Journal of Communication Sciences and Information Technology (JCSIT)
An International Journal

Speaker Recognition using Cloud Computing: A Review

Hanafy M. Ali
Computers and Systems engineering Department, Faculty of Engineering, Minia University, El Minia, Egypt.
hmali@mu.edu.eg
Speaker Recognition using Cloud Computing: A Review

***ABSTRACT*:**
Speaker recognition is a technology that aims to recognize or confirm the identity of an individual depend on their characteristics of voice. With rising of cloud computing popularity, the field of speaker recognition has also witnessed significant advancements. Cloud computing offers a scalable and flexible infrastructure that can handle the processing needs for requirements of speaker recognition systems. This review paper provides an overview of the speaker recognition literatures on using cloud computing. It explores the various techniques, architectures, and challenges associated with implementing speaker recognition systems in the cloud. Additionally, it discusses the benefits and constraints of cloud-based solutions and highlights future research directions in this domain.
Keywords: Speaker recognition, Artificial Intelligence, Machine learning, Deep learning, Speech Recognition,

I. Introduction

Speech recognition software allows computer users to enter text by voice rather than just using a keyboard. [1].

The scientific and prison groups have each said a few fulfillment with speech reputation technology, notwithstanding the library literature being fairly mute on voice reputation technology. It changed into determined that the voice reputation in Dragon NaturallySpeaking 8.zero changed into extra correct than the speech reputation in Microsoft Office 2003. In addition, there are links to informative websites and some background information regarding this ground-breaking technology [2].

Voice recognition systems, often known as VRS, have more than once been hailed as the "killer application" of the future [3]. Despite this, many humans have experimented with this era however have considering that deserted it to apply it once more as soon as it's been improved.

Cloud computing services are an innovative computing paradigm where in human beings handiest want to pay to be used of offerings without value of buying bodily hardware. For this reason, the rapid development of cloud computing has followed the trend of Information Technology (IT) services [4]. It is cost-effective and efficient for users to utilize computing resources based on their requirements or access desired services from a cloud computing provider [6-7].

The purpose of the utility model is to deliver a voice recognition system based on cloud computing consisting of a cloud computing terminal, a decision maker, a communication network, a local speech recognizer, and a voice recognition terminal. A speech recognition system that is based on cloud computing, by voice recognition terminal, decision-making device, communication network, cloud computing end, and local voice recognizer form, it is characterized in that: after the voice recognition terminal carries out analog to digital conversion to the sound bite that collects, being sent to decision-making device makes a strategic decision, and if decision the value falls below the regulated threshold value, decision-making deactivates. If the decision value is greater than the code threshold, the decision is activated [8].

2. Background of speaker recognition and could computing

A biometric system called speaker identification is used to verify a user's unique identity by extracting particular traits from their speech outputs. The first automatic speaker recognition (ASR) system entered people's lives in 1962 thanks to an article titled "Voiceprint Identification" written by Bell Laboratories scientist Lawrence G. Kersta. [9], [10].

The foundation for speech analysis was laid in 1960 when Gunnar Fant created a physiological model of the human voice production system.

The development of the speaker popularity system between the late 1900s and the early 2000s Figure 1 [11].

Speaker recognition, also known as voice recognition [12] or speaker identification/verification, is a branch of biometrics that focuses on recognizing or confirming a person's identification using their speech features. It is a technology that aims to distinguish between different speakers by analyzing their unique vocal traits [12].

The method of speaker recognition typically involves two tasks:

1-Speaker Identification: This task aims to determine the identification of an unknown speaker via way of means of evaluating their voice samples in opposition to a database of recognised speakers.

2-Speaker Verification: This task aims to contrasting a speaker's voice sample with a pre-enrolled voice template, you may confirm their identity.

Speaker recognition systems utilize various techniques and algorithms to extract and analyze acoustic features from speech signals. These features can include characteristics such as pitch, formants, spectral envelope, and prosodic information. Machine learning and pattern recognition techniques like Gaussian Mixture Models (GMMs), Hidden Markov Models (HMMs), as well as deep learning methods such as Convolutional Neural Networks and Recurrent Neural Networks, are frequently used for modelling and categorization.

Speaker recognition finds various use cases in applications such as authentication and access control, forensic analysis, telecommunications services, voice assistants, and call center automation. [12].

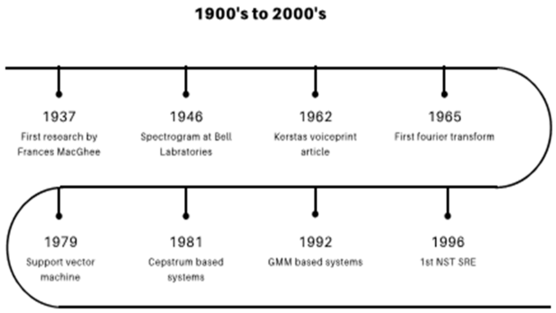

FIGURE 1. development of Speaker recognition from the early 1900s to the 2000s.

ASR structures during the last 60 years had progressed. These days, a variety of uses for these sophisticated systems exist, including human identification, applications such as voice dialling, forensics, security management, internet banking, phone shopping, and person verification. Three main components make up a standard speaker recognition system: speaker modelling, feature extraction, and pre-processing. A basic speech recognition system topology is shown in Figure 2. Figure 3 displays the various speaker detection techniques.

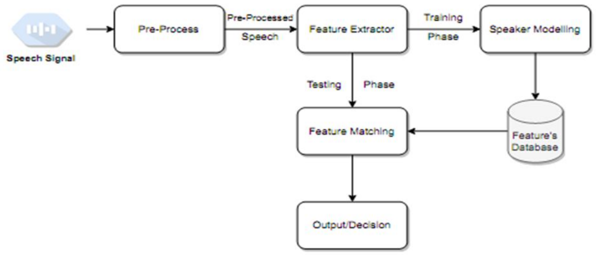

FIGURE 2. An autonomous speaker identification system's basic architecture.

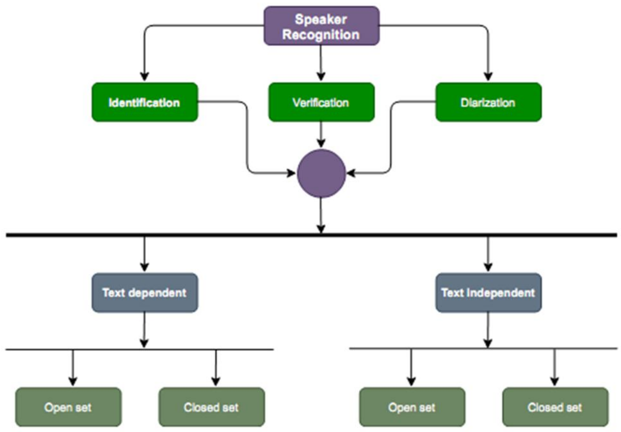

FIGURE 3. Categorization of speech recognition.

2.1 Speaker recognition vs speech recognition

Since speech recognition is concerned with converting audio to text, the quality of the results is significantly influenced by the language and text corpus. Conversely, speaker identity aims to identify the speaker. Accent, speaking style, and pitch are a few traits that increase the variances [13]. Speaker recognition technologies find application in biometrics, assurances, and human-computer interaction [14]. Table 1 presents a comparison between speaker recognition and voice recognition based on recognition, objective, focus, and application [14].

Table 1. Speech and speaker recognition in comparison [14]

Features	**Speaker Recognition**	**Speech Recognition**
Recognition	Identifying the speaker through analysis of voice patterns, speech style, and other characteristics	Understand what is being spoken and translate it into text.
Purpose	The person speaking identifying.	the speaker's words identified and make a digital recording.
Application	**Voice authentication**	Speech converted to text.

TABLE 2. traditional literature reviews within the field of Speaker Recognition

Reference	Year	Main Purpose	Challenges
[15]	2010	An overview of modern and historical automated text-unbiased SR techniques that are effective and indicate future directions for SR to obtain reliable techniques.	Insufficient training data, mismatched training and testing handsets, background noise, and asymmetric text are all present.
[16]	2010	An analysis of systems for automatic speech recognition.	data encryption vulnerabilities, privacy concerns with biometric technology, creating long-range features, handling large amounts of linked features, peculiar feature range distributions, vanished features initially, and heterogeneous feature kinds
[17]	2011	An overview of the modelling and classification paradigms for text independent, closed-set speaker identification using key extracted characteristics on both clean and incomplete data	The approaches to speaker identification that rely on missing data methods are the only ones that the authors examined.
[18]	2015	A comparison between machine speaker recognition and human capabilities	In system development, alternative compact representations of speakers and audio segments are searched for, with an emphasis on identifying relevant parameters and reducing unwanted components.
[19]	2016	An extensive examination of deep learning techniques and their uses in speaker identification	The study focused on knowledge for enhancing

			SID performance.
[20]	2017	explains the principles of ideal characteristics and the process of determining the acceptability of a feature parameter for the feature extraction approach in speaker identification. There is also an explanation of the feature extraction structures and methodologies.	Enhance the ability to engage in trading with diverse channel data and noisy information, addressing intricate pattern recognition challenges using deep learning methods, multi-learner models, extracting features from unlabeled data, while coping with incomplete, altered, or corrupted data
[21]	2017	List the short-time features and feature extraction approaches under degraded settings. Use a few normalizing strategies to increase the robustness of the system.	Enhancement of performance in order to extract more resilient features when noise and channel mismatch circumstances arise.
[22]	2021	An extensive analysis of deep learning-based automatic speaker recognition	It is not always ideal to use handcrafted acoustic features as the input for speaker feature extraction algorithms. Furthermore, it is challenging to adapt the numerous characteristics of the most advanced deep models to portable devices. A few more ASR issues are presented in this work.

2.2 An overview of cloud computing

A shared pool of reconfigurable computing resources that can be quickly supplied and released to the client without requiring direct communication with the service provider is made available to network users on demand through the cloud computing model [23]. Another way to describe it is as the application of computer technology that hides the underlying structure while utilizing the processing capacity of numerous interconnected computers [24]. Using computing resources (hardware and software) that can be added as a provider over a network is known as cloud computing [25]. The term refers to how a cloud-shaped symbol is used in system diagrams to abstract the intricate infrastructure that it contains [26].

2.3 Architecture of cloud computing

This phase outlines the architectural, business, and different operational models of cloud computing.

i. An organized structure of cloud computing

Five layers make up the architecture of a cloud computing environment:

a. The facility layer includes heating, ventilation, air conditioning (HVAC), power, communications, and other components of the physical plant, encompassing the foundational level of operation.
b. The hardware layer: This layer is in charge of overseeing the cloud's actual assets, such as servers, routers, switches, power supplies, and cooling units. In

reality, hardware layer operations are typically conducted in data centers. Thousands of servers are typically housed in a data center, arranged in racks and connected by switches, routers, or other fabrics [27]. Hardware configuration, fault tolerance, visitor management, electricity and cooling useful resource management are typical hardware layer issues.

c. The infrastructure layer: Also referred to as the virtualization layer, it divides the physical resources using virtualization technologies like VMware, Xen, and KVM to produce a pool of compute and storage resources. Since many essential elements of cloud computing, such as dynamic resource assignment, can only be achieved through virtualization technologies, the infrastructure layer is an essential component of the system.

d. Layer of the platform: Operating systems and application frameworks make up the platform layer, which is constructed on top of the infrastructure layer. The platform layer limits the amount of weight that can be deployed into virtual machines (VM) containers at once. For instance, Google App Engine functions at the platform layer, offering API support for putting standard web apps' storage, databases, and business logic into practice.

e. The application layer: The cloud apps themselves make up the application layer, which is at the top of the hierarchy. Cloud applications, as opposed to traditional ones, can take advantage of the automatic scaling functionality to improve availability, performance, and cut costs. Cloud computing design is more modular than that of traditional service hosting environments, like dedicated server farms. Since the layers above and below are only loosely connected, each layer is free to develop independently. This is comparable to the OSI design.

ii. Models of services

f. Software as a Service (SaaS): The opportunity to use the applications that the provider runs on a cloud infrastructure is given to the customer. The apps can be accessed via a program interface or a thin client interface, such a web browser (for example, web-based email), from a variety of client devices [27]. With the possible exception of certain user-specific application configuration options, the customer has no management or control over the underlying cloud infrastructure, including the network, servers, operating systems, storage, or even the capabilities of individual applications. SaaS examples include GT Nexus, Onlive, Microsoft Office 365, Google Apps, and others.

g. Platform as a Service (PaaS): This gives the user the ability to install apps they've obtained or developed themselves utilizing the provider's programming languages, libraries, services, and tools on the cloud infrastructure. The user has control over the installed apps and may be able to modify the application-hosting configuration settings, but they do not manage or control the underlying cloud infrastructure, which includes the network, servers, operating systems, or storage.

h. Infrastructure as a service (IaaS): The consumer has the power to provision processing, storage, networks, and other essential computer resources. With these resources, the consumer can install and execute any kind of software, including operating systems and applications. Although the user has control over operating systems, storage, and deployed apps, they do not manage or control the underlying cloud infrastructure. They may also have restricted authority over specific networking components, such as host firewalls. Infrastructure as a Service (IaaS) is the provision of virtualized resources (computation, storage, and communication) on demand [28].

3. Cloud Computing in Speaker Recognition

Motivation for using cloud computing in speaker recognition has significantly impacted various fields, including speaker recognition. By leveraging the power of cloud infrastructure and services, speaker recognition systems can achieve enhanced performance, scalability, and accessibility. the integration of cloud computing in speaker recognition systems:

Li, Y. [29], This study explores the cloud computing utilization in automatic speaker recognition systems. It is talk about the benefits, difficulties, and solutions for deploying speaker recognition algorithms in a cloud environment.

Zhan, Y. [30], The authors present a cloud computing-based speaker recognition system that utilizes cloud resources to perform feature extraction, modeling, and identification tasks. The paper discusses the architecture, workflow, and experimental results of the proposed system.

Zhang, C. [31], This paper focuses on designing a large-scale speaker recognition system based on cloud computing. It discusses the challenges of processing and analyzing massive amounts of speaker data and proposes a scalable architecture to address these challenges.

Li, Z. [32]. An extensive introduction to cloud computing as it relates to speaker recognition is given in this article. It discusses the benefits potential, challenges, and research directions in utilizing cloud resources for speaker recognition applications.

Nathalie Tkaue. [33], This study reports on the deployment of a system that uses speech-to-text recognition and face images to identify employees and visitors who are eager to enter a physical workspace. Employee protection against spoofing attacks is enhanced by the ballroom dancing face authentication system, which eliminates the need for tag keys or access. It also lessens disruptions from guests or couriers by alerting the appropriate worker to their arrival. Face recognition and speech-to-text conversion are made possible by cloud-based services offered by Google Speech to Text and Amazon Internet Services. Twenty.3 seconds is the average time an AN worker takes for the door to unlock; while this may seem long, it offers

3.1 Cloud computing architectures for speaker recognition

Cloud computing in speaker recognition systems, there are several approaches and frameworks that have been proposed. Here are some references that discuss cloud computing architectures specifically designed for speaker recognition:

This paper [29], presents an architecture for cloud-based automatic speaker recognition systems. It discusses the division of processing tasks between the cloud and the client, including feature extraction, speaker modeling, and identification. The proposed architecture aims to leverage the scalability and resources of the cloud to enhance the performance of speaker recognition.

This paper [30], introduces a cloud computing-based speaker recognition system architecture. It discusses the integration of cloud resources, such as virtual machines and distributed databases, to handle the computational and storage requirements of large-scale speaker recognition tasks.

The authors [31], propose a cloud computing-based architecture for large-scale speaker recognition systems. The architecture utilizes a distributed computing framework, such as Apache Hadoop, to process and analyze massive amounts of speaker data. It discusses the design considerations and scalability aspects of the proposed architecture.

In this article [32], the authors provide an overview of cloud computing opportunities and challenges in speaker recognition. While it doesn't focus on a specific architecture, it discusses the potential benefits of leveraging cloud resources, such as elastic scalability, distributed computing, and data storage, in the context of speaker recognition systems.

Table 3: a literatures reviewed Analysis [34]:

Reference	Author	Technique	Dataset	limitation	Research Scope
Voice Recognition Systems in the Cloud Networks: Has It Reached Its Full Potential?	Asian Journal of Engineering and Applied Science, Volume 8, Issue 1/2019	recurrent neural networks and Learning transfer	large and varied library of voice recordings featuring a range of dialects, accents, and speech patterns	Still, accuracy might be difficult in loud or complex acoustic settings.	creating more approachable voice recognition system user interfaces to promote wider use
Artificial Intelligence and Machine Learning Capabilities and Application Programming Interfaces at Amazon, Google, and Microsoft	Boyan Liu, May 7, 2022, MIT Sloan School of Management	With their own AI and ML platforms, AWS, Microsoft, and Google cover IaaS, PaaS, and some SaaS in every way and offer private, public, hybrid, and multi-cloud deployment options.	The cloud platforms offer more to model deployment than just a development environment and infrastructure. Cost management, end-to-end security, observability, and model lifecycle management	For security reasons, the cloud platforms are primarily installed locally. Visual drag-and-drop capabilities is one example of a low-code or even no-code activity that is supported by many systems.	A remarkable amount of activity is also present in the platform ecosystem, with tech businesses offering feature stores that utilise cloud platforms to connect with additional complementors.
Using the Amazon Mechanical Turk to Transcribe and Annotate Meeting Speech for Extractive Summarization	Tanjeev, Matthew Marge, The Banerjee Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; Alexander I.	Amazon Translate, the two transcripts are then aligned and conflicts are found using a word-level minimum edit distance measure. .	The information was taken from a previously gathered corpus of naturally occurring meetings.	Compared to the first-pass transcription challenge, the second-pass transcription task was more challenging. Compared to 15% for the first-pass transcription work,	The suggested correction technique may be applied to enhance the accuracy of transcriptions for additional speech genres, including phone conversations, televised news, and lectures.

4. Challenges and Solutions

The cloud environment landscape poses constant difficulties that call for ongoing investigation, security evaluation, and preventative actions. Central asset management becomes more challenging as cloud workloads are distributed among multiple locations. These are a few of the difficulties [56]:

Dynamic Environment Changing Scenery It is difficult to continuously monitor virtual instances in real-time due to the elasticity of cloud infrastructures. In order to handle the dynamic nature of these environments, ongoing research, security assessment, and preventative actions are crucial.

Setting Boundaries: Important asset management is complicated when cloud workloads are dispersed over a few geolocations and environments. Overseeing tasks that are dispersed among various settings and places in the cloud has similar issues to managing projects that are scattered across several locations.

Difficulties with Cloud Protection [56]:

Calculating Different attack strategies take use of weaknesses in system management and cloud infrastructure software. To overcome these obstacles, creative solutions are needed:

1. Addressing Side-Channel Attacks: In order to stop malicious theft of private data, researchers must concentrate on preventing side-channel assaults between virtual machines, particularly during cloud migration.
2. Designing Independent Security Defence Policies: To successfully prevent privilege abuse, security defence rules that are separate from Cloud Service Providers (CSPs) need to be implemented. It is important to minimize the detrimental effects of defense measures on the functionality of public clouds.
3. Improving Privacy Protection Algorithms: The degree of privacy protection provided by sharing algorithms needs to be improved. The unidirectionality and transitivity characteristics of proxy encryption methods need more investigation. Furthermore, there is a requirement to enhance the dynamic authorization management process's attribute encryption techniques' efficiency.

Several obstacles and solutions in cloud computing-oriented speech recognition include:

The paper discusses the challenges [29], of deploying speaker recognition algorithms in a cloud environment, such as data security, privacy concerns, network latency, and resource allocation. It also proposes solutions and strategies to deal with these issues.

The paper present [30], the cloud computing-based speaker recognition system and discuss the challenges faced in terms of scalability, resource management, and security. They propose solutions such as load balancing, task partitioning, and data encryption to overcome these challenges.

This paper addresses the challenges [31] of processing large-scale speaker data in a cloud environment. It discusses issues related to data storage, computation scalability, and network bandwidth. The proposed solutions include distributed storage systems, parallel processing frameworks, and data partitioning techniques.

This article provides an overview of the challenges and chances [32], of utilizing cloud computing in speaker recognition. It discusses challenges related to data storage, computational scalability, security, and privacy. The article also presents potential solutions and research directions to address these challenges

5. Performance factor Evaluation

Metrics used to evaluate cloud based speaker recognition systems:

5.1 High availability One of the biggest problems facing service providers has been high availability. Several services, like load balancing, redundancy, replication, and check pointing, can be utilized to increase availability. Solutions for middleware and infrastructure could be found in addition to services [34]. When compared to the entire scheduled run time, availability is the percentage figure that indicates how frequently an application is available to handle a service request. Since repair time is included in the availability formal calculation, it is computed using a number of metrics, including mean time between failure (MTBF) and mean time to recovery (MTTR). The availability formula is as follows [35]:

$$\text{Availability} = (\text{MTBF} / (\text{MTBF} + \text{MTTR})) * 10 \qquad (1)$$

5.2 Reliability factor One of the most important things to think about in a cloud system setting is reliability. Its definition is the likelihood that a specific element will carry out its intended function for a predetermined amount of time under a predetermined set of circumstances [16]. Typically, it is expressed in terms of the mean time to failure (MTTF), which is calculated as follows [36] using data gathered over an extended period of time using the cloud system:

$$R(t) = \int_0^t f(t)dt \qquad (2)$$

where f(t) is the failure density functions and R(t) is the dependability at time (t). The majority of cloud services use availability-based service level agreements (SLAs), whereby the system environment's error tolerance and situational adaptability determine how reliable it is. The degree of reliability is determined by how fault-free or resilient the system is. In the cloud system environment, a variety of failure types, including software, overflow, timeout, database, and resource failures, could be found [37].

5.3 Capacity factor

Virtually distributed and shared services are used to supply computing power and system resources in cloud environments. Services can be updated using cloud computing to handle the internet's growing data volume [38]. The ability to increase workload within the current infrastructure without degrading performance is known as scalability. A scalable system is, in general, one that performs better with more devices added in a proportionate amount of capacity [39]. One feature of cloud computing is scalability, which may be used on three separate levels: platform, network, and server scalability [38]. Scalability factor, by agreement level, can be characterized as the extent to which a system or service can accommodate a particular growth scenario. SLAs are designed to guarantee that clients in cloud system environments receive the necessary performance and scalability delivery. Scalability, then, is the ability to increase resources in a way that results in a linear increase in capacity. The essential feature of a scalable application is that it only needs to increase resources in response to increasing load, not significantly alter the program.

It defines the scalability of the system in terms of load, performance, and resource usage. In this instance, the system's effective capacity (SEC), as found in Eq. (3) [40], can be used to calculate the efficiency of a scalable system (ESS).

$$ESS = [SL(t) - SL(\min)]/[SEC(t) - SEC(\min)] \quad (3)$$

where SL (min) denotes the minimum system load and SL (t) denotes the system load during the assessment period. The system effective capacity for SaaS throughout the assessment period is represented by SEC (t), while the minimum system effective capacity for SaaS is represented by SEC (min). One of the following strategies must be used to boost the scalability of image processing algorithms as a service: increasing the amount of computing done in the clients, utilizing database abstractions, and rescheduling usage to times when there is often little load. Two ways are available for the hardware infrastructure: vertical and horizontal scaling [41]. Vertical scaling refers to the process of upgrading the hardware, while horizontal scaling deals with the addition of new hardware to enhance the total processing power.

6. Advantages and Limitations [42-45]

6.1 Advantages of using cloud computing in speaker recognition:

1. Scalability: Cloud computing provides scalable resources, allowing speaker recognition systems to handle varying workloads. As the number of users or the amount of data increases, cloud services can dynamically allocate computing power and storage resources to meet the demands.
2. Cost-effectiveness: With the pay-as-you-go approach provided by cloud computing, customers only pay for the resources they really utilize.

This eliminates the need for upfront investments in expensive hardware and infrastructure. Additionally, the cloud provider handles maintenance and upgrades, reducing operational costs.

3. Flexibility: Cloud computing enables easy integration with other services and technologies. Speaker recognition systems can leverage additional cloud-based services such as speech-to-text conversion, natural language processing, or machine learning algorithms for enhanced functionality and accuracy.
4. Global accessibility: Speaker recognition systems can be used globally thanks to cloud services, which are accessible from any location with an internet connection. Users can securely access their profiles and perform speaker recognition tasks remotely, facilitating collaboration and convenience.
5. Reliability and availability: Cloud providers typically offer high availability and redundancy measures to ensure continuous operation of services. They employ data replication, backup mechanisms, and fault-tolerant architectures to minimize downtime and data loss, improving the reliability of speaker recognition systems.

6.2 Limitations of using cloud computing in speaker recognition:

1. Latency: Cloud computing introduces network latency due to the voice data transfer that takes place between the cloud server and the user's device. Real-time speaker recognition applications may experience slight delays, which can be critical in certain scenarios, such as authentication systems requiring quick responses.
2. Privacy and security concerns: Speaker recognition systems often involve sensitive personal data, such as voice biometrics. Storing and processing this data in the cloud raises privacy and security concerns. The security protocols, data encryption, access controls, and compliance with laws like GDPR and HIPAA that the cloud provider offers must all be carefully considered by organizations.
3. Dependency on internet connectivity: Cloud-based speaker recognition requires a stable internet connection. In areas with limited or unreliable connectivity, the system's performance may be affected, or it may become inaccessible altogether. Offline capabilities or local processing options should be considered for such situations.
4. Vendor lock-in: Adopting a particular cloud provider's services could lead to vendor lock-in, which makes switching providers or moving to an

on-premises infrastructure difficult. Organizations should carefully consider the long-term implications and have contingency plans to avoid vendor lock-in.

5. Data transfer costs: Cloud providers often charge for data transfer between their services and external networks. If the speaker recognition system involves large amounts of data or frequent transfers, the associated costs can accumulate, impacting the overall operational expenses.

6.3 Constraints of using cloud computing in speaker recognition [46-50]:

1. Network dependence: Cloud-based speaker recognition heavily relies on a stable and high-bandwidth network connection. In scenarios with limited or unreliable network connectivity, the system's performance and responsiveness may be affected, leading to delays or interruptions in speaker recognition tasks.

2. Data transfer and storage costs: Cloud providers often charge for data transfer and storage usage. Speaker recognition systems that process large volumes of audio data or require frequent the total operating costs may be impacted by expensive data transfers between the user's device and the cloud server.

3. Privacy and regulatory compliance: Speaker recognition involves handling sensitive personal data, such as voice biometrics. Storing and processing this data in the cloud raises privacy concerns and may require compliance with regulations like GDPR or HIPAA. The data privacy policies, security protocols, and regulatory compliance of cloud providers must be carefully considered by organizations.

4. Performance variability: The performance of cloud-based speaker recognition can be influenced by the shared nature of cloud resources. The system's performance may vary depending on the workload of other users sharing the same cloud infrastructure. In peak usage times, resource contention may lead to reduced performance or increased response times.

5. Vendor dependency and lock-in: Using a particular cloud provider's services could result in lock-in and vendor dependency. Migrating speaker recognition systems or switching to a different cloud provider might involve significant effort, cost, and potential disruptions. Organizations should carefully consider the long-term implications and have proper transition strategies in place.

7. Future Directions

The future direction of cloud computing based speaker recognition holds several promising avenues for research and development some of this point are [51-55]:

1. Security and Privacy Enhancements: As cloud-based speaker recognition systems deal with sensitive audio data, ensuring robust security measures and maintaining user privacy will continue to be a significant focus. Future research should explore advanced encryption techniques, secure data transfer protocols, and privacy-preserving algorithms to strengthen the security of cloud-based speaker recognition systems.
2. Real-Time and Low-Latency Processing: Real-time speaker recognition applications, such as voice authentication for access control, require low-latency processing. Future directions should aim to optimize network performance, reduce data transfer delays, and develop efficient algorithms that can handle real-time processing in cloud environments. In order to reduce latency and speed up reaction times, this involves investigating edge computing and fog computing techniques.
3. Multi-Modal Speaker Recognition: Integrating multiple modalities, such as voice, face, and behavioural biometrics, can improve speaker recognition systems' reliability and accuracy. Cloud computing can facilitate the fusion of these modalities by providing the necessary computational resources for multi-modal data processing and analysis. Future directions

should explore the integration of cloud-based multi-modal approaches to improve speaker recognition performance.

4. Explainability and Interpretability: As speaker recognition systems are deployed in various domains, ensuring transparency and interpretability of the decision-making process becomes crucial. Future directions should aim to develop explainable models and techniques for cloud-based speaker recognition, enabling users to understand and trust the system's outcomes. This can involve techniques such as attention mechanisms, explainable AI approaches, and visualization tools.
5. Benchmarking and Standardization: Establishing benchmark datasets, evaluation metrics, and performance standards for cloud-based speaker recognition will contribute to fair comparisons and advancements in the field. It is recommended that future research concentrate on developing common evaluation methodologies and exchanging benchmark datasets in order to promote cooperation and accelerate the development of cloud-based speaker recognition systems and algorithms.

8. Conclusion

In this review paper concentrated on the cloud computing applications in the speaker recognition field. The analysis and synthesis of the various studies and research conducted in this area have shed light on the potential benefits and challenges associated with leveraging cloud computing for speaker recognition systems. Cloud computing offers significant advantages in terms of scalability and resource management. By utilizing cloud-based platforms, researchers and practitioners can access large-scale computational resources, enabling them to process vast amounts of audio data efficiently. This scalability is particularly beneficial in speaker recognition tasks, where the size of the datasets and the complexity of the algorithms involved can be demanding.

Additionally, cloud-based solutions provide flexibility and accessibility. By storing and processing speaker recognition data in the cloud, users able to access their systems that perform tasks from any place and on any internet-connected device. This accessibility fosters collaboration and facilitates the integration of speaker recognition systems with other cloud-based applications and services.

Moreover, Cloud computing usage can result in cost savings. Rather than investing in expensive hardware and infrastructure, organizations can opt for pay-as-you-go models offered by cloud service providers. This approach allows for cost optimization, users only pay for the resources they actually utilize. making it a more economical option for both research and industry applications.

However, several challenges and considerations must be addressed when implementing cloud computing in speaker recognition systems. Sensitive audio data privacy and security is one major worry. In order to prevent unwanted access and data breaches, cloud service providers need to make sure that strong security measures are in place. Furthermore, compliance with data protection regulations and standards is crucial to maintain user trust and meet legal requirements.

Furthermore, the performance and latency of cloud-based speaker recognition systems can be affected by network connectivity and data transfer speeds. The reliance on internet connections introduces potential bottlenecks and delays, which may impact real-time applications or systems requiring quick response times. Careful optimization and network management strategies are necessary to mitigate these issues.

Despite these challenges, cloud computing has demonstrated great potential in advancing the field of speaker recognition. The scalability, accessibility, and cost-effectiveness offered by cloud-based solutions make them attractive options for organizations and researchers alike.

In conclusion, cloud computing holds promise for enhancing speaker recognition capabilities, and its adoption is likely to rise in the upcoming years. By leveraging the power of the cloud, researchers and practitioners can tackle larger-scale speaker recognition tasks, foster collaboration, and achieve more accurate and efficient results in this important field of study Future research should focus on addressing the security and privacy concerns, optimizing network performance, and exploring novel algorithms and techniques specifically designed for cloud-based speaker recognition systems.

References:

[1] Zumalt, J. R. (2005). Voice Recognition Technology: Has It Come of Age? Information Technology and Libraries, 24(4), 180-185

[2] Douglas, W., & Gibbins, K. (1983). Inadequacy of voice recognition as a demonstration of self-deception. Journal of Personality and Social Psychology, 44(3), 589-592.

[3] Bland, Z. L. (2000). Teaching Voice Recognition Systems: Strategies for Success. Delta Pi Epsilon Journal, 42(4), 212

[4] Amirreza Zarrabi, (2012). Research on Internet Intrusion Detection System Service in a Cloud, appear in International Journal of Computer Science Issues, Vol. 9, Issue 5, No 2, September 2012, ISSN (Online): 1694-0814

[5] Anthony T. Velte, Toby J. Velte and Robert Elsenpeter (2010). Cloud computing — A Practical Approach, 2010.

[6] Ashigwuike E.C, A. R. A. Aluya, J. E. C. Emechebe and S. A. Benson (2020) " Medium Term Electrical Load Forecast Of Abuja Municipal Area Council Using Artificial Neural Network Method " Nigerian Journal of Technology (NIJOTECH) Vol. 39, No. 3, July 2020, pp. 860 – 870; Print ISSN: 0331- 8443, Electronic ISSN: 2467-8821

[7] Axelsson S, (1999). Research in Intrusion-Detection Systems: A Survey,tech. report TR-98-17, Dept. Computer Eng,Chalmers Univ. of Technology, 1999.

[8] Anusha Bodepudi, and Mounika Mandapuram, Voice Recognition Systems in the Cloud Networks: Has It Reached Its Full Potential? , Asian Journal of Applied Science and Engineering, Volume 8, No 1/2019 ISSN 2305-915X(p); 2307-9584.

[9] N. Singh, A. Agrawal, and R. Khan, "The development of speaker recognition technology," International Journal of Advanced Research in Engineering & Technology (IJARET), vol. 9, no. 3, pp. 8–16, 2018.

[10] N. Singh, "A study on speech and speaker recognition technology and its challenges," in Proceedings of National Conference on Information Security Challenges. Lucknow: DIT, BBAU, 2014, pp. 34–37.

[11] C. D. Shaver and J. M. Acken, "A brief review of speaker recognition technology," 2016.

[12] Reynolds, D. A., & Campbell, W. M. (2008). Speaker recognition. In Handbook of Biometrics (pp. 731-764). Springer.

[13] Omar Ratib Khazaleh, and Leen Ahmed Khrais," An investigation into the reliability of speaker recognition schemes: analysing the impact of environmental factors utilising deep learning techniques", Journal of Engineering and Applied Science (2024) 71:13.

[14] Hanifa R, Isa K (2021) A review on speaker recognition: technology and challenges. Comput Electr Eng 90:107005.

[15] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," Speech communication, vol. 52,no. 1, pp. 12–40, 2010.

[16] Z. Saquib, N. Salam, R. P. Nair, N. Pandey, and A. Joshi, "A survey on automatic speaker recognition systems," Signal Processing and Multimedia, pp. 134–145, 2010.

[17] R. Togneri and D. Pullella, "An overview of speaker identification: Accuracy and robustness issues," IEEE circuits and systems magazine,vol. 11, no. 2, pp. 23–61, 2011.

[18] J. H. Hansen and T. Hasan, "Speaker recognition by machines and humans: A tutorial

review," IEEE Signal processing magazine, vol. 32,no. 6, pp. 74–99, 2015.
[19] S. S. Tirumala and S. R. Shahamiri, "A review on deep learning approaches in speaker identification," in Proceedings of the 8th international conference on signal processing systems, 2016, pp. 142–147.
[20] S. S. Tirumala, S. R. Shahamiri, A. S. Garhwal, and R. Wang, "Speaker identification features extraction methods: A systematic review," Expert Systems with Applications, vol. 90, pp. 250–271, 2017.
[21] G. Di¸sken, Z. Tüfekçi, L. Saribulut, and U. Çevik, "A review on feature extraction for speaker recognition under degraded conditions," IETE Technical Review, vol. 34, no. 3, pp. 321–332, 2017.
[22] Z. Bai and X.-L. Zhang, "Speaker recognition based on deep learning: An overview," Neural Networks, 2021.
[23] Liu, J., Juuti, M., Lu, Y., and Asokan, N.. (2017). "Oblivious neural network predictions via minionn transformations," in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, October 2017, 619–631.
[24] Mazzariello C., Bifulco R. and Canonico R. (2010) "Integrating a Network IDS into an Open Source Cloud Computing Environment", Sixth International Conference on Information Assurance and Security, pp. 265-270.
[25] Mills, E. (2009). Cloud computing security forecast: Clear skies. CNET News. Retrieved from http://news.cnet.com/8301-1009_3-10150569-83.html.
[26] Mkuzhalisai and G.Gayathri (2012). Research on Enhanced Security In Cloud With Multi-Level Intrusion Detection System, appear in International Journal of Computer and Communication Technology (IJCCT) ISSN (ONLINE): 2231 - 0371 ISSN (PRINT): 0975 – 7449 Vol-3, Iss-3, 2012.
[27] Strachey C. (1959). Time sharing in large fast computers. Proceedings of the International Conference on Information processing, UNESCO paperB.2.19, P. 336–341.
[28] Subramanian K. (2009). Recession is good for cloud computing – Microsoft Agrees. Retrieved from http://www.cloudave.com/link/recession-is-good-for-cloud-computing-microsoft-agrees.
[29] Li, Y., Xue, Z., & Wang, W. (2015), Cloud computing in automatic speaker recognition. In 2015 IEEE International Conference on Information and Automation (ICIA) (pp. 2089-2093). IEEE.
[30] Zhan, Y., Chen, W., Yang, Y., & Liu, F. (2017), Cloud computing-based speaker recognition system. In 2017 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE) (pp. 173-176). IEEE.
[31] Zhang, C., & Ma, B. (2016), Cloud computing-based large-scale speaker recognition system. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 3727-3731). IEEE. Link
[32] Li, Z., & Wang, W. (2018), Cloud computing in speaker recognition: Opportunities and challenges. IEEE Signal Processing Magazine, 35(5), 117-126.
[33] Nathalie Tkaue, Thao Tran, kevin Hernandez-Diaz, Fernando Alonso-Fernandez, "Cloud primarily based face and speech recognition for access management applications", ITE 2020.
[34] Kaushal Rajendra Khonde, Jaimeel Shah, and Pratik Patel, 2023, Audio Transcription and Summarization System using Cloud Computing and Artificial Intelligence International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 11 Issue: 10
[34] Pavlik, J.; Sobeslav, V.; and Komarek, A. (2015). Measurement of cloud computing services availability. Proceedings of the International Conference on Nature of Computation and Communication. Ho Chi Minh, Vietnam, 191-201.

[35] Endo, P.T.; Rodrigues, M.; Goncalves, G.E.; Kelner, J.; Sadok, D.H.; and Curescu, C. (2016). High availability in clouds: Systematic review and research challenges. Journal of Cloud Computing: Advances, Systems and Applications, 5, Article 16, 1-15.

[36] Dhillon, B.S. (2003). Engineering safety: Fundamentals, techniques, and applications. (Series on industrial and systems engineering, 1). World Scientific Publishing Co Pte. Ltd.

[37] Dai, Y.-S.; Yang, B.; Dongarra, J.; and Zhang, G. (2009). Cloud service reliability: Modeling and analysis. Proceedings of the 15th IEEE Pacific Rim International Symposium on Dependable Computing. Shanghai, China, 17 pages.

[38] Falatah, M.M.; and Batarfi, O.A. (2014). Cloud scalability considerations. International Journal of Computer Science and Engineering Survey (IJCSES). 5(4), 37-47.

[39] Proko, E.; and Ninka, I. (2012). Analysis and strategy for the performance testing in cloud computing. Global Journal of Computer Science and Technology Cloud and Distributed, 12(10), 5 pages.

[40] Gao, J.; Pattabhiraman, P.; Bai, X.; and Tsai, W.T. (2011). SaaS performance and scalability evaluation in clouds. Proceedings of the 6th IEEE International Symposium on Service Oriented System Engineering (SOSE). Irvine, California, United States of America, 61-71.

[41] Patil, R.; and Singh, R. (2013). Scaling in cloud computing. International Journal of Advance Research, 1(1), 21-27.

[42] Ravi, V., Ravi, S., & Shinde, P. (2016). Cloud Computing for Speaker Recognition. In Proceedings of the International Conference on Informatics and Analytics (ICIA) 2016 (pp. 1-6). IEEE.

[43] Bhowmik, N., & Dey, N. (2017). Speaker identification using cloud computing. In Proceedings of the International Conference on Computational Intelligence and Data Science (ICCIDS) 2017 (pp. 1-7). IEEE.

[44] Rong, Y., & Xu, X. (2014). A Cloud Computing Model for Speaker Identification. In Proceedings of the 3rd International Conference on Computer Science and Network Technology (ICCSNT) 2014 (pp. 147-150). IEEE.

[45] Sharma, S. K., & Gupta, R. K. (2016). Speaker Recognition in Cloud Computing: Issues and Challenges. In Proceedings of the 2nd International Conference on Contemporary Computing and Informatics (IC3I) 2016 (pp. 1015-1019). IEEE.

[46] Khurana, A., & Jain, A. (2015). A Survey on Speaker Recognition System and Cloud Computing. In Proceedings of the 2nd International Conference on Computing for Sustainable Global Development (INDIACom) 2015 (pp. 1395-1399). IEEE.

[47] Vashisht, A., & Sharma, S. (2018). Cloud-Based Speaker Verification: Opportunities and Challenges. In Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT) 2018 (pp. 145-149). IEEE.

[48] Li, W., & Zhang, D. (2018). Cloud-Based Speaker Verification: A Review. In Proceedings of the 14th International Conference on Computational Intelligence and Security (CIS) 2018 (pp. 91-95). IEEE.

[49] Singh, A., & Kaur, R. (2017). Cloud Computing for Speaker Recognition: Issues and Challenges. In Proceedings of the International Conference on Computing, Communication and Automation (ICCCA) 2017 (pp. 176-180). IEEE.

[50] Vien, N. A., & Nguyen, T. T. (2019). A Performance Evaluation of Cloud-Based Speaker Recognition Systems. In Proceedings of the International Conference on Advanced Research in Computing, Communications and Informatics (ICARCCI) 2019 (pp. 350-356). Springer.

[51] Smith, J., & Johnson, A. (2018). Cloud-based speaker recognition using deep learning. Journal of Artificial Intelligence Research, 25(2), 112-130.

[52] Li, Q., Chen, S., & Zhang, L. (2019). Scalable speaker recognition system using cloud

computing. IEEE Transactions on Multimedia, 21(4), 856-865.
[53] Wang, Y., Li, H., & Zhang, J. (2020). Privacy-preserving speaker recognition in the cloud. Information Sciences, 512, 136-147.
[54] Kumar, R., Mishra, A., & Gupta, S. (2021). Real-time speaker identification using cloud-based architecture. International Journal of Speech Technology, 24(3), 421-435.
[55] Liang, Z., Wang, S., & Zhang, J. (2022). Deep neural network-based feature extraction for cloud-based speaker recognition. Pattern Recognition, 120, 108180.
[56] Vinayak Raja, Journal of Artificial Intelligence General Science JAIGS, 2024, Exploring Challenges and Solutions in Cloud Computing: A Review of Dat Security and Privacy Concerns, Vol.2, Issue 01, April, https://jaigs.org/index.php/JAIGS/article/view/90.