A Secure and Reliable Mobile Authentication Alternative Utilizing Hand Structure

Sahil Chatiwala*, Reva Hajarnis*, Alexei Korolev*, Danielle Park*, David Shenkerman*
Yingying Chen, and Yilin Yang****

*Governor's School of New Jersey Program in Engineering & Technology, Rutgers University–New Brunswick, NJ, USA
**Corresponding author

Emails: {sahilchatiwala, reva.hajarnis, alexeikorolev7, danielle.kim.park, shenkermandavid}@gmail.com
{yingche, yy450}@scarletmail.rutgers.edu

†*Sahil Chatiwala, Reva Hajarnis, Alexei Korolev, Danielle Park & David Shenkerman all contributed to this project equally.*

***Abstract*—As technology advances, methods to protect information on personal devices are shifting towards forms of biometric user authentication. When assessing these new modes of authentication, it is pertinent to consider their implementation of security, reliability, privacy, and convenience. Current facial and fingerprint identification technologies compromise privacy since they require the collection of personal data. On the other hand, non-biometric mobile authentication lacks security due to a preference for convenient and reliable techniques such as short passwords. Hand structure-based authentication is a solution that incorporates all of these criteria. With privacy and convenience inherent to this mode of authentication, the proposed experimental procedure leverages hand biometrics and the K-Nearest Neighbors, Support Vector Machines, and Bagged Decision Trees machine learning algorithms to examine the security-reliability trade-off. The results show that there are statistically significant properties of structure-borne signals that enable differentiation between individuals and hand positions. These properties, along with manipulation of algorithmic learning parameters, are used to enable the model to specialize either in security or reliability. While all of the developed models excelled in security, the Bagged Decision Trees model outperformed the others by rejecting false users at a peak accuracy of 99.9% and rejecting the wrong hand positions with a peak accuracy of 99.4%.**

***Keywords*— Mobile authentication, biometrics, structure-borne sound**

I. Introduction

Overview: With over 80% of the world's population using smartphones, it is crucial to explore and optimize authentication techniques that would best suit mobile users. The ideal authentication system would incorporate the following aspects: consistent blockage of false users (security), consistent authentication of the true user (reliability), minimal collection of personally identifiable data to the authenticating entity (privacy), and minimal active participation from users (convenience).

Motivation: Currently implemented methods of mobile authentication are all deficient in at least one of the four aforementioned categories, compromising the integrity of the authentication process [1]. Password/knowledge-based authentication is completely insecure if a false user acquires the password to another's device. Facial and fingerprint identification pose a privacy risk with the data they require to function and are often unreliable, such as when a user is trying to authenticate using facial identification while wearing a mask [2]. Past studies have revealed that artificial intelligence can effectively identify users from structure-borne vibrations of their hands (see Figure 1) [3]–[7]. The methodology being examined suggests that

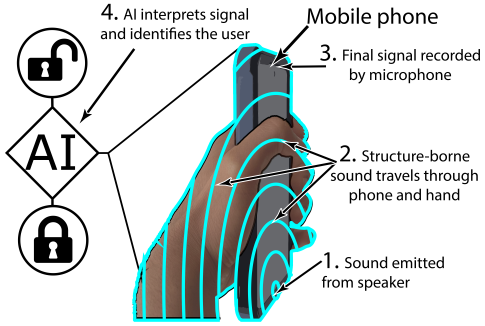

Fig. 1: A step-by-step overview of the authentication process.

hand biometrics are feasible since structure-borne signals propagate predictably even with variations in the environment [2].

Contribution: A hand structure-based authentication system that transmits high frequency audio signals from the mobile speaker and analyzes the way they interact with a user's hand is an ideal solution to the proposed problem [3], [4]. Specifically, the structure-borne signals created as the audio signals propagate through a user's hand will be perceived by the phone's microphone [6]–[8]. Privacy is built into the data collection process since users' hands are never explicitly mapped like finger contours in fingerprint identification or faces in facial identification. The system is also maximally convenient as it solely requires users to pick up their mobile device with the grip that they are accustomed to. The current research addresses the security-reliability tradeoff in hand structure-based authentication by analyzing structure-borne signals in a user's hand with artificial intelligence. To make the design resilient to various conditions and environments, we investigate the differences in users' hands and hand positions to develop models that will perceive measurable differences in the transmissions [9]. Overall, the proposed technique aims to simulate a high fidelity authentication system which maintains a reasonable degree of reliability while maximizing security.

II. Background

A. Artificial Intelligence

The role of artificial intelligence (AI) and machine learning (ML) in biometric user authentication is to create distinctions across different users' data. In the context of this research, ML can be used to analyze features of structure-borne signals to enable classification of users and hand positions. The classification algorithms used in the experiment are the supervised learning algorithms Weighted

K-Nearest Neighbors (Weighted KNNs), Gaussian Support Vector Machines (Gaussian SVMs), and Bagged Decision Trees (BDTs).

The crux of supervised learning is in the loss function which quantifies the error of a model's predictions. To minimize this loss function effectively, the learning rate, or the rate at which the model reduces it's error, must be calibrated for different scenarios. Before the models are trained, certain numerical costs are assigned for each wrong classification that the AI makes during training in MATLAB's Classification Learner's Application. These costs help inform how the learning rate will change based on the predictions made by the model at each epoch. Manipulation of this cost hyperparameter is used for experimentation with the security-reliability trade-off.

B. Statistical Processes

1) Signal Enveloping: Speakers transmit audio signals by causing a diaphragm to vibrate at certain frequencies. One physical limitation present in creating high frequency signals is that the diaphragm cannot be instantaneously accelerated to high frequencies. Similarly, the diaphragm is also unable to instantaneously decelerate between the end of one signal and the start of the next. This causes interference in a phenomena known as frequency leakage. These problems are resolved using signal enveloping, which refers to controlling the frequency range of a signal at a certain point in time. By gradually increasing the frequencies in the signal's envelope to the desired frequency, the diaphragm functions predictably and prevents frequency leakage. The specific type of enveloping used in this research is the Hann Window function (Hanning window) [10].

2) Cross-Correlation: Cross-correlation quantifies the similarity between two discrete signals by comparing each data point in one signal to each data point in the other. The cross-correlation calculation provides two outputs: firstly, the correlation coefficients calculated for each comparison, and secondly, the lag in time between points being compared. Higher correlations imply that the two points being compared are similar, while a lower correlation implies the opposite. Lag refers to the difference in time for when a data point in one signal occurs from a data point in the other signal. Using a comparison of both correlation and lag values, the points of maximum similarity between two signals can be found by identifying the highest calculated correlations. Then, using the lag values, the exact time that a similarity occurs is located [11].

3) Fast-Fourier Transform: The Fast-Fourier Transform (FFT) is an optimization of the Discrete Fourier Transform (DFT), a function that converts a discrete signal from the time domain to the frequency domain. The equation for the Fourier Transform of a signal $f(x)$ in the time domain is given by its frequency content $F(\xi)$ in the equation

$$F(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx \tag{1}$$

Specifically, the FFT makes this conversion using the constructive/destructive interference of sinusoidal waves of various frequencies that sum to the original signal [12]. The power of these constituent frequencies, or how often they occur in the complete signal, are then graphed as a function of the frequency itself to form the spectrogram that represents the Fourier transformed signal.

4) Mel Frequency Cepstrum Coefficients: Mel Frequency Cepstrum Coefficients (MFCCs) are a set of values that provide a compact representation of a signal on a subdomain of the frequency domain, known as the Mel frequency scale [13]. This scale is a modification of the frequency scale that is based on the listeners' perception on the difference between frequencies. The Mel scale is based on the notion that the difference between lower frequencies can be perceived by humans more easily than the difference in higher frequencies. MFCCs are calculated by calculating the FFT of a signal on the Mel frequency scale, and then calculating the Discrete Cosine Transform of the real logarithm of the resulting spectrogram [13]. The MFCCs of a signal are useful since they include information about both the frequency and Frequency Band Energy (FBE) of the signal.

III. Experimental Procedure

Experimentation was completed in four main phases: signal design, data collection, data preprocessing/analysis, and AI model development. The MATLAB script chirps.m contains the signal design specifications and the scripts cross-correlate.m, feature-extraction.m, and learning.m are using for data preprocessing/analysis. The MATLAB Classification Learners Application is used to devise, train, and test models. The authors of this paper participated as the five subjects in this study; their hand biometric data was collected on a single Samsung Galaxy S20 FE 5G smartphone with a Dexnor case. The subjects had varying hand shapes, sizes, and positions.

A. Signal Design

During the signal design process, the term chirp refers to an enveloped audio transmission that is 0.025 seconds long. Each chirp has a target frequency that it broadcasts for about 0.0125 seconds after the accelerative Hanning window. This target frequency is known as the frequency of the chirp. A segment is used to refer to a series of chirps with ascending frequencies, a signal is a combination of segments, and a transmission refers to the entire series of signals broadcasted. The frequency of range of the tranmission was determined using the converse of the Nyquist-Shannon Theorem, which states that accurate discretization of a continuous audio signal can be achieved by sampling at a frequency that is at least twice the greatest frequency in the signal [14]. Since the maximum sampling capability of the device used was 48 kHz, the maximum possible frequency present in a tranmission would be 24 kHz. To account for sampling errors due to possible microphone deformation, a maximum frequency of 22 kHz was assigned to each transmission with a minimum frequency of 18 kHz to guarentee inaudibility.

Using the 18kHz-22kHz frequency range and the values of the *reps_per_freq* and *reps_per_signal* variables from chirps.m, the transmission was created. For the data collected in this experiment, each segment consisted of 10 chirps with two at each frequency (*reps_per_freq* = 2) ranging from 18 kHz to 22kHz, incremented by 1 kHz. Each of these segments was also repeated four times (*reps_per_signal* = 4) meaning that each signal consisted of 40 chirps, with eight at each frequency. Finally, each signal was repeated about 20 times per transmission. The resulting transmission was about 20 seconds long by the computation below:

$$\frac{.025 \text{ s}}{\text{chirp}} \cdot \frac{2 \text{ chirps}}{\text{freq.}} \cdot \frac{5 \text{ freq.}}{\text{seg.}} \cdot \frac{4 \text{ segs.}}{\text{sig.}} \cdot \frac{10 \text{ sig.}}{\text{tran.}} = \frac{20 \text{ s}}{\text{tran.}} \tag{2}$$

The length of these transmissions was chosen to increase the ease of data collection. In reality, the authentication process would take significantly shorter since at most one signal (40 chirps) would need to be transmitted from when the user picks up their mobile device.

B. Data Collection

The Android mobile application, Amazing Voice Recorder (AVR), was used to record structure-borne data in two main parts.

1) User Data: For the user distinction part of the experiment, data was collected from four subjects who gripped the phone in the same position. For each sample collected, subjects firmly gripped the device while a microphone recording was started and the transmission was played. During the transmission, the subject maintained a firm grip and did not adjust their hand at all to reduce variability in structure-borne vibrations. This process was repeated for 10 transmissions per subject, with no adjustment of grip in between transmissions.

2) Hand Position Data: Data for three hand positions was collected from a single subject (see Figure 2). Samples were collected in an identical manner to Part 1 except that data was collected for 10 transmissions per hand position as opposed to 10 transmissions per individual. No adjustments were made in hand position or grip in between transmissions for the same hand position.

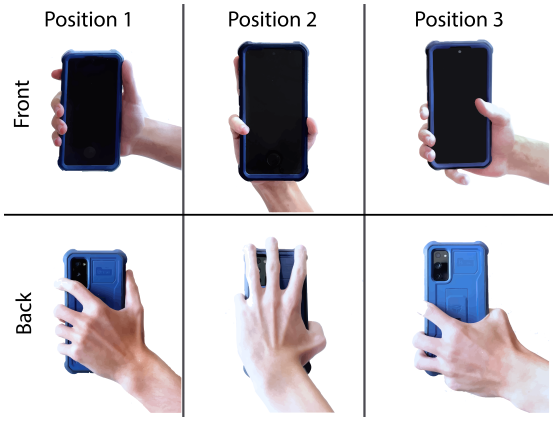

Fig. 2: Varying hand positions for Subject 5.

C. Data Preprocessing and Analysis

Data from both stages of the collection phase was preprocessed and analyzed identically. The left stereo channel file for the recording of each structure-borne transmission was first loaded into the MATLAB workspace. Then, a bandpass filter was used to eliminate frequencies outside of the 18kHz-22kHz range. This denoising step was possible because sound compressions do not change frequency when they change medium; thus the 18kHz-22kHz transmission emitted from the mobile speaker maintains its frequency as it propagates through a user's hand and into the microphone. After this filtering, the structure-borne transmissions were cross-correlated with the original transmitted audio to determine the point in the recorded audio where the structure-borne waves began to enter the microphone. Then, the beginning point of the structure-borne signal was determined from these cross-correlation coefficients. If cross-correlation overestimated the start point, the start of the structure-borne signal was determined manually. With the airborne data eliminated, the remaining structure-borne data was then segmented back into individual signals, segments, and chirps. After this post-recording segmentation, chirp data for each user and hand position was aggregated into a separate statistical profile. Each profile included 315 statistics that were either extracted or calculated from the structure-borne data, including mean amplitude, FFT variations, and MFCCs.

D. Training and Testing the AI Model

The creation of unique statistical profiles aided the machine learning process by creating clear numerical signatures for each user and hand position. Some model architectures considered included Naive Bayesian Classifiers, linear SVM, quadratic SVM, cubic SVM, Gaussian SVM, KNN, Weighted KNN, and BDTs. Ultimately, the majority of testing was done using Gaussian SVMs, Weighted KNNs, and BDTs due to their high baseline accuracies.

Authentication of users and authentication of hand positions were tested separately with separately developed models. To simulate a realistic authentication scenario, the machine learning task was framed as a binary classification problem. In each model, one statistical profile was chosen to represent the true user or true hand position, the entity that should be authenticated, while the other profiles were pooled together to simulate impersonators. Security was quantified by the accuracy of the model in denying authentication to impersonators. This value is alternatively called the false validation rate. Similarly, reliability was quantified by the ability of models to successfully authenticate the true user or hand position. For all models, the dataset for users and hand positions was split into 90% for training and 10% for testing. Additionally, k-Folds Cross-Validation was implemented on the training dataset to prevent overfitting of the model. This guaranteed that the model did not see the same data for the entire training phase and also provides the model with "new" data to evaluate on after each epoch.

IV. Results

A. User Distinction

Of the three main model architectures, BDTs performed best in distinguishing the hands of the true user as opposed to the impersonators. Optimal results with the highest security-reliability balance were achieved using the default false validation costs. The specific accuracies of this model are shown in Figure 3 below.

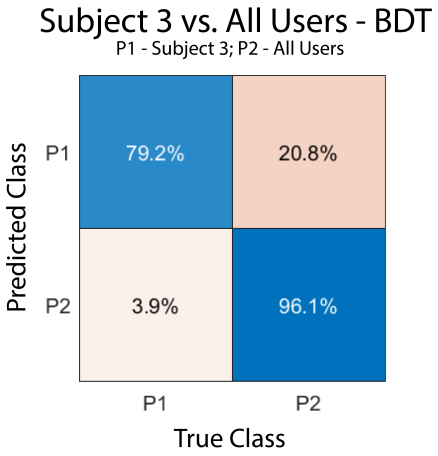

Fig 3: Confusion matrix of the optimal BDT model for user distinction. Subject 3's statistical profile was assigned as the true user (P1) and other profiles were combined to represent impersonators (P2).

When the cost proportion was skewed to more heavily penalize a false validation, the algorithm was able to improve its security. Its false validation percent fell from 3.9% to 0.1% when the false validation cost was scaled six-fold. However, the algorithm began to misclassify an authorized user more often, with the false negative rate increasing from 25.4% to 70.7%. Both properties are inversely proportional: as the security of a model increases at an approximately logarithmic rate, its accuracy drops linearly (see Figure 4).

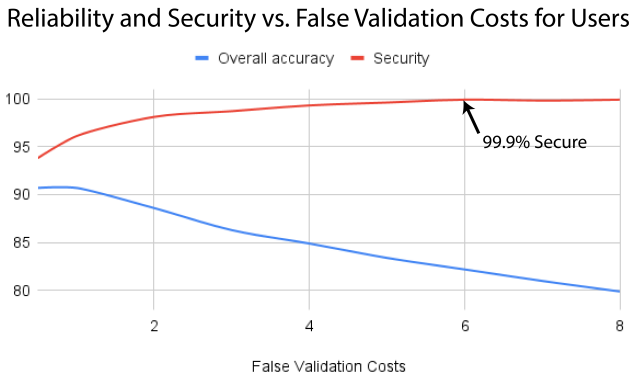

Fig. 4: Comparison of security and overall accuracy for varying false validation costs. As false validation costs increase, security increases and accuracy decreases, implying a decrease in reliability.

B. Hand Position Distinction

BDTs were also the most effective in the distinction of hand positions. Through similar adjustment of false validation costs, the model specialized in security, denying the wrong hand positions with a maximum of 99.4% accuracy. This was an increase over the 96.1% accuracy achieved with lower validation costs, where reliability of the model was higher. Figure 5 shows the specific breakdown of security and reliability in these models over both iterations.

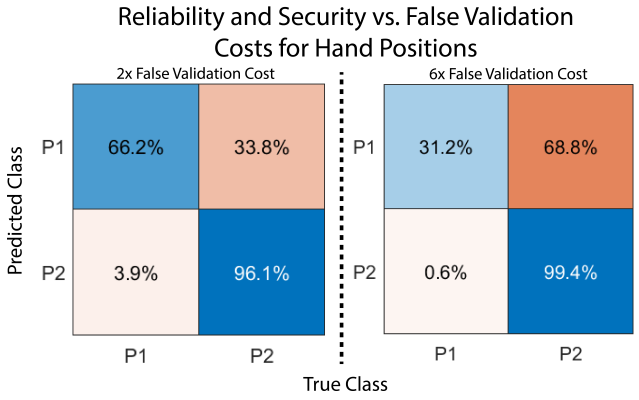

Fig 5: Confusion matrices for iterative BDT models of varying false validation costs. Hand Position 1 represents the true position (P1) while others are impersonators (P2); reliability decreases inversely with security.

Accuracy of the model overall suffered significantly with increases in learning rate adjustment for false validation costs. This is attributed to several factors, namely that structure-borne signals of different hand positions for users possess less unique characteristics than structure-borne signals of different hands. Thus, increasing security of the model resulted in a scenario where even the true hand position could not successfully authenticate the device reliably due to overfitting on these few statistical differences (see Figure 6).

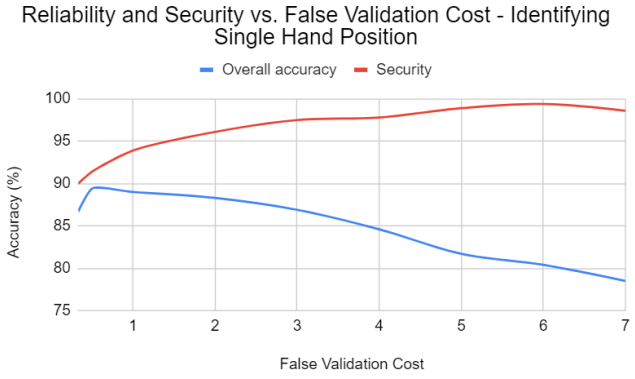

Fig 6: Line plot representing variability in security and overall accuracy for increasing false validation costs with hand positions. Security and reliability of the model diverged more aggressively for different hand positions than with different users.

C. Dataset Size Manipulation

A lot of testing was also done to determine the optimal dataset size for each model's performance in the user distinction portion. While having a larger dataset usually benefits ML models in general, there were noticeable increases in accuracy for both BDTs and Weighted KNNs due to overfitting of the models on less data. Training the models with statistical profiles in which each profile consisted of data from only one structure-borne transmission resulted in an 80.0% overall accuracy for Gaussian SVM model. This accuracy steeply declined when three transmission were used and the models overfit less, but rebounded to an even higher accuracy of >80.0% when seven transmissions were used. On the other hand, reducing the dataset size led to a significantly increased overall accuracy for both the BDT and Weighted KNN models to 93.8% and 88.8%, respectively (see Figure 7). Therefore, the Gaussian SVM model may be more suitable for applications where there is a substantial amount of diverse data per profile. Alternatively, Weighted KNNs and BDTs are more practical for implementations where there are limitation on the quantity of data that can be stored or gathered.

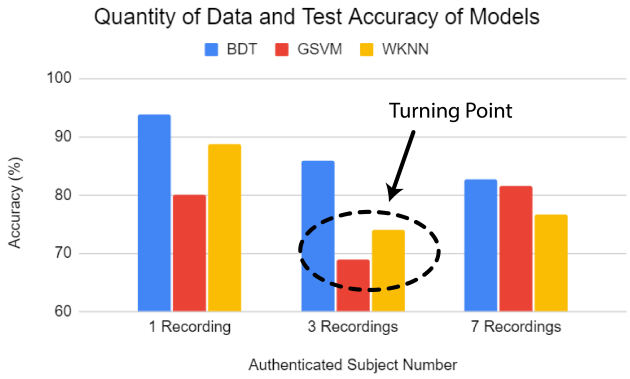

Fig. 7: Accuracy of the Gaussian SVMs (GSVM), Weighted KNNs (WKNN), and BDTs models in terms of the quantity of transmissions per profile.

D. Model Efficiency

The training time and corresponding accuracy of the models was also analyzed to determine each model's practicality and efficiency (see Figure 8). The BDTs models required the highest training time but also produced the most accurate across both parts tested. The Gaussian SVMs model was able to analyze the data quickly and provided accurate predictions about identifying users with particularly high security but low reliability. Weighted KNNs were able to achieve an accuracy slightly higher than Gaussian SVMs with a low training duration, but required a higher amount of time while testing since the entire dataset would have to be queried for every prediction. For practical purposes, the implementation of this authentication software would be best served with a model that could make quick predictions without a high storage requirement.

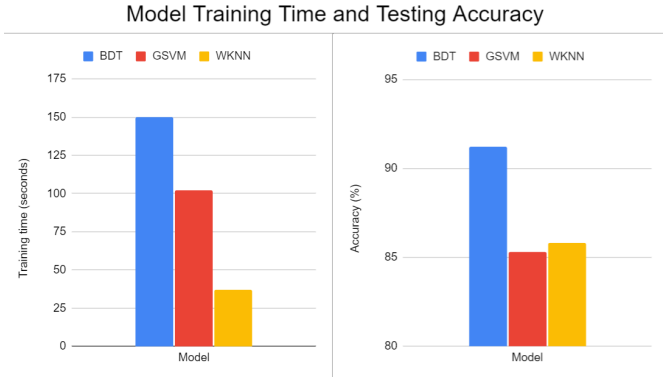

Fig. 8: BDT, Gaussian SVM (GSVM), and Weighted KNN (WKNN) models based on their training time and accuracy.

V. CONCLUSION

The research resulted in significant evidence that showed hand biometric authentication based on different users and different hand positions to be both secure and accurate. Structure-borne data was collected from five subjects using an Android smartphone, preprocessed, and analyzed. Then, this data was used to train ML models using the Gaussian SVM, Weighted KNN, and BDT algorithms. These models were able to take advantage of differences between the structure-borne signals in both of the aforementioned categories to create an effective mode of authentication. Specifically, BDTs provided high security against the authentication of false users in both the different users and hand position parts of the experiment. In contrast, Weighted KNN was a faster-performing algorithm that achieved high reliability, but declined in accuracy when a bigger data set was used. The model's learning rate was manipulated by assigning different training cost combinations; the resulting models were better suited to specialize in security as false validation costs were increased. A lower security version of the model achieved with low false validation costs would be ideal for situations with lower privacy risks, such as personally-managed system. Alternatively, for systems that serve the public, the high security models trained with higher false validation costs would be ideal to protect against malicious activity. The software repository for the collection, preprocessing, and analysis of data as well as the machine learning models is available at https://github.com/sahil485/HandStructureAuth. Further research into models that achieve higher reliability with maintained security levels for both parts would be beneficial and necessary before implementation of this authentication technique in mobile devices. Ultimately, hand structure-based mobile authentication with these improvements is a viable alternative to current modes of biometric mobile authentication.

ACKNOWLEDGMENTS

The authors of this paper gratefully acknowledge the following: Rutgers School of Engineering; Rutgers University; The NJ Office of the Secretary of Higher Education; Governor's School Alumni for their continued participation and support; Dean Jean Patrick Antoine, the Director of the Governor's School of New Jersey Program in Engineering and Technology 2022 (GSET) for his management and guidance; project mentors Dr. Yingying Chen and Teaching Assistant Yilin Yang for their direction and expertise; Residential Teaching Assistant Benson Liu for his invaluable assistance; and Research Coordinator June Lee for her advice on conducting proper research.

REFERENCES

[1] S. Arora and M. P. S. Bhatia, "Fingerprint Spoofing Detection to Improve Customer Security in Mobile Financial Applications Using Deep Learning," Arabian Journal for Science and Engineering, vol. 45, no. 4, pp. 2847–2863, Oct. 2019, doi: 10.1007/s13369-019-04190-1.

[2] L. Lu et al., "Lip Reading-Based User Authentication Through Acoustic Sensing on Smartphones," IEEE/ACM Transactions on Networking, vol. 27, no. 1, pp. 447–460, Feb. 2019, doi: 10.1109/tnet.2019.2891733.

[3] Yang, Y. Chen, Y. Wang, and C. Wang, "Echolock: Towards Low-effort Mobile User-Identification Leveraging Structure-born Echos," in ASIA CCS '20: Proceedings of the 15th ACM Asia Conference on Computer and Communications Security, Taipei, Taiwan, Oct. 2020, pp. 772–783. Accessed: Jul. 04, 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3320269.3384741

[4] C. Qiu and M. W. Mutka, "Silent Whistle: Effective Indoor Positioning with Assistance from Acoustic Sensing on Smartphones," in IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Macao, China, Jun. 2017, pp. 1–6. doi: 10.1109/WoWMoM.2017.7974312.

[5] L. Lu et al., "Lip Pass: Lip Reading-based User Authentication on Smartphones Leveraging Acoustic Signals," in IEEE INFOCOM 2018, Honolulu, HI, USA, Oct. 2018, pp. 1466–1474.

[6] M. Goel et al., "SurfaceLink: Using Inertial and Acoustic Sensing to Enable Multi-Device Interaction on a Surface," in CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Ontario, Toronto, Canada, Apr. 2014, pp. 1387–1396. Accessed: Jul. 10, 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/2556288.2557120

[7] L. Lu, J. Yu, Y. Chen, and Y. Wang, "VocalLock," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 2, pp. 1–24, Jun. 2020, doi: 10.1145/3397320.

[8] R. K. Rowe, U. Uludag, M. Demirkus, S. Parthasaradhi, and A. K. Jain, "A Multispectral Whole-Hand Biometric Authentication System," Oct. 2007, pp. 1–6. Accessed: Jul. 10, 2022. [Online]. Available: https://www.researchgate.net/publication/4311442_A_Multispectral_Whole-Hand_Biometric_Authentication_System

[9] N. Kim, J. Lee, J. J. Whang, and J. Lee, "SmartGrip: grip sensing system for commodity mobile devices through sound signals," Personal and Ubiquitous Computing, pp. 643–654, Nov. 2019, doi: 10.1007/s00779-019-01337-7.

[10] Yang, Yanli. "A Signal Theoretic Approach for Envelope Analysis of Real-Valued Signals." IEEE Access, vol. 5, 2017, pp. 5623–5630, ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7891054, 10.1109/ACCESS.2017.2688467. Accessed 31 July 2022.

[11] M. Krumin and S. Shoham, "Generation of Spike Trains with Controlled Auto- and Cross-Correlation Functions," Neural Computation, vol. 21, no. 6, pp. 1642–1664, Jun. 2009, doi: 10.1162/neco.2009.08-08-847.

[12] S. Bakheet, A. Al-Hamadi, and R. Youssef, "A Fingerprint-Based Verification Framework Using Harris and SURF Feature Detection Algorithms," Applied Sciences, vol. 12, no. 4, p. 2028, Feb. 2022, doi: 10.3390/app12042028.

[13] F. Zheng and G. Zhang, "Integrating The Energy Information Into MFCC," in International Conference on Spoken Language Processing, Beijing, China, Oct. 2000., pp. 1–4. Accessed: Jul. 14, 2022. [Online]. Available: https://doi.org/10.1109/wcica.2004.1342302

[14] Z. Song, B. Liu, Y. Pang, C9 Hou, and X. Li, "An Improved Nyquist–Shannon Irregular Sampling Theorem From Local Averages," IEEE Transactions on Information Theory, vol. 58, no. 9, pp. 6093–6100, Sep. 2012, doi: 10.1109/tit.2012.2199959.