Lips Recognition

Passwords are phrases, words, or personal identification numbers (PINs) used for the purpose of securing information. The process of authentication is simple; he who knows the correct password is considered to be a legitimate user and he does not is an imposter. An archaic but effective method in a pre-industrial revolution era, but a terrible method to adapt to a 21st century use case - smartphones. In modern times, the user cannot be completely trusted because users tend to do things the easy way. Users for one cannot be trusted to use safe and cryptographically challenging passwords or be expected to not write them down in form or another [1]. Here the abilities of biometric authentication can be shown to great effect. As biometric authentication relies on who you are i.e. something that cannot practibly be lost or stolen. There exists a variety of biometric authentication modalities, methods, and frameworks here I propose a different one.

My modality will be lips recognition. Specifically speaking I will be elaborating on an authentication method using a combination of voice, facial movement, and vibration. The intent here is firstly to remove the risk of password being stolen and social awkwardness. Since each user will have the same spoken password, that is all users will use the same phrase the same way. The spoken phrase is meant to be taking similarly to an activation phrase for the phone. Meaning saying the password will cause the phone to authentication you and the phone will be listening for your voice. Very
similar to modern IoT devices like the Amazon Alexa which can be prompted to listen with the activation phrase “Alexa” [2]. The voice data will be used in authentication, but will not have as much weight as the next two characteristics. The next of which is captured facial movement using a front-facing camera. The camera will be placed very close to the face while the user recites the password. The captured movement will then be used for authentication. And lastly using the phone’s movement sensor specifically the magnetometer the slight magnetic field changes caused by motion of the lips. Experiments have shown that a small magnetometer is all that is needed to measure the magnetic field changes from something as small as eyelid movement [3]. Magnet sensing systems (magnetometers) have in the last decade risen to popular use in the medical field for their accuracy and I intend to apply the same effect using the built-in sensors of smartphone [4]. The other intent is to be secure, for that a combination of all three will be used in the decision making process.

This modality, has been introduced around the premise that it will be used for authentication on mobile devices. Furthermore, mobile devices being a highly personal object tend to only need authentication, but that is not to say that identification is not also possible. Technically speaking, the system can be enlarged and used en masse for commercial identification in the same way fingerprint is. Identification will however be very off putting to an user, given that they will likely be putting their face very close to a sensor that many people have already used. In light of the expected negative reaction from users, the modality will therefore be confined to authentication only.
The workflow will consist of 4 steps: data capture, feature extraction, matching, and decision. Data capture will happen as described above. Feature extraction will comprise of two steps. First noise reduction must be applied to the magnetometer, microphone, and the camera input. Then a machine learning method will extract the pertinent features from each respective input separately. Afterwards matching will take place. The users query will be matched against the preexisting templates. There are multiple templates since the users facial movements and voice can change in response to their level of awareness. There will be three separate scores produced in the matching phase. Lastly, a decision will be made using the score-level fusion of the three prior scores with more weight given to the magnetometer followed by the facial movement followed by the voice input.

The system is cooperative and user friendly. The user must participate but they are allowed variance at no security cost. This does also mean the system is overt and non supervised given the exclusive and private nature of smartphones. The system is somewhat controlled as specific movements are required for authentication to happen i.e. speaking to and holding the phone in a particular way. Also the user will need some habitation due to the uniqueness of the design as any bold new system mandates.

My modality admiintly is not as universal as some other biometric modalities. People who cannot move their face, or talk cannot use it, but this still does leave the vast majority of the population. Uniqueness is expected to be high because the inherent uniqueness of the face but will require some testing to prove. The modality is expected to be permanent and not affected too much by damage done to the face. Measure
Ability and performance are questionable given the sensitivity of the magnetometer needs to be high and since three scores computed the performance on lower end phones will be low as well [5]. However, the system should prove very difficult to circumvent given that it requires a 3D face with a moving mouth just the same way as a specific human and a voice replay attack to be even approachable to intrusion.

References

2. López, Gustavo, Quesada, Luis, Guerrero, Luis, Nunes, Isabel L., Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. URL: https://link.springer.com/chapter/10.1007/978-3-319-60366-7_23