Preface

Waleed H. Abdulla

Department of Electrical and Computer Engineering,
The University of Auckland,
Private Bag 92019, Auckland, New Zealand
E-mail: w.abdulla@auckland.ac.nz

Sadaoki Furui

Department of Computer Science,
Tokyo Institute of Technology,
2-12-1 Ookayama, Meguro-ku, Tokyo,
152-8552 Japan
E-mail: furui@cs.titech.ac.jp

Kuldip K. Paliwal

Signal Processing Laboratory,
School of Engineering,
Griffith University,
Brisbane, QLD 4111, Australia
E-mail: K.Paliwal@griffith.edu.au

Biographical notes: Waleed H. Abdulla holds a PhD from the University of Otago, New Zealand. Since 2002, he has been working in the Department of Electrical and Computer Engineering/The University of Auckland. His main research interests are in human biometrics, speech and audio signal processing. He has been serving as Editorial Board member of several journals and Vice-President of Member Relations and Development of APSIPA Association. He published more than 80 refereed publications including a patent and a book, and supervised more than 25 postgraduate students. He has been recently awarded JSPS. He received the award for Excellent Teaching for the year 2005. He is a member of ISCA, IEEE and IET.

Sadaoki Furui received BS, MS and PhD in Mathematical Engineering and Instrumentation Physics from Tokyo University, Tokyo, Japan, in 1968, 1970 and 1978, respectively. He is engaged in a wide range of research on speech analysis, speech recognition, speaker recognition, speech synthesis and multimodal human–computer interaction and has authored or co-authored over 800 published articles. He has received Paper Awards and Achievement Awards from the IEEE, the IEICE, the ASJ, the ISCA, the Minister of Science and Technology, and the Minister of Education, and the Purple Ribbon Medal from the Japanese Emperor.

Kuldip K. Paliwal received BS, MS and PhD from Agra University (in 1969), Aligarh University (in 1971) and Bombay University (in 1978), respectively. He has been carrying out research in the area of speech processing since 1972 and has published more than 250 papers. He has served as an Associate Editor

of the *IEEE Transactions on Speech and Audio Processing* and the *IEEE Signal Processing Letters*. He has received IEEE Signal Processing Society's best (senior) paper award in 1995 for his paper on LPC quantisation. He is currently serving the Speech Communication journal (published by Elsevier) as its Editor-in-Chief.

The 2001 MIT Technology Review indicated that *biometrics is one of the emerging technologies that will change the world*. Human Biometrics is automated recognition of a person using adherent distinctive physiological and/or involuntary behavioural features.

Human voice biometrics has gained significant attention in recent years. The ubiquity of cheap microphones, human identity information carried by voice, ease of deployment, natural use, telephony applications dispersion, and non-obtrusiveness have been big motivations for developing biometrics based on speech signal. The robustness of speech biometric is sufficiently good. However, there are significant adversaries with respect to conditions that cannot be controlled easily. These issues include changes in acoustical environmental conditions, respiratory and vocal pathology, age, channel, etc. The goal of speech biometric research is to solve and/or mitigate these problems.

This special issue is an attempt to tackle some of the difficulties through selective papers presented by leading researchers and investigators in speech research for security applications.

There are six papers in this special issue. The first paper is: 'The effect of correlation on strength of evidence estimates in Forensic Voice Comparison: uni- and multivariate Likelihood Ratio-based discrimination with Australian English vowel acoustics' by Phil Rose. He shades light on the importance correlation between features in traditional forensic speaker recognition and the consequences of ignoring this attribute. Two likelihood ratio-based discrimination experiments on the same multivariate formant data are described. One is taking correlation into account and the other one not. The discrimination is performed using Naïve Bayes univariate, and multivariate generative likelihood ratios as discriminant functions, exemplified with Tippett plots and evaluated with the C_{llr} cost function. It is shown that ignoring within-segment correlation can result in considerable over- or underestimation of the strength of evidence when traditional features are used, and poorer overall discrimination between same-speaker and different-speaker pairs. The use of logistic-regression fusion to handle between-segment correlation is also demonstrated.

The second paper entitled 'Voice biometric feature using Gammatone filterbank and ICA' by Waleed H. Abdulla and Yushi Zhang introduces a new feature to be used for speaker identification systems. The presented paper proposes a robust feature extraction technique for the purpose of speaker identification. The technique is based on processing monaural speech signal using Gammatone auditory filterbank (GTF) and Independent Component Analysis (ICA). The proposed feature is evaluated in real environments under varying noisy conditions. The proposed feature is benchmarked against the commonly used features such as: MFCC, PLCC, and PLP, and it outperforms them in different noisy environments.

The third paper entitled 'Using MMSE to improve session variability estimation' by Gang Wang and Thomas Fang Zheng deals with the mismatch problem caused by session variability in speaker recognition. The authors improve the Session Variability Subspace Projection (SVSP) method based on model compensation for speaker verification by using a training method based on the Minimum Mean Square Error (MMSE) criterion. The issue of SVSP is that session-independent supervector of a speaker is approximated by the average of the speaker's all session-dependent GMM supervectors when estimating the SVSP matrix. The research goal is to minimise the error based on the MMSE criterion when estimating the SVSP matrix. Compared with the original SVSP, the proposed method could achieve an error rate reduction of 6.7% for Equal Error Rate (EER) and 5.3% for Minimum Detection Cost Function (Min-DCF) over the NIST SRE 2006 single-side one conversation training, single-side one conversation test.

The fourth paper is 'On the use of perceptual Line Spectral Pairs Frequencies and higher-order residual moments for Speaker Identification' by Md. Sahidullah, Sandipan Chakroborty and Goutam Saha. The paper investigates extracting the Line Spectral Frequency (LSF) feature from perceptually modified speech signal for speaker identification task. It is an attempt to integrate perceptual analysis with LSF to obtain improved speaker identification performance. The paper also presented a novel feature set extracted from the residual signal using Higher-Order Statistical Moments (HOSM). Residual signal contains information related to vocal cord characteristics which is complementary to spectral characteristics in identifying speakers. HOSM based feature is fused with the spectral feature to improve the identification performance. The proposed system is benchmarked using different speech databases.

The fifth paper entitled 'Text-Independent speaker identification in phoneme-independent subspace using PCA transformation' by Haoze Lu, Masafumi Nishida, Yasuo Horiuchi and Shingo Kuroiwa discusses the implication of variation of phonetic information on the performance of speaker identification systems. The authors propose a text-independent speaker identification method that suppresses the phonetic information by a subspace method. It depends on the valid assumption that a subspace with large variance in the speech feature space is a "phoneme-dependent subspace" and a complementary subspace of it is a "phoneme-independent subspace". Principal Component Analysis is employed to construct these subspaces. GMM-based speaker identification experiments using both a new feature vector of the proposed method and the conventional MFCC are demonstrated. The results show that the proposed method is effective for decreasing the identification error rates by suppressing the phonetic information.

The last paper carrying the title 'Neuro-fuzzy-based biometric system using speech features' by Anupam Shukla, Ritu Tiwari and Chandra Prakash Rathore investigate using a multitude of features to improve the performance of speaker identification systems. In this paper, a biometric system is simulated using several speech features, which identifies speakers along with their gender and mental status. The work is broadly classified into two main parts. The first part deals with the extraction of speech features; namely pitch, amplitude, number of zero-crossing, and average power spectral density. In the second part a simulation model has been developed which is based on the technique of Adaptive Neuro-Fuzzy Inference System for speaker identification along with their gender and mental status. The recognition score varies depending upon the different input-output membership functions.