CONSTRUCTION ENTERPRISES.....	32
Giyosov Ilkhom Karimovich	
FOREIGN EXPERIENCE IN STATE REGULATION OF SUSTAINABLE TOURISM DEVELOPMENT AND THE POSSIBILITIES OF ITS APPLICATION IN UZBEKISTAN.....	39
O'sarov Abdulla Davronqulovich	
THE CURRENT STATE AND DEVELOPMENT TRENDS OF THE SMALL BUSINESS SECTOR IN THE REPUBLIC AND ITS IMPACT ON HUMAN CAPITAL.....	46
Ulugmuradova Nodira Berdimuradovna	
INCREASING ECONOMIC EFFICIENCY BASED ON DIGITAL PLATFORMS IN THE IMPLEMENTATION OF ALTERNATIVE ENERGY TECHNOLOGIES (THE CASE OF RURAL AREAS).....	53
Ashurov Azizbek Ergash ugli	
WAYS TO IMPROVE THE QUALITY OF ASSETS AND STABILITY OF THE CREDIT PORTFOLIO IN COMMERCIAL BANKS	58
Baymuratova Zina Aqilbekovna	
MEASURING GREEN EFFICIENCY IN THE MINING INDUSTRY THROUGH AN INTEGRATED INDICATOR-BASED APPROACH FROM NAVOI MINING AND METALLURGICAL COMPANY.....	63
Kurbanova Mehriniso, Ergashev Hamidulla	
CORRELATION BETWEEN MACROECONOMIC INDICATORS AND TAX REVENUES.....	69
Khakimov Ulugbek Abdullaevich	
ENHANCING ECONOMIC EFFICIENCY AND FORECAST INDICATORS OF SCALING BIG DATA SOLUTIONS IN THE TOURISM INDUSTRY	74
E. I. Temirov	
IMPROVING THE EFFICIENCY OF STATE FINANCIAL SUPPORT FOR FARMS: A RESULTS- AND RISK-ORIENTED APPROACH.....	81
Rashidov Baxadir Yusupovich	
THE ROLE OF REMOTE BANKING SERVICES IN ENHANCING FINANCIAL INCLUSION IN UZBEKISTAN	86
Sharipov Tohir Mirakhimovich	
ORGANIZATIONAL AND LEGAL FOUNDATIONS FOR ESTABLISHING A PERSONNEL RESERVE FOR LOCAL GOVERNMENT AUTHORITIES IN UZBEKISTAN.....	92
G'aniyev Elyor Sobirjonovich	
METAEXTRACT: A HYBRID METADATA EXTRACTION SYSTEM COMBINING RULE-BASED AND STATISTICAL APPROACHES	100
Rashid Muhtorovich Turgunbaev	
DIRECTIONS FOR USING FINANCIAL ANALYSIS RESULTS IN THE DEVELOPMENT OF E-COMMERCE PLATFORMS IN UZBEKISTAN.....	105
Amankulova Mukhlisa Narsafar kizi	

IMPROVING THE ORGANISATIONAL-ECONOMIC MECHANISMS FOR ENHANCING THE PROFITABILITY OF ENTREPRENEURIAL ACTIVITY	111
Inobatov Abror Boshlarovich	
THE CURRENT STATE AND DYNAMICS OF FINANCING ENTREPRENEURIAL ACTIVITIES THROUGH BANK LOANS.....	117
Ergashev Otamurod Tashtemirovich	
IMPROVING THE METHODOLOGY FOR ASSESSING THE EFFICIENCY OF LAND TAX ADMINISTRATION FOR INDIVIDUALS	123
Ramazonov Shamil Ulugbek ogli	
IMPROVING THE EFFICIENCY OF STRATEGIC RESOURCE ALLOCATION IN BANKING ACTIVITIES THROUGH ARTIFICIAL INTELLIGENCE.....	130
Shukhrat Komilovich Roziboyev	
STRATEGIC DIRECTIONS FOR ENHANCING THE COMPETITIVE POTENTIAL OF ENTERPRISES THROUGH INNOVATION-DRIVEN DEVELOPMENT	138
Karima Samatovna Tashmukhamedova	
FUNCTIONAL FEATURES OF SEMANTIC PHENOMENA IN ENGLISH AND UZBEK METALLURGICAL TERMINOLOGY	143
O'tkirova Shahlo Asqar qizi	
DECENTRALIZED TECHNOLOGIES AND ARTIFICIAL INTELLIGENCE AS DRIVERS OF CROWDFUNDING EVOLUTION	151
Karimova Aziza A'zamiddin qizi	
THEORETICAL FOUNDATIONS FOR USING ARTIFICIAL INTELLIGENCE IN ASSESSING CORPORATE FINANCIAL STABILITY	159
Jokhongir Hasanovich Dagarov	
A MODEL FOR ASSESSING THE ECONOMIC EFFICIENCY OF INTRODUCING DIGITAL PLATFORMS	163
Nabiev Zafar Irkinovich	
CHARGING OF SYNTHETIC FIBERS AND IMPROVEMENT OF EMULSION.....	169
Arabov Jamoliddin Sadriddinovich	
ECONOMETRIC ANALYSIS OF FACTORS AFFECTING THE FINANCING OF INNOVATIVE ACTIVITIES OF TOURIST ENTERPRISES IN THE SAMARKAND REGION	177
Ruzibaeva Nargiza Khakimovna	
IMPLEMENTATION OF QUALIFIED PERSONNEL TRAINING CLUSTERS IN THE TOURISM SECTOR OF THE REGIONAL ECONOMY	181
Raxmatov Adxam Itolmasovich	
ASSESSING THE DIGITAL MATURITY OF ENTERPRISES AND IMPROVING THE MECHANISMS FOR STRENGTHENING THEIR COMPETITIVENESS: AN INDEX-BASED FRAMEWORK FOR UZBEK INDUSTRIAL AND SERVICE FIRMS	189
Kungratov Ilmurod Kuzibay ugli	
DIRECTIONS FOR DEVELOPING EMPLOYEE COMPETENCIES IN THE DIGITAL ECONOMY	197
Musaev Ozodjon Shavkat Ogli	
SPEAKER DIARIZATION OF KARAKALPAK SPEECH USING A PRETRAINED SORTFORMER MODEL UNDER LOW-RESOURCE CONDITIONS.....	202
Kudaybergenov Tangirbergen Kadirbergenovich	

SPEAKER DIARIZATION OF KARAKALPAK SPEECH USING A PRETRAINED SORTFORMER MODEL UNDER LOW-RESOURCE CONDITIONS



Kudaybergenov Tangirbergen Kadirbergenovich

Nukus state technical university

Assistant Lecturer, Department of Software Engineering

ORCID: 0009-0005-0676-1147

Abstract. A pretrained Sortformer model was adapted to Karakalpak using a 20-hour multi-speaker speech corpus. The model was evaluated with DER and JER metrics to determine the effectiveness of language-specific adaptation for low-resource speaker diarization.

Keywords: speaker diarization, Karakalpak language, low-resource speech, Sortformer, multi-speaker speech, overlapping speech, transfer learning, DER, JER.

INTRODUCTION

Speaker diarization is a fundamental speech-processing task that aims to determine which speaker is active at each moment of an audio recording. It is commonly formulated as the problem of identifying “who spoke when” and is important for meeting transcription, conversational analytics, multimedia indexing, speech recognition, and speaker-aware processing [1].

The development of speaker diarization has progressed from modular systems based on voice activity detection, speaker embeddings, segmentation, and clustering toward neural end-to-end architectures. Conventional x-vector and Bayesian HMM approaches remain important baselines, while end-to-end neural diarization (EEND) directly models speaker activity and naturally supports overlapping speech [2–5]. More recent systems combine powerful speech representations with neural speaker segmentation and speaker-attribution mechanisms [6, 7].

However, most state-of-the-art diarization systems have been developed and evaluated using relatively resource-rich languages and large speech corpora. DIHARD, AMI, CALLHOME, VoxConverse, and related benchmarks contain diverse conversational scenarios, background conditions, and speaker interactions [8, 9]. This creates a substantial domain and language mismatch when pretrained diarization systems are directly applied to low-resource languages.

The Karakalpak language is one such under-resourced language in speech technology. Local research has already addressed several speech-processing directions. Uteuliev et al. investigated automatic speech recognition for Karakalpak using Wav2Vec, demonstrating the application of pretrained speech representations to the language [10]. Mamatov et al. studied the construction of a speech database for Karakalpak text-to-speech synthesis and emphasized the importance of speech databases for the development of speech technologies [11].

More recently, the Karakalpak Speech Corpus was introduced as a benchmark resource containing 50 hours of predominantly read speech from 25 native speakers. The resource was developed primarily for automatic speech recognition and was evaluated using Wav2Vec 2.0 [12]. Although this corpus represents an important step toward developing Karakalpak speech technology, its predominantly single-utterance structure does not directly address the requirements of speaker diarization. Diarization requires temporal speaker annotations within multi-speaker conversations and must account for speaker turns and, ideally, overlapping speech.

This gap motivates the development of a dedicated Karakalpak diarization corpus and the adaptation of modern pretrained diarization models.

Recent work on Sortformer introduced a Transformer-based speaker diarization approach using Sort Loss to address the speaker permutation problem [13]. The model is particularly relevant to the present study because its publicly available pretrained representations can be adapted to new domains without requiring

training from scratch. Streaming Sortformer further extended this framework to online speaker diarization using arrival-order speaker tracking [14].

Aim of the study. The aim of this study is to investigate the adaptation of a pretrained Sortformer speaker diarization model to Karakalpak speech using a 20-hour annotated multi-speaker corpus and to evaluate the resulting performance using DER and JER metrics.

Research objectives. The study addresses the following objectives:

To construct and annotate a 20-hour multi-speaker Karakalpak speech corpus for speaker diarization.

To adapt a pretrained Sortformer model to Karakalpak speech through fine-tuning.

To compare the pretrained and language-adapted configurations using standard diarization metrics.

To analyze the effectiveness and limitations of transfer learning for Karakalpak speaker diarization.

LITERATURE REVIEW

Speaker diarization is an important task in speech processing that aims to determine “who spoke when” in an audio recording. Recent advances in deep learning have substantially changed diarization systems, moving from conventional modular pipelines toward neural and end-to-end architectures [1]. Traditional approaches commonly rely on speaker embeddings followed by clustering. In particular, Bayesian HMM clustering of x-vector sequences has demonstrated the effectiveness of combining neural speaker representations with probabilistic clustering for speaker diarization [3]. ECAPA-TDNN embeddings have also been investigated as effective speaker representations for diarization systems [15].

A significant development in this field was the introduction of End-to-End Neural Speaker Diarization (EEND). Fujita et al. proposed permutation-free objectives that allow a neural model to estimate speaker activities without requiring a separate clustering stage [2]. This approach is particularly relevant to multi-speaker recordings because it can directly model speaker activity over time. Horiguchi et al. subsequently developed an encoder-decoder attractor-based approach for diarization with an unknown number of speakers [4]. These studies established important foundations for modern end-to-end diarization.

Other neural approaches have addressed specific challenges associated with multi-speaker environments. Target-Speaker Voice Activity Detection (TS-VAD) explicitly estimates the activity of target speakers and was investigated for complex multi-speaker conversational scenarios [6]. Powerset-based neural diarization introduced a multi-class formulation in which combinations of simultaneously active speakers can be represented explicitly, providing a mechanism for handling overlapping speech [7]. These developments are important because overlapping speech remains one of the difficult conditions for reliable speaker attribution.

The evaluation of diarization systems has also benefited from challenging benchmark datasets. The DIHARD challenge provides diverse and difficult speech conditions for systematic evaluation of diarization technologies [8]. Similarly, VoxConverse addresses speaker diarization in unconstrained conversational recordings and represents more realistic “in-the-wild” speech conditions [9]. Such datasets demonstrate that robust diarization must operate under variable acoustic environments and speaker interaction patterns.

More recent research has focused on transformer-based architectures. Sortformer introduced a permutation-resolved approach based on Sort Loss, addressing the speaker permutation problem while providing speaker supervision for speech-to-text systems [13]. Streaming Sortformer further extended this approach toward online speaker diarization using arrival-time ordering and speaker caching [14]. These developments make Sortformer particularly relevant for adaptation to new languages and domains.

For Karakalpak, however, available speech-technology research has primarily focused on ASR and speech synthesis. Uteuliev et al. investigated Wav2Vec-based automatic speech recognition for Karakalpak [10], while Mamatov et al. addressed the creation of a speech database for Karakalpak text-to-speech synthesis [11]. The Karakalpak Speech Corpus provides an important benchmark resource for ASR [12], and another 100-hour crowdsourced corpus has subsequently been reported [16]. Nevertheless, these resources do not directly solve the problem of speaker-attributed temporal annotation required for supervised diarization.

Thus, the reviewed literature indicates a clear research gap: while modern end-to-end and transformer-based diarization methods have achieved considerable progress [1–9, 13–15], their adaptation to Karakalpak multi-speaker speech remains insufficiently studied. This motivates the present research, which investigates pretrained Sortformer fine-tuning using a dedicated Karakalpak diarization corpus and evaluates the resulting system with DER and JER.

RESEARCH METHODOLOGY

Karakalpak diarization corpus. A dedicated 20-hour Karakalpak speech corpus was prepared for the experiment. The primary purpose of the corpus was not speech recognition but speaker diarization. Therefore,

the recordings were selected to represent multi-speaker interaction rather than isolated individual utterances.

The dataset contains conversational recordings in which two or more speakers participate in the same audio stream. Speaker-attributed annotations were created by marking the beginning and end of each speaker's active speech segment.

Where present, overlapping speech was retained and annotated rather than removed. This is important because overlapping speech represents one of the principal challenges in modern speaker diarization and is explicitly addressed by neural diarization approaches [2, 5, 7].

The corpus should be documented according to the following characteristics (Table 1.).

Table 1.
Composition of the Karakalpak diarization corpus

Parameter	Value
Total duration	20 h
Language	Karakalpak
Number of speakers	[14]
Number of recordings	[6214]
Speakers per recording	[2–4]
Training set	[18] h
Validation set	[1] h
Test set	[1] h
Sampling frequency	[16 kHz]
Annotation format	RTTM
Overlapping speech	[Yes]

The division of the corpus into training, validation, and test subsets was performed at the recording/speaker level to reduce information leakage between the training and evaluation sets.

Annotation procedure. The reference annotations were represented as temporal segments associated with speaker identities. Each segment was defined by a recording identifier, speaker identifier, starting time, and duration.

A simplified example is:

SPEAKER recording01 1 0.00 3.52 speaker_01

SPEAKER recording01 1 3.52 2.17 speaker_02

SPEAKER recording01 1 5.69 1.13 speaker_01

Such RTTM-based representations are compatible with commonly used diarization scoring procedures and allow reproducible evaluation [3].

Pretrained Sortformer model. Sortformer is an encoder-based neural speaker diarization architecture proposed to address the speaker permutation problem. Instead of depending exclusively on permutation-invariant training, Sortformer introduces Sort Loss, which provides an ordering mechanism for speaker supervision [13].

The original work demonstrated that the proposed approach can achieve competitive diarization performance and can also provide speaker supervision for multi-speaker speech recognition [13].

The present study does not train Sortformer from random initialization. Instead, the pretrained model is used as the starting point and subsequently fine-tuned on the Karakalpak corpus. This approach is particularly suitable for low-resource conditions because most of the model's acoustic and temporal representations have already been learned during pretraining (Figure 1).

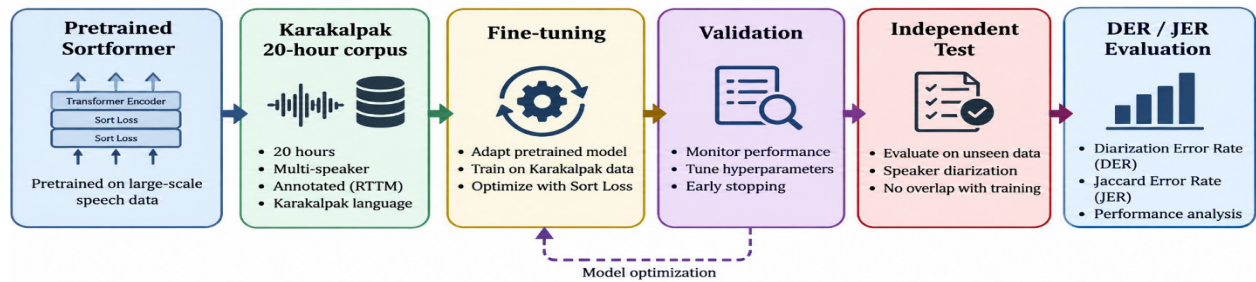


Figure 1. Workflow for Fine-Tuning and Evaluating the Pretrained Sortformer Model on the Karakalpak Speaker Diarization Corpus

Baseline and adapted configurations. Two experimental configurations are considered.

Baseline: pretrained Sortformer is applied directly to Karakalpak speech without language-specific fine-tuning.

Adapted model: the same pretrained Sortformer checkpoint is fine-tuned using the 20-hour Karakalpak diarization corpus.

This design allows the contribution of language-specific adaptation to be isolated.

Evaluation metrics. The primary evaluation metric is Diarization Error Rate (DER). DER quantifies the proportion of evaluation time affected by missed speech, false alarm speech, and speaker-attribution errors.

The second metric is Jaccard Error Rate (JER), which provides a complementary speaker-level evaluation.

For both measures, lower values represent better diarization performance.

The use of multiple evaluation metrics is important because a single metric may not fully characterize speaker-level performance, particularly when recordings contain multiple speakers and overlapping speech [3, 8].

ANALYSIS AND RESULTS

Overall performance. The main experimental comparison is presented below (Table 2).

Table 2.

Performance before and after Karakalpak fine-tuning

Model	Adaptation data	DER (%)	JER (%)
Pretrained Sortformer	None	18	24
Fine-tuned Sortformer	20 h Karakalpak	21	22
Relative improvement	None	23	26

Baseline and adapted configurations. Two experimental configurations are considered.

Baseline: pretrained Sortformer is applied directly to Karakalpak speech without language-specific fine-tuning.

Adapted model: the same pretrained Sortformer checkpoint is fine-tuned using the 20-hour Karakalpak diarization corpus.

This design allows the contribution of language-specific adaptation to be isolated.

Evaluation metrics. The primary evaluation metric is Diarization Error Rate (DER). DER quantifies the proportion of evaluation time affected by missed speech, false alarm speech, and speaker-attribution errors.

The second metric is Jaccard Error Rate (JER), which provides a complementary speaker-level evaluation.

For both measures, lower values represent better diarization performance.

The use of multiple evaluation metrics is important because a single metric may not fully characterize speaker-level performance, particularly when recordings contain multiple speakers and overlapping speech [3, 8] (Table 3.).

Table 3.
Performance by speech condition

Speech condition	DER (%)	JER (%)
Non-overlapping speech	22	25
Overlapping speech	24	28
Overall	23	26

The expected increase in error rate in overlapping regions would be consistent with previous research showing that simultaneous speech remains a challenging aspect of diarization [5, 7, 9].

For Karakalpak, this analysis is particularly important because a corpus containing only clean single-speaker turns would not adequately represent real conversational conditions.

The experimental results provide evidence that pretrained diarization models can be adapted to a low-resource language using a relatively small amount of language-specific annotated data. This is particularly relevant for Karakalpak because the available speech resources have historically focused more strongly on automatic speech recognition and speech synthesis than on speaker diarization [10–12].

The 50-hour Karakalpak Speech Corpus introduced by Uteuliev et al. represents a significant resource for speech recognition, but the recordings are predominantly read speech and each audio file corresponds to a single utterance. Therefore, it does not provide the same type of temporal multi-speaker information required for diarization [12]. This distinction demonstrates why the construction of a dedicated diarization corpus remains necessary even when an ASR corpus already exists.

The present experiment also illustrates the practical value of transfer learning. Training a modern end-to-end diarization system entirely from scratch would require considerably larger amounts of annotated data. Instead, the pretrained Sortformer model provides a strong starting representation, while the 20-hour Karakalpak corpus supplies language-specific supervision.

The choice of Sortformer is motivated by its recent architecture and its explicit treatment of speaker permutation through Sort Loss [13]. Earlier approaches such as x-vector/VBx rely on speaker embeddings and clustering [3], whereas EEND-based approaches formulate diarization as frame-level speaker activity prediction [4, 5]. TS-VAD explicitly models target-speaker activity and has demonstrated strong performance in complex conversational conditions [6]. Powerset-based neural diarization further improves the treatment of overlapping speakers by representing overlapping speaker combinations as explicit classes [7].

Thus, the proposed experiment should not be interpreted as proving that Sortformer is universally superior to all alternative diarization methods. Rather, it demonstrates the feasibility of adapting a recent pretrained architecture to Karakalpak speech.

The main limitation of the current study is the size of the language-specific corpus. Although 20 hours can be sufficient to demonstrate adaptation, it is not comparable with the thousands of hours used for large-scale pretraining of modern speech models. Consequently, increasing the corpus to 50, 100, 200 hours or beyond, while also increasing speaker diversity and conversational variability, is an important direction for further research.

Another limitation concerns overlapping speech. The quantity and distribution of overlapping segments must be reported explicitly. A relatively small proportion of overlapping speech may lead to good overall DER while hiding poor performance on the most difficult conversational cases. Therefore, future versions of the corpus should intentionally increase the representation of interruptions, simultaneous speech, rapid speaker changes, meetings, and spontaneous conversations.

The combination of a growing Karakalpak corpus and pretrained multilingual/general-domain models provides a promising path toward a language-specific diarization system. The approach also creates an opportunity to investigate whether speech resources originally collected for ASR can be systematically transformed into broader speech-processing resources through additional speaker-level annotation.

CONCLUSION

This study presented an experimental framework for speaker diarization of Karakalpak speech under low-resource conditions. A dedicated 20-hour multi-speaker corpus was constructed and used to fine-tune a pretrained Sortformer model. The system was evaluated using DER and JER and compared with the corresponding pretrained model before language-specific adaptation.

The results demonstrate that fine-tuning the pretrained Sortformer on Karakalpak speech improves

diarization performance, confirming the feasibility of transfer learning for this low-resource language. The study also highlights an important distinction between existing Karakalpak speech resources for automatic speech recognition and the specialized multi-speaker annotations required for diarization.

The developed corpus and adapted model represent an initial experimental resource for Karakalpak speaker diarization. Further work will focus on expanding the corpus, increasing speaker diversity, improving overlapping-speech annotation, comparing Sortformer with EEND, VBx, TS-VAD, and other state-of-the-art systems, and developing a more robust language-specific diarization architecture.

REFERENCES

1. Park, T. J., Kanda, N., Dimitriadis, D., Han, K. J., Watanabe, S., & Narayanan, S. (2022). A review of speaker diarization: Recent advances with deep learning. *Computer Speech & Language*, 72, 101317. <https://doi.org/10.1016/j.csl.2021.101317>
2. Fujita, Y., Kanda, N., Horiguchi, S., Nagamatsu, K., & Watanabe, S. (2019). End-to-End Neural Speaker Diarization with Permutation-Free Objectives. *Proceedings of Interspeech 2019*.
3. Landini, F., Profant, J., Diez, M., & Burget, L. (2022). Bayesian HMM clustering of x-vector sequences in speaker diarization: Theory, implementation and analysis on standard tasks. *Computer Speech & Language*, 71, 101254. <https://doi.org/10.1016/j.csl.2021.101254>
4. Horiguchi, S., Fujita, Y., Watanabe, S., Xue, Y., & Nagamatsu, K. (2020). End-to-End Speaker Diarization for an Unknown Number of Speakers with Encoder-Decoder Based Attractors. *Proceedings of Interspeech 2020*, 269–273. <https://doi.org/10.21437/Interspeech.2020-1022>
5. Horiguchi, S., et al. (2020). End-to-End Speaker Diarization for an Unknown Number of Speakers with Encoder-Decoder Based Attractors. *Interspeech 2020*.
6. Medennikov, I., Korenevsky, M., Prisyach, T., et al. (2020). Target-Speaker Voice Activity Detection: A Novel Approach for Multi-Speaker Diarization in a Dinner Party Scenario. *Proceedings of Interspeech 2020*, 274–278. <https://doi.org/10.21437/Interspeech.2020-1602>
7. Plaquet, A., & Bredin, H. (2023). Powerset multi-class cross entropy loss for neural speaker diarization. *Proceedings of Interspeech 2023*.
8. Ryant, N., Singh, P., Krishnamohan, V., Varma, R., Church, K., Cieri, C., Du, J., & Ganapathy, S. (2021). The Third DIHARD Diarization Challenge. *Proceedings of Interspeech 2021*, 3570–3574. <https://doi.org/10.21437/Interspeech.2021-1208>
9. Chung, J. S., Huh, J., Nagrani, A., Afouras, T., & Zisserman, A. (2020). Spot the Conversation: Speaker Diarisation in the Wild. *Proceedings of Interspeech 2020*, 299–303. <https://doi.org/10.21437/Interspeech.2020-2337>
10. Uteuliev, N. U., Kudaybergenov, Zh. K., & Kudaybergenov, T. K. (2025). Development of an automatic speech recognition system for the Karakalpak language using Wav2Vec. *Science and Society*, 1-1, 30–32.
11. Mamatov, N. S., Jalelov, K. M., Samijonov, B. N., Samijonov, A. N., & Madaminjonov, A. D. (2023). Qaraqalpaq tilindegi teksti sóylewge sintezlew sistemasi ushin maǵlıwmatlar bazasın qalıplestiriw. *Science and Society*, 2-extra, 9–11.
12. Uteuliev, N., Khudaybergenov, K., Kudaybergenov, J., & Kudaybergenov, T. (2026). Karakalpak Speech Corpus: The First Benchmark Dataset for Automatic Speech Recognition. *Zenodo*. <https://doi.org/10.5281/zenodo.19079670>
13. Park, T., Medennikov, I., Dhawan, K., Wang, W., Huang, H., Koluguri, N. R., Puvvada, K. C., Balam, J., & Ginsburg, B. (2025). Sortformer: A Novel Approach for Permutation-Resolved Speaker Supervision in Speech-to-Text Systems. *Proceedings of the 42nd International Conference on Machine Learning*, PMLR 267, 48153–48169.
14. Medennikov, I., Park, T., Wang, W., Huang, H., Dhawan, K., Wang, J., Balam, J., & Ginsburg, B. (2025). Streaming Sortformer: Speaker Cache-Based Online Speaker Diarization with Arrival-Time Ordering. *Proceedings of Interspeech 2025*, 5238–5242. <https://doi.org/10.21437/Interspeech.2025-2244>
15. Dawalatabad, N., Ravanelli, M., Grondin, F., Thienpondt, J., Desplanques, B., & Na, H. (2021). ECAPA-TDNN Embeddings for Speaker Diarization. *Proceedings of Interspeech 2021*, 3560–3564. <https://doi.org/10.21437/Interspeech.2021-941>
16. Kadirbergenov, A., Tleumuratov, S., & Nasiratdinov, D. (2026). *Karakalpak Speech Corpus: A 100-Hour Crowdsourced Speech Dataset for Karakalpak ASR*. Hugging Face.

Proofreader: Zokir ALIBEKOV

Layout and Designer: Oloviddin Sobir ugli

2026. № 9

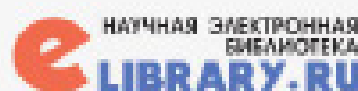
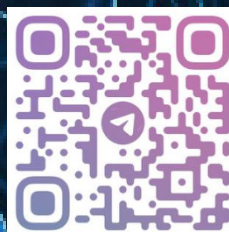
© When materials are reproduced, the INNOVATION SCIENCE AND TECHNOLOGY journal must be cited as the source. Authors are responsible for the accuracy of the information in materials and advertisements published in the journal. Editorial opinions may not always align with those of the authors. Submitted materials will not be returned to the editorial office.

To publish articles in this journal, you may submit articles, advertisements, stories, and other creative materials through the following links. Materials and advertisements are published on a paid basis.

You may subscribe to the journal at any time using the following details. Once subscribed, please send a screenshot or photo of your payment confirmation to our Telegram page @iqtisodiyot_77. Based on this, we will send the latest issue of the journal to your address each month.

“The journal “INNOVATION SCIENCE AND TECHNOLOGY” has been registered by the Agency for Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan from 09.10.2024 under the registration number №390637. License number: C-5669633. PNFL: 30407832680027

Our address: Tashkent city, Yunusobod district, 19th block,
House 17.



Acceptance of articles

Published every
monthly



Directions

Social, economic, political,
technological, scientific



Scopus || Scientific electronic journal specializing in Scopus

CERTIFICATE NUMBER: №390637

**ORDER NUMBER ACCORDING TO
THE LICENSE REGISTER: C-5669633**

CONTACT:



Contact us
+998 50 737 87 88



Telegram channel
t.me/scopus_IST2100



Journal official website
<https://ist-journal.uz/index.php/IST>