Choosing your passphrase

Introduction

If you wish to use a passphrase for your speaker verification implementation, the following advice will help you select a suitable phrase.

The first thing to say is that an oft quoted phrase such as *My voice is my password*[1] perhaps isn't the best option. Its very popularity suggests that hackers are likely to target it for spoofing.

Using the same passphrase across all your applications, contact centre solutions and IVR systems could be said to be like having the same password for all of your accounts, which is never a good thing. However, there is a fundamental difference between the nature of voice biometrics and passwords, which makes that less of a concern.

Nevertheless, with corporate identity in mind, your clients are likely to demand their own, unique passphrase. Business units are unlikely to accept being forced to use the same passphrase as their neighbour. With Aculab's VoiSentry, individual users can be permitted to choose their own passphrase.

What makes a good passphrase?

In a text dependent system i.e., one reliant on a specific phrase or sequence of words, what you should be looking for is a unique arrangement, which takes about 2-3 seconds to utter.

An easy to say and remember phrase containing a minimum of 4 syllables is recommended. That's sufficient to create a viable voiceprint when enrolment involves analysing several repetitions of that passphrase.

In a text dependent system, very few phonemes are needed as long as you have enough samples. And, on the basis that you 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.

Note 1: The phrase *My voice is my password* has been trademarked (EU IPO trade mark number: EU012768156), albeit in conjunction with a logo.

VoiSentry

Enrolment vs. verification

If you are using a text dependent passphrase for verification, it makes good sense to use the same phrase for enrolment.

However, if you are implementing a text independent system, it can be said that the ideal enrolment would mean getting high phonemic content, repeated in many different contexts. In that case, the enrolment data will be sufficient, in terms of coverage of speech sounds or phonemes, to match the range of sounds expected to be encountered during verification.

Recordings with lots of syllables will generally produce more precise models and thus better verification accuracy. However, as it is rare for one speaker in a telephone dialogue to speak continuously for more than 10 to 20 seconds, enrolment recordings should be multiple, shorter passages captured throughout the duration of a conversation.

Verification is achieved during extended dialogue between the caller and an agent or through spoken responses to an IVR system, rather than by repeating a specific passphrase.

If you are implementing a text prompted system, where the response will also be recognised using ASR, several examples of each possible prompt will be required for enrolment.

If your prompt is to be a random, 4-digit sequence of numbers between 1 and 9, enrolment should consist of repeating each number several times. Counting up and down from 1 and 9 in separate recordings will suffice.

A determining factor is likely to be what is practical to do at the time of enrolment. A passive enrolment, where sufficient audio is captured during a conversation, may be less feasible than enrolling with multiple renditions of a fixed, 3-second, active passphrase or repeating a number sequence.