On estimation of the strength of evidence in forensic speaker recognition

Anil Alexander and Andrzej Drygajlo
Institut de Police Scientifique et de Criminologie, Universite de Lausanne, Switzerland ; Signal Processing Institute, Swiss Federal Institute of Technology Lausanne, Switzerland

Recently, automatic speaker recognition for forensic applications has begun to rely on the estimate of a likelihood ratio of two competing hypotheses (i.e. the suspect is the source of a questioned recording or he/she is not) as a measure of the strength of evidence. In forensic speaker recognition, statistical modeling techniques are based on the distribution of various features pertaining to the suspect's speech and its comparison to the distribution of the same features in a reference population with respect to the questioned recording. However, speaker recognition based on statistical modeling techniques such as Gaussian Mixture Modeling (GMM) has a useful property that it directly returns a likelihood of whether an utterance can come from the statistical model created for a speaker.

As a consequence, in order to calculate the likelihood ratio we propose to follow two approaches, one directly using the likelihoods returned by the GMMs, and the other by modeling the distribution of these likelihood scores and then deriving the likelihood ratio on the basis of these score distributions. The former approach is used implicitly in speaker verification systems, although in forensic speaker recognition the latter is preferred as it does not depend on the automatic speaker recognition technique used. In this paper we compare these two approaches and establish a link between them.

Since likelihood ratios calculated across the different systems may vary as a consequence of different methodologies and due to their sensitivity to the case conditions, we consider establishing measures that can be used to calibrate results across the two systems. We consider two techniques, the first minimizing errors when testing the hypotheses, and the second using Tippett plots to estimate how well the system is able to discriminate between the hypotheses. These methods are analyzed in the context of the two approaches presented in the paper to calculate likelihood ratios in forensic speaker recognition.