WHITEPAPER

VOICE BIOMETRIC REVOLUTION:

WHY VOICE ID IS NOW SECURE ENOUGH FOR DEVICE UNLOCK

IDR&D

INTRODUCTION

With the ever-increasing reliance on smartphones, laptops, and smart IoT devices comes a growing need to protect personal data and applications. Consequently, our devices must implement suitably strong measures to guard against fraudulent access.

Until now, the security standard for unlocking a mobile device or laptop exceeded what was possible to achieve with spoken voice, relegating voice to a useful convenience for some functions but not sufficient for completely unlocking a device with full access at the same level as a 4-digit PIN, a fingerprint, or a face match.

However, there are numerous circumstances where a user needs hands-free access and cannot touch or look at the device directly. Examples include driving, cooking, exercising, or even working in an environment that requires gloves and other personal protective equipment. And, there are several cases where voice is the only means of interaction. Therefore, using spoken voice to unlock a device and enable a full set of voice commands is highly desirable.

This paper describes how ID R&D's latest voice biometrics technology delivers a breakthrough level of accuracy that meets or exceeds the required security standard, enabling a new level of secure, frictionless voice interaction with personal and IoT devices.

DEVICE UNLOCK SECURITY OVERVIEW

There are different ways to protect a personal device against fraudulent access. The most common are PIN codes and biometric-based methods such as fingerprints and facial recognition. The probability of accepting an impostor with a 4-digit PIN code is 1 in 10,000. The accuracy requirements when using biometrics for mobile device unlock are similarly high. The Android Compatibility Definition Document (CDD) requires the false acceptance rate of fraudsters to not exceed 1 in 50,000 with a maximum false reject rate of 10%, as well as requiring support for spoofing detection.

An example of a biometric that meets or exceeds this standard is face matching. The technology that enables Apple's Face ID utilizes advanced hardware and software. The iPhone camera creates a depth map of your face while also capturing an infrared image while Apple puts the probability of a random person looking at your iPhone and unlocking it using Face ID at approximately 1 in 1,000,000.

DISADVANTAGES OF CURRENT SOLUTIONS

While face and fingerprint biometrics offer strong device-based security, there are cases when a user would benefit from hands-free access to their locked device. An obvious example is while driving. In this case, typing a PIN or using facial or fingerprint biometrics is not safe as require the driver to interact with the device's touch screen or to position their face in front of the camera, diverting attention from the task of driving. Voice biometrics offer a passive interaction without diversion. The user can unlock the device and perform a task with a short spoken phrase such as «Ok, Google, read my last text message.»

However, the advantages of offering a voice-based option extend beyond the need for hands-free access for safety. First, unlike alternative methods of unlocking a device, only voice offers the ability to unlock a secure device and execute a command such as «read my last text message», in one simple step. Second, the user environment may be better suited to voice, such as in the case of poor lighting conditions for face capture and wet or dirty fingers for fingerprint capture. Voice now provides a secure and convenient alternative.

Finally, unlike other biometric modalities, voice biometrics doesn't require sophisticated-device-specific fingerprint or camera sensors and will work even with low-end devices. Voice unlock can now be used with billions of existing devices without additional hardware costs. The net result is that voice extends the value of the device to more situations.

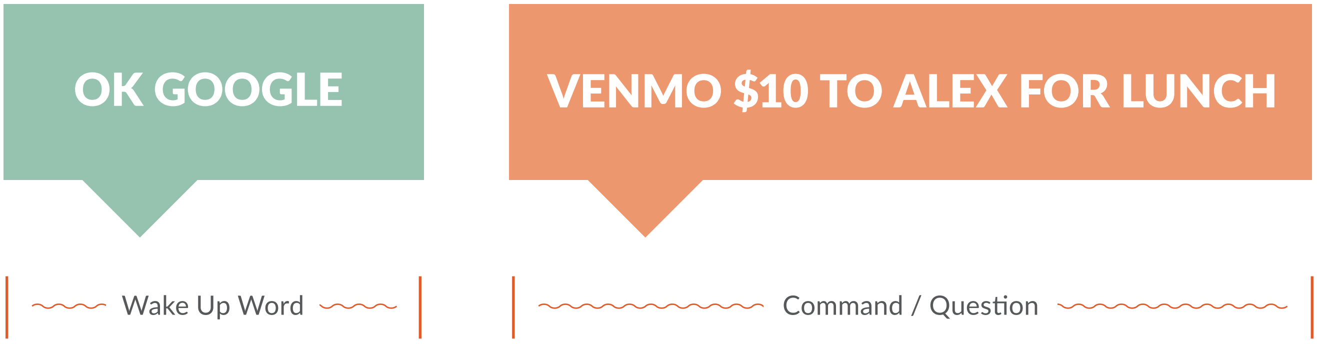
THE ID R&D SOLUTION

The scientific community has recently made advancements in the voice recognition space, more precisely called the speaker verification space. The voice modality offers a desirable approach to handling user authentication for device unlock, payments, and other activities that require high security. However, there are currently no commercially available solutions on the market that enable secure device unlock using voice. Until now, the accuracy of voice biometrics has not met accepted standards.

ID R&D made significant progress in the use of voice biometrics for device unlock use cases. ID R&D accomplished this breakthrough by combining its advanced algorithms with multiple speaker verification methods and its unique voice anti-spoofing technology. The approach is described in the following paragraphs.

The main methods for authenticating a person's unique voiceprint can be divided into two categories: text-dependent and text-independent. In a text-dependent approach, the analyzed phrases are fixed and known beforehand. Conversely, the text-independent approach places no constraints on the words which the user is allowed to speak for authentication. Both approaches have pros and cons. but unique capabilities arise when combining the two approaches in a natural voice user interface interaction.

For instance, voice interactions with personal electronic devices commonly start with a fixed wake-up word such as «Ok, Google», «Hi, Alexa», or «Hey, Siri». A wake-up word alerts the device to listen for a command phrase. The actual command phrase is not fixed and thus will not be handled by the text-dependent approach. The command phrase could be something like: «What time is my next meeting?» or «Venmo 10 dollars to Alex for lunch».



OK GOOGLE

VENMO $10 TO ALEX FOR LUNCH

Wake Up Word

Command / Question

Text-Dependent Verification
Utterance length less 1 sec

Text-Independent Verification Based on Free Speech
Utterance length 1 - 3 sec

© 2021 ID R&D

The proposed solution applies text-dependent speaker recognition for the wake-up word, text-independent speaker recognition for the command/question, and voice anti-spoofing technology for the entire utterance. The matching results are combined to provide an authentication decision.

The voice anti-spoofing algorithm protects the system from spoofing attacks. The types of voice spoofing attacks covered are:

- **Text-to-Speech attacks.** This attack is made by generating synthesized voice and presenting it directly to the voice biometric system.
- **Voice Conversion attacks.** This attack is made using a voice conversion tool that converts the voice of one subject into a voice similar to another (target) subject.
- **Replay attacks.** This attack is made by recording and playing the subject's voice through a dictaphone or any speakers.
- **Mixed attacks.** Text-to-Speech or Voice Conversion attacks replayed through speakers.

The combination of speaker verification methods and anti-spoofing results in a high level of authentication accuracy with a False Acceptance Rate below 1 in 50,000, a False Rejection Rate below 10%, and a Spoofing Acceptance Rate as low as 3%.

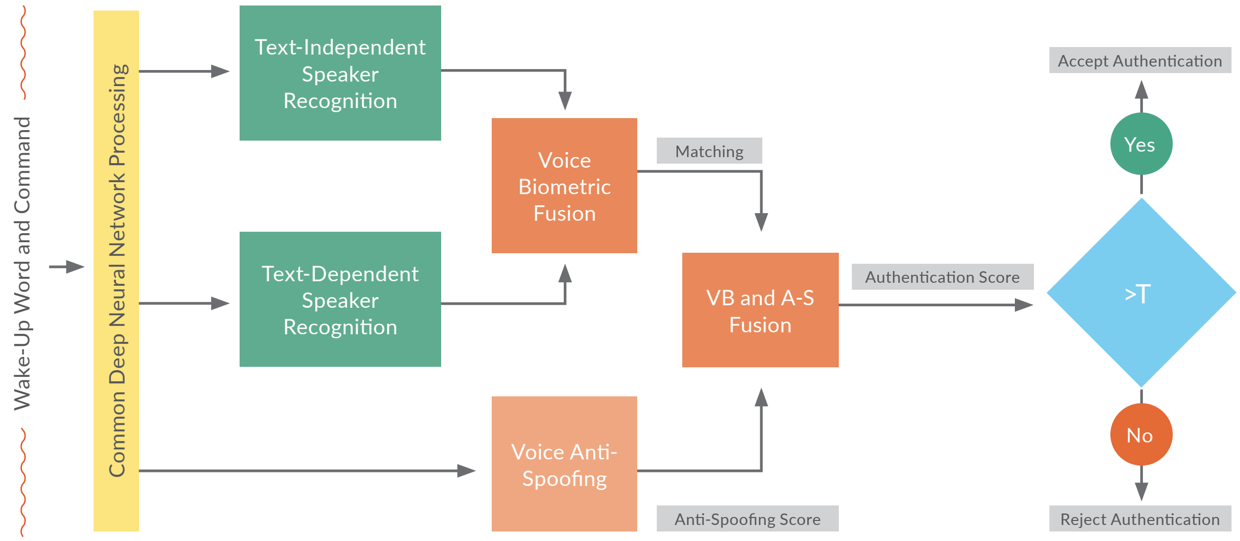


Figure. ID R&D voice authentication scheme.

The uniqueness of the technology lies in using a Common Deep Neural Network processing step that enables the extraction of robust features from the voice for text-dependent, text-independent, and anti-spoofing within a single network. Combining datasets that were previously used separately for these tasks doubles the training set and enables a synergistic effect for each task and the authentication task in particular.

© 2021 ID R&D

DIVERSITY OF EVALUATION DATA

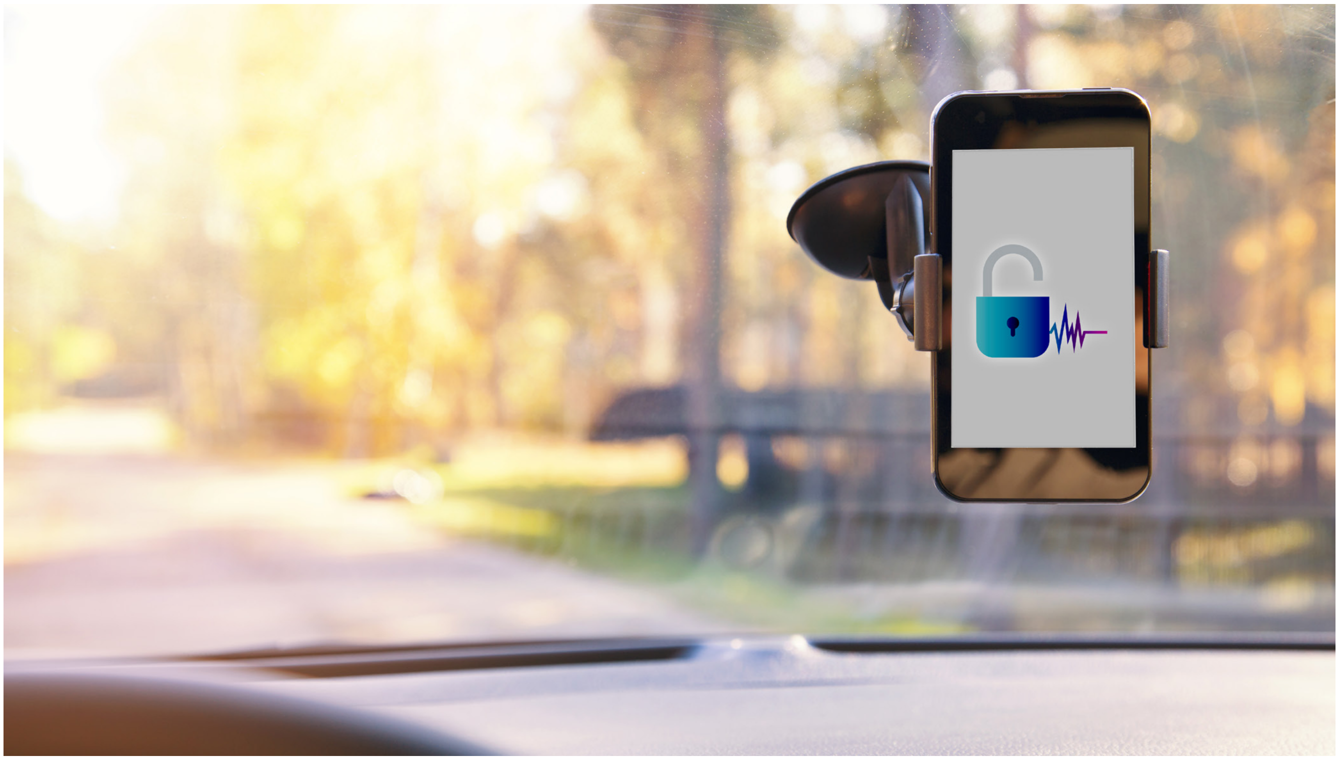
ID R&D pays significant attention to the diversity of data and is guided by industry standards. Our evaluation methodology and accuracy metrics were defined according to the ISO standards: ISO/IEC 19795-1 (biometric performance testing and reporting), ISO/IEC 30107-3 (biometric presentation attack detection), and ISO/IEC TR 19795-3 (biometric performance testing and reporting).

The following principle factors are accurately measured and taken into account:

- Biological factors: age distribution, gender
- Social factors: language
- Environmental factors: noise level and type, transmission channel, reverberation

Additionally, one more biological factor is taken into account - inter-day voice variability. This is the variability in voice characteristics caused by changes in a user's emotional and physical state across different days.

ID R&D's data comprises up to five different data sources: data collection services with fully controlled conditions of data gathering, individual subcontractors, crowd collection services with uncontrolled conditions, our data collection department, and data from partners. There are 10 different languages, including European and Asian languages, 10k speakers, 5 environments, and near-, mid- and far-field conditions.



EVALUATION RESULTS

ID R&D recently participated in a blind voice biometrics accuracy evaluation conducted by a third party. That is, the third party conducted the test independently of ID R&D, presenting inputs to the ID R&D voice biometric system and measuring results. The test measured authentication using a wake-up word combined with a random command.

As with all biometric authentication systems, measuring the two types of errors, the False Accept Rate (FAR), the error of letting an impostor through, and the False Reject Rate (FRR), the error of blocking a valid person, constitute the basis for measuring accuracy. A Detection Error Trade-off (DET) plot illustrates the trade-off between these two types of errors for a biometric matching system.

RESULTS FOR AUTHENTICATION USING WAKE-UP WORD AND RANDOM COMMAND

In the evaluation scenario, the voice commands were not predetermined. The parameters follow:

- Commands of random length, consisting of 3 to 9 syllables, most being between 3 and 5 syllables
- Distances from 0.3-1 meter
- Normal vocal effort and pronunciation speed
- Voice enrollment consisting of 3 repetitions of the wake-up word and 5 randomly chosen commands
- Use of an Android smartphone as the recording device
- A single wake-up word and command was used for the test utterance
- Asian language
- Enrollment and test speech are recorded on different days
- User environments for Driving and Indoors

The results of this evaluation are as follows: ID R&D achieved a 1 in 50k False Acceptance Rate (FAR) in both test environments: Driving and Indoors. The averaged False Reject Rate (FRR) was 9.9%. Testing for the Indoor environments resulted in a FRR of 6.8%.

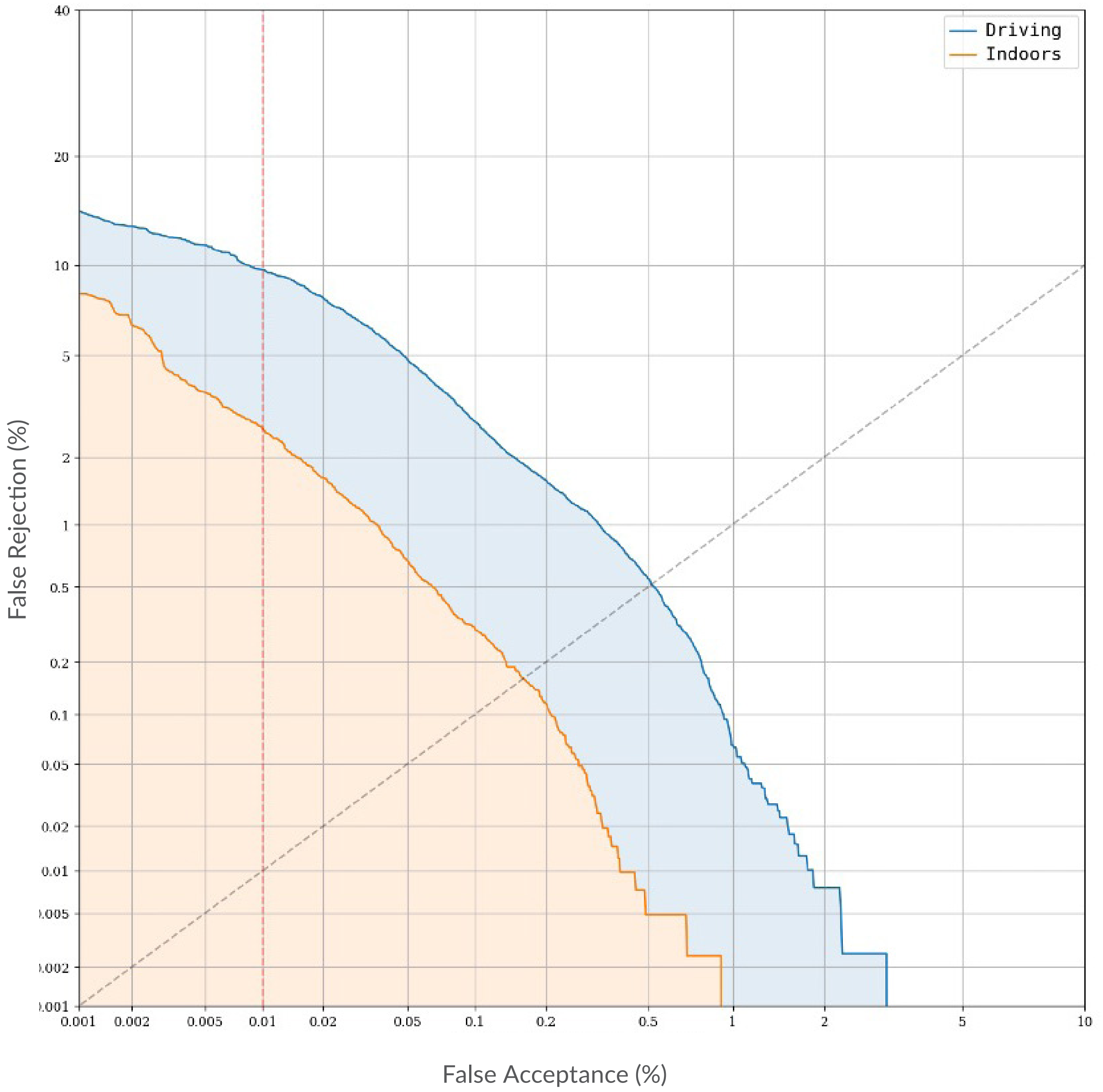


Figure. DET plot for voice authentication while Driving and Indoors, interacting with voice assistants using wake-up and short random command.

Table. FAR/FRR metrics for voice authentication while Driving and Indoors, interacting with voice assistants using a wake-up word and short random command.

ENVIRONMENT	FAR	FRR
Driving	1/50,000 (0.002%)	12.9%
Indoors	1/50,000 (0.002%)	6.8%
Averaged	1/50,000 (0.002%)	9.9%

CONCLUSION

This whitepaper illustrates a method for using voice biometrics to enable a sufficiently secure level of accuracy combined with an entirely passive user experience for authentication. No extra effort is necessary on the part of the user to securely make requests of a device. In situations where other authentication modalities are inconvenient or unsafe, voice offers a convenient alternative as a way to gain full access to the capabilities of a device.

Therefore, voice biometrics open up new possibilities for secure, hands-free access to information and applications on a variety of voice-enabled devices

Examples of use cases for hands-free voice biometric unlock:

- Listening to a new text message
- Accessing your calendar
- Sending a meeting invite
- Giving a command to make a payment
- Activating or deactivating a smart security system
- Bypassing parental controls without a PIN
- Authenticating spoken commands in an automobile

To implement voice biometrics in your product, contact ID R&D to learn more about our SDKs for Android, iOS, Linux, and Windows.

ID R&D

1441 Broadway, Suite 6019
New York, New York 10018 USA
info@idrnd.ai