2013 IEEE 8th International Conference on Industrial and Information Systems, ICIIS 2013, Aug. 18-20, 2013, Sri Lanka

A Comparison of Single-Stage and Two-Stage Modelling Approaches for Automatic Forensic Speaker Recognition

Tharmarajah Thiruvaran, Eliathamby Ambikairajah, Julien Epps, Ewald Enzinger

***Abstract*—In automatic forensic speaker recognition research two frameworks, namely single-stage and two-stage modelling, are used. Although both have their own strengths and limitations, performance is an important attribute that needs to be considered when selecting these methods for forensic research. This paper compares a calibrated single-stage system with a calibrated two-stage system using a common database in terms of different performance metrics. Neither of the systems provides a consistent advantage over the other in terms of all performance measures, raising the question of whether the use of a two-stage system, which requires additional data and effort, is warranted.**

I. Introduction

Automatic forensic speaker recognition (FSR) systems can be used by a forensic scientist to produce a meaningful estimate of the strength of the evidence (information extracted from a questioned recording) if an expert opinion is required about a recording, e.g. from a crime scene relating to the suspect in judicial proceedings. Among the possible biometrics used in forensics, such as DNA and fingerprinting, the main advantage of speech is that acquiring a sample is a non-intrusive process, and speech is a common form of evidence in telephone based crimes.

Generally a court is interested to know the strength of the evidence, i.e. how likely the questioned recording was produced by the suspect, compared with how likely it was produced by someone who 'sounds like' the suspect. The likelihood ratio, according to the Bayesian interpretation, is a scientific way to measure the strength of the evidence [1]. The likelihood ratio (LR) is given in equation (1), where Pr(.) denotes the probability, H_0 is the hypothesis that the suspected speaker is the source of the questioned recording, $H_1$is the hypothesis that the suspected speaker is not the source of the questioned recording and E is the information extracted from the questioned recording.

$$LR = \frac{\Pr(E \mid H_0)}{\Pr(E \mid H_1)} \qquad (1)$$

A two-stage statistical approach ('double-statistical')was proposed to calculate the likelihood ratio based on the Bayesian interpretation in [2] and it is widely used [2-5]. Another attractive property of this approach is that the second stage is modelled as a univariate probability distribution, so it is easy to articulate in layman's terms for the judge, the attorneys and jury[6]. This increases the transparency of the system, which is an essential attribute for the admissibility of scientific evidence in court, and transparency is one of the reasons for the unanimous acceptance of DNA [7].

On the other hand, a single-stage modelling approach is also used extensively in forensic speaker recognition to calculate the likelihood ratio [7-12]. Even though the two-stage approach has been specifically designed to consider legal requirements, the wide use of single-stage systems in forensic research may be due to one or more of the following reasons: (i) lack of a comprehensive database suitable for two-stage modelling, (ii) the availability of highly developed single-stage systems mainly driven by NIST evaluations and (iii) Brummer's introduction of calibration as a mapping from the output of any single-stage system to log likelihood ratio that can be assessed using any forensic performance metrics [9].

The system used in acoustic–phonetic based forensic speaker recognition,which employs a multivariate kernel density estimation approach [12, 13], can also be considered as a single-stage system.

As there are two approaches to estimate LR in automatic FSR, choosing a method is a crucial decision. In order to facilitate this decision process, this paper presents a comparison, mainly based on performance, between these two approaches. In particular, the purpose of this paper is to objectively answer a possible question that can arise in a court case where a calibrated signal-stage system is used, of why such system is used while a calibrated two-stage system is available. So this paper compares the results when calibrated single-stage and two-stage systems are used on the same database with two different front-ends. Previously single-stage and two-stage systems were applied to the same database of NIST 2004 and the DET curves of both systems are given in [4], however, the single-stage system is an un-calibrated system. Further, the main objective of that work was to assess the proposed robust LR computation procedures for the two-stage approach, and not to assess whether the single-stage or two-stage is better. Presumably this is because the single-stage

T. Thiruvaran is with the school of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, Australia (corresponding author thiru@ee.unsw.edu.au)

E. Ambikairajah, J. Epps, E. Enzinger are with both school of Electrical Engineering and Telecommunications, The University of New South Wales, and 2ATP Research Laboratory, National ICT Australia (NICTA), Sydney Australia. (email: ambi@ee.unsw.edu.au, j.epps@unsw.edu.au, e.enzinger@student.unsw.edu.au)

978-1-4799-0910-0/13/$31.00 ©2013 IEEE