
Abstract

Nowadays, state-of-the-art automatic speaker recognition systems show very good performance in discriminating between voices of speakers under controlled recording conditions. However, the conditions in which recordings are made in investigative activities (e.g., anonymous calls and wire-tapping) cannot be controlled and pose a challenge to automatic speaker recognition. Differences in the phone handset, in the transmission channel and in the recording devices can introduce variability over and above that of the voices in the recordings. The strength of evidence, estimated using statistical models of within-source variability and between-sources variability, is expressed as a likelihood ratio, i.e., the probability of observing the features of the questioned recording in the statistical model of the suspected speaker's voice, given the two competing hypotheses: the suspected speaker is the source of the questioned recording and the speaker at the origin of the questioned recording is not the suspected speaker. The main unresolved problem in forensic automatic speaker recognition today is that of handling mismatch in recording conditions. Mismatch in recording conditions has to be considered in the estimation of the likelihood ratio.

The research in this thesis mainly addresses the problem of the erroneous estimation of the strength of evidence due to the mismatch in technical conditions of encoding, transmission and recording of the databases used in a Bayesian interpretation framework.

We investigate three main directions in applying the Bayesian interpretation framework to forensic automatic speaker recognition casework. The first addresses the problem of mismatched recording conditions of the databases used in the analysis. The second concerns introducing the Bayesian interpretation methodology to aural-perceptual speaker recognition as well as comparing aural-perceptual tests performed by laypersons with an automatic speaker recognition system, in matched and mismatched recording conditions. The third addresses the problem of variability in estimating the likelihood ratio, and several new solutions to cope with this variability are proposed.

Firstly, we propose a new approach to estimate and statistically compensate for the effects of mismatched recording conditions using databases, in order to estimate

parameters for scaling distributions to compensate for mismatch, called ‘scaling databases’. These scaling databases reduce the need for recording large databases for potential populations in each recording condition, which is both expensive and time consuming. The compensation method is based on the principal Gaussian component in the distributions. The error in the likelihood ratios obtained after compensation increases with the deviation of the score distributions from the Gaussian distribution. We propose guidelines for the creation of a database that can be used in order to estimate and compensate for mismatch, and create a prototype of this database to validate the methodology for compensation.

Secondly, we analyze the effect of mismatched recording conditions on the strength of evidence, using both aural-perceptual and automatic speaker recognition methods. We have introduced the Bayesian interpretation methodology to aural-perceptual speaker recognition from which likelihood ratios can be estimated. It was experimentally observed that in matched recording conditions of suspect and questioned recordings, the automatic systems showed better performance than the aural recognition systems. In mismatched conditions, however, the baseline automatic systems showed a comparable or slightly degraded performance as compared to the aural recognition systems. Adapting the baseline automatic system to mismatch showed comparable or better performance than aural recognition in the same conditions.

Thirdly, in the application of Bayesian interpretation to real forensic case analysis, we propose several new solutions for the analysis of the variability of the strength of evidence using bootstrapping techniques, statistical significance testing and confidence intervals, and multivariate extensions of the likelihood ratio for handling cases where the suspect data is limited.

In order for forensic automatic speaker recognition to be acceptable for presentation in the courts, the methodologies and techniques have to be researched, tested and evaluated for error, as well as be generally accepted in the scientific community. The methodology presented in this thesis is viewed in the light of the Daubert (USA, 1993) ruling for the admissibility of scientific evidence.

Version abrégée

A l'heure actuelle, les systèmes de reconnaissance automatique de locuteur possèdent d'excellentes performances lorsqu'il s'agit de discriminer entre des voix de locuteurs acquises dans des conditions contrôlées. Pourtant, les conditions dans lesquelles la police effectue ses enregistrements (dans des cas d'appels anonymes ou d'écoutes téléphoniques), ne peuvent être contrôlées et sont donc un défi difficile pour la reconnaissance automatique de locuteurs. Des différences dans les combinés téléphoniques, dans le canal de transmission et dans l'appareil d'enregistrement peuvent introduire des variations plus importantes que celles des voix sur les enregistrements. La force probante de la preuve, estimée à l'aide de modèles statistiques des intra- et inter-variabilités de la source, est exprimée sous la forme d'un rapport de vraisemblance, i.e., est-ce plus probable que l'enregistrement de question (trace) a été produit par un locuteur suspecté ou par toute autre personne appartenant à la population pertinente. Le principal problème non résolu dans la reconnaissance automatique de locuteurs en sciences forensiques est la manière de traiter les conditions d'enregistrement différentes. Les conditions d'enregistrement différentes doivent être considérées dans l'estimation du rapport de vraisemblance.

Ce travail de thèse traite essentiellement du problème lié à l'estimation erronée de la force probante de la preuve due aux conditions différentes d'encodage, de transmission et d'enregistrement des bases de données utilisées dans un canevas d'interprétation bayésienne.

Nous avons approfondi trois directions principales en appliquant le canevas d'interprétation bayésienne à des cas forensiques de reconnaissance automatique de locuteurs. La première direction traite du problème des conditions différentes d'enregistrement des bases de données utilisées lors des analyses. La deuxième direction concerne l'introduction de la méthodologie d'interprétation bayésienne à la reconnaissance auditive de locuteurs et la comparaison entre les capacités de reconnaissance vocale de personnes sans connaissance en phonétique et d'un système automatique, dans des conditions semblables et différentes. La troisième direction traite du problème de l'incertitude dans l'estimation du rapport de vraisemblance, et plusieurs approches novatrices pour faire face à cette incertitude sont proposées.

Premièrement, nous avons proposé une approche novatrice pour estimer et compenser statistiquement les effets des conditions différentes d'enregistrement des bases de données, afin d'estimer les paramètres pour balancer les distributions, dans le but de compenser pour les conditions différentes appelées 'scaling databases'. Ces scaling databases réduisent le besoin d'enregistrer de grandes bases de données représentant les populations potentielles dans toutes les conditions d'enregistrement, ce qui demande beaucoup de temps et d'argent. La méthode de compensation est basée sur la gaussienne principale dans les distributions. L'erreur dans le rapport de vraisemblance obtenu après compensation augmente avec la déviation des distributions des scores de la distribution gaussienne. Nous proposons des directives pour la création d'une base de données pouvant être utilisée afin d'estimer et de compenser les conditions différentes, et avons créé un prototype de cette base de données dans le but de valider notre méthodologie de compensation.

Deuxièmement, nous avons analysé les effets des conditions différentes d'enregistrement sur la force probante de la preuve, pour la reconnaissance vocale de personnes sans connaissance en phonétique et d'un système automatique. Nous avons introduit la méthodologie d'interprétation bayésienne à la reconnaissance auditive par des profanes et nous avons estimé les rapports de vraisemblance correspondants. Nous avons pu observer expérimentalement que dans des conditions semblables pour les enregistrements de question et de référence, le système automatique obtient de meilleures performances par rapport à l'approche auditive. Cependant, dans des conditions différentes, le système automatique de référence a obtenu des performances comparables ou légèrement moins bonnes par rapport à l'approche auditive. En adaptant le système de reconnaissance de référence à des conditions différentes, le système obtient des performances comparables, voire meilleures, par rapport à l'approche auditive.

Troisièmement, dans l'application de l'interprétation bayésienne à des analyses de cas forensiques réels, nous proposons plusieurs approches novatrices pour l'analyse de la variabilité de la force probante de la preuve, en utilisant des techniques de 'bootstrapping', des tests statistiques de signification et des intervalles de confiance, ainsi que des extensions multivariées du rapport de vraisemblance pour traiter des cas dans lesquels les données du suspects sont limitées.

Afin que la valeur probante de la preuve, obtenue dans l'analyse de cas de reconnaissance automatique de locuteurs en sciences forensiques, puisse être présentée à la cour, les méthodologies et les techniques utilisées doivent pouvoir être testées, le taux d'erreurs engendré par la méthode doit être connu, et la méthode doit faire l'objet d'une acceptation générale de la communauté scientifique. La méthodologie présentée dans cette thèse est évaluée à la lumière des critères de l'affaire Daubert (USA, 1993), qui fixe les normes et standards au niveau de l'admissibilité des preuves scientifiques dans un tribunal.