VOICE BIOMETRIC IDENTITY AUTHENTICATION MODEL FOR IOT DEVICES

Salahaldeen Duraibi[1], Frederick T. Sheldon[2] and Wasim Alhamdani[3]

[1]Computer Science Department University of Idaho Moscow, ID, 83844, USA
[1]Computer Science Department Jazan University Jazan, 45142, KSA
[2]Department of Computer Science University of Idaho, Coeur d'Alene, ID 83814, USA
[3]Department of Computer and Information Sciences, University of the Cumberlands, Williamsburg, KY, 40769, USA

ABSTRACT

Behavioral biometric authentication is considered as a promising approach to securing the internet of things (IoT) ecosystem. In this paper, we investigated the need and suitability of employing voice recognition systems in the user authentication of the IoT. Tools and techniques used in accomplishing voice recognition systems are reviewed, and their appropriateness to the IoT environment are discussed. In the end, a voice recognition system is proposed for IoT ecosystem user authentication. The proposed system has two phases. The first being the enrollment phase consisting of a pre-processing step where the noise is removed from the voice for the enrollment process, the feature extraction step where feature traits are extracted from user's voice, and the model training step where the voice model is trained for the IoT user. And the second being the phase verifies whether the identity claimer is the owner of the IoT device. Based on the resources limitedness of the IoT technologies, the suitability of text-dependent voice recognition systems is promoted. Likewise, the use of MFCC features is considered in the proposed system.

KEYWORDS

Internet of Things, Authentication, Access control, Biometric, Voice recognition, Security, Cybersecurity

1. INTRODUCTION

Biometrics based authentication is about the automatic verification of an identity claimer using his/her physiological and behavioral traits. Using biometric authentication for securing the IoT ecosystem is a promising approach [1]. In general, biometric authentication systems involve two steps. These are the enrollment and verification of user. The two steps are discussed in Section 2. Due to the portability, stability, and privacy of the voice features, voice recognition authentication has attracted extensive attention and application in recent years [2]. Voice recognition systems are versatile, simple to use, and non-intrusive by nature. It is considered accurate and does not require specialized tools, just a smartphone is enough for remote authentication to various services. Likewise, among other biometric authentication parameters voice is the simplest and easiest unimodal to require and use for user authentication [3].

As a result, in recent years, voice recognition has attracted various technology leading companies. For example, Google has provided Android-based Trusted Voice to allow users unlock their smartphones. Saypay's mPayment consumers use a voice password to conduct transactions [2]. In addition, Google has promoted the employment of automatic speaker recognition for authenticating users in the IoT [4]. Regarding the nature of the IoT ecosystem, especially its mobile remote control, the use of voice recognition for user authentication may give

DOI : 10.5121/ijsptm.2020.9201

an overwhelming advantage [5]. In addition, the IoT ecosystem related advantage of voice biometric include requiring of small storage, ease of transmission, and non-intrusiveness [6].

In this paper, a voice recognition authentication system to be used in the IoT ecosystem is proposed. The resource limitedness of the IoT devices and remote access are taken into consideration. For example, the proposed system uses of MFCC to extract features, and Support Vector Machines (SVMs) for user verification which is fundamental to Remote Speaker Identification [7].

Voice features that can be extracted from acquired voice data can be of high-level or low-level attributes. Low-level attributes, related to the vocal tract, are derived from spectral measurements, while the high-level is derived from behavioral cues such as dialect, word usage, conversation patterns, etc. High-level attributes are difficult to extract but are less sensitive to noise [8, 9]. In this light, extraction of low-level cues is necessary for IoT user authentication.

The rest of the paper is as follows: Section 2 is the background, Section 3 presents the related work, and Section 4 discusses the research gap, Section 5 the proposed system is presented, Section 6 presents the limitations and assumptions, and Section 7 is the conclusion.

2. BACKGROUND

Interconnected environments such as machine to machine (M2M), Machine to Individual, or Individual to Individual are what make up the IoT ecosystem. In the IoT ecosystem smart objects can communicate between themselves, things can detect each other, and everything can interact with each other and with the local environment. These interconnections are facilitated with remote sensing and tracking capabilities, and every entity is provided with data transfer through the internet, Wi-FI, ZigBee, or Bluetooth. In particular, organizations may need such data for business, social, or research analytics [10]. For this reason, a vast variety of information is stored, managed, and processed. Access to these private data needs to practice secure access control. Employing conventional authentication mechanisms such as passwords is reported to have fallen short of the IoT ecosystem. Thus, biometric technology is considered as a better substitute for the protection of IoT private data [11].

Biometrics are either physiological or behavioral. The voice recognition authentication mechanism is part of the behavioral biometric schemes [12]. Based on the personal particulars of voices, researchers have proposed a number of authentication schemes that employ voice recognition. Voice recognition is “the process of automatically recognizing who is speaking based on the signals of the voice” [13]. However, voice recognition schemes commonly suffer from the issues of the owner’s voice change and the use of a recorded owner’s voice. That is, the voice of the owner may change because of environmental reasons, such as fatigue, cold, or flu. Likewise, attackers could voice record the legitimate voice owner and later use it for illegal authentications [14].

There are two main steps taken in the voice recognition process. That is voice enrollment and voice verification as depicted in Fig. 1. The former is required for determining whether the voice is a sample in the database, and the latter identifies which sample it is in the database. In the voice enrollment process, some papers claim that the process consists of four steps including data collect, feature extraction, feature template creation, and template storage. There are also some researchers who have added one more step that comes after data collection and before the feature extraction steps. They call it as pre-processing, and it aims at removing noise from the collected data. Likewise, the verification process consists of steps such as data collecting, feature extraction, template matching and matching decision. These are discussed in the following subsections:

2.1. Data Collection/Acquisition

The process of collecting voice is nothing but the digitization of the speaker's voice. This is usually accomplished by using a microphone that captures the voice at a sampling rate. Subsequently, these data are later sent to a computing device for processing. Some of the researchers refer to this process as dataset generation or data sample collection [15]. There are two main ways of collecting voice, that is, fixed text and random number string. However, systems usually use the latter where each string of numbers is set to have 8 Arabic numbers with a range of 0 to 9 [16]. This process may include a pre-processing step where noise is removed from the original voice [17].

2.2. Feature Extraction

Features are extracted from the voice data collected and pre-processed in the preceding process. These features must be robust to intrinsic variability that may cause to user’s voice distortion due to stress or diseases. In general, there is a number of techniques that are involved in extracting features from the user's voice. These may include but not limited to, Linear Prediction Cepstral Coefficients (LPCCs) and Mel-Frequency Cepstral Coefficients (MFCC) [16, 17]. The second has been employed in some research in order to overcome the issue of constrained resources and uncontrolled operating conditions that are similar to the nature of IoT technologies [18].

It is after this process where the enrollment and verification processes take different routes. For example, for the enrollment process, template creation and template storage come after the feature extraction, while in the verification process, template matching and taking decision follow the feature extraction.

2.3. Template Creation and Storage

This process involves the creation of templates from common features that correspond to its owner. Subsequently, the templates are stored in a voice recognition database. There are several databases such as VidTimit database and MEEI database.

2.4. Template Matching and Analysis

This process tries to find an exact or near-exact match between the identity claimer’s voice and the previously stored voice templates. This can be accomplished by using Fourier transforms or linear predictive coding (LPC) [19]. Subsequently, after templates are created, the system is trained on the templates. Training methods include vector quantization (VQ) which is based on LindeBuzoGray (LBG) algorithm. In addition, Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM) are also used for feature training [17].

2.5. Matching Decision

This process is to determine whether the identity claimer corresponds to the claimed identity based on the similarities of the two voices. Subsequently, the match is either rejected or accepted. In this process, two kinds of errors may happen, false-negative or false positive. False-negative means the system has failed to identify a genuine claimer, while false positive refers to granting access to a non-authorized user [20].

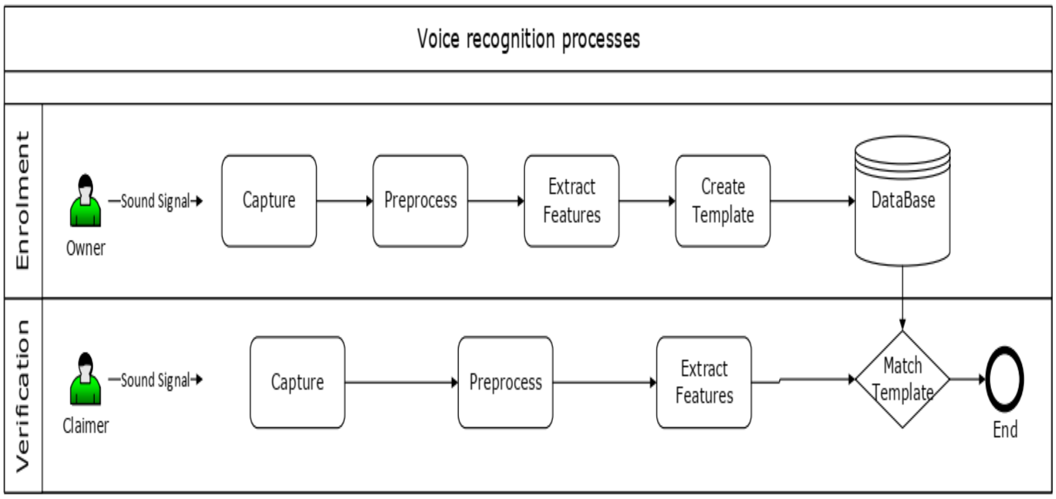


Figure 1. Voice recognition processes.

2.6. Text-Dependent

On the other hand, based on the textual contents of the speech data, voice recognition systems can be classified into two categories; Text-dependent, the identity claimer is expected to produce the same words as those pronounced during the enrollment; in this method, the speaker has to satisfy two conditions, knowing the word and being the rightful owner of the voice [2]. Text-independent, the user can speak freely during enrollment and verification phases [21]. Most of the research claims that Text-dependent recognition systems have better performance and are simpler compared to the Text-independent systems [22]. Thus, in connection to the resource limitedness of smart devices Text-dependent voice recognition approach would better fit for the IoT authentication.

2.7. Evaluation Metrix

To measure the effectiveness of voice recognition systems a number of parameters are studied. These parameters include the false acceptance rate (FAR) which is the number of attacks being incorrectly labelled as authentic by the system. False rejection rate (FRR) refers to the number of authentic interactions being incorrectly rejected as attacks. Relative operating characteristics (ROC) represents a compromise between FAR and FRR. It helps systems minimize both FAR and FRR [23].

3. RELATED WORK

In general, biometric based authentication systems employed in the IoT ecosystem are of two types. Human physiology for instance, face, eyes, fingerprint or electrocardiogram. And behavioral features such as signature, voice, gait, or keystroke. For example, the researchers in [24] introduced a gaze feature based model which is secure against iterative and side channel attacks. Likewise, in [25] the researchers used electrocardiogram for the development of their method in which they proved the good candidacy of biometric features for authentication of IoT devices. The result of the implementation of this scheme reveal that it has 1.41% FAR and 81.82% TAR for 4 seconds of signal acquisition. One of the main strengths of the scheme is that it conceals the biometric features during authentication, but privacy preservation mechanism is not taken into consideration.

The researchers argue in [26] the suitability of signature based authentication systems for the IoT devices whereby they presented three categories of signature based scheme namely, offline, online and behaviour. Some Gait recognition based authentication systems proposed for IoT devices are also in the literature [27]. A touch screen based authentication scheme is proposed in [28].

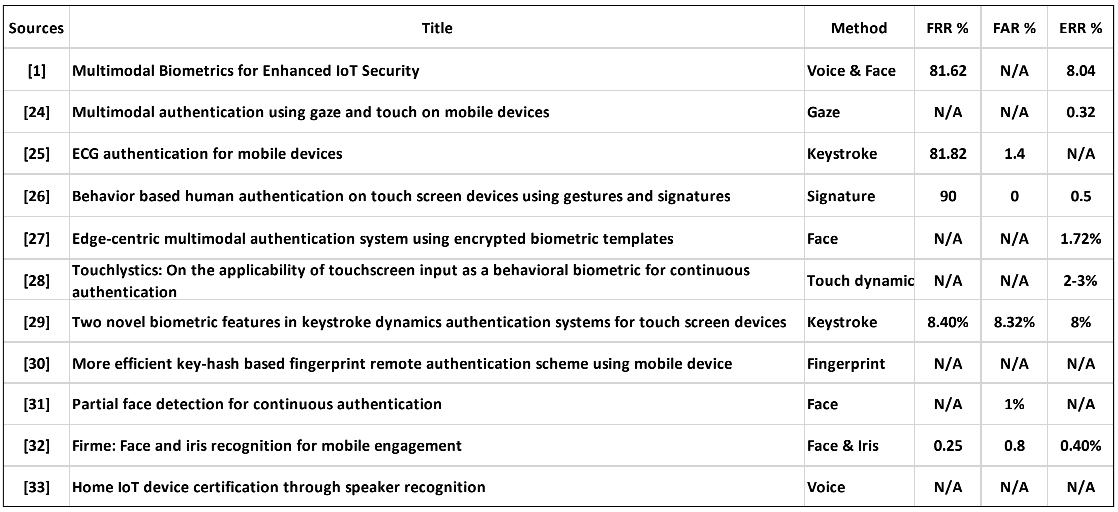
In [29], a keystroke dynamics based authentication scheme with three steps enrollment, classifier and user authentication is proposed. Similarly, in [30] a fingerprint based authentication system is provided. In [31], the researchers have introduced an authentication and authorization scheme that uses face recognition which can be used for IoT ecosystem. Iris based authentication system used for unlocking mobile IoT devices is proposed in [32].

To the best of our knowledge, there are only two researchers who adopted voice biometrics as an authentication mechanism for the IoT ecosystem. Shin and Jun [33] have implemented voice recognition technology to verify authorized users for controlling and monitoring an automated home environment. The researcher proposed a voice recognition system that is divided into server and device parts. The role of the server part of the system is for user preregistration, user recognition, and control command analysis. The role of the device part is device command reception and device controlled then response. The type of models and techniques employed in this research is not discussed. Likewise, the implementation of the model is not reported.

Oscar et al. [1] have proposed a multimodal biometric approach for IoT based on face and voice modalities. The researchers have designed their system in order to scale to the limited resources of IoT technologies. For the voice recognition part of the system, the researchers were able to extract MFCC features from voice with the use Fourier transform. In the light of this, the filter banks are decorrelated with the application of a discrete cosine transform. This system is not fully utilizing voice recognition. Although, it has been implemented in a case study, yet the end result of this model cannot be compared to a system that fully utilizes voice recognition.

The overall advantages of such biometric based schemes are that cannot be lost, they are very difficult to copy, they are hard to distribute, and they cannot be easily guessed. Conversely, conventional password-based authentication methods are suffering from a number of drawbacks and can be easily guessed, hacked and cracked. The performance of the reviewed biometric systems is shown in Table 1.

Table 1. Summary of Systems performance



Sources	Title	Method	FRR %	FAR %	ERR %
[1]	Multimodal Biometrics for Enhanced IoT Security	Voice & Face	81.62	N/A	8.04
[24]	Multimodal authentication using gaze and touch on mobile devices	Gaze	N/A	N/A	0.32
[25]	ECG authentication for mobile devices	Keystroke	81.82	1.4	N/A
[26]	Behavior based human authentication on touch screen devices using gestures and signatures	Signature	90	0	0.5
[27]	Edge-centric multimodal authentication system using encrypted biometric templates	Face	N/A	N/A	1.72%
[28]	Touchlystics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication	Touch dynamic	N/A	N/A	2-3%
[29]	Two novel biometric features in keystroke dynamics authentication systems for touch screen devices	Keystroke	8.40%	8.32%	8%
[30]	More efficient key-hash based fingerprint remote authentication scheme using mobile device	Fingerprint	N/A	N/A	N/A
[31]	Partial face detection for continuous authentication	Face	N/A	1%	N/A
[32]	Firme: Face and iris recognition for mobile engagement	Face & Iris	0.25	0.8	0.40%
[33]	Home IoT device certification through speaker recognition	Voice	N/A	N/A	N/A

4. RESEARCH GAP

Only few researches have been done on the areas related to deployment voice recognition systems to the IoT ecosystem for access control and user authentication. Building a working voice recognition system or integrating it to the IoT ecosystem is lacking in the literature. However, there are some sufficient projects done in the area of voice recognition in general. Some are adopted to the mobile and cloud computing paradigms. The challenges of IoT devices' restricted computational, storage and power resources are threatening development of sophisticated authentication systems. Hence, a novel biometric approach has to be proposed [11, 34].

5. OUR WORK

We envision an automatic voice biometric authentication system that would be suitable for managing and monitoring IoT devices from remote. As discussed previously, our model will have a training or an enrollment phase and a verification or authentication phase (Fig.2). The following section broadly discuss different components of the model.

5.1. Enrollment Phase

Sound capture

This step captures the sound or voice of the IoT device owner for training. This is expected to be done by the smartphone where the owner uses control apps of the IoT devices. The output of this step is converted files with a suitable file format.

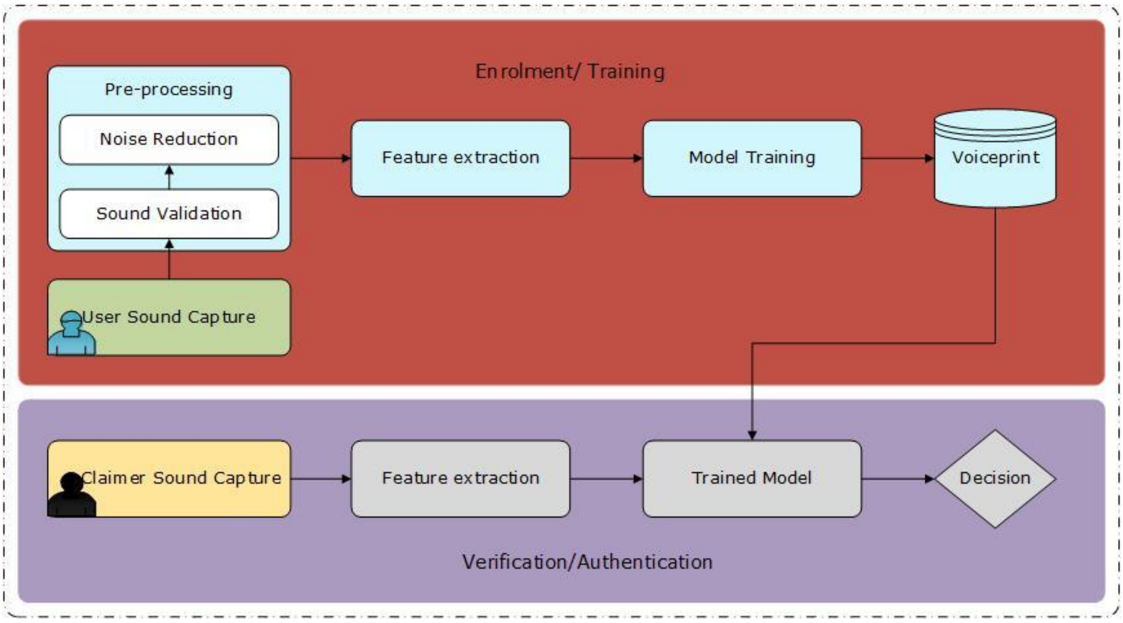


Figure 2. IoT voice authentication model.

Pre-processing

In this step, the collected voice data is validated for defects. This is accomplished by decomposing the data at different frequencies at different scales. And the resulting wavelet are checked for existence of any clipping. Subsequently, identified noise wavelets are removed and the noise free data is obtained. There are two ways of removing noise from collected sound data. That is with the use threshold based de-noising method [35, 36, 37] and recursive least squares adaptive filtering method [38, 39]. However, the first is adopted in our work for its suitability of smart devices.

Feature extraction

Voice features deemed important for the system are extracted in this step. The extraction and selection of such feature vectors adds to the quality of the of the voice recognition system. Feature traits extracted from owner's voice are expected to be different that of others, must be robust to noise and distortion, should be easily extractable, difficult for playback attacks, and should not change with the change of environment or health of the owner. As such, the most appropriate features used in this model are MFCC features. The MFCC coefficient is selected for its computing simplicity which is suitable for resource constrained characteristics of the IoT ecosystem. And it's mimicking nature of human auditory of the human ears. Likewise, the MFCC divides the voice signal into frames by subsequently applying a hamming window for every frame [40].

There are two well-known signal analysis tools that are used in existing voice recognition systems. Discrete cosine transformation and Hidden Markov Model Toolkit (HTK). Hence, in this proposed system these tools will take care the details of obtaining the cepstral features of each frame.

Model Training

After extraction of the MFCC features, the voice model is trained for the IoT owner. The HMM model is adopted for this system. The reason is HMMs are considered very effective for phones because the system app is to be used on a smartphone. Finally, the Voiceprint are stored in database.

5.2. Verification/recognition phase

Once the user enrollment phase is accomplished the system is now expected to verify whether the identity claimer is the owner of the IoT device. The same steps of voice data collection and feature extraction are conducted to the claimer's sound via smartphone. Subsequently, the extracted MFCC features are tested against the trained model for verification. Support vector machines (SVMs) are used in this step for training classifiers in order to provide a good generalization to automatically determine the verification data from the enrollment data. And finally, the decision is made for either rejection or acceptance. The authentication is rejected if the claimer's voice features fail to pass the test against the trained model.

6. LIMITATIONS AND ASSUMPTIONS

One of the main limitations of this work is that the model is conceptual and not yet implemented in its intended environment. Most of the tools and algorithms proposed or promoted for the use in the system are not technically evaluated too. Authors focused on the resource constrained nature of the IoT technology and proposed different tools for that aspect. The robustness and resilience of the tools are not thoroughly studied scientifically as well. Nevertheless, all these limitations will be handled in our upcoming research contributions.

7. CONCLUSIONS

To prevent unauthorized users from accessing IoT ecosystem, behavioral biometrics authentication systems are considered most. Through voice recognition, it is believed that IoT user authentications will be more secure, accurate and robust. Hence, in this paper we proposed a text dependent voice recognition system for IoT ecosystem. The system consists of two phases: the enrollment phase where the user is supposed to enroll the voice, and the verification of authentication phase where the identity claimer is expected to utter the voice and subsequently compared with enrolled one. In the future, we plan to develop and test the system for its security and performance in comparison to other biometric schemes proposed for the same area. Furthermore, combining various techniques that have been reviewed in this paper, we will optimize the usability for voice feature extraction and recognition. We will also consider using them in cloud for improved computational requirements.

REFERENCES

[1] Olazabal, O., et al. Multimodal Biometrics for Enhanced IoT Security. in 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). 2019. IEEE.

[2] Ren, Y., et al., Replay attack detection based on distortion by loudspeaker for voice authentication. Multimedia Tools and Applications, 2019. 78(7): p. 8383-8396.

[3] Nainan, S. and V. Kulkarni. Performance evaluation of text independent automatic speaker recognition using VQ and GMM. in Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies. 2016. ACM.

[4] McLaren, M., Automatic Speaker Recognition for Authenticating Users in the Internet of Things. August 26, 2016.

[5] Gupta, S. and S. Chatterjee, Text dependent voice based biometric authentication system using spectrum analysis and image acquisition, in Advances in Computer Science, Engineering & Applications. 2012, Springer. p. 61-70.

[6] Kolkata, C., About Voice biometric and Speaker Recognition. 2014.

[7] Thakur, A.S. and N. Sahayam, Speech recognition using euclidean distance. International Journal of Emerging Technology and Advanced Engineering, 2013. 3(3): p. 587-590.

[8] Petrovska-Delacrétaz, D., A. El Hannani, and G. Chollet, Text-independent speaker verification: state of the art and challenges, in Progress in nonlinear speech processing. 2007, Springer. p. 135-169.

[9] Rosenberg, A.E., F. Bimbot, and S. Parthasarathy, Overview of speaker recognition, in Springer Handbook of Speech Processing. 2008, Springer. p. 725-742.

[10] Yassine, A., et al., IoT big data analytics for smart homes with fog and cloud computing. Future Generation Computer Systems, 2019. 91: p. 563-573.

[11] Ferrag, M.A., L. Maglaras, and A. Derhab, Authentication and Authorization for Mobile IoT Devices Using Biofeatures: Recent Advances and Future Trends. Security and Communication Networks, 2019. 2019.

[12] Hamidi, H., An approach to develop the smart health using Internet of Things and authentication based on biometric technology. Future generation computer systems, 2019. 91: p. 434-449.

[13] 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.

[14] Su, X., et al., Study to improve security for IoT smart device controller: drawbacks and countermeasures. Security and Communication Networks, 2018. 2018.

[15] Li, D., J. Wang, and Y. Yang. PVD: A new pathological voice dataset for intra-speaker recognition research interest. in 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP). 2016. IEEE.

[16] Zhang, X., et al. Single Biometric Recognition Research: A Summary. in Proceedings of the 6th International Conference on Information Technology: IoT and Smart City. 2018. ACM.

[17] Brunet, K., et al. Speaker recognition for mobile user authentication: An android solution. 2013.

[18] Gofman, M., et al. Multimodal biometrics via discriminant correlation analysis on mobile devices. in Proceedings of the International Conference on Security and Management (SAM). 2018. The Steering Committee of The World Congress in Computer Science, Computer.

[19] Gbadamosi, L., Voice Recognition System Using Template Matching. International Journal of Research in Computer Science, 2013. 3(5): p. 13.

[20] Thullier, F., B. Bouchard, and B.-A. Menelas, A Text-Independent Speaker Authentication System for Mobile Devices. Cryptography, 2017. 1(3): p. 16.

[21] Anwar, M.U., Design of an enhanced speech authentication system over mobile devices. 2018.

[22] Khitrov, A. and K. Simonchik, System for text-dependent speaker recognition and method thereof. 2019, Google Patents.

[23] Oak, R., A Literature Survey on Authentication Using Behavioural Biometric Techniques, in Intelligent Computing and Information and Communication. 2018, Springer. p. 173-181.

[24] Khamis, M., et al. Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. in Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016.

[25] Arteaga-Falconi, J.S., H. Al Osman, and A. El Saddik, ECG authentication for mobile devices. IEEE Transactions on Instrumentation and Measurement, 2015. 65(3): p. 591-600.

[26] Shahzad, M., A.X. Liu, and A. Samuel, Behavior based human authentication on touch screen devices using gestures and signatures. IEEE Transactions on Mobile Computing, 2016. 16(10): p. 2726-2741.

[27] Ali, Z., et al., Edge-centric multimodal authentication system using encrypted biometric templates. Future Generation Computer Systems, 2018. 85: p. 76-87.

[28] Frank, M., et al., Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. IEEE transactions on information forensics and security, 2012. 8(1): p. 136-148.

[29] Tasia, C.J., et al., Two novel biometric features in keystroke dynamics authentication systems for touch screen devices. Security and Communication Networks, 2014. 7(4): p. 750-758.

[30] Khan, M.K., S. Kumari, and M.K. Gupta, More efficient key-hash based fingerprint remote authentication scheme using mobile device. Computing, 2014. 96(9): p. 793-816.

[31] Mahbub, U., et al. Partial face detection for continuous authentication. in 2016 IEEE International Conference on Image Processing (ICIP). 2016. IEEE.

[32] De Marsico, M., et al., Firme: Face and iris recognition for mobile engagement. Image and Vision Computing, 2014. 32(12): p. 1161-1172.

[33] Shin, D.-G. and M.-S. Jun. Home IoT device certification through speaker recognition. in 2015 17th International Conference on Advanced Communication Technology (ICACT). 2015. IEEE.

[34] Atwady, Y. and M. Hammoudeh. A survey on authentication techniques for the internet of things. in Proceedings of the International Conference on Future Networks and Distributed Systems. 2017. ACM.

[35] Sahoo, T.R. and S. Patra, Silence Removal and Endpoint Detection of Speech Signal for Text Independent Speaker Identification. International Journal of Image, Graphics & Signal Processing, 2014. 6(6).

[36] Zhang, X., et al. Voice Biometric Identity Authentication System Based on Android Smart Phone. in 2018 IEEE 4th International Conference on Computer and Communications (ICCC). 2018. IEEE.

[37] Moussa, A.N., N.B. Ithnin, and O.A. Miaikil. Conceptual forensic readiness framework for infrastructure as a service consumers. in 2014 IEEE Conference on Systems, Process and Control (ICSPC 2014). 2014. IEEE.

[38] Dhakal, P., et al., A Near Real-Time Automatic Speaker Recognition Architecture for Voice-Based User Interface. Machine Learning and Knowledge Extraction, 2019. 1(1): p. 504-520.

[39] Moussa, A.N., et al. A Consumer-Oriented Cloud Forensic Process Model. in 2019 IEEE 10th Control and System Graduate Research Colloquium (ICSGRC). 2019. IEEE.

[40] Moussa, A.N., N. Ithnin, and A. Zainal, CFaaS: bilaterally agreed evidence collection. Journal of Cloud Computing, 2018. 7(1): p. 1.