VoiSentry speaker verification

Implementation guide

Authentication solution tailored for your user groups

Notwithstanding the accuracy and reliability of the speaker verification system, the design of your authentication solution is critical to its acceptance by users and achieving your implementation objectives.

The user interface design involves more than simply deciding on a passphrase. Authentication solutions require a good deal of thought and careful consideration.

That includes the method of enrolment, how you handle identity claims and verifications, and your implementation of multiple authentication factors. Furthermore, effective handling of retries and confidence results will ensure a positive and encouraging experience for your user groups.

The best solution for your specific user scenarios results from careful implementation.

Application examples	Impact	Key benefits
• IVR and self-service portals • Transactional services requiring authentication • Privilege services requiring access credentials • Contact centres	• Transform your user interface with widespread customer acceptance • Improve your identification and authentication metrics • Mitigate fraud issues	• User convenience • High acceptance • Fraud reduction • Staff motivation • Cost-effectiveness • Rapid ROI

APB0460V | Issue 2 | 2019

How is your voice biometrics implemented?

Voice samples

The first step is to determine how you will get voice samples from your customers, in order to create their individual, reference voiceprints. You need a method of collecting audio from each person, and you need to make sure you get that one voice, and that voice only, in each file i.e., you must avoid capturing others' voices.

The most common method of getting audio is over the telephone, which is perfect, because that's where voice biometrics comes into its own. In fact, some vendors will have optimised their speech verification system for telephone speech. The telephone can be a landline phone, mobile phone, VoIP phone, or a PBX handset, it can be a SIP client, WebRTC in a browser, or it could be Skype.

You will need to record audio from the caller, with just the caller's voice, and pass that to the voice biometrics system. Integration with an IVR platform is the best method of achieving that, on the basis that future authentications are likely to be triggered by calls to the IVR.

Other methods of recording audio include using a laptop with a microphone, a native 'app' on a mobile phone, or a studio microphone.

You just need to ensure the recording format is suitable for the voice biometric system, and to minimise background sounds and noise.

If you are thinking of using database recordings of contact centre customers, most likely you will need to process the audio to separate the customer's voice from that of any other person on the call e.g., an agent or supervisor.

Enrolment

Enrolment is the process of creating a reference voiceprint for the individual. The enrolment step is important as you need to consider what each person will have to say or speak in order to be enrolled in the system.

To have enough audio to analyse for adequate enrolment in the system, you will need several (typically, at least three) distinct samples of each person's voice. That requirement applies regardless of the source of the audio.

There are a number of options for enrolment (and later verification), under the labels of active and passive methods.

Passive method

This method gets its name from the idea that speech samples for enrolment are recorded passively while the speaker is in conversation with e.g., a call taker. Consequently, the speaker doesn't have to say anything specific, such as a passphrase.

However, the label is somewhat of a misnomer, because the speaker must actively consent to the recording. Since the introduction of the European Union's General Data Protection Regulation (GDPR), biometric data (a 'special category of personal data'), such as voiceprints, cannot be created and stored without the explicit consent of the 'data subject'.

Because biometrics is a special category under the GDPR, it requires more protection in the sense that, in order to lawfully process voiceprints, you need to identify and document a lawful basis under Article 6 and a separate condition for processing under Article 9. The lawful basis for processing voiceprints may be e.g., consent, or the legal (regulatory) obligation of the controller/processor, or in order to protect the vital interests of the data subject by e.g., verifying their identity prior to transacting on their behalf.

Regulations notwithstanding, the benefit of this "passive" approach is that you do not need to train the speaker or ask them to say anything specific in order to enrol. It is perhaps ideal for the user, because they don't have to remember a special passphrase.

A potential downside of this approach is that the caller may not say much, or what they do say isn't of sufficient 'phonetic diversity'. For an adequate enrolment, the recording samples need to contain enough unique spoken sounds. Ideally, that means the person speaking all the phonemes in their language twice, which for English means 44 (x2) phonemes.

As it happens, a caller is likely to speak all phonemes twice or more in a two- to three-minute phone conversation. However, as it is rare for one speaker in a telephone dialogue to talk continuously for more than 10 to 20 seconds, recordings should be multiple, shorter passages captured throughout the duration of the call. If you can get from three to five samples of usable speech, each of say 10 to 20 seconds duration, with silence removed and exclusively from the caller, you will get a more precise reference model, and subsequently, better verification accuracy.

Active method

This method is so-called, because the speaker must knowingly speak a specific sequence of words in order to enrol a voiceprint. Typically, a passphrase with a unique arrangement of about three seconds duration (about six to 10 words), repeated three to five times, is needed for an effective enrolment. Note that consent is needed regardless of the method used. It's simply more overt with this method, as the speaker is fully engaged in repeating the passphrase to enrol.

An advantage of the active method is that the process of authentication is faster and more efficient. That's because the same passphrase is used for enrolment and verification, and the length of speech sample that's needed is short. An easy to say and remember phrase, containing a minimum of four syllables, is recommended.

Very few phonemes are needed as long as you get enough samples. On the basis that users will enrol by repetition, repeating sounds or phonemes within the passphrase can be a good thing, because they are never said in exactly the same way.

Active participation means the speaker has to devote time to the process of enrolment, which may be considered a downside. However, as the speaker is aware of the process and its purpose in either case, that's unlikely to be a real issue. And considering the benefits to the user, having to follow 'repeat after me' instructions is an investment in future convenience. It's also worth adding that if the system allows autonomous passphrase selection per user, such objections disappear.

Other active methods

As alternatives or in addition to the classic passphrase option, there are other active things you can ask your customers to do for enrolment, and verification. With a primary purpose of voice biometrics being to enhance security and mitigate fraud, how you manage enrolment and verification will have a bearing on the effectiveness of any solution.

A fixed or static passphrase, whether based on words (text) or numbers (digits), has the disadvantage of being relatively easy for a determined, high-tech fraudster to record a voice and use that to try to fool the system.

Vendors go to great lengths to be able to detect recorded playbacks and other methods of spoofing. However, in terms of risk assessment, in a given application, the designer must consider how likely it is that an attack will be attempted, and what the consequences of any resultant security breach will be.

If the likelihood of a breach is significant, or the consequences severe, then a simple, fixed passphrase verification is unlikely to be suitable. The best practice is always for applications to be designed to use strong, multi-factor authentication, usually including text-prompted speaker verification with different randomly selected prompts for each access attempt. In that way, your solution will be at its most effective.

Adding a second prompt to get the caller to speak a non-static, random word or number sequence give two benefits:

1. it adds an additional authentication factor;
2. it serves as a form of liveness detection i.e., if the caller responds with the correct sequence, it's more likely to be a human than a machine.

To be able to verify using such active methods, enrolment must involve getting the customer to speak enough words and numbers to be able to generate a viable voiceprint, rather than simply repeating a passphrase.

For example, in order to enrol, you could get the customer to speak the numbers zero to nine in ascending order, followed by speaking them in descending order, followed by speaking two, five-digit sets. That will give you the spoken diversity to have customers verify by repeating, say, one or two sets of randomly generated four-, five-, or six-digit strings, which are of course different every time. That means you will collect a robust speech sample in a simple way, which is an excellent approach that will give you high security and convenience.

If your voice biometric system includes speech recognition and DTMF recognition, you get the added advantage of being able to verify what is said in addition to who said it. At the end of the day, the methods you choose will depend on the level of security and user experience you wish to provide.

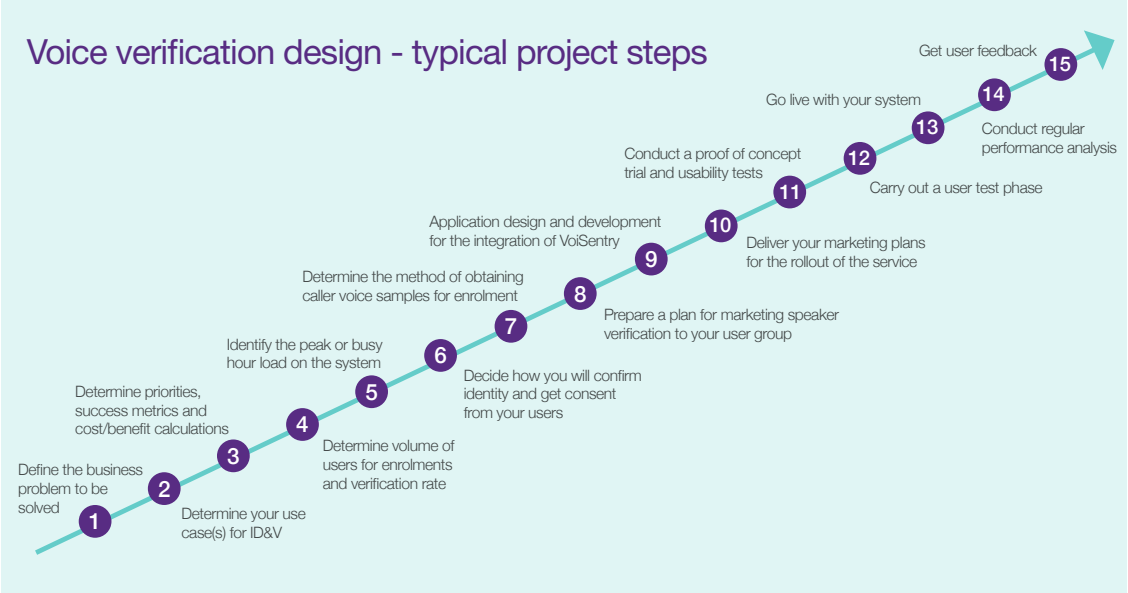
Verification

Voice biometric systems employ a method of scoring to express the similarity between a user's voice and their reference voiceprint. Therefore, it follows that the higher the score, the greater the probability of the voice belonging to the correct person. It also follows that authentication depends on the score being higher than a pre-determined threshold.

In real-world systems, it is a fact that the errors of FAR and FRR have to be managed.

Because a voice biometric system is a statistical system based on probabilities, there is a trade-off between those error rates. They are inversely separable, and if you were to plot them against a threshold axis, you'd see that they overlap.

Setting a threshold within that area of overlap means some impostors will undoubtedly score higher than the threshold, and some genuine users will score lower. Therefore, it is unavoidable that some classification errors will occur.



If you set a higher confidence level, such that no impostor scores higher, there will be no false acceptances. However, at that same threshold, some genuine users may fail to score high enough and be falsely rejected. Conversely, if you set a low threshold, such that no user is falsely rejected, some impostors will score high enough to be falsely accepted. If you choose an optimum threshold between those two points, it is inevitable that both false rejections and false acceptances will occur.

The necessary trade-off is a balance of security versus convenience. Set a high threshold to block impostors and you will inconvenience some genuine users. Therefore, the threshold setting depends on the application and the relative importance of those two considerations.

To maintain both high security and convenience, the best practice is to set the threshold sensitivity high enough to strongly reject impostors. It also makes sense to enable user retries. Retries are commonplace in any IVR, contact centre, or self-service platform, and by allowing retries you increase the chances of a genuine user being authorised. It works on the principle of 'if we're sure, we'll let you in, if not, we'll ask for another sample'. However, you should not enable more than two or three attempts as, well… that would be silly.

It is industry standard security practice to lock accounts after two or three failed attempts at verification. Furthermore, adding several layers of security, acting in tandem with a multi-factor authentication, is also considered best practice when implementing voice biometrics for high-risk transactions.

VoiSentry benefits

- **Self-contained virtual appliance** – easy to install on your platform of choice
- **REST APIs** – quickly and easily add voice verification on any customer interaction channel
- **Scalable architecture** – facilitates high capacity systems whilst maintaining system response times
- **Resilience** – database replication offers greater resilience and persistence than a single, back-end database
- **Multi-tenant capability** – enables distinct applications to be presented to individual user groups
- **Multiple verification modes** – enable you to implement a solution tailored to the needs of your users
    - Options for multi-factor authentication and liveness detection
    - Implement passphrase, prompted or passive verification methods
- **Language independence** – lets you provide one solution across a multi-lingual population

Consultative approach

Voice biometrics is a means of enhancing the delivery of and access to the services you provide via your applications. To achieve your goals, you need to plan your project. We offer consultative support for the implementation of VoiSentry in your authentication solutions. Our focus is on helping you achieve the best results and meeting your business objectives.

With support from our specialist team, you will be able to build a high-performance, voice access interface, with multi-factor authentication, that delivers a high degree of user satisfaction through automating the process of caller identification and verification (ID&V).

The result will be your success: callers get quick and convenient access to their desired service, and can carry out transactions without the intrusive interrogation of agent-led ID&V; customer touchpoints are enhanced, because callers can get down to business right away; and your business will benefit overall from a more effective and efficient authentication process.