Crafting a Speaker Authentication System for Online Proctored Exams

Sherif Welsen[0000-0002-2614-813X], Zhuoyi Chen and Richard Rankin

The University of Nottingham Ningbo China, Ningbo, Zhejiang, China
Sherif.welsen@nottingham.edu.cn

Abstract.

Authenticating the identity of examinees in online proctored examinations is a significant challenge, particularly when invigilators are unfamiliar with the students. The traditional method of identity verification, involving visual inspection of the examinee and their student identification card, is inadequate in the online environment. A speaker authentication system, designed for use in combination with the aforementioned visual checks in online proctored exams, was developed. The advanced speech recognition technologies incorporated result in accurate verification. The system offers a universal solution for exam invigilators, eliminating the need for prior acquaintance with the examinee, and its Graphical User Interface contributes to its user-friendliness for the proctor. It is also user-friendly for the examinee as it does not require them to have any additional hardware or install any additional software. The proposed robust and reliable speaker authentication system addresses the challenge of preventing contact cheating by way of impersonation in online proctored examinations.

Keywords: Online Exams, Examinee Authentication, Speech Recognition.

1 Introduction

In response to the global pandemic, the University of Nottingham Ningbo China (UNNC) introduced proctored online examinations to accommodate international students unable to return to campus. This shift, while necessary, has presented significant challenges in verifying the identities of online examinees, particularly when invigilators are unfamiliar with the students. Traditional methods, which rely heavily on the visual inspection of the examinee and their student identification card (ID), have proven insufficient in the online environment. This inadequacy underscores the urgent need for a more reliable system of authentication, a system that can effectively address the complex challenges we are facing. The proposed Speaker Authentication System, with its innovative approach, is not just a potential solution, but a necessary one to ensure the integrity and credibility of online exams.

Online proctoring at UNNC ensured that assessments remained invigilated, mirroring the controlled conditions of on-campus exams and maintaining fairness across the

student body, regardless of their physical location due to COVID-19 restrictions. However, the current system has its limitations. Most UNNC students were present on campus for these assessments; however, those who were quarantined or unable to enter China were monitored online. While faculty members typically invigilated exams for on-campus students, online proctoring for remote students was managed by external services or other university employees, potentially leading to a lack of personal recognition in verifying student identities [1]. This issue was further complicated in situations where space constraints forced the separation of students from the instructing faculty during computer-based tests. These limitations underscore the pressing need for a more robust authentication system. Although UNNC does not currently utilize online proctoring, the possibility of its reimplementation in future scenarios remains [2]. The challenge is managing the logistics of such assessments and ensuring they are conducted with integrity. Academic misconduct, a potential risk even under supervised conditions, can still occur if the identity verification process is compromised [3]. This reiterates the need for a more secure authentication system, such as the proposed Speaker Authentication System, to prevent misconduct and maintain online exams' credibility. Variations in appearance due to factors like makeup or hairstyle, along with potential limitations in video quality due to poor internet connections, can further obscure the identities of examinees [4]. In these cases, traditional visual verification methods fall short, necessitating additional measures to confirm that the person being examined is the student enrolled in the course [5]. The gravity of these challenges cannot be overstated, highlighting the necessity of a more reliable authentication system.

This paper advocates for a novel solution: a speaker authentication system specifically designed for the unique requirements of proctored exams in a remote learning context. The proposed solution is an outcome of a final year project-based learning [6]. This system is not merely an adjunct but a comprehensive tool that bolsters the integrity of online examinations. It introduces a dual-authentication mechanism that supplements traditional student ID checks. This is particularly crucial in cases where physical IDs are damaged or lost. Notably, the proposed system is designed to be user-friendly, requiring no additional hardware or software installations on the examinee's part, thus ensuring accessibility and ease of use. By implementing a speaker authentication system, institutions can enhance the security and reliability of their remote assessments, ensuring that each student's identity is verified through robust, technologically advanced methods. This system not only supports the academic fairness and integrity required during these examinations but also adapts to the evolving landscape of education in the face of ongoing global challenges.

2 Literature Review

The transition to online education, necessitated by the COVID-19 pandemic, has propelled the adoption of online examinations as a critical assessment method [7, 8]. Recognized as an effective alternative for student evaluation, online exams have surged in popularity due to their inherent conveniences and flexibilities [9, 10]. The benefits

of online examinations include reliable grading, cost-efficiency, time savings, and their suitability for formative assessments [11]. These exams offer unparalleled flexibility and convenience for learners and institutions, making them highly favored in academic settings [2]. Despite these advantages, online examinations are not without challenges. Students often encounter technical difficulties, which impede their ability to perform optimally during exams. Moreover, concerns about privacy invasion and the potential for academic dishonesty are significant [2, 7, 9]. The prevalence of academic dishonesty has emerged as a formidable challenge, with common cheating practices threatening the integrity of online assessments [7, 12]. Ensuring the enforcement of academic integrity is, therefore, crucial, as the credibility of online examinations hinges on robust mechanisms to mitigate dishonest behaviors [1].

Introducing remote proctoring services, which utilize lockdown browsers and webcam monitoring, has responded to these challenges. Although these services have led to a noticeable reduction in average exam scores, suggesting a positive impact on academic integrity, they also raise concerns about their implications on student well-being and equity [3, 13-15]. The potential adverse effects on mental health and the creation of inequitable assessment conditions have been highlighted as significant drawbacks [16]. Additionally, the trade-offs between student privacy and the effectiveness of proctoring services, especially concerning the collection of identifiable information, have been debated among educators [4]. To mitigate these issues, integrating various proctoring techniques, such as live webcam monitoring and biometric verification, has been suggested to enhance security and fairness [17].

In biometric verification, voice biometrics has emerged as a promising solution for authenticating identities in an online proctoring context. Voice biometrics, or Automatic Speaker Verification (ASV), provides high accuracy in identifying and authenticating individuals based on voice samples [18]. ASV systems are secure and can be adapted for continuous authentication, making them a preferred choice in various security-sensitive environments, including educational settings [19]. The application of speaker authentication in online exams confirms students' authenticity, thereby enhancing the integrity of the examination process [20]. Moreover, it supports scalable interactions between students and tutors while ensuring consistent validation of student identities over time, effectively addressing privacy concerns and institutional costs [5].

3 Methodology

The methodology employed in this research harnesses a speaker recognition framework developed in MATLAB, utilizing the Mel-Frequency Cepstral Coefficient (MFCC) for feature extraction [21] and the Gaussian Mixture Model-Universal Background Model (GMM-UBM) [22] for subsequent modeling and classification. While the GMM model generally performs well in speaker recognition, its effectiveness diminishes when faced with limited training data and short speech samples. To address this limitation, the Gaussian Mixture Model-Universal Background Model (GMM-UBM) was proposed [23]. This approach involves collecting a large volume of speech data from non-target users, including both male and female speakers, which is then

combined to train a comprehensive GMM model. The Universal Background Model (UBM) is balanced regarding gender representation [24], enabling the adaptation of speaker characteristics not adequately covered by limited training data. The UBM, a large-scale GMM model, is trained using the Expectation-Maximization (EM) algorithm. During training, the Maximum A Posteriori (MAP) algorithm adapts the GMM model for each speaker.

The performance evaluation of the proposed system is based on the Equal Error Rate (EER), calculated using the speaker False Rejection Rate (FRR) and False Acceptance Rate (FAR) [25]. The FRR is determined by testing the GMM model with its corresponding speaker feature vector against a threshold range of -0.5 to 2.5 with intervals of 0.01. Similarly, the FAR is calculated by testing the GMM model with unmatched speaker feature vectors against the same threshold range. The FRR and FAR values are then used to generate two curves, with the EER indicating the point where these curves intersect. A lower EER signifies better system performance, as it represents the balance between false acceptance and false rejection rates at a specific threshold.

The training and testing of the model were conducted using custom MATLAB programs developed for this purpose, utilizing the Google Speech Commands Dataset [26] as the source for training speech data. The implementation leveraged the Audio Toolbox and the Statistics and Machine Learning Toolbox. The proposed system employs speech-dependent recognition, emphasizing comprehensively evaluating authentication accuracy. Initial system performance was evaluated using a designated dataset for offline processing. At the same time, real-time authentication efficacy was assessed using voice samples from a diverse group of 10 volunteer students, ensuring gender representation. This phase aimed to enhance text-dependent recognition's accuracy and operational efficiency, which has the potential for greater authentication precision. Before participation, all volunteers provided informed consent, which was approved by the faculty research ethics committee.

4 Results

The audio feature extractor function in MATLAB was utilized to extract Mel-Frequency Cepstral Coefficients (MFCC) as depicted in Fig. 1. Training speech data for the Universal Background Model (UBM) was sourced from the Google Speech Commands Dataset. Initially, a speech-dependent speaker recognition system was employed, with each speaker trained on their corresponding Gaussian Mixture Model-Universal Background Model (GMM-UBM). The system's performance was evaluated using speech material from public databases, and the False Rejection Rate (FRR) and False Acceptance Rate (FAR) values were measured and plotted across thresholds ranging from -0.5 to 2.5. These values were used to generate the Detection Error Rate (DER) curve.

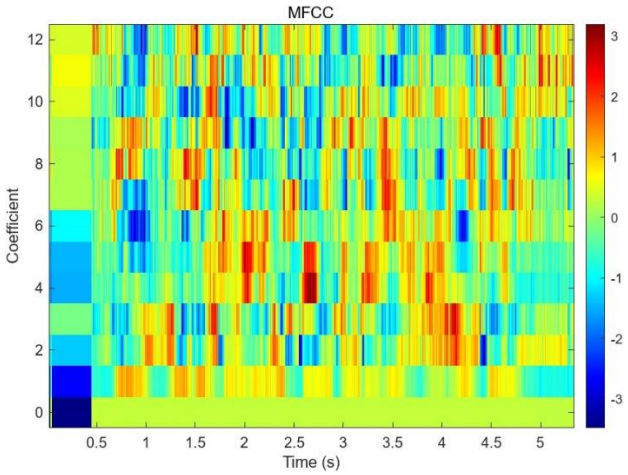

Fig. 1. Extracted MFCC feature.

Fig. 2 illustrates the FRR and FAR curves, with their intersection representing the Equal Error Rate (EER) and its corresponding threshold. The accuracy of the system, calculated as (1– EER) ×100, reached up to 97% in this pilot project. A user-friendly Graphical User Interface (GUI) was developed in MATLAB for system management. University administrative staff responsible for overseeing exams in the remote learning environment can utilize the GUI to add or delete speakers who, in this context, are students taking exams. The GUI allows admin staff to record and add users or import (.wav) format audio files into the system. This feature is handy for remote speech recognition, where the admin can record a sample speech from the remote student and upload it to the cloud for processing. The system can then train a corresponding GMM-UBM model for each speaker and correctly identify the speaker, with the recognition results displayed on the GUI.

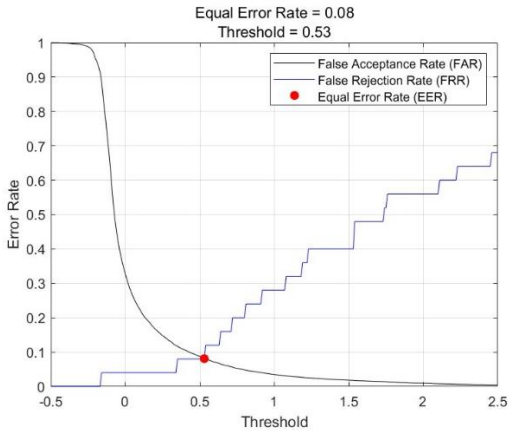

Fig. 2. Equal Error Rate (EER) for specific threshold.

The GUI has two primary functions: recognizing speakers and adding new speakers. Users can choose between real-time recording or importing audio. The Import Audio feature is designed for remote speech recognition. Admin staff can upload the speech file to the cloud after recording, and the user imports the file into the cloud to complete the recognition or addition of the speaker. The GUI's initial design layout is illustrated in Fig. 3. Admin staff can set the threshold based on the illustration in Fig. 3.

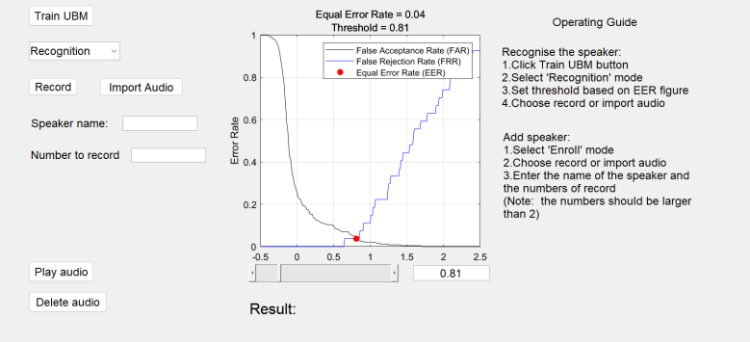

Fig. 3. GUI layout illustrating threshold setting and EER.

Optimizing the Equal Error Rate (EER) threshold proved to be a nuanced task. For instance, when speakers possess a substantial amount of speech material, opting for a higher threshold can effectively reduce the false acceptance rate, consequently enhancing accuracy. Despite an increase in the false rejection rate due to the larger number of speech files per speaker, the likelihood of rejection remains largely unchanged for individual speakers. To capture a speaker's (student's) speech sample, the system prompts the speaker to say a predefined word (e.g., "stop"), agreed upon for this project. Upon completing the recording, the Graphical User Interface (GUI) confirms the process as "complete" and displays the recognition result, as shown in Fig. 4.

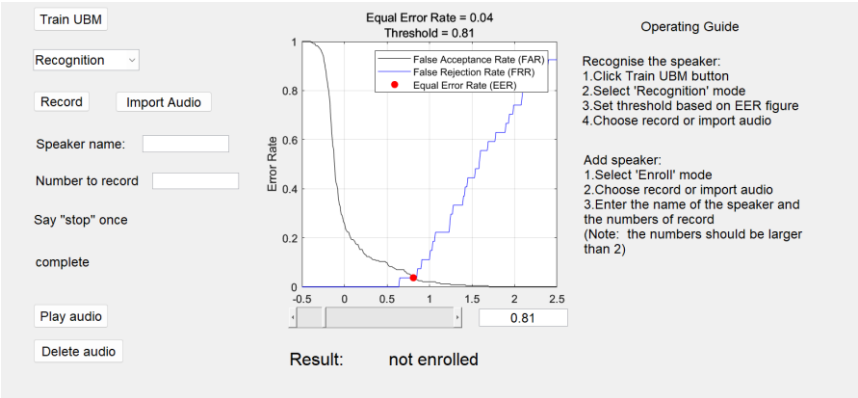

Fig. 4. GUI (Speaker Recognition).

To maintain privacy and security, the system automatically deletes the recording files used for identification purposes. In cases where administrators need to import an offline speech sample, they can utilize the "Import Audio" function to select the previously stored sample. Additionally, administrators can add new speakers by clicking "Enroll" and entering the speaker's name, as illustrated in Fig. 4. When adding a new speaker, administrators must specify the number of speech recordings, which should exceed 2, to ensure adequate enrollment and validation. Administrators can also utilize the "Play audio" function to review recorded speech and delete unsatisfactory recordings using "Delete audio" for re-recording.

5 Conclusion

The proposed speaker authentication system offers a reliable and robust solution for the dual authentication of students participating in online proctored exams for the purpose of mitigating contract cheating by way of impersonation. The authentication system is accurate and is user-friendly to both the proctor and the examinees. By incorporating advanced speech recognition technologies, this system addresses immediate challenges posed by remote examination processes and sets a precedent for future educational integrity and security advancements. While designed for use in online proctored exams, the system developed could also be employed to enhance the candidate authentication process for in-person examinations.

References

1. Manoharan, S. and X. Ye. *On upholding academic integrity in online examinations*. in *2020 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*. 2020. IEEE.
2. Aristeidou, M., et al. *Examining university student satisfaction and barriers to taking online remote exams*. in *9th International Conference on Higher Education Advances (HEAd'23)*. 2024. Editorial Universitat Politècnica de València.
3. Whitlow, E. and S. Metts, *Assuring Academic Integrity of Online Testing in Fundamentals of Accounting Courses.* Online Learning, 2024. **28**(1): p. 87-105.
4. Balash, D.G., et al. *{Educators'} Perspectives of Using (or Not Using) Online Exam Proctoring*. in *32nd USENIX Security Symposium (USENIX Security 23)*. 2023.
5. Renzella, J., A. Cain, and J.-G. Schneider, *Verifying student identity in oral assessments with deep speaker*. Computers and Education: Artificial Intelligence, 2022. **3**: p. 100044.
6. Welsen, S., M. Zhang, and Y. Chu. *Project-Based Network Simulation of Campus Remote Seat Booking System*. in *2022 IEEE 2nd International Conference on Educational Technology (ICET)*. 2022. IEEE.
7. Dayananda, D., et al., *A Systematic Literature Review on Online Exams in COVID-19 Pandemic: Assessment Methods, Students' Preferences, Dishonest Behaviors and Challenges in Online Exams.* 2021 From Innovation To Impact (FITI), 2021. **1**: p. 1-6.
8. Clark, T.M., D.A. Turner, and D.C. Rostam, *Evaluating and improving questions on an unproctored online general chemistry exam.* Journal of Chemical Education, 2022. **99**(10): p. 3510-3521.

9. Nguyen, N.T., et al., *Student experience of online exams in professional programs: current issues and future trends.* Digital Teaching, Learning and Assessment: The Way Forward, 2023: p. 161.
10. Ilgaz, H. and G. Afacan Adanır, *Providing online exams for online learners: Does it really matter for them?* Education and Information Technologies, 2020. **25**(2): p. 1255-1269.
11. Shraım, K., *Online examination practices in higher education institutions: Learners' perspectives.* Turkish Online Journal of Distance Education, 2019. **20**(4): p. 185-196.
12. Tong, A.C.Y., et al., *The age of academic integrity in COVID-19: new normal changes to the health professional education.* International Journal of Mobile Learning and Organisation, 2023. **17**(1-2): p. 45-57.
13. Balash, D.G., et al. *Examining the examiners: Students' privacy and security perceptions of online proctoring services*. in *Seventeenth symposium on usable privacy and security (SOUPS 2021)*. 2021.
14. Marano, E., et al., *What is the student experience of remote proctoring? A pragmatic scoping review.* Higher Education Quarterly, 2023: p. e12506.
15. Davis, A.B., R. Rand, and R. Seay, *Remote proctoring: The effect of proctoring on grades*, in *Advances in accounting education: Teaching and curriculum innovations*. 2016, Emerald Group Publishing Limited. p. 23-50.
16. Pokorny, A., et al., *"Out of my control": science undergraduates report mental health concerns and inconsistent conditions when using remote proctoring software.* International Journal for Educational Integrity, 2023. **19**(1): p. 22.
17. Sinjini, M. and M. Gofman. *Towards greater integrity in online exams*. in *Proceedings of the Americas Conference on Information Systems*. 2016.
18. Bekkanti, N., L. Busch, and S. Amman, *Evaluation of Voice Biometrics for Identification and Authentication*. 2021, SAE Technical Paper.
19. Choudhury, S.P., et al. *Comparative analysis of two different system's framework for text dependent speaker verification*. in *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*. 2015. IEEE.
20. Gadkar, S., et al. *Online Examination Auto-Proctoring System*. in *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*. 2023. IEEE.
21. Abdul, Z.K. and A.K. Al-Talabani, *Mel frequency cepstral coefficient and its applications: A review.* IEEE Access, 2022. **10**: p. 122136-122158.
22. Xiong, Z., et al., *A tree-based kernel selection approach to efficient Gaussian mixture model–universal background model based speaker identification.* Speech communication, 2006. **48**(10): p. 1273-1282.
23. Reynolds, D.A., T.F. Quatieri, and R.B. Dunn, *Speaker verification using adapted Gaussian mixture models.* Digital signal processing, 2000. **10**(1-3): p. 19-41.
24. Kumari, T.J. and H. Jayanna. *Comparison of LPCC and MFCC features and GMM and GMM-UBM modeling for limited data speaker verification*. in *2014 IEEE International Conference on Computational Intelligence and Computing Research*. 2014. IEEE.
25. Cheng, J.-M. and H.-C. Wang. *A method of estimating the equal error rate for automatic speaker verification*. in *2004 International Symposium on Chinese Spoken Language Processing*. 2004. IEEE.
26. Warden, P., *Launching the speech commands dataset.* Google Research Blog, 2017.