Look who's talking:

The essentials of speaker verification

White paper

VoiSentry by Aculab

APB0424V | Issue 2 | 2019

Contents

Executive summary

There are clearly audible differences between people's voices, sufficient for them to be used to identify an individual by means of an analysis of their voice. In fact, the use of voice as an aid to verifying the identity of an individual and/or detecting imposters, who might otherwise succeed in making false identity claims, has achieved commercial viability and user acceptance in a variety of use cases. Furthermore, the use of the technology is on the increase.

Authenticating a person making an identity claim by a process of analysing their voice patterns is a technology commonly referred to as speaker verification. That's because the primary purpose is to verify that a speaker is who they claim to be. Another application of the technology is to assign an identity to an individual speaker from amongst a number of possible or potential speakers i.e., speaker identification.

The process involves two essential activities; enrolment and verification. Enrolment involves providing samples of your voice to the system, which then creates a reference model. Verification takes place when the system is required to analyse a passage of newly input speech compared to the reference model, in order to confirm or deny a match.

Using a voiceprint[1] presents a key advantage in cases where the claim is to be made remotely over the telephone. The ubiquity of the telephone and the fact that telephone quality speech contains a wealth of information, not only allowing for a caller's speech to be understood, but also for individual speaker characteristics to be identified, means voiceprints are a simple, practical and secure method of identification.

Across all use cases for speaker verification, there are three key benefits; security, speed, and convenience. On the basis that businesses must introduce stronger identity verification methods, voice biometrics will become a preferred method of preventing fraud, data breaches, and the compromise of credentials.

In relation to speed, if agents don't need to spend time authenticating callers, the process is faster and time/cost is naturally saved. And as far as both users and agents are concerned, automating authentication is infinitely preferable to interminable security questions.

[1] For the purpose of this document, the term 'voiceprint' is used as a synonym for an individual reference model.

Introduction

Secure personal identification will be eternally linked to the personal identification number (PIN), developed by Scottish engineer James Goodfellow[2]. Since its debut in 1965, and quite apart from its original purpose related to currency dispensing systems, PIN use has become ubiquitous in systems such as interactive voice response (IVR) and for a myriad of on-line verification purposes. The combination of IVR and PIN has had a profound influence on many industries, notably those deploying inbound contact centres.

PINs and IVRs made use of another technology introduced in the 1960s, namely dual-tone multi-frequency (DTMF) signalling. Later on, the 1990s saw the first introduction of commercially successful speech recognition technologies[3], which ultimately brought automatic speech recognition (ASR) to IVR applications. The advent of ASR meant that customers were able to speak their PINs, as an alternative to keying touch-tone DTMF.

A key element of PIN-based customer verifications is that they are accomplished remotely, without human intervention. That fact applies whether a PIN is keyed or spoken. However, in relation to the latter, it led to the obvious thought, which was to use speech technologies, not merely to detect and decipher the spoken PIN, but to determine who it was who said it. Developing that idea brought in a new, biometric dimension to secure personal identification and verification. Whereas speech recognition is designed to identify that a caller said, “It’s me!” speaker verification is designed to establish that it was me who said, “It’s me!”

Of course, biometric identification is not a new thing. It can be said to have begun in the late 1870s[4], when it was first proposed that fingerprints could be used for identification. Long after that revolutionary introduction, fingerprints are no longer associated just with being on the wrong side of the law, as every smartphone user knows. In turn, speaker verification has achieved a level of commercial viability and user acceptance in a variety of use cases.

A significant reason for that is the rise of identity theft and associated fraud. Financial fraud losses in the UK totalled £768.8m in 2016, up 2 per cent on 2015, with impersonation and deception scams amongst the main elements in financial fraud[5].

Many businesses continue to use PINs, passwords and security questions to authenticate their customers. However, it’s clear that stronger forms of authentication are needed, because the traditional methods offer weak protection against social engineering, identity theft and fraud.

Speaker verification can be used to counter that threat, and, in conjunction with multi-factor authentication, it helps to reduce fraud whilst making identity verification far more convenient for customers.

This paper takes a look at the technology behind voiceprint analysis, how it can be used for the purpose of identifying an individual speaker, and why it offers a more secure verification solution than PIN- or password-based authentication.

[2] Evidence from the Patent Record on the Development of Cash Dispensing Technology: https://mpra.ub.uni-muenchen.de/9461/

[3] PCWorld: Speech Recognition Through the Decades: How We Ended Up With Siri: http://bit.ly/2lrnBNd

[4] Henry Faulds: the Invention of a Fingerprinter: http://www.galton.org/fingerprints/faulds.htm

[5] From the Financial Fraud Action UK (FFA UK) 2016 report: https://www.financialfraudaction.org.uk/wp-content/uploads/2016/07/2016-Year-end-fraud-bulletin.pdf

Voice identity

Due to the physical characteristics of a person's vocal tract and the way they habitually articulate the sounds when they speak, it is possible to differentiate between speakers by analysing their voices.

Thanks to advances in computational speech technology – four decades on from the first applicable international patent, filed in Italy in 1983[6] – and the ever improving performance of computing resources, it is now feasible to identify an individual by means of an analysis of the acoustic and behavioural features of their voice. In that sense, it is similar to fingerprints and DNA 'fingerprinting', and as with those other biometric methods, the use of voice as a method of verifying or identifying the identity of an individual has become an established practice.

The acoustic characteristics of an individual's voice are determined in part by the anatomy of that person's vocal tract, which consists of an airway, and the vocal folds and other structures within and without the larynx from which vocal sounds originate.

In articulating speech, those anatomical components work in combination with the physical size and movement of the organs of speech (e.g., the jaw, lips, and tongue), and resonances in the nasal passages, under control of learned behavioural patterns or personality traits.

It is those idiosyncratic articulatory movements that have been found to be the most influential factor in speaker uniqueness[7], however, all of those elements combined affect voice pitch and speaking style, including pronunciation and the acoustic characteristics of accent or dialect.

Verifying the identity of an individual speaker by analysing their voice patterns is a technology commonly referred to as speaker verification, essentially because the primary purpose, similar to that of a PIN, is to verify that a speaker is who they claim to be.

Another application of the technology is to assign an identity to an individual speaker from amongst a number of speakers. In the former case, a comparison is done between a sample of live speech from the subject making an identity claim and a reference biometric voiceprint. In the latter, multiple comparisons are attempted in order to match a unique voiceprint.

[6] Device for speaker verification patent US4752958A: http://www.google.com/patents/US4752958?hl=it&cl=en

[7] Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors, by Prof. Dr. Volker Dellwo et alia at the University of Zurich Phonetics Laboratory, published in The Journal of the Acoustical Society of America, March, 2015: http://asa.scitation.org/doi/abs/10.1121/1.4906837

Voiceprints

The data contained in a voiceprint is not a waveform, or a spectrogram, nor is it a recording of a person's voice. A voiceprint is the result of a statistical analysis of the voice patterns in a passage of speech, comparing characteristics such as frequency, intensity, duration, dynamics and pitch that produces a mathematical, reference model of the digital speech signal.

That statistical representation of the sounds can be used to estimate the likelihood of an unknown speaker being the original creator of the voiceprint.

Those statistics represent the underlying variations and temporal changes that are characteristic of the physiology and behaviour of the individual, but they do not define how those changes should be invoked to produce any particular sequence of words or sounds.

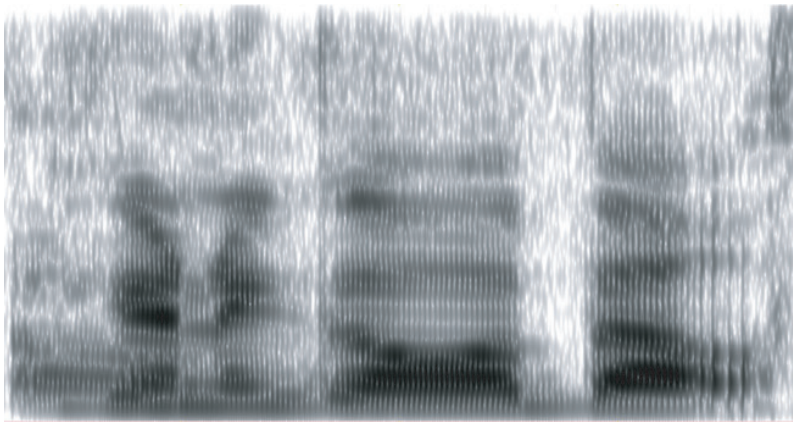

Figure 1 - Generic spectrogram

With the result being stored in binary form, it is not possible for the spoken phrase to be reconstructed from the discrete voiceprint file created from the speaker's voice input. A reference model cannot be reverse engineered to create the original enrolment waveforms and input to a system by an impostor. Thus a voiceprint has no value to a hacker.

Furthermore, from a security standpoint, the statistical models are embedded within the speaker verification system, and they are not externally accessible. Thus, voiceprints are much less susceptible to the sorts of data breaches and credentials theft that compromise millions of identity records.

Additionally, reference voiceprints themselves cannot be considered personally identifiable information, in the context of HIPAA[8], for example. Provided they are anonymized within the framework, they cannot be used to identify the individual.

[8] Health Insurance Portability and Accountability Act: https://www.hhs.gov/hipaa/for-professionals/index.html

Voiceprint analysis

The process of voiceprint analysis involves two essential activities; enrolment and verification (or identification). Enrolment is in reality no more intrusive or complex than confirming a chosen PIN, or changing a provider issued password. It involves giving samples of your voice to the system, which then creates a reference model of your voice. Verification takes place when the system is required to statistically analyse a passage of newly input speech, convert it to a set of features, compare those with the reference model and confirm or deny a match (see Figure 2 below).

To enrol in a system, for example, an end user will be prompted to speak and repeat a number of times, in their natural voice, either a passphrase or a sequence of numbers repeated several times in random order, which can be used for subsequent verification attempts in text dependent or text prompted modes respectively.

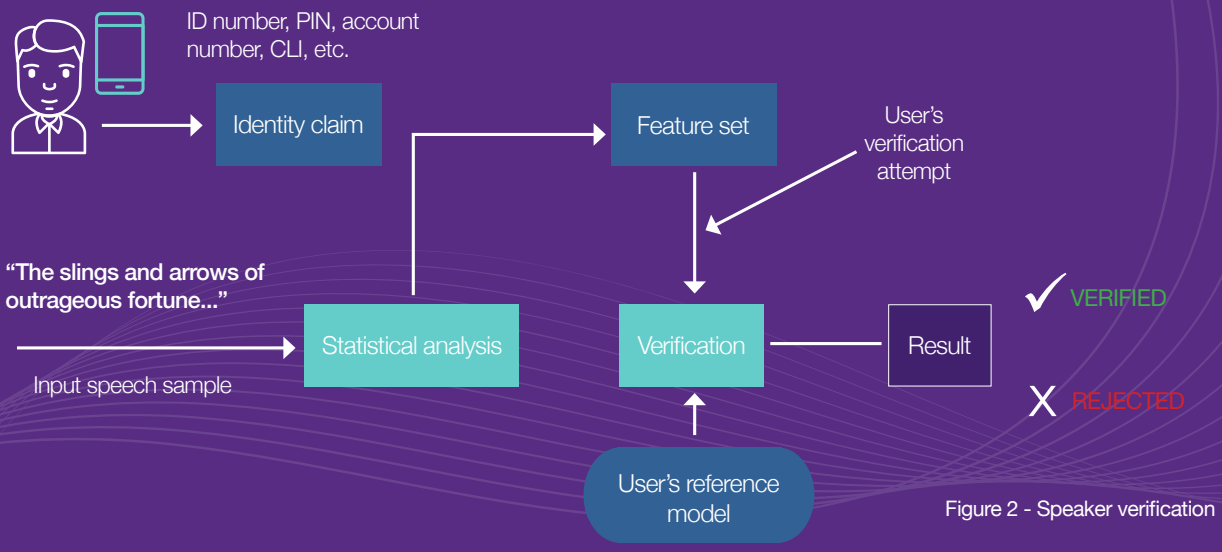

Figure 2 - Speaker verification

A verification attempt will be initiated by the caller making an identity claim. The speaker is 'voice verified' by the system processing the sounds offered to it, which are simultaneously compared against the enrolled reference model. The result indicates the statistical likelihood of the input sounds deriving from the same source as the enrolled voiceprint.

The latest verification software systems can produce a comparison result and near-certain authentication in near real-time using, for example, a specific passphrase, or a sequence of words or numbers. Compare that with the time taken for an agent to validate identity, which can take up to a minute at the onset of a call, and it is easy to see how an automated system can save time and cost in the contact centre in addition to offering enhanced fraud protection.

Strictly speaking, voice biometrics is not concerned with what is said, only with who is saying it. However, when speaker verification is deployed in practice, it is often combined with a means of identifying what was said i.e., incorporating ASR, thus the two technologies are highly complementary and present a means of offering strong authentication.

The myth of accuracy

The essential purpose behind speaker verification is a demand to know if the person making an identity claim is who they claim to be. Therefore, if the claimant truly is that person, an accurate system should be expected to verify every such claim. If the person is a fraudster, the system should be expected to deny all such claims.

Unfortunately, no system can be that precise on a single pass, other than theoretically or in highly controlled conditions that are unrealistic and unrepresentative of real-world scenarios. Only if the input audio was identical on each verification attempt, and identical in turn to the audio signal used to enrol and produce a voiceprint for an individual, is it possible for a system to be 100 percent accurate 100 percent of the time.

Additionally, as this is a technology that is fundamentally useful when used remotely i.e., the predominant use case is over the phone, many factors can influence the audio signal and hence the result. Some of those factors include: network conditions, such as latency, signal-to-noise ratios, CODECS, and input audio technique (e.g., recording or streaming audio); variability in devices e.g., different telephones, microphone sensitivity, and volume; acoustic and environmental conditions, including background noise levels, absorption, resonance, and reverberation; and physical effects as variable as microphone positioning, and the individual's wellbeing 'on the day'.

For those reasons, and because what is undertaken is a statistical analysis, system results for a single verification attempt are presented with a probability or confidence score.

In practice, a confidence threshold is needed beyond which claims can be confidently accepted. However, in setting such a threshold, it must be acknowledged that some small proportion of valid claims may be rejected. That is the cost of the guarantee. It's a trade-off between high security and user convenience.

In the case of a PIN, you'd be forgiven for thinking the input digits, if entered correctly, couldn't fail to be recognised. Nevertheless, despite the accuracy of DTMF detection being to all intents and purposes 100 percent, retries are anticipated. That is for reasons including incorrect input and false detection. It is the same with biometrics. Retries make sense. Therefore, implementation is the key to getting the right result, particularly if the input is variable.

The real value of voice biometrics lies not in the delivery of absolute and unfailing accuracy. It lies in mitigating fraud, because the technology makes it fundamentally more difficult for criminals to steal and use the credentials needed to access a user's account. Identity fraud is the number one fraud threat in the UK. It has reached epidemic levels and now accounts for more than half of all reported fraud, with identities being stolen at a rate of almost 500 a day[9]. It appears to be very easy to fraudulently obtain someone's PIN or password. It is very difficult to get hold of someone's voiceprint.

[9] Cifas, the UK's not-for-profit fraud prevention service organisation: https://www.cifas.org.uk/newsroom/identity-fraud-soars-to-new-levels

Authentication security

Notwithstanding claims of accuracy, best practice suggests that relying on a single method of identification is rarely sufficient. There are many techniques for ensuring tight accuracy and security.

Multilayer security systems combine a biometric method with something like a key-card to verify that somebody is an authorised user. Multimodal systems combine multiple biometric methods, such as a fingerprint scanner and a voiceprint system. Both of those can create strong authentication methods.

Another approach, relevant to telephone based verification, is to use a form of multi-factor authentication. That is to use speaker verification to confirm the speaker's identity while simultaneously using speech recognition to validate the words spoken, such as a PIN or password.

More complex strong authentication systems involve the threefold mantra of asking for what you've got, what you know and who you are. That means: i) claimants need to have some kind of token, such as an identity card with a magnetic strip; ii) they need to know a password or PIN; and iii) they need to satisfy a biometric system with physical evidence, such as a fingerprint or voiceprint. The characteristics of various methods of authentication are illustrated in Figure 3, for comparison.

Given always that it is necessary to achieve two things when employing speaker verification i.e., make an identity claim and verify the claimant, best practice dictates the claim is used as an additional level of strong authentication security.

That means multiple factors of authentication are recommended, where the system asks for additional information or applies another factor in order to grant access.

That could mean using e.g., an individual account number or a PIN as a means of making the claim, input using DTMF or validated using ASR, followed by voiceprint analysis and verification of identity. If the passphrase is text prompted, using words or numbers selected at random at the point of validation, rather than being text dependent on a generic, pre-determined passphrase, the system is inherently more robust against fraudulent attacks.

	Secret knowledge	Personal possession	Biometrics
	What you know	***What you have***	***Who you are***
Examples	Password; PIN	ID card; pass	Fingerprint, DNA, voiceprint
Copied	Easy to difficult	Easy to very difficult	Easy to impossible
Lost	Easily forgotten	Easy	Practically impossible
Stolen	Easy to difficult	Easy to difficult	Difficult to impossible
Circulated	Easy	Easy	Easy to impossible
Changed	Easy	Easy	Easy to impossible

Figure 3 - Characteristics of authentication methods

Authentication security cont.

In certain use cases, notably where the identity claim is to be made remotely, some methods are rendered less useful by virtue of necessitating the claimant to be present.

Using a voiceprint presents a key advantage in cases where the claim is to be made remotely over the telephone, and where strong authentication can be achieved readily by requesting additional what you know information.

The ubiquity of the telephone and the fact that telephone quality speech contains a wealth of information, not only allowing for a caller's speech to be understood, but also for individual speaker characteristics to be identified, means speaker verification is a simple, practical and secure method of aiding remote identification.

That holds true, notwithstanding that smartphones have accelerated the use of fingerprints as a localised identification method.

Anti-spoofing

Fingerprints can be stolen and DNA can be faked or planted at a crime scene[10], and fraudsters can record a person's voice. That being the case, another aspect of security is the susceptibility of a system to spoofing or presentation attacks. That is, a system should be secure against intentional circumvention using faked audio recordings.

In order to be proof against so called replay attacks, where a fraudster seeks to assemble recordings to recreate copies of an intended victim's passphrase, a system must be capable of discriminating fake audio from authentic audio.

A well designed system may use text prompted passphrases or request a string of several, randomly chosen words or numbers, such that the input audio is never the same. In that way, it is impossible to pre-record the exact, requested audio and extremely difficult to assemble it on the fly during a call. Couple that with the use of ASR as described above and the result will be an extremely secure, strong authentication system.

Furthermore, good core verification engines will employ technology that detects artefacts created in recording and playback i.e., as used in presentation or replay attacks.

[10] Authentication of forensic DNA samples: http://www.fsigenetics.com/article/S1872-4973(10)00004-9/fulltext

Use cases and benefits

Voice biometrics can be implemented to improve the security of a variety of business processes, across a wide range of market sectors. Customers and employees can be verified or authenticated more securely and faster than ever before, with solutions benefitting the user experience and total cost of ownership metrics of many businesses.

Voice naturally lends itself to remote authentication scenarios and no additional hardware is needed, unlike other biometric methods such as iris scans or fingerprints.

Systems can be delivered through hosted or on-premise models and examples include, but are clearly not limited to the following:

- Fraud and identity theft prevention in any vertical market
- Authentication in customer care and contact centres across all markets
- Voice signatures for transaction confirmation and PCI-DSS compliance
- Proof of life for public services and correctional establishments
- HIPAA compliance in the healthcare and telehealth sectors

Mitigating fraud

Stemming the increase of fraud and identity theft is one of the key benefits of expanding the use of biometric identity verification. With the pervasiveness of passwords for every conceivable purpose, it's unsurprising to find personal credentials are the target of criminal activity. And with a seeming widespread, collective human reluctance to using strong passwords, it's no surprise to find fraud costs industry billions each year.

Financial fraud losses in the UK alone, for example, totalled £768.8m in 2016, a figure that was up 2 per cent on the previous year[11].

Speaker verification can provide the highest level of security when used in conjunction with other authentication methods. Across a range of susceptibilities, from data breaches and theft, to social engineering and hacking, multi-factor authentication incorporating voice biometrics presents a low risk, whereas traditional PINs and passwords alone are split between medium and high.

Voice biometrics is an effective method of strong authentication, enabling businesses to protect themselves and their customers from attacks by fraudsters. Identity theft is high-profile, and businesses have an obligation to provide, and be seen to provide, stronger protection. Fraud prevention is a brand issue as well as a regulatory one, and the risk of losing customers' confidence by being seen as lackadaisical about security is at least as great a risk as losses due to fraud and regulatory fines[12].

[11] Financial Fraud Action UK (FFA UK), 2016 data: https://www.financialfraudaction.org.uk/wp-content/uploads/2016/07/Fraud-the-Facts-A5_24.11_ LR.pdf

[12] The 2017 UK Contact Centre Decision-Makers' Guide, from Contact Babel: http://www.contactbabel.com/reports.cfm

Use cases and benefits cont.

Contact centres

Helpdesks and contact centres are a segment that is increasingly susceptible to identity theft and associated fraudulent activity. In those environments, voice biometrics offers a higher level of protection, and its intuitive simplicity and effectiveness can reduce average verification time by up to two thirds, simply by replacing the time consuming and intrusive inquisition of security questions. In addition, the motivational influence on staff that no longer need to pose the same, tedious questions, day after day, cannot be underestimated.

Furthermore, the double benefit of security and convenience makes for a far better authentication experience for customers, clients and subscribers – a fickle demographic – which is why customer care is so important. No business can be immune to the attendant bad publicity if customers' accounts are compromised through a data breach.

Faster authentication means convenience for the caller and time/cost savings for the contact centre. That's a clear win-win. Although how much can be saved depends on the business circumstances, estimates range from up to 80 percent quicker or 20 seconds faster, to saving over £0.40p per call[13]. In the UK contact centre market, it takes an average of 34 seconds to manually verify a customer's identity[14]. If that duration can be reduced, the organisation becomes more cost-effective, and more customer friendly.

From a security perspective, in 2017, surveys[15] have determined that, across the US and UK contact centre industries, a mean average of approaching 70 percent of calls require caller identity verification. Of those, 94 percent are authenticated through purely human means i.e., via a live contact centre agent.

That means the opportunity for cost saving is extremely significant, and illustrates very well the need for effective security measures.

Public services

In the public service domain, the use of voice biometrics to verify identity can be applied in obvious ways. That is also true of establishments overseen by governmental agencies, such as correctional facilities. Proving an identity claim can be critical in preventing identity theft associated with benefit fraud, particularly where transactions are conducted remotely via telephone. The same can be said for online transactions using voice authentication e.g., during a click-to-call session via a browser, perhaps using WebRTC.

The value of the technology lies, of course, in its ability to 'prove' that a person is both alive and who they claim to be. That is obviously important in relation to benefits claimants. Saving a visit to a government bureau by authenticating over the phone lessens queues, improves back-office efficiency, mitigates fraud, and is overall more cost-effective.

Perhaps less obvious, but equally important, is the benefit of using speaker verification to establish that prisoners, say those granted certain privileges, are entitled to those rights. Felons can be kept on a short leash at low cost, with automated checks requiring a voice ID response as confirmation of presence and identity.

[13] The UK contact centre decision-maker's guide 2017: http://bit.ly/2lyJw1V

[14] The UK contact centre decision-maker's guide 2017: http://bit.ly/2lyJw1V

[15] The UK and US contact centre decision-maker's guides 2017: http://www.contactbabel.com/reports.cfm

Use cases and benefits cont.

Voice signatures

In certain jurisdictions, a voice signature can be used as a legally binding means of underwriting documents such as life insurance applications and similar fiscal artefacts. During a conversation with a contact centre agent, when a document needs to be authenticated or underwritten, the remote caller can be asked to provide their pre-registered voice signature, which, on verification, the system associates with the document file.

In a similar manner, for authorising financial transactions where PCI-DSS compliance is needed, integrating speaker verification ensures a highly secure method of meeting strict industry regulations, such as the GDPR, for privacy and data protection.

Healthcare

For healthcare applications such as those related to patient management and advisory (PMA) and electronic health record (EHR) systems, a layer of authentication that meets the unique security requirements of HIPAA and HITECH[16], is undoubtedly beneficial. Voice biometrics provides a layer of security that prevents unauthorised, patient portal logins, reduces exposure to data breaches, enhances privacy and security in line with regulatory demands, and mitigates the economic consequences of medical identity theft.

The issue of privacy, especially in the healthcare vertical market, is a powerful driver for using right-party authentication to facilitate personal information sharing. That is also the case when using speech-enabled, automated outbound calling, it being necessary to make sure that the person answering or receiving the call is the one to whom the healthcare provider needs to leave a message or talk.

Summary

Security is important, because the risk of fraud and identity theft is high. Data breaches and unauthorised access are expensive; not only damaging brand reputations, adding to cost and impacting sales, they can also result in costly litigation. It is now clear that what used to be considered best practice i.e., frequently changed, strong, complex passwords, isn't practiced. The reason for that is human nature.

From top to bottom in most businesses, people are prone to ignoring good password security. Because passwords are inconvenient in so many ways, it should be no surprise to read that 63 percent of confirmed data breaches involve the use of weak, default, or stolen credentials[17].

As a result, it should be anticipated that many organisations will be encouraged to adopt new approaches. If the use of passwords, agent-led security questions and generic, two-factor SMS authentication are already compromised, there has to be a better method. If businesses are to introduce stronger identity verification methods, and even consider abolishing passwords, well managed voice biometrics is likely to become a preferred method of preventing misuse or compromise of credentials.

Unlike iris recognition and fingerprint scanning, biometric speaker verification requires no special memory skills or hardware. As voice biometrics is completely natural and instinctive for the user, security no longer has to be at the expense of convenience.

[16] Health information privacy: https://www.hhs.gov/hipaa/for-professionals/index.html

[17] Verizon's 2016 Data Breach Investigations Report (DBIR): http://vz.to/1NTb7I8

About VoiSentry

VoiSentry is a voice biometrics product from Aculab, a leading provider of advanced speech technology systems for telecommunications related applications. VoiSentry is part of a software-based advanced speech, enabling technology, and developer API portfolio that serves the evolving needs of automated and interactive telephony-based systems; whether on-premise, data centre hosted, or cloud-based, across a wide variety of markets and business-critical services and solutions.

Development APIs are offered for voice biometrics (speaker verification and identification), advanced speech (ASR and TTS), voice, data, fax, and SMS, on hardware, software or cloud-based platforms, giving users the choice between capital investment and cost-effective, 'pay as you go' alternatives.

Company offices are located in both the UK and USA.

Copyright and other notices

Follow Aculab: Twitter | Blog | LinkedIn

+44 (0) 1908 27 38 00 | www.aculab.com

VoiSentry by Aculab