



A Guide To Speaker Verification

produced by
ICR Speech Solutions & Services



**ICR Speech Solutions
& Services Ltd**

Merchants Quay
Ashley Lane
Saltaire

West Yorkshire
BD17 7DB

T (0)1274 821111

F (0)1274 821177

www.icr3s.co.uk

e info@icr3s.co.uk

1 Introduction

What is it? - Speaker Verification is an automatic process that uses human voice characteristics to determine whether a speaker (or caller) is the person he or she claims to be.

Speaker Verification solutions offer a highly accurate and efficient method of authenticating a person's identity by analysing their voice. Voice is essentially a biometric measurement very much like others which are well known such as the fingerprint pattern, the iris structure, the retina structure etc. However, in real-life scenarios, voice has a number of advantages in regards of its ease of implementation, cost performance and ease of use.

Speaker verification adds another layer of security by verifying, over the phone, that the caller is who he/she claims to be. Security is increased as speaker verification measures voice characteristics that are difficult to duplicate even if obtained from a recorded voice or intentional impersonation.

Why use it? - As e-commerce transactions become more commonplace over the telephone with automated speech recognition (ASR) systems, it is important to provide security and ensure privacy/fraud protection, whilst at the same time, providing callers with non-intrusive, pleasant interaction with the automated application.

Top-level business benefits include:

- Automate the Authentication process and free up call center agents for other activity
- Reduce staff and call costs
- Offer enhanced e-business applications with security and convenience
- Improve customer service by reducing caller hold time.

Top-level caller benefits include:

- Higher confidence in protected and sensitive account information
- Bypass live agents for "speed-to-information"
- Easy to remember access requirements
- Improve customer service by reducing verification time
- Option to eliminate passwords by associating voiceprint with; e.g., account I.D.

Additional benefit details include:

- *Speaker Convenience* - Many call centers are concerned with cutting costs yet maintaining quality customer care. Asking callers to speak digit strings and/or phrases is easier and more convenient for them than manual entry.
- *Cost Effectiveness* - Speaking account numbers or phrases and verifying on the voiceprint vs. a PIN or password will shorten the transaction time. The shortened transaction time will result in a reduction of trunk/facility costs.
- *Efficiency* - Combined with ASR the possibilities of an accurate match are increased by using both the account number or some other identifying string/phrase and the voiceprint
- *Security* - Anyone can enter or speak a stolen account number and PIN. The addition of speaker verification adds another level of security by adding a voiceprint comparison to the approval process.
- *Automatic Updates of Voiceprints* - An argument against speaker verification is that human voice constantly changes and depends on external factors. Advancements in this technology provide a solution to this problem by updating the voiceprints with voice data from successful verifications allowing satisfactory levels of accuracy.
- *Secure Financial Transactions* - The increased use of ASR applications for accessing financial information and accomplishing banking/brokerage transactions results in the need for increased security.
- *High Technology Image* - Speaker verification is regarded as leading edge security. Companies, especially financial institutions, will consider deploying to boost their high technology image among customers.

How does it work? - To perform speaker verification, the biometric system checks a speaker's identity by comparing speech samples to an existing voiceprint for that speaker. The system actually performs three functions:

- **Enrolment**, or training, of voiceprints for new users. During training, a user provides enough speech samples to allow the system to "learn" the voice. The system creates a voiceprint (also called voice signature), stores it in a database, and uses it later to verify that user's identity. Typically users need to train voiceprints only the first time they use an application. Note - free-speech applications will likely require more speech and more calls. The voiceprint is not a recording of the voice but rather a small file (approx 15-20k bytes) containing the person's selected voice characteristics represented in a numerical format. The creation of these features is a complex mathematical process. There is no particular standard and

suppliers tend to claim competitive advantage in this area. Note: The customer must be securely identified and verified prior to being issued with the identification token that will be used in the enrolment process.

- **Verification** of the user's voice. The user asserts a particular identity, and the biometric system retrieves the voiceprint associated with that identity (called the *claimant* voiceprint) from the database. After one or more utterances, the system authenticates or rejects the identity of the speaker by verifying live speech samples against that voiceprint. The use of an authentication policy can add different levels of security to the verification process depending on the individual caller or the transaction type.
- **Adaptation** - training sessions can produce a new voiceprint or refine a previously existing one. If the voiceprint does not exist, the system creates a new voiceprint in the verification database. If the voiceprint already exists, the data gathered during the training session is used to improve, or *adapt*, that voiceprint. Adaptation is the process of enhancing an existing voiceprint using new data.

Speaker Verification can be deployed in a number of different ways:

- **Text-Dependent Mode**– verification is performed on a password or pass phrase. This is less convenient for the user as they have to remember their password. There are also overheads in changing the password etc.
- **Text-Independent Mode** - verification that can process an unconstrained utterance.
- **Text-Prompted Mode** – verification process that asks the user to repeat random numbers and/or words. This is easier for the user in that no passwords or PINs need to be remembered.
- **Free speech Mode** – This is similar to Text-Independent but is described separately as the distinction is important. It allows verification to be instantly performed in the background during the course of a normal and natural conversation, probably with a call center agent. The verification is carried out independently of the spoken content. This mode can also be used throughout the duration of the call thereby increasing security.

Where is it being used? - Although originally used in covert intelligence gathering Speaker Verification is now a mainstream technology with several real-world business deployments.

Currently there are only a few Banks worldwide using it in customer-facing situations for caller identification and verification, however, many more are evaluating the technology especially with a view to deploying it internally first to gain experience.

The two most popular applications being considered for internal deployment are the Helpdesk and HR functions. A high proportion of calls into Helpdesks are for password resets – Speaker Verification can assist in automating this process in an efficient and cost-effective way. In many large organizations managers need to identify themselves when calling into the HR department. Again, Speaker Verification can be used to automate the process whilst at the same time adding to the level security.

2 Benefits of Speaker Verification

Voice has a number of advantages over other biometric methods and these include:

- ✓ **Voice is ubiquitous** – almost everyone has a voice and can therefore be authenticated.
- ✓ **Voice is non-intrusive** – the authentication of the individual can take place as part of a normal telephone call. There is no inconvenience to the user while they are being authenticated and the user does not need to touch anything or learn a new operating device. Negative connotations prevalent with other biometrics, for example fingerprinting, are not as pronounced with voice (this is not to say that there will be no issues regarding privacy etc).
- ✓ **Voice is easy to use** – everybody can use a telephone (any of the many types) to communicate. For the user this is all that is required for speaker verification. There is no positioning of fingers, heads, or eyes to gain access.
- ✓ **Flexible** – voice is a very flexible means of authenticating a caller's identity. Callers can be authenticated with or without needing to remember passwords or PINs. However, adding knowledge-based questions to the process can lead to significant improvements in security.
- ✓
- ✓ **No new infrastructure required for the user** – most people either own or have access to a telephone. The telephone greatly outnumbers the PC.
- ✓ **Cost** – primarily because of the lack of user infrastructure hardware the rollout of speaker verification is relatively cheap compared to other methods (especially for large implementations). When the traditional call centre task of authentication is replaced by an automated speaker verification process significant cost savings can be expected.

3 Issues

The accuracy levels alone do not measure successful implementations of speaker verification systems. There is wide range of issues that need to be considered before deployment:

1. Robustness

- A valid speaker verification system needs to perform well in real-world scenarios.

2. Natural Voice Variations

- How well will the system cope with colds, daily voice variations and long-term voice variations caused by factors such as ageing?

3. Mix of Communication Channels

- How well will the system handle different phone types - landlines, mobiles, etc?

4. Background Noise

- How well will the system handle background noise - talking, street noise, road noise in cars, airports etc?

5. Scalability and High Availability

- The system requires scalability and 100% availability.

6. Accuracy

- This may well be trial and error during the stages of an internal deployment until the true performance of the system in a real-world environment is realised.

7. Rejections

- A policy is required for handling rejections – both true and false.

8. Voiceprint Database

- Who owns it and are Voiceprints stored centrally or are they distributed?
- How long should Voiceprints be kept for?

9. Elapsed Time

- What impact does elapsed time have on the effectiveness of Voiceprints?

10. Integration

- With existing systems – telephony, ASR systems and agent desktop applications.

4 Risks

There are a number of identifiable risks which need to be recognized, analysed and assessed when considering Speaker Verification.

1. The “Betamax Factor”

- Currently there are about a dozen credible suppliers in the speaker verification field each of which has their own proprietary solution.
- There are biometric standards but they are not enforceable.
- The government has not, as yet, issued any guidelines.
- RISK = Your chosen solution may not be the standard adopted by other financial institutions (either by choice or legal enforcement). Therefore it becomes redundant.

2. Privacy

- Currently there are only a few known deployments of speaker verification and the general public is probably not aware of them. This technology is still seen as something “out of a James Bond film”.
- The general public is mainly apathetic about privacy.
- There have been no significant outcries about privacy. It’s debatable whether the greater threat is from public organisations (e.g. the government) or private company misuse.
- RISK = Any misuse or abuse relating to any privacy could tarnish all other implementations.

3. Press Misinterpretation

- Currently there are only a few known deployments of speaker verification and the success of these is not known in the public domain.
- The claimed accuracy of systems is mainly from trials not the real-world.
- The press could pick upon any performance issues.
- RISK = Performance however satisfactory for the Bank could be misinterpreted by the press and thereby lead to customer concerns.

4. Customer/User Acceptance

- We assume customers and staff will appreciate speaker verification as it removes the need to remember PINs, TINs, passwords etc.
- However, both customers and staff need to be sold the benefits of speaker verification.
- RISK = Without being told the organisation’s intentions or the potential benefits customers could be put-off Speaker Verification.

5 Summary

ICR believes that the introduction of Speaker Verification will have a measurable positive effect on customer care. The key reasons for this are:

1. Improving the ID&V Process.

- Customers are currently required to remember too many identification and verification tokens.

Speaker Verification will remove the need for tokens.

- The existing ID&V processes are lengthy and inconvenient for the caller.

Speaker Verification is a faster and more natural process.

- There is currently no consistency to the way customers are identified and verified.

Speaker Verification will provide a more consistent interface.

2. Better Management of Resources.

- Existing processes are lengthy and use up valuable telephony resources such as ACD ports.

Speaker Verification is a faster process therefore frees up capacity for other more valuable customer services.

- Existing processes are lengthy and use up valuable agent time.

Speaker Verification is a faster process therefore frees up agent time for more productive effort.

3. Improving Agent Satisfaction.

- A significant part of every call is just spent carrying out ID&V.

Speaker Verification will remove (or shorten) the time taken.

- Agents find the existing process unrewarding.

Agents can spend their time more productively by helping or selling to the caller.

4. Improving the Security of the ID&V Process.

- Existing ID&V processes are open to fraud.

Provided the initial enrolment process is secure Speaker Verification results in a more long-term secure process.

6 Who are ICR?

ICR is the UK's leading independent provider of speech based solutions. We provide business solutions – not technology for its own sake.

Starting always with a clear appreciation of our client's business issues, we design and implement solutions harnessing the most appropriate technology drawn from our portfolio of partners.

Some solutions are pragmatic implementations of long established IVR techniques such as touch-tone input, others are truly leading edge, and draw on speech recognition, text-to-speech and speaker verification.

We also have an extensive portfolio of products, which can be implemented rapidly, and tailored to meet individual requirements.

We can create applications to run on equipment based on our customer's own premises, and also provide hosted services.

7 **Why Choose ICR for Speaker Verification deployments**

ICR has real-world experience of implementing Speaker Verification solutions, its vendor independence, experience of designing speaker verification implementation processes and delivery set us apart from the competition.

Speaker Verification initiatives we have worked on include:

- Deployment for a mobile payment application using Nuance Verifier
- Speaker Verification trial using Persay (free speech) technology for one the big 4 UK banks