National Journal of Multidisciplinary Research and Development
ISSN: 2455-9040
Impact Factor: RJIF 5.22
www.nationaljournals.com
Volume 3; Issue 1; January 2018; Page No. 1296-1299

A predictive approach for speaker verification by machine learning and MFCC

Shipra Gupta[1], Dr. Amit Sharma[2]
[1] M.Tech Scholar, Department of Computer Science & Engineering, Vedant College of Engineering & Technology, Bundi, Rajasthan, India
[2] Professor, Department of Computer Science & Engineering, Vedant College of Engineering & Technology, Bundi, Rajasthan, India

Abstract
The project proffers an intelligent two factor speaker authentication system. The word intelligent is used as the project aims to combine Artificial Intelligence with Feature Extraction Techniques (MFCC). In first stage voice authentication system that incorporates Mel Frequency Cepstral Coefficients and Vector Quantization performs Speaker Identification and Verification based on an audio signal. In second stage further verification and authentication of speaker take place using an algorithm of artificial intelligence-machine learning (AI System) namely decision tree. The AI system asks random questions related to the authorized user and verify that the reply is relevant and came from the same user as determined by the voice authentication system, i.e., AI system checks that the voice signal is not breached.

Keywords: Mel frequency cepstral coefficients (MFCC), vector quantization, artificial intelligence, decision tree, machine learning

1. Introduction
In Speaker Verification there are two phases involved, training and identification. In training phase, the speech samples are collected, and a reference model for that speaker is constructed, and in identification phase, the input speech is compared with the reference and speaker is recognized. Machines are becoming more competent even in case of complex problems it handles the situation and gives output within seconds. The machine can learn faster although it is human programmed code; Machines are getting a high level of intelligence as humans, a bit worried about the future. Natural language is an artificial intelligence specialization that processes the human natural language and prepares computers to give the response. In this decade AI has become less artificial but more intelligent. One of the most exciting areas of the AI is machine learning which turns the concept very interestingly to think that how a machine can learn so fast. In the computer science world, AI is creating another different world for the humans.

2. Existing spoofing attacks and their countermeasures
2.1 Impersonation: Impersonation or human mimicking is a spoofing attack in which target speaker's timbre and prosody are imitated by the attacker without computer-aided technologies. This is the most common spoofing attack. However, it does not pose a real threat to the ASV system as the imposter cannot mimic the spectral features of the target speech. Lau *et al*. Have shown that the impersonator can overcome ASV provided that his voice is naturally similar to the target speaker. As the threat of the attack is not fully understood there are fundamentally no prior studies relating to countermeasures.

2.2 Replay Attack
In replay attack, an imposter plays a prerecorded sample of the target voice as an input to the speaker verification system. The voice may be recorded superficially or can even be concatenation of a number of smaller samples. Replay attack is simple and doesn't require specific knowledge of speech processing. Moreover, the availability of high quality and low cost recording devices have further aggravated the situation. Figure depicts the huge similarity between the voice spectrum of the target and the Replay speech.

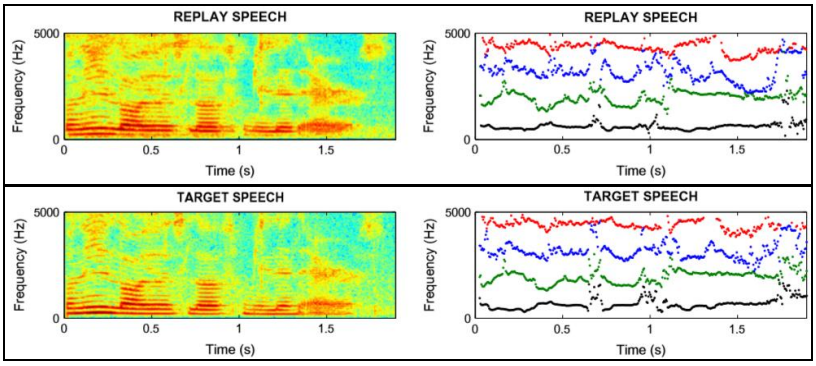

Fig 1

2.2.1 Countermeasures

Various countermeasures have been reported in the literature. Using fixed pass-phrase in which the input signal is compared against a dataset of predefined replay speech and the access is denied if the similarity is more than a predefined threshold value.

1. Channel noise comparison is another technique as the legit input contains channel noise only from the ASV system, but the replay voice consists of other noise in addition to ASV channel noise, for instance, a noise of the loudspeaker.
2. Noise and reverberation analysis are an alternative countermeasure as noise and reverberation is increased while playing recordings, especially far field recordings.
3. Combination of ASV System with other biometric technologies, for instance, face recognition can also combat spoofing attacks.

3. Issues in existing countermeasures

Though extensive research has been conducted in the voice spoofing and countermeasures, but certain issues as mentioned below must be addressed in future for effective realization of a robust speaker verification system.

Generalized Countermeasure

Past researches deal with only a specific type of spoofing attack and suggest countermeasures that are effective for a particular specific attack. However, the future research must be dedicated in pursuit of the countermeasures that can be generalized to stop more than one spoofing attacks and even combat the combination of these attacks.

Countermeasure under acoustic mismatch

Future research should concentrate evaluating countermeasures in acoustically degraded and channel mismatched conditions. For instance, in various transmission channels additive noise and other imperfections are expected which have potential to mask the processing artifacts key for spoofing detection.

Scalability

The proposed approach should be scalable such that it can be easily implemented in devices without constraints.

Computationally Simple

For commercial application of any security device compact size and light weight are essential features. Thus algorithms that don't require much processing power and memory should be constructed for commercial and practical realization of Speaker verification systems.

4. Methodology

The first authentication stage consists of text independent speaker recognition system based on Mel frequency cepstral coefficients and vector quantization. The above mentioned step takes voice signal as input. The input may be from a genuine user or a malicious user or a pre recorded voice signal. The degree of closeness or proximity with the registered user is computed by MFCC based Automatic Speaker Recognition System. The ASR is developed in MATLAB 2015a.

If the degree of closeness is equal or exceeds the threshold value which is 80 percent in our case, the control is directed to second authentication phase, otherwise access is denied. The second authentication phase is developed in Microsoft Visual Studio 2016. The techniques involved in this phase are Decision tree algorithm of Machine learning and Microsoft speech recognition engine. In this stage random question from the database is asked. If the reply is one of the expected answers the counter is incremented else decremented. If the user answers all three questions correctly, i.e., if value of counter is three, access is allowed.

4.1 Process flow algorithm

Step 1: Input signal (Recorded voice/voice input from genuine/malicious user).
Step 2: Pre-emphasis performed on voice signal.
Step 3: Calculate Mel frequency Cepstral coefficients.
Step 4: Perform Vector Quantization to check degree of proximity to the registered user.
Step 5: If (degree of proximity >= threshold value), direct to second authentication phase.
Else deny access.
Step 6: Random Questions based on decision tree.
Step 7: if (User Answer== Expected Answer [])
Count = count+1s
Step 8: Repeat step 6 and 7 two times.
Step 9: if count>= specified threshold, Allow Access
Else Access denied.

5. Proposed algorithm results

To validate the efficiency of the developed system, A database was created consisting of five users. Five samples of 10 seconds duration were taken from each user and the obtained MFCC or vectors were used for training. To establish the accuracy of the system, a test and train user was selected at random, and test vector was given as input to the first user authentication phase for twenty-five times. Despite the dynamic nature of sound the proposed model is efficient with False Acceptance Rate 0.16 and False rejection rate of 0.8. The second authentication state considerably improves the accuracy of ASV system from 84 to 96 percent.

Table 1

Sr No.	Actual USER	Test USER	Access Allowed after first authentication state	Difference
1	AD2	P3	N/A	0.2683
2	AD2	P2	N/A	0.1073
3	AD2	P1	N/A	0.1093
4	AD2	M3	N/A	0.1172
5	AD2	M2	N/A	0.2382
6	AD2	M1	N/A	0.2672

7	AD2	AD1	YES	<0.1
8	AD2	AD3	YES	<0.1
9	AD3	P3	N/A	0.28
10	AD3	P2	N/A	0.1339
11	AD3	P1	N/A	0.1355
12	AD3	M3	N/A	0.1419
13	AD3	M2	N/A	0.2513
14	AD3	M1	N/A	0.279
15	**M4**	**M6**	**N/A**	**0.1876**
16	**M4**	**M5**	**N/A**	**0.2359**
17	N1	N2	YES	<0.1
18	N1	N3	YES	<0.1
19	N2	N3	YES	<0.1
20	N1	P4	YES	<0.1
21	N1	P5	YES	<0.1
22	N1	P6	YES	<0.1
23	N1	M4	N/A	0.2355
24	N1	M5	YES	<0.1
25	N1	M6	N/A	0.1425

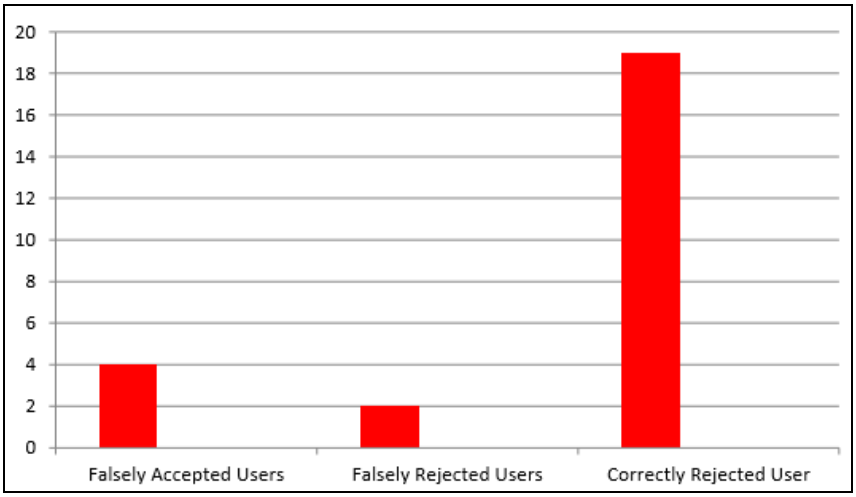

Fig 2

FAR_1 =False Acceptance Rate (After first authentication state) $= \frac{\text{No of falsely accepted users}}{\text{Total Number of users}} = \frac{4}{25} = 0.16$

FRR_1 = False Rejection Rate (After first authentication state) $= \frac{\text{No of falsely rejectted users}}{\text{Total Number of users}} = \frac{2}{25} = 0.08$

FAR_2 =False Acceptance Rate (After first authentication state) $= \frac{\text{No of falsely accepted users}}{\text{Total Number of users}} = \frac{1}{25} = 0.04$

Percentage improvement in efficiency by two factor authentication system$= \frac{FAR_1 - FAR_2}{FAR_1} * 100 = 75\%$

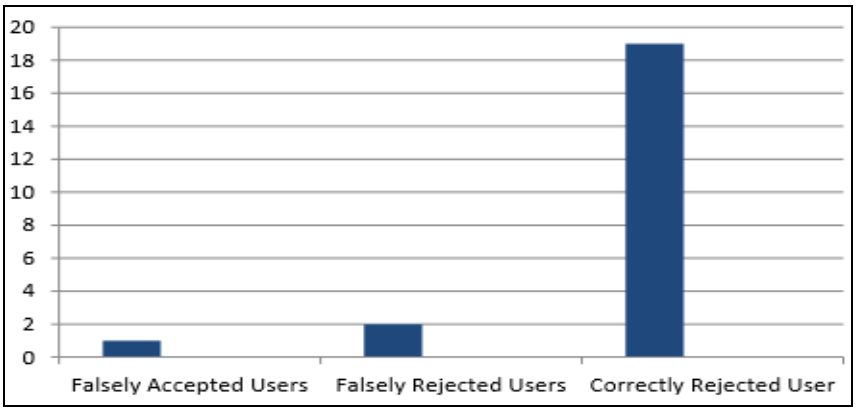

Fig 3

Table 2

Sr No.	Actual USER	Test USER	Access Allowed after Second authentication state	Difference
1	AD2	P3	N/A	0.2683
2	AD2	P2	N/A	0.1073
3	AD2	P1	N/A	0.1093
4	AD2	M3	N/A	0.1172
5	AD2	M2	N/A	0.2382
6	AD2	M1	N/A	0.2672
7	AD2	AD1	YES	<0.1
8	AD2	AD3	YES	<0.1
9	AD3	P3	N/A	0.28
10	AD3	P2	N/A	0.1339
11	AD3	P1	N/A	0.1355
12	AD3	M3	N/A	0.1419
13	AD3	M2	N/A	0.2513
14	AD3	M1	N/A	0.279
15	**M4**	**M6**	**N/A**	**0.1876**
16	**M4**	**M5**	**N/A**	**0.2359**
17	N1	N2	YES	<0.1
18	N1	N3	YES	<0.1
19	N2	N3	YES	<0.1
20	N1	P4	NO	<0.1
21	N1	P5	YES	<0.1
22	N1	P6	NO	<0.1
23	N1	M4	N/A	0.2355
24	N1	M5	NO	<0.1
25	N1	M6	N/A	0.1425

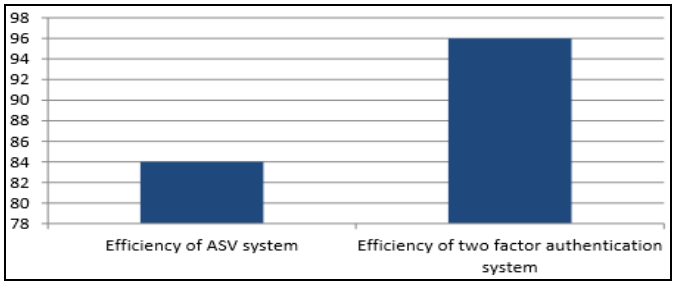

Fig 4: Improve the accuracy of ASV system

6. Conclusion

The proposed approach of enhancing the security of Automatic Speaker Verification System based on MFCC by addition an Artificial Intelligence as Two-Factor Authentication is successfully implemented. The proposed two factor authentication considerably improves the accuracy of ASV system. The developed application is tested in real-time environment. The achieved accuracy is 96 percent. The developed model can stop multiple voice spoofing attacks.

7. References

1. Raghvendra Priyam, Rashmi Kumari, Dr. Prof Videh Kishori Thakur. Artificial Intelligence Applications for Speech Recognition. In Conference on Advances in Communication and Control Systems, 2013.
2. Trevor Agus R, Simon Thorpe J, Clara Suied, Daniel Pressnitzer. Characteristics of human voice processing. In Circuits and Systems (ISCAS), Proceedings of IEEE International Symposium, 2010.
3. Shaik Shafee, Dr. Anuradha B. Speaker Identification and Spoken word Recognition in Noisy Background using Artificial Neural Networks. In International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016.
4. Chandwadkar DM, Dr. Sutaone MS. Role of Features and Classifiers on Accuracy of Identification of Musical Instruments. In Computational Intelligence and Signal Processing (CISP), 2012 2nd National Conference.
5. Chao Wang, Ruifei Zhu, Hongguang Jia, Qun Wei, Huhai Jiang, Tianyi Zhang *et al*. Design of Speech Recognition System. In Third International Conference on Information Science and Technology, 2013.
6. Niladri Sekhar Dey, Ramakanta Mohanty, LChugh K. Speech and Speaker Recognition System using Artificial Neural Networks and Hidden Markov Model. In International Conference on Communication Systems and Network Technologies, 2012.
7. Md. Afzal Hossan, Sheeraz Memon, Mark Gregory A. A Novel Approach for MFCC Feature Extraction. In Signal Processing and Communication Systems (ICSPCS), 4th International Conference, 2010.
8. Qingyang Hong, Caihong Zhang, Xiaoyang Chen, Yan Chen. Embedded Speech Recognition System for Intelligent Robot. In Mechatronics and Machine Vision in Practice, 2007.
9. Yuan Xue Luping Wang, Linxuan Li Zhiqi Liu, Jialin Liu. Matlab-based Intelligent Voiceprint Recognition System. In Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control, 2016.
10. Rozeha Rashid A, Nur Hija Mahalin, Mohd Adib Sarijari, Ahmad Aizuddin, Abdul Aziz. Security System Using Biometric Technology: Design and Implementation of Voice Recognition System (VRS). In Proceedings of the International Conference on Computer and Communication Engineering, 2008.
11. Reynolds DA, Heck LP. Integration of speaker and speech recognition systems. In Acoustics, Speech, and Signal Processing, 1991.