

RESEARCH

Open Access

# An investigation of supervector regression for forensic voice comparison on small data

Chee Cheun Huang<sup>1,2\*</sup>, Julien Epps<sup>1,2</sup> and Tharmarajah Thiruvaran<sup>1</sup>

## Abstract

Automatic forensic voice comparison (FVC) systems employed in forensic casework have often relied on Gaussian Mixture Model - Universal Background Models (GMM-UBMs) for modelling with relatively little research into supervector-based approaches. This paper reports on a comparative study which investigates the effectiveness of multiple approaches operating on GMM mean supervectors, including support vector machines and various forms of regression. Firstly, we demonstrate a method by which supervector regression can be used to produce a forensic likelihood ratio. Then, three variants of solving the regression problem are considered, namely least squares and  $\ell_1$  and  $\ell_2$  norm minimization solutions. Comparative analysis of these techniques, combined with four different scoring methods, reveals that supervector regression can provide a substantial relative improvement in both validity (up to 75.3%) and reliability (up to 41.5%) over both Gaussian Mixture Model - Universal Background Models (GMM-UBMs) and Gaussian Mixture Model - Support Vector Machine (GMM-SVM) results when evaluated on two studio clean forensic speech databases. Under mismatched/noisy conditions, more modest relative improvements in both validity (up to 41.5%) and reliability (up to 12.1%) were obtained relative to GMM-SVM results. From a practical standpoint, the analysis also demonstrates that supervector regression can be more effective than GMM-UBM or GMM-SVM in obtaining a higher positive-valued likelihood ratio for same-speaker comparisons, thus improving the strength of evidence if the particular suspect on trial is indeed the offender. Based on these results, we recommend least squares as the better performing regression technique with gradient projection as another promising technique specifically for applications typical of forensic case work.

**Keywords:** Forensic voice comparison (FVC); Likelihood ratio; Reliability; Supervector regression; Validity

## 1 Introduction

Forensic voice comparison (FVC) systems have often employed Gaussian Mixture Model - Universal Background Models (GMM-UBMs) [1-3] for modelling in forensic casework, in which it is common that only a very small speech database is available for the entire system development. Other approaches, such as the supervector-based regression techniques prevalent in numerous face and speaker recognition studies [4-6], have received little attention in this context. This therefore motivates a comparative study of the effectiveness of other modelling approaches in FVC system performance.

The likelihood ratio is defined as the likelihood that the evidence would be observed if the same-origin

hypothesis were true divided by the likelihood that the evidence would be observed if the different-origin hypothesis were true [7-11].

$$LR = \frac{p(E|H_{so})}{p(E|H_{do})} \quad (1)$$

where  $LR$  is the likelihood ratio,  $H_{so}$  is the same-origin hypothesis and  $H_{do}$  is the different-origin hypothesis.  $E$  is the evidence or the observed property of a speech sample.  $p(E|H_{so})$  denotes conditional probability density of the evidence given same-origin hypothesis. An FVC system typically relies on statistical evaluation of input speech utterances that first involves training or modelling of the speaker identity based on an input speech utterance A and a subsequent testing of the trained model based on an input speech utterance B.

The initial output of the FVC system is defined as a score  $s$ . A higher valued score can be interpreted as

\* Correspondence: cheecheunh@hotmail.com

<sup>1</sup>School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, NSW 2052, Australia

<sup>2</sup>National ICT Australia (NICTA), Australian Technology Park, Sydney, NSW 1430, Australia