

# Linking biometric voice identity with self-monitoring health data as a temporal-spatial event stored in a mobile device

Bon Sy<sup>1,2,3</sup> <sup>a</sup>

<sup>1</sup>Graduate Center/City U. of NY, 365 5<sup>th</sup> Ave, NY NY 10016, USA

<sup>2</sup>Queens College/City U. of NY, 65-30 Kissena Blvd, Queens NY 11367, USA

<sup>3</sup>SIPPA Solutions, 42-06A Bell Blvd, Queens NY 11361, USA

[bsy@sippasolutions.com](mailto:bsy@sippasolutions.com)

**Keywords:** Biometric authentication; Self-health monitoring; Secure computation in mobile devices.

**Abstract:** The goal of this research is to investigate a biometric solution that links biometric personal identity to self-monitoring data, with time and location information, as a temporal-spatial event in a personal health record stored in a mobile device. The proposed biometric solution is based on a secure computation technology that reconstructs a cryptographic key for (un)locking personal health record in real time when a verification sample is sufficiently similar to the enrollment sample --- whereas the verification process is based on a secure two-party security computation that compares the enrollment and verification samples without either party sharing the data with each other, nor relying on a trusted third party. The contribution of this research is to demonstrate the practical feasibility of the approach in a resource constrained mobile computing environment. The significance of this research is its potential application for enabling a safe bubble space for social interaction among individuals who have self-monitoring data showing lack of Covid-19 symptoms at a specific time and location.

## 1 INTRODUCTION

The main research goal is to investigate a scheme for linking biometric identity to self-monitoring health data with time and location information in a mobile computing environment. The contribution of this research is to demonstrate --- in an edge resource constrained mobile computing environment --- the feasibility of (a) biometric voice feature extraction and verification, and (b) a secure computation technique for cryptographic key (re)generation based on personal biometrics with privacy protection. The significance of this research includes (i) a biometrically enabled cryptographic solution that guarantees security and privacy assurance since neither the cryptographic key nor personal biometric information is stored/shared at rest or in-transit, and (ii) a m-Health solution promoting individual health self-monitoring via IoMT (Internet of Medical Things) in a mobile computing environment that also enables a safe bubble space for work place re-opening in the event of Covid-19.

Covid-19 has caused lockdown and has taken economy down with it in many countries (Nicola,

2020). At the same time, mental health has increasingly been a concern due to the public health practice on social distancing, isolation, and quarantine (Pfefferbaum, 2020). While contact tracing (Yap, 2020) could be a good incidence response safeguard, it is a reactive approach. To streamline operational workflow process, reopening from a Covid-19 lockdown requires proactive self-health monitoring for public health safety; e.g., health monitoring is explicitly stated in the guidance on returning to work by Occupational Safety and Health Administration in the United States (OSHA, 2020). In order to create a safe bubble space, an individual should satisfy three criteria:

- (1) A self-health monitoring result in a personal health record showing lack of Covid-19 symptoms such as fever and low oxygen saturation level;
- (2) The result of self-monitoring should be timely; e.g., a self-monitoring record is valid if it contains consecutive negative test results for the most recent 7 (or 14) days;
- (3) Location self-reporting, together with a verifiable self-monitoring record with a

<sup>a</sup> <https://orcid.org/0000-0001-8827-2702>

date/time stamp, constitutes a temporal-spatial event for contact tracing purpose.

There are two security and privacy questions related to the three criteria just mentioned:

- (a) Self-health monitoring data should be automatically captured, time stamped, and transferred from an electronic monitoring device such as Bluetooth enabled thermometer or Pulse Oximeter to update a personal health record. Using electronic monitoring devices removes the uncertainty on subjective self-assessment. But how do we ascertain data sharing with security and privacy protection?
- (b) In linking an identity to a self-reporting record with self-monitoring data for creating a temporal-spatial event, how do we preserve the integrity and assure non-repudiation in data sharing?

In section 2 we will discuss the related work in biometric voice authentication and present a real world use case scenario to motivate this research formulation, as well as the assumption on the operational environment of an end user. The security and privacy risk will be discussed, as well as the state-of-the-art and the best practice. In section 3 a secure computation technology to enable privacy preserving biometric verification reported elsewhere (Sy, 2012) will be summarized. In section 4 the design and implementation of the proposed system in a mobile environment will be given. In section 5 the result of a preliminary evaluation for informing the feasibility of biometric voice will be shown. This will be followed by a discussion on the lesson learned in section 6, which include comparative analysis and security analysis. In section 7 this paper will be concluded with our future research plan.

## 2 RELATED WORK

### 2.1 Literature review

In terms of security and privacy, this research draws on biometric and cryptographic technologies. Hao et al. (Hao 2005) were among the pioneers in successfully melding biometrics with cryptography. Clarke et al (Clarke, 2002) is among the first to survey the performance of biometric authentication on a mobile device. An interesting finding in their survey is that biometric voice is the second most preferred biometric modality (next to fingerprint) to achieve the

desired level of security for mobile devices. Parthasarathy et al (Parthasarathy, 2017) reported a study on speaker verification performance with expressive speech. It was found that the error rates strongly depend on the duration of the sentence. In particular, the error rate increases for shorter sentences (i.e., less than four seconds). Their performance result is based on i-vector scheme. I-vector scheme reduces a high dimensional Gaussian super vector into a low-dimensional vector that retains most of the high-level information of a speech segment. 39-dimensional MFCC (Mel Frequency Cepstrum Coefficient) feature vectors are then extracted from i-vector of 200 dimensions as a basis for verification based on Probabilistic Linear Discriminant Analysis (PLDA). It reports an excellent performance of ERR (Equal Error Rate) of 0.5% when speech duration is greater than 5 seconds in a laboratory environment. Sathiamoorthy et al (Sathiamoorthy 2018) reports a performance study based on speech recorded using a Close Speaking Microphone (CSM) and Throat Microphone (TM). By applying auto-associative neural network, it could achieve an EER of 7% on laboratory based clean speech, and an EER of 40% on noisy speech. A common drawback on most of the performance studies is the lack of information regarding the fail-to-acquire rate during an enrollment phase as well as in the verification phase. Fail-to-acquire (biometric sample) could occur frequently in mobile device, especially when the real world operating environment is typically noisy. In this paper, biometric voice is applied to protect self-monitoring health data stored in a mobile device for a Covid-19 use case.

### 2.2 Covid-19 scenario use case

In the United States, the policies and requirements for business and school re-opening after Covid-19 lockdown vary from time to time, as well as from one state to another (Angulo, 2020). Nonetheless, one common emphasis is safety. Currently a “quick fix” solution being adopted is a self-assessment survey to be completed and self-reported by an individual. This is primarily an honest system and it assumes the self-assessment survey response is reliable. For example, one may rely on recollection and subjective belief in answering a question “Did you have fever or experience shortness of breath in the last 14 days?” A more reliable approach is to actually conduct temperature and SPO2 measurements rather than relying on a self-assessment survey. For example, a building owner or an organization may conduct

contactless body temperature measurement for visitors and employees returning to work on-site.

In the traditional approach shown in Figure 1 (Azra, 2017), the vital signs and body temperature of an individual may be measured by oneself or a third party. The data are then (emailed or) shared with a medical profession. Under regulatory compliance, medical professions are not allowed to share the data of an individual. Therefore, an individual needs to repeat the monitoring process for each medical profession, or during a visit to a store and a building.

In the scenario where an individual is capable of self-monitoring, the scenario shown in Figure 2 is attractive because it removes the “choke point” on the workflow process and the data privacy is under the control of the individual.

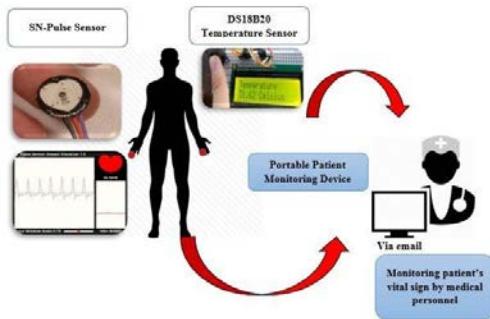


Figure 1: Traditional approach.

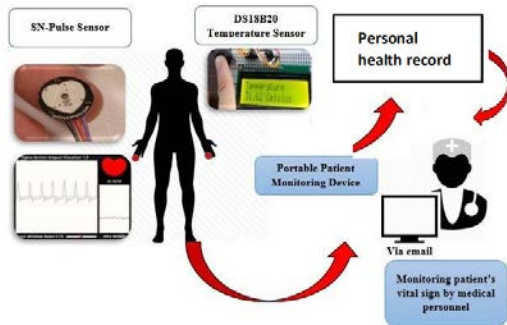


Figure 2: User owned self-monitoring & sharing.

In Figure 2, the individual retains the self-monitoring data in a personal health record, and shares the personal health record when needed. Therefore, the individual does not need to repeat the monitoring process until the self-monitoring data are expired. For the purpose of contact tracing, each self-monitoring is associated with the time stamp and the location. This constitutes a temporal-spatial event. A personal health record is a collection of temporal-spatial events. Each temporal-spatial event could then be further labelled as “active” or “expired” if such an

event may be used for determining compliance on health monitoring.

Nowadays consumer grade monitoring devices with FDA approval or CE mark are available. Many such devices support Bluetooth 4.0 or above – referred to as smart Bluetooth technology. The significance of smart Bluetooth technology is the health device profile defined in ISO/IEEE 11073-20601 (ISO/IEEE 11073-20601). This provides a common standard and interoperability for data exchange based on health characteristic profiles. In addition, Bluetooth technology also supports data encryption/decryption using a common link key derived from the pairing process between two devices.

In our research, link layer encryption for data transfer between a monitoring device and the software application implemented for a mobile device is generally acceptable because self-monitoring is performed by a user in private rather than in a public space. The challenge is the security and privacy protection of the monitoring data tagged with a time stamp and location information as a temporal-spatial event stored in a personal health record.

### 3 SECURE COMPUTATION

The technology for security and privacy protection is based on a secure computation technique, referred to as Secure Information Processing with Privacy Assurance – SIPPA.

SIPPA is a two-party secure computation for two untrusted parties to compare private data without sharing it (Prakash, 2012). The key technical properties of SIPPA are outlined below:

*There are two parties P1 (Client) and P2 (Server). P1 and P2 have private data D1 and D2 respectively. Without the presence of a trusted third party, P1 and P2 would like to know whether D1 and D2 are sufficiently similar. And if so, P1 could derive an estimate of D2 under the following two conditions:*

1. *P1 and P2 have to first find out whether D1 and D2 are sufficiently similar without either party disclosing the private data to another party.*
2. *If D1 and D2 are sufficiently similar, P1 can derive an estimate of D2 (call it D2'), without P2 ever sending D2. The only data that P2 will send P1 is some helper data with negligible overhead, where P2 can control the level of accuracy in D2' through the helper data that it sends to P1.*

The specific use case of SIPPA in this research is to provide these security and privacy properties:

- For privacy protection, biometric identity of a user is never stored in plain.
- A cryptographic key for security protection is never stored. It is regenerated in run-time when a user could produce a biometric sample sufficiently similar to the enrollment sample.

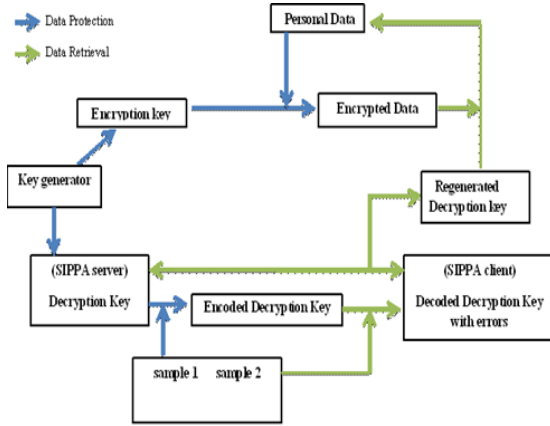


Figure 3: SIPPA workflow.

The basic concept behind the application of SIPPA is the following. During the “enrollment” process, a user  $P1$  will generate a random seed  $N$  and a cryptographic key  $K$ , and will present a biometric sample  $T$ .  $N$ ,  $K$  and  $T$  are in form of a vector of values from a finite integer field.  $N+K$  is computed and is sent to  $P2$  for enrollment.  $P1$  retains only  $N$  and  $N+K+T$ . Note that  $P2$  could not derive  $K$  from  $N+K$  without knowing  $N$ . Similarly,  $P1$  does not possess  $K$  or  $T$ .  $P1$  can only derive  $K+T$  using  $N$  and  $N+K+T$ . In other words, neither  $K$  nor  $T$  is stored; thus eliminating the security risk on the server side ( $P2$ ), and the risk of privacy leak on the client side ( $P1$ ) even if the device storing  $N$  and  $N+K+T$  is stolen.

For reconstructing the cryptographic key  $K$ , a user will present a biometric sample  $T'$  and compute  $(N+K+T - T')$ . Then the user will engage  $P2$  in SIPPA secure computation for a private comparison between  $(N+K+T - T')$  of  $P1$  and  $(N+K)$  of  $P2$ . If the user is  $P1$ ,  $T'$  will be sufficiently similar to  $T$ . Thus,  $(N+K+T - T')$  and  $(N+K)$  will be sufficiently similar. In such a case,  $P1$  could use the helper data (condition 2 stated previously) provided by  $P2$  to reconstruct  $(N+K)$ . Upon perfect reconstruction of  $N+K$ ,  $P1$  can reconstruct  $K$  from  $(N+K)-N$ . If the user is an impersonator,  $T'$  will be different, rendering the helper data to reconstruct an error laden term  $(N+K+error)$  that prevents the reconstruction of  $K$ .

The operational workflow is shown in Figure 3. The implementation details on enabling SIPPA for the proposed use case is presented in the next section.

## 4 PROPOSED SYSTEM

### 4.1 System procedure

From an end user perspective, the operational workflow process of the proposed system consists of the following steps:

1. A user self-monitors body temperature and other vital signs such as body temperature and SPO2 through Internet of Medical Things such as Bluetooth Low Energy enabled health devices.
2. Data captured by health devices are transferred through Bluetooth link layer end-to-end encryption to the personal health record managed by this proposed system in user’s mobile device.
3. The user self-verifies against a unique patient ID through biometric voice verification. The verification result is time-stamped & tagged with a geo-location to create a temporal-spatial event.
4. The user provides a voice sample to generate a cryptographic key to encrypt the personal health record updated in step 2.

Linking biometric identity with self-monitoring health data as a temporal-spatial event is a process of composing a record consisting of four pieces of information below:

- (1) Self-monitoring data in a personal health record.
- (2) Location of self-monitoring and verification.
- (3) Date/time stamp.
- (4) Verified biometric identity.

During the process of creating a temporal-spatial event, the location information will be extracted from the GPS service of a mobile device, together with date time stamp information. Below is an example:

Temperature: 98.6 F  
Longitude: -118.3097981      Latitude: 33.8019404  
Date time stamp: 2020-07-03 15:13:07  
ID: 56491905408240

During the biometric verification process, the identity being verified will be used for generating a cryptographic key to encrypt the self-monitoring data in a personal health record. This links the identity information with the health data, and ascertains the confidentiality, integrity and ownership of the data. The end result of the linking process is a temporal-

spatial event that represents a record of the four pieces of information just discussed.

## 4.2 System design

Our focus on this research is to explore the feasibility of biometric voice verification in a mobile device. The purpose is to enable a linking process to associate the biometric identity of an individual with self-monitoring health data, and timestamp and location information as a temporal-spatial event. The SIPPA process described in the previous section could be applied to other biometric modalities as was demonstrated before (Sy, 2012). In this research the proof-of-concept prototype is implemented using biometric voice. It is because every mobile device has at least one audio channel. On the other hand, high quality fingerprint scanner and/or camera depends on the phone models. Furthermore, biometric data access is possible only if it is made available by manufacturers for integration.

### 4.2.1 Biometric modality consideration

Biometric voice is generally less accurate in comparison to other modalities such as fingerprint. To compensate this, the design and implementation strategy is to incorporate content dependent voice verification. In other words, a user could opt to rely on only acoustic signature via SIPPA secure computation, or acoustic signature with the speech content, for verification. If speech content is incorporated for verification, Google speech service is utilized to perform speech-to-text conversion, and the text content matching could be either precise, or approximate using Levenshtein distance function by normalizing the distance as an error tolerance between 0 and 1.

## 4.3 System implementation

### 4.3.1 Speech signature extraction

The process of extracting biometric voice signature from a speech sample consists of the following steps:

1. Zero crossing detection algorithm (Freeman 1989) is applied to identify and remove the silent region before and after the recording.
2. The time frame for speech processing is a 16-ms non-overlapping timeframe under the short-term stationary assumption (vlab.amrita.edu, 2011).
3. For each 16-ms time frame, the signal is pre-emphasized using a hamming window filter (Smith, 2011).

4. Mel filter bank (Sahidullah, 2012) consisting of 20 Mel filters  $S'(l)$  is used to aggregate the frequency spectrum obtained from the output of Fast Fourier Transform FFT; i.e., for each 16-ms time frame, the following is computed:

$$S'(l) = \sum_{k=0}^{N/2} S(k)M_l(k) \text{ where}$$

- $l=0 \dots 19$  is the index of the filter bank;
  - $k \rightarrow (kf_s/N)\text{Hz}$  with  $f_s$  = sampling frequency;
  - $N$  being the size of FFT,
  - $S(k)$  being the output of FFT of discretized speech samples in a 16-ms timeframe;
  - $M_l(k)$  being the  $l^{\text{th}}$  band-pass triangular filter with Mel scale that defines the center frequency and the bandwidth of the band-pass filter, and the 20 Mel filters cover the frequency range between 0 and 4000 Hz.
5. Derive the  $20 \times 1$  mean vector consisting of the mean of  $S'(l)$  for each  $l=0 \dots 19$ ; and the  $20 \times 20$  covariance matrix.

Biometric voice signature is modelled by multivariate Gaussian distribution; more specifically the mean vector and covariance matrix in step 5 above. Comparing two biometric voice signatures (S1 and S2) is then reduced to computing the average of Kullback-Leibler distance (Kullback, 1951) between S1 and S2, and that between S2 and S1. This is required because Kullback-Leibler (KL) distance is asymmetrical. In encoding the cryptographic key  $K$  as described before, only the mean vector is used for computational efficiency. But when calculating the KL distance, both the mean vector and covariance matrix are used in comparing the multivariate Gaussian models of the enrollment and the sample.

## 5 PRELIMINARY STUDY

Generating a secured temporal-spatial event relies on biometric verification using SIPPA secure computation with the process described in Figure 3. At the present time, a prototype developed to support this research is available for Android platform. The implementation in the Android platform consists of the following configuration:

Sampling rate: 8000 HZ (mono channel)  
Time frame for data processing: every 16ms  
Number of bits per sample: 16  
Compression and format: PCM, WAV  
Dynamic threshold adaptation: Enabled/Disabled as determined by user

Dynamic threshold adaptation refers to an automatic calibration process; i.e., the threshold for biometric verification will be adjusted based on the consistency of enrollment samples when there are multiple enrollment samples. In case of high inconsistency (large intra-variation) among the multiple enrollment samples, the threshold will be relaxed to lower the risk on false rejection. In case of high consistency, the threshold will be tightened to lower the risk on false acceptance (small intra-variation) among the multiple enrollment samples.

## 5.1 Experimentation design

This preliminary study is conducted by three users of different ethnicities. All three speak fluent English and one of the three is a native (American) English speaker. All three were provided a Samsung Galaxy phone for this study.

This study did not attempt to recruit a large number of participants. It is because the test environment consisted of a personal mobile device that was under the custody of a participant. The security safeguard in the event of a stolen phone will be discussed in the next section.

## 5.2 Experimental setup

Three user subjects, referred to as S1, S2 and S3, participated in the study. Each subject participated in three sessions. In the first study session, a subject was asked to enroll once and enable content dependent verification. When content dependent verification was enabled, verification was accepted only if (1) the acoustic signature is sufficiently similar, and (2) the content of a speech sample for verification matches the content of that for enrollment. In addition, each user could opt for precise match or approximate content match.

In this study, only precise match was chosen by all three subjects. Therefore the default threshold value (zero) was used as the error tolerance for comparing the enrollment and the verification samples during the verification. Thirty-two speech samples were recorded for testing true acceptance (TA) and false rejection (FR). In 16 of the 32 samples, the text content of each sample must be identical to that of the enrollment sample, while the subject is free to choose any content for utterance in the other 16 samples.

During the verification phase, those 16 samples with matching text content as that of the enrollment sample were used for content dependent verification. An additional 16 samples from the other two subjects were randomly selected for testing true rejection (TR) and false acceptance (FA). In this study, we assumed

the enrollment phrase (i.e., fixed message content) is known to the impersonator. Therefore, the study result reports the lower bound of the true rejection.

Each subject was then asked to repeat the procedure for content independent verification. In content independent verification, the verification was based on only the similarity of the acoustic signature. The similarity of the text content between the enrollment and verification was not considered.

In the second session, each subject was asked to enroll three times. This is the minimum number of enrollments that will trigger dynamic threshold adaptation. In other words, a dynamic threshold was automatically derived based on the intra-variation of the enrollment samples. The verification process similar to that of the first session was then repeated. In the event that the contents were different among the multiple enrollment samples, the content of the most recent enrollment was used during the content dependent verification.

In the third session, the procedure was identical to that of the second session. The only difference was that each subject was asked to enroll four times. By enrolling four times, a second dynamic threshold was obtained for each subject.

During a verification, a separated third-party mobile app was used to record the level of background noise since the signal-to-noise ratio could be a factor that affects the verification result.

## 5.3 Result and discussion

The results of the study are summarized in the plots. Figure 4 shows the overall performance in terms of false rejection rate (FRR) and false acceptance rate (FAR) under two different scenarios: content dependent verification and content independent verification. By aggregating the verification results of all three subjects, nine pairs of (FRR, FAR) data points under different thresholds are expected for content dependent verification, and another nine for content independent verification. It is because one data point per threshold per subject could be derived from each session and each of three subjects participated in three sessions of the experiment described before. However, figure 4 shows only seven data points. It is because there are only five distinct thresholds (instead of nine) for content dependent verification, and seven distinct thresholds (instead of nine) for content independent verification. In addition, there are five overlapping data points of content dependent and content independent verification at (FAR=1, FRR=0), and of which two overlapping at the origin (FAR=FRR=0).

Figure 5 shows the relationship between the threshold value and the false acceptance rate. In the case of content independent verification, the false acceptance rate in general increases as the threshold value increases. Since the threshold value is related to error tolerance for false acceptance, the result is expected. The result of content dependent verification also shows a similar pattern.

Furthermore, there seems to be an outlier at threshold = 0.115 at a first glance as one expects a monotonic trend. But it is noted that the relationship between FAR and threshold guarantees monotonic behavior only if the plot is for one single user. Yet Figure 5 shows the aggregated result of three users.

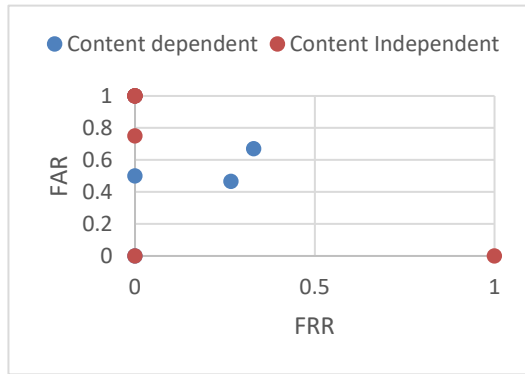


Figure 4: Overall performance.

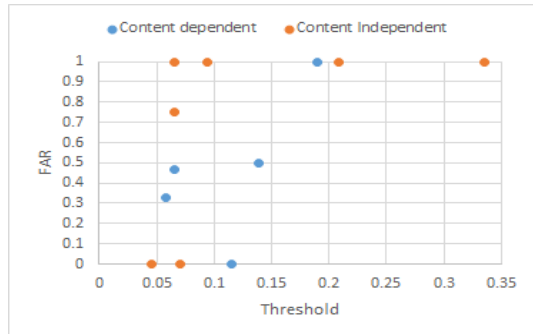


Figure 5: Threshold vs FAR.

Figure 6 shows the relationship between threshold and FRR. Figure 6 is roughly a mirror of figure 5 as expected. At approximately same threshold around 0.05, the false rejection rate of content dependent verification is better than that of content independent verification.

Figure 7 shows the relationship between threshold and FRR. But it shows the break down with respect to each user rather than showing the aggregated result as in Figure 6. It is noted that the false rejection rate is reduced to zero for subject 1 and subject 3 when threshold is increased to 0.335. However, reducing false rejection rate to zero for subject 2 occurs only when the threshold is 1. This suggests a greater intra-

variation in subject 2 when comparing to that in the other two subjects.

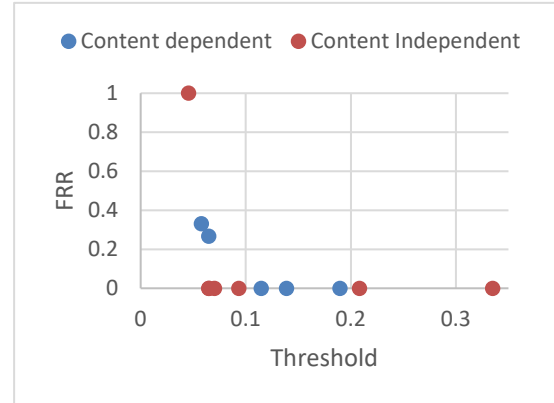


Figure 6: Threshold vs FRR.

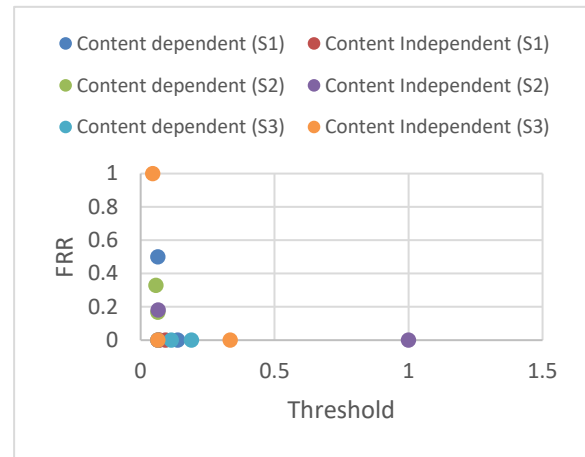


Figure 7: Threshold vs FRR per user.

Figure 8 shows the effect of the background noise and the verification error. The distribution shown in figure 8 does not show the background noise affecting the performance in terms of error.

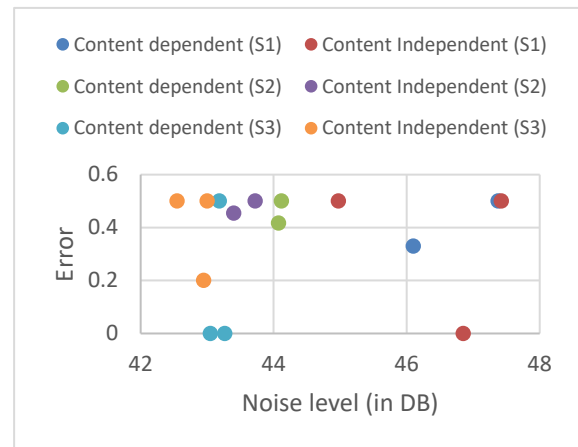


Figure 8: Relationship between noise and error.

## 6 LESSON LEARN

### 6.1 Comparative analysis

In order to compare the biometric voice verification result obtained in this research against others, the following factors should be considered:

1. Experimentation environment; i.e., the enrollment and verification are conducted based on speech samples of a laboratory-based noise-free environment or a real-world noisy environment.
2. Computing platform and biometric sensors; i.e., the experimentation is conducted on a desktop or mobile device with varying computing powers and sensors.
3. Feature extraction and verification techniques; i.e., choice of feature representation such as Mel scale or Bark scale for feature representation, matching techniques such as PLDA or associative neural network approach, as well as distance functions such as Euclidean or Kullback-Liebler distance functions.
4. Types of verification; i.e., content dependent or content independent verification.

To the best of our knowledge, available results for comparison are all based on different setups. For example, Parthasarathy et al (Parthasarathy, 2017) reported performance analysis on desktop-based environment using laboratory-based noise-free samples of expressive speech with Mel-frequency with 39 dimensions to achieve low Equal Error Rate on speech with specific durations. Since their evaluation is on given speech samples, the analysis did not include considerations on fail-to-enroll or fail-to-verify due to noisy environment exceeding the capability of microphones. Sathiamoorthy et al (Sathiamoorthy 2018) reported performance analysis by providing explicit information on the types of microphones. However, it applied associative neural network approach as opposed to PLDA, and again the verification is conducted on the “back-end” desktop environment.

In contrast to the two just discussed, our study is on mobile environment with Mel-frequency with only 20 dimensions to cover only the frequency range of normal speech conversation and the input device was the microphone of a mobile device rather than an externally added on microphone. Nonetheless, Clarke et al (Clarke, 2002) has reported performance analysis in an environment matching the environment of our experimentation. For mobile environment, it reported an ERR of 33%.

In our study, the best performance is an ERR of 0%. While the ERR is similar to that of Parthasarathy et al (Parthasarathy, 2017) and Sathiamoorthy et al (Sathiamoorthy 2018). Direct comparison is not appropriate because our approach allows personal tuning via dynamic threshold that takes into the consideration of individual inter- and intra-variations. In comparison to Clarke et al (Clarke, 2002), our performance in a noisy environment is consistent to that reported by them and others; i.e., an ERR of about 33%. Despite the experimentation is under a similar environment, one should refrain from a direct comparison since the evaluation of Clarke et al was conducted more than 15 years ago.

### 6.2 Security analysis

This proposed research on linking biometric identity with self-monitoring health data as a temporal-spatial event is secured and private under the semi-honest model. Under the semi-honest security model, a user will not deviate from the expected procedure in both the measurement and linking processes.

Without the assumption on semi-honest security model, there are two vulnerabilities. First, self-health monitoring assumes a user to not use a faulty instrument, and to not ask another person to impersonate during self-monitoring, say, temperature reading. If this assumption does not hold, the data integrity in terms of data source could be compromised. Second, creating a temporal-spatial event requires location services such as GPS for network-based location discovery. Location spoofers (Chandler, 2019) are available to fake GPS location for privacy protection. Fortunately, successful exploit on these two vulnerabilities will require a user to act maliciously, which is not an expected behaviour under the semi-honest security model.

Regarding security analysis, SIPPA secure computation is secure and private with the security and privacy properties already discussed in section 3. Recall that the linking process for generating a temporal-spatial event involves the encryption of the self-monitoring data stored in a personal health record using the cryptographic key that is (re)generated using a verified biometric identity. Since both the cryptographic key and the biometric signature are never stored in plain, the risk of such information being stolen from either the mobile device or back end server does not exist. Even if the mobile device that stores such information is stolen, one would still need a biometric sample that is sufficiently similar to the enrollment sample for recovering the cryptographic key. Therefore, both security and privacy protection are still intact.

## 7 CONCLUSION

A method for linking biometric identity to self-monitoring health data stored in a mobile device was presented. It demonstrated how SIPPA secure computation could be applied for biometric verification that guaranteed private data comparison. Verified biometric identity was then used to encrypt a record consisting of self-health monitoring data, location and time/date information. A preliminary study was conducted to gain insights into its feasibility for deployment to a mobile device. When user behaviour could be modelled as semi-honest, security and privacy assurance could be analysed and verified. Our future research will focus on an architectural solution that could extend user behaviour assumption beyond semi-honest for use cases beyond personal mobile computing environment.

## ACKNOWLEDGEMENTS

The author is indebted to the reviewers for their valuable comments that help to improve this paper. This research is conducted under the support of NSF phase 2 grant 1831214 in the United States.

## REFERENCES

- Angulo, F.J., Finelli, L., Swerdlow, D.L., 2020. Reopening Society and the Need for Real-Time Assessment of COVID-19 at the Community Level. *JAMA*. 2020; 323(22):2247–2248. doi:10.1001/jama.2020.7872
- Azra, H.A., Muhammad, M.A.J., and Radzi, A., 2017. Design and Development of Patient Monitoring System, *IOP Conference Series: Materials Science and Engineering*, Vol 226, IOP Publishing. (last access: Sept 13 2020) <https://doi.org/10.1088/1757-899x/226/1/012094>
- Chandler, N., 2019, How to Fake a GPS Location on Your Phone, Nov 2019. (last access: Sept 13 2020) <https://electronics.howstuffworks.com/cell-phone-apps/fake-gps-phone.htm>
- Clarke N.L., Furnell, S.M., Reynolds, P.L., 2002. Biometric Authentication for Mobile Devices, *Proc. of 3<sup>rd</sup> Australian Information Warfare and Security Conference*, 2002.
- Freeman, D. K., 1989. The voice activity detector for the Pan-European digital cellular mobile telephone service. *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP-89)*. pp. 369-372. doi:10.1109/ICASSP.1989.266442.
- Hao, F., Anderson, R., and Daugman, J., 2005. Combining cryptography with biometrics effectively, *University of Cambridge, Tech. Rep. UCAMCL-TR-640*.
- Kullback, S., Leibler, R.A., 1951. "On information and sufficiency". *Annals of Mathematical Statistics*. 22 (1): 79–86. doi:10.1214/aoms/1177729694. JSTOR 2236703. MR 0039968.
- ISO/IEEE 11073-20601, "11073-20601: health informatics-personal health device communication, application profile optimized exchange protocol," *IEEE 11073-20601*. <http://www.iso.org>
- Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Losifidis, C., Agha, M., Agha, R., 2020. The socio-economic implications of the coronavirus pandemic (COVID-19): A review, *International Journal of Surgery*, Vol 78, June 2020, Pages 185-193, Elsevier.
- OSHA, 2020. Guidance on Returning to Work, OSHA 4045-06-2020. (last access: Sept 13 2020) <https://www.osha.gov/Publications/OSHA4045.pdf>
- Parthasarathy, S., Zhang, C., Hansen, J., Busso, C., 2017. A study of speaker verification performance with expressive speech. *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP-17)*.
- Prakash Kumara Krishnan A., Sy, B., 2012. SIPPA-2.0 - Secure Information Processing with Privacy Assurance (version 2.0), *Proc. of the 9<sup>th</sup> Annual Conference on Privacy, Security, and Trust*, Paris, France, July 2012.
- Pfefferbaum, B., North, C. S. North, 2020. Mental Health and the Covid-19 Pandemic, *New England Journal of Medicine*, 383:510-512, August 6, 2020.
- Sahidullah, M., Saha, G., 2012. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition, *Speech Communication*, V 54, # 4, Pages 543-565, Elsevier.
- Sathiamoorthy S., Ponnusamy, R., Visalakshi R., 2018. Performance of Speaker Verification Using CSM and TM, *Asian Journal of Computer Science and Technology*, ISSN: 2249-0701 Vol.7 No.2, 2018, pp. 123-127.
- Smith, J. O. III, 2011. *Spectral Audio Signal Processing*, W3K Publishing, ISBN 978-0-9745607-3-1.
- Sy, B., Prakash Kumara Krishnan A., 2012. Generation of Cryptographic Keys from Personal Biometrics: An Illustration based on Fingerprints, *New Trends and Developments in Biometrics*, ISBN 980-953-307-576-6, InTech.
- Yap, K.Y., Xie, Q., 2020. Personalizing symptom monitoring and contact tracing efforts through a COVID-19 web-app. *Infect Dis Poverty* 9, 93. <https://doi.org/10.1186/s40249-020-00711-5>
- vlab.amrita.edu, 2011. Short Term Time Domain Processing of Speech. Retrieved 14 September 2020, <https://vlab.amrita.edu/?sub=3&brch=164&sim=857&cnt=1>