Speaking of bank details...

Voice authentication is an effective tool, but must be used in conjunction with other security measures

BY STEPHEN WILSON

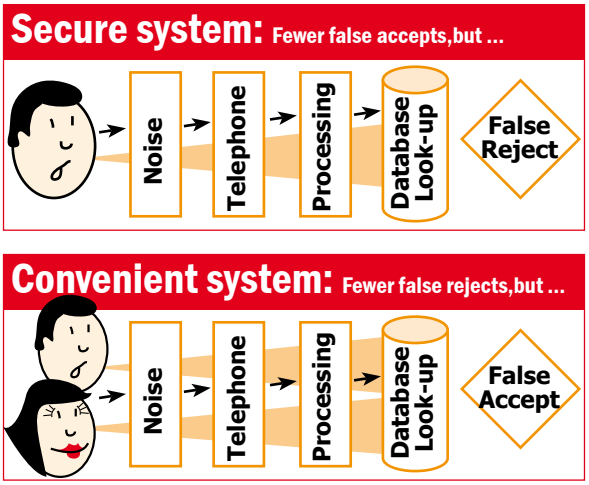
Voice authentication is one of the more interesting biometrics and probably the only technique in the class that so far makes sense for retail banking. Occasionally we hear of iris, fingerprint or face recognition being proposed for ATMs but they remain too problematic, especially for unattended teller machines.

Biometric voice authentication seems a natural fit for IVR (interactive voice response) and call centre automation. Capturing a customer's voice pattern requires no new infrastructure, and can be done remotely (provided you can first make sure the real customer is on the other end of the phone).

Voice recognition as a means of authentication must always be implemented with some sort of 'transaction dependence'; that is, each time a customer presents their spoken voice, it should be with a different phrase that in some way changes from one transaction to the next, so as to thwart replay attacks. This is similar to how an SMS one-time Internet banking code ideally must incorporate the transaction amount so that each challenge-response event is different and therefore useless if replayed.

Speech recognition algorithms are complex. You need to be able to characterise the quality of a person's voice across a wide range of speech samples, not just the one-off spectral fingerprint of a set piece of speech. We can tell this must be hard in practice, because the real world experience of commercial speech recognition software remains pretty poor. Anyone who's used these tools knows they fall well short of 100 per cent success in simply picking out words; clearly, computer modelling of how sound waves map back onto words is a research topic still in its infancy. Therefore the task of accurately recognising individual speakers by their sonic characteristics cannot be straightforward. Good security practice dictates that we should approach with caution any technique where the theoretical foundations are still being worked out.

As with any biometric, the most important design decision is probably getting the trade-off right between "false accepts" and "false rejects". This is especially critical in retail settings where customer convenience is paramount, and can even outweigh security considerations. It's vital to appreciate that any biometric system will always suffer to some extent from both types of error. On some occasions, the system will confuse the person being presented with someone else enrolled in its database (i.e. a false accept). On others, it will fail to recognise an enrolled person when they present again (i.e. a false reject). The reason that both false accepts and false rejects are inevitable has to do with the fuzziness inherent in the measurement of any biological parameter. Body features change over time, and the performance of any measurement apparatus varies as well.



Consider the many factors that interfere with the quality of the voice received: background noise, the possibility of a head cold or allergy, fatigue, the telephone equipment itself, and the transmission network, to name a few. So the voice that one presents today is always going to differ from the voice that was digitised some time ago during enrolment. The voice detector needs to be de-tuned somewhat to allow for variations that make the one individual sound different from one day to the next. But in the process, the fuzzier we make the detector, the more prone it is to confusing two different speakers as one. The need to allow for variations is illustrated in the figure, along with the unintended consequence of being too accommodating.

In retail authentication settings, where customer convenience can be as important as security, striking the right balance can be a challenge. If the false accept rate (FAR) is too high, then security suffers. But strengthening the system too far will raise the false reject rate (FRR), causing excessive demands on customers to try again, delaying service and increasing queue lengths.

Reported figures for FAR and FRR are notoriously hard to obtain from biometric vendors. Independent testing suffers from a lack of agreed standards, and results vary wildly. For instance, the US National Institute of Standards and Technology (NIST) in 2000 reported voice recognition FRRs of 10-20 per cent and FARs of two to five per cent. Testing by the University of Canberra in 2005 showed improved performance for one particular product, with FAR and FRR of less than one per cent under noise free conditions. Failure to enrol for voice recognition is typically around two per cent.

One of the lessons of all of this is to not rely on any biometric as the one-and-only authenticator. They can be most useful as an adjunct to other security methods. In call centres, voice authentication is probably best used to streamline people identifying themselves, and as a supplement to human operators. It might make it quicker to reach a non-critical function, such as an account balance query, but if the customer 'fails the test' they still get to speak with a human who can take them through the traditional security questions.

Stephen Wilson is a leading international authority on identity management and information security. In early 2004, Stephen established Lockstep Consulting to provide independent security advice and to develop new smartcard solutions to identity theft.

swilson@lockstep.com.au