



# 17<sup>th</sup> Interpol International Forensic Science Managers Symposium, Lyon

8<sup>th</sup> - 10<sup>th</sup> October 2013

## Review Papers

Edited by Prof. Niamh Nic Daéid

Centre for Forensic Science, University of Strathclyde,

Glasgow, UK

## TABLE OF CONTENTS

<b>PREFACE</b>	<b>3</b>
Mr Nelson Santos, Chairman of the Organising Committee	
<b>CRIMINALISTICS</b>	<b>5</b>
Technical co-ordinator :- Prof. Kimmo Himberg	
Examination of Firearms	6-43
Examination of Firearms – Gun Shot Residue	44-66
Forensic Examination of Marks	67-128
Examination of Paint	129-174
Forensic Examination of Fibres and Textiles	175-205
Forensic Geology	206-229
<b>FORENSIC CHEMISTRY</b>	<b>230</b>
Technical co-ordinator :- Prof Niamh NicDaeid	
Fire Cause Investigation and Fire Debris Analysis	231-279
Analysis and Detection of Explosives	280-435
Drug Evidence	436-524
Toxicology	526-610
<b>MEDIA EVIDENCE</b>	<b>611</b>
Technical co-ordinator :- Dr. Peter Pfefferli,	
Forensic Audio Analysis	612-637
Forensic Video Analysis	638-651
Imaging	652-687
Digital Evidence	688-743
<b>IDENTIFICATION SCIENCES</b>	<b>744</b>
Technical co-ordinator :- Dr. Kevin Sullivan	
Fingermarks and other Impressions	745-820
Body Fluid Identification and DNA Typing	821-853
Questioned Documents	854-897
<b>Forensic Science Management</b>	<b>898-923</b>

# **Forensic Audio Analysis**

## **Review: 2010–2013**

Catalin Grigoras, Ph.D.<sup>1</sup>

Jeff M. Smith, M.Sc.<sup>1</sup>

Geoffrey Stewart Morrison, Ph.D.<sup>2</sup>

Ewald Enzinger, M.Phil.<sup>2</sup>

<sup>1</sup> National Center for Media Forensics, University of Colorado Denver

<sup>2</sup> Forensic Voice Comparison Laboratory, School of Electrical Engineering and Telecommunications, University of New South Wales

## TABLE OF CONTENTS

<b>1 Introduction</b>	614
<b>2 Authentication (Catalin Grigoras)</b>	614
2.1. <i>Individual Techniques</i>	614
2.2. Forensic Audio Authentication Framework	615
<b>3. Forensic Speech Science (Geoffrey Stewart Morrison &amp; Ewald Enzinger)</b>	616
3.1 <i>Reviews And Introductions</i>	616
3.2 Survey Of Approaches And Interpretive Frameworks Used By Forensic-Voice-Comparison Practitioners	616
3.3 <i>Paradigm Change</i>	617
3.4 Empirical Research On Forensic Voice Comparison Conducted Within The New Paradigm	618
3.5 Empirical Research On Forensic Voice Comparison Not Conducted Within The New Paradigm	621
3.6 <i>Disputed Utterance Analysis</i>	622
3.7 <i>Voice-Based Lie Detectors</i>	623
<b>4. Audio Enhancement (Jeff M. Smith)</b>	623
4.1 <i>Introduction</i>	623
4.2 <i>Reference Works</i>	624
4.3 Enhancement Of Monaural And Binaural Recordings And Future Areas Of Research	624
4.4 Speech Quality Vs. Speech Intelligibility	625
<b>5. Organizations</b>	626
<b>6. Acknowledgements</b>	627
<b>7. References</b>	627

## 1 Introduction

This report is a follow up to the review prepared for the 16th Interpol International Forensic Science Symposium in October 2010, and catalogues the research, advances, and application of scientific methodologies and techniques relating to the forensic examination of audio evidence. This report primarily consists of a literature review of published articles in forensic science journals and the proceedings of various working groups and forensic conferences between July 2010 and July 2013. It also contains references from other sources such as the Internet.

## 2 Authentication (Catalin Grigoras)

Forensic audio authentication research focused on two major topics: individual techniques to assess authenticity and a general methodology to authenticate digital audio recordings.

### *2.1. Individual techniques*

File structure including header analysis is an important stage in digital audio authentication. Koenig & Lacey [1] presented a methodology to investigate the Olympus WMA (Windows Media Audio format) headers, showing that numerous differences between original and edited WMA files can be found for forensic purposes.

Compression history identification is another important issue in digital audio authentication. Shen et al [2] presented a method to discriminate between D-compressed AMR audio recordings and S-compressed, showing that more research is need to apply this technique on real data. Luo et al [3] proposed a method to assess the compression history of WAV files that have been previously compressed by MP3 or WMA and estimating hidden compression rates.

Other digital recording characteristics can also be used for digital audio authentication. Direct Current (DC) analysis is another technique presented by Koenig et al [4], showing its limits and the possibility to use it in forensic audio.

Malik & Miller [5] proposed a statistical framework for microphone identification explaining the effectiveness using ambient noise recordings, while in another paper, Malik and Zhao [6] propose a method to analyze acoustic reverberations. Pan et el. [7] described a fast and blind local noise level estimation method that can be employed to detect digital audio forgeries.

Chen et al [8] introduced a singularity analysis of the wavelet packet decomposition to detect speech audio forged operations in time domain.

Boss [9] presented the results of using ripple signals for digital audio analysis, showing that ripple signals are of high inter-local and very low intra-local variability.

The ENF analysis continued to garner attention and more studies have been run to extract, analyze, and compare ENF. Bao [10] proposed a new method to extract ENF using fractional Fourier Transform, while Su et al [11] explained a solution to separate the ENF components from recaptured audio recordings. Nicolalde et al [12][13] and Coetzee [14] presented detailed methods to investigate ENF amplitude and phase continuity of digital audio recordings. A correction algorithm for the effect of oscillator errors on ENF was proposed by Yuan et al [15]. Liu et al [16] presented a study of the accuracy and precision of quadratic frequency interpolation for ENF estimation. Yuan et al [17] described how simple Monte Carlo techniques and a database of grid reference data can be used to determine the operational parameters of the ENF matching process. Archer [18] analyzed the effects of lossy compression algorithms on ENF showing that hum is robust against the investigated codecs. Grigoras et al [19] presented a detailed database configuration for forensic analysis of ENF, while Jenkins & Steinhour [20] proposed developments to minimize ENF database corruption and system errors. Grigoras & Smith [21] reported advances in forensic ENF analysis including techniques to extract it, statistical tools for automatic search and analysis against a database, possible problems and proposed solutions to minimize measurement errors, database validation testing, and sample cases.

## ***2.2. Forensic audio authentication framework***

Korycki [22][23] presented different techniques for time-frequency investigations, tampering detection, and discussed the main methods used for authenticity analysis, including ENF and MP3 compression.

Gupta et al [24] explained recent developments in the audio authentication field including basic, preliminary audio analysis and advanced audio authentication techniques that exploit audio recording conditions and compressed audio features.

Grigoras et al [25] proposed an analytical framework for digital audio authentication including a neutral methodology to interpret and report the results.

Koenig & Lacey [26] presented an inconclusive digital audio authenticity examination, concluding that the recordings could not be scientifically authenticated through accepted forensic practices.

### **3. Forensic Speech Science (Geoffrey Stewart Morrison & Ewald Enzinger)**

This section of the review focuses primarily on recent research on forensic voice comparison, but also briefly discusses recent research on disputed utterance analysis and on voice-based lie detectors. The review aims to be relatively comprehensive, but has not attempted to include all papers published in the area, and has for the most part ignored conference presentations not associated with archived proceedings papers.

#### ***3.1 Reviews and introductions***

In 2010 Jessen [27] presented an overview of speaker profiling and of acoustic-phonetic approaches to forensic voice comparison, and Morrison [28] published an introduction to forensic voice comparison and evaluation of evidence intended to be accessible to a broad audience including lawyers. The latter also included a review of research on speaker identification by naïve listeners. In 2012 Amino et al [29] published a historical review of forensic voice comparison, and Perrot & Chollet [30] published a paper including a review of voice disguise techniques and their detection.

#### ***3.2 Survey of approaches and interpretive frameworks used by forensic-voice-comparison practitioners***

Gold & French [31] surveyed forensic-voice-comparison practitioners as to their approaches and methodologies for evaluating the strength of evidence and the interpretive frameworks which they employed. The results were published in 2011 and included responses from 36 practitioners from 13 countries. The most striking finding, although not unexpected for those familiar with the field, was the lack of consistency across practitioners.

In terms of approach, the majority of practitioners (25 practitioners) based their evaluations on a mixture of acoustic-phonetic measurements and listening (although it was not clear how they combined the two), three used listening only (auditory approach), one used an acoustic-phonetic approach without an auditory component, and eight used an approach characterized as human-assisted automatic speaker recognition (although it was not clear what the human-assisted part consisted of and how it combined with the automatic part). Although not mentioned in Gold & French [31], the aural-spectrographic approach is also still practiced in multiple parts of the world (Morrison [32]).

In terms of interpretive framework, two practitioners made binary decisions (which logically require the imposition of a threshold on a posterior probability), and the largest group (14 practitioners) used verbal expressions from an ordinal posterior-probability scale (not necessarily the same scale across different practitioners). The second largest group (11 practitioners) used the so-called UK framework (French & Harrison [33]), and most of the remainder (7 practitioners) used the likelihood-ratio framework (4 presenting

numeric likelihood-ratio values, and 3 presenting verbal expressions). The use of posterior probabilities by forensic scientists has been criticized (e.g., Balding [34], Buckleton [35], Evett [36], Robertson & Vignaux [37]) as it logically requires consideration of prior probabilities, e.g., the trier of fact's belief as to the relative probabilities of the same-speaker versus the different-speaker hypothesis before the strength of the forensic-voice-comparison evidence is presented. The trier of fact's prior probabilities with respect to the forensic-voice-comparison evidence will usually be influenced by other evidence already presented in the trial. The forensic scientist cannot know what the trier of fact's prior probabilities will be, and they should not be exposed to other (task-irrelevant) evidence in the trial which could bias their estimate of the strength of the particular evidence which they have been asked to assess. The UK framework has also been criticized as being logically inconsistent, overly vague, and suffering from cliff-edge effects (Rose & Morrison [38], Morrison [28] [39]; see also the response in French et al [40]).

It would seem unlikely that there will be a substantial decrease in the fragmentation of the field in the short term, particularly at the practitioner level, but there are ongoing trends affecting both forensic science in general and forensic voice comparison in particular which are reflected in much of the research conducted over the last three years.

### **3.3 Paradigm change**

One ongoing trend in the literature on the evaluation and interpretation of forensic evidence in general is the call to adopt the likelihood-ratio framework as the only logically correct framework. This has been strengthened over the last three years due to the response to the 2010 England & Wales Court of Appeal Ruling in *R v T* [2010 EWCA Crim 2439], e.g., Evett et al [41], Berger et al [42], Redmayne et al [43], Robertson et al [44], Morrison [45]. It was also at the core of the Royal Statistical Society's first practitioner guide for judges, lawyers, forensic scientists and expert witnesses published in 2010 (Aitken et al [46]), and a major focus of the National Institute of Standards and Technology and National Institute of Justice (NIST/NIJ) 2012 report on latent fingerprint analysis [47]. Forensic voice comparison conducted within the likelihood-ratio framework has a history going back to the mid-to-late 1990s as can be traced through earlier Interpol Forensic Science Symposium review papers: Broeders [48][49], Bijhold et al [50], Kriigel et al [51]. Morrison [39] presented a history up to 2009 of the adoption of the likelihood-ratio framework for forensic voice comparison. Researchers and practitioners in forensic speech science fully committed to the use of the likelihood-ratio framework are, however, probably still a minority of those working in the field.

Another ongoing trend affecting forensic science in general is pressure to assess the validity and reliability of analytic approaches and methodologies. Calls for this have recently been published in the 2009 National Research Council Report on Strengthening Forensic Science in the United States (NRC [52]), and in the aforementioned 2012 NIST/NIJ fingerprint report [47].



Morrison [32] reviewed calls, from the 1960s onward, for empirical testing of the validity and reliability of forensic-voice-comparison approaches and methodologies under conditions reflecting those of the case under investigation. Morrison and colleagues [28] [32] [39] [45] [53] have proposed that the field of forensic voice comparison is undergoing a paradigm shift (also affecting forensic science in general), and that the use of the likelihood-ratio framework and the empirical testing of the validity and reliability of approaches and methodologies under conditions reflecting those of the case under investigation are two essential elements of the new paradigm.

Morrison and colleagues have also proposed the use of relevant data, quantitative measurements, and statistical models as a highly preferred element of the new paradigm because such an approach is more transparent, more easily replicated, and more easily tested than an approach in which the output of the system is based directly on the subjective experience-based judgment of a human expert. This last element is presented as highly preferred rather than essential because it must be subservient to the testing element – whichever system performs the best under the conditions of the case at trial should be employed. Morrison and colleagues have described concrete procedures for collecting and selecting relevant data [53] [54] and concrete procedures and metrics for testing validity and reliability [55] [56] [57] (see also Ramos & González-Rodríguez [58]).

### ***3.4 Empirical research on forensic voice comparison conducted within the new paradigm***

Papers reviewed in this section describe empirical studies which were, to a greater or lesser extent, conducted within the new paradigm, i.e., to a greater or lesser extent likelihood ratios were calculated on the basis of data, quantitative measurements, and statistical models, and the validity and reliability of the system was tested, and the training and test data were representative of the relevant population and reflective of the recording conditions of some real or imagined forensic case.

Becker et al [59] and Solewicz et al [60] compared the performance of several automatic forensic-voice-comparison systems on a test database of recordings taken from actual forensic cases (the data suffered from a degree of heterogeneity). Systems tested were two in-house systems employed by the German Federal Criminal Police Office (BKA) (see Becker et al [61] [62] [63] for detailed descriptions of these systems, SPES and VoCS), three in-house systems employed by the Israeli National Police, and two commercial systems employed by the French Police Technique et Scientifique. The performance of the different systems was broadly similar, although relative to the other systems one system had a bias towards good performance on same-speaker trials at the cost of poor performance on different-speaker trials and another had a bias towards good performance on different-speaker trials at the cost of poor performance on same-speaker trials. The authors discussed the importance of selecting appropriate data for modeling the population, including language spoken, and/or compensation techniques to

account for mismatches between the training and test data, including mismatches in recording duration.

In 2012 Rose [64] described how likelihood ratios had been calculated from fundamental-frequency and formant-frequency measurements made on the word “yes” and the phrase “not too bad” in an actual forensic case for which the analysis was conducted in 2007. The offender recording was from a telephone call and the suspect recordings from police interviews. Rose also commented on advances made in forensic-voice-comparison research since that time.

Enzinger [65] published a preliminary report on a study based on the conditions of an actual forensic case. The case was somewhat atypical: There were two speakers on a single mobile-to-landline telephone recording. The identity of the speaker of a two-second-long utterance within the recording was in question, but it had to be one of the two aforementioned speakers. In most of the training data, one speaker was relatively far from the microphone and one relatively close, but the questioned utterance was close. The paper illustrated procedures for calculating a likelihood ratio under the conditions of this case using relevant data, quantitative measurements (cepstral coefficients in this case), and statistical models, and procedures for testing the validity and reliability of the forensic-voice-comparison system under the conditions of this case, i.e., it provided an example of how to implement the new paradigm under actual casework conditions.

A number of forensic-voice-comparison studies have investigated the effectiveness of extracting acoustic information by fitting parametric curves to the formant trajectories (and for tonal languages fundamental-frequency trajectories) of tokens of selected vowel phonemes (e.g., Chen & Rose [66], Enzinger [67], Hughes [68], Jialin & Rose [69], Li & Rose [70], Morrison [71] [72], Pingai et al [73], Rhodes [74]) and assessing whether adding these features to a baseline system (e.g., mel frequency cepstral coefficients, MFCCs, fitted to the entire speech-active sections of the recordings) leads to improvement in performance over the baseline system (e.g., Zhang et al [75] [76] [77] [78]). Initial results using high-quality voice recordings were promising, but studies using various combinations of landline- and mobile-telephone-transmitted voice recordings found little or no meaningful improvement in performance over a much cheaper baseline system, especially when mobile telephones were involved (Zhang et al [77] [78]). The latter is an important finding given the popularity of the use of formant measurements by acoustic-phonetic forensic-voice-comparison practitioners and the propensity for forensic casework to involve telephone-transmitted (especially mobile-telephone-transmitted) speech.

A number of studies investigated the effectiveness for forensic voice comparison of extracting information from glottal features. Kinoshita & Ishihara [79] and Zheng & Rose [80] tested the use of features based on the distribution of fundamental-frequency measurements made across all voiced speech in recordings, but neither compared their system’s performance with that of a baseline system. As mentioned above, several studies (Chen & Rose

[66], Jialin & Rose [69], Li & Rose [70], Zhang & Enzinger [78]) tested the use of fundamental-frequency trajectories for selected vowels in tone languages (Cantonese and Mandarin, see also Wang & Rose [81]). Enzinger et al [82] tested a number of glottal-source measurements (jitter, shimmer, and many more) extracted using commercial software, but did not obtain substantial improvement over a baseline system.

Kavanagh [83] [84] and Yim & Rose [85] investigated the effectiveness for forensic voice comparison of extracting acoustic information from the spectra of selected nasal phonemes. They did not compare the performance of their systems with a baseline system. Rose [86] [87] [88] investigated the effectiveness for forensic voice comparison of using cepstral coefficients to measure the spectra of tokens of a selected fricative phoneme and tokens of selected vowel phonemes. The data were read speech recorded over landline telephone systems. The last of these studies found an improvement in performance over a baseline system based on the entire speech-active portion of the recordings when the fricative-spectra system was fused with the baseline.

The use of long-term-formant (LTF) distributions for forensic voice comparison has been discussed in previous Interpol Forensic Science Symposium reviews (Bijhold et al [50], Kriigel et al [51]). Gold et al [89] tested the performance of an LTF forensic-voice-comparison system but did not compare the results with the performance of a baseline system. Becker [63] did not find substantial improvement over an MFCC baseline system when an LTF system was fused with the baseline system.

Rose & Winter [90], Morrison [71], and Zhang et al [75] tested the effectiveness of the Gaussian Mixture Model - Universal Background Model procedure (GMM-UBM, e.g., Reynolds et al [91]) versus the Multivariate Kernel Density procedure (MVKD, Aitken & Lucy [92]) for calculating likelihood ratios based on formant measurements. Which of the two procedures works best appears to depend on bias-variance tradeoffs related to the number of variables and number of data points used to train the models.

Rhodes [74] investigated the effect of large time differences between suspect and offender recordings on the performance of formant (including formant-trajectory) based forensic-voice-comparison systems and on the performance of a commercial forensic-voice-comparison system. Testing was conducted on recordings of eight speakers made at seven-year intervals between age 21 and 49 (a total of five time points per speaker). Performance for both systems decreased with increased time span.

A number of the studies reported above did not calibrate the forensic-voice-comparison systems employed. Calibration can ameliorate what would otherwise be very misleading results, and in some circumstances it is essential if one wishes to interpret system output as likelihood ratios. Morrison [93] published a tutorial on logistic-regression calibration and fusion including examples taken from forensic voice comparison as well as fingerprint

comparison. A number of the studies reported above tested on contemporaneous data, i.e., same-speaker test pairs were created by dividing a single recording. Apart for exceptional cases (such as in Enzinger [65]) if the recordings of known and questioned origin are in fact from the same speaker, they are non-contemporaneous recordings of that speaker. Enzinger & Morrison [94] reported on a study which empirically illustrated that testing on contemporaneous data gives an overly optimistic impression of system performance compared to testing on non-contemporaneous data.

### ***3.5 Empirical research on forensic voice comparison not conducted within the new paradigm***

Papers reviewed in this section describe empirical studies which were not conducted within the new paradigm. A number of the studies mentioned in the previous section included multiple analyses some of which were more or less compatible with the new paradigm and some of which were clearly incompatible with the new paradigm, those studies are not re-reviewed in this section.

Schwartz et al [95] described the United States Secret Service - Massachusetts Institute of Technology Lincoln Laboratory (USSS-MITLL) forensic-voice-comparison system applied to the National Institute of Standards and Technology's Human Assisted Speaker Recognition Evaluation (NIST HASR). An auditory-acoustic-phonetic system whose ultimate output was based on a human expert's judgment was fused with an automatic system. The relative weighting of the two systems in the fusion was also subjectively decided. The HASR Evaluation required that the system provide a same-speaker or different-speaker decision.

Mendes & Ferreira [96] obtained improvement in correct-identification rate when they fused a baseline MFCC system with a system based on normalized relative delays of source harmonics from selected vowels. High-quality audio recordings were used.

Thaitechawat & Foulkes [97] investigated the effectiveness for forensic voice comparison of extracting acoustic information from formant and fundamental-frequency trajectories in a tone language (Thai). Classifications were performed using discriminant analysis.

Künzel [98] tested the performance of a commercial forensic-voice-comparison system on cross-language compared to same-language test pairs. Transmission conditions tested were landline telephone, mobile telephone, and voice over Internet protocol. Cohorts of recordings in the same language and same transmission condition as the suspect recording were used to normalize system scores (Z-norm). False-alarm rates for different-speaker trials were only slightly worse for the cross-language trials than for the same-language trials.

### **3.6 Disputed utterance analysis**

Three papers looked at issues related to the disputed utterance in the 2009 New Zealand Supreme Court case *Bain v R* [2009 NZSC 16]. This was a very high profile case in New Zealand. Innes [99] discussed the background to the case and the expert opinions with respect to the disputed utterance. The prosecution contended that the words spoken were “I shot the prick”, an admission of guilt, whereas the defense contended that these were not the words spoken. Some of the experts consulted thought they heard the words “I can’t breathe” (and this was actually what Bain claimed to have said, although this was not revealed at the time). Most of the experts (French, Harrison, Cawley, Foulkes, Innes) based their opinions on what they heard and some opined that it was even uncertain as to whether the disputed utterance was speech or simply breathing. None came down in support of the “I shot the prick” hypothesis. The Supreme Court ruled that the jury in the trial proper should not be allowed to hear the disputed utterance, or any reference to the prosecution hypothesis, or expert testimony relating to the disputed utterance.

Fraser et al [100] experimented on what jury members might have heard had they been asked to listen to the disputed utterance. A total of 190 listeners were tested in two conditions. On initial listening the most common response from the listeners as to what they heard was “I can’t breathe” (from 60 of the 190 listeners). Only three heard “I shot the prick” (one of these had previous knowledge of the case and the other two were police officers). After one group heard mock expert testimony in support of the hypothesis that the words spoken were “I shot the prick” the number of listeners reporting this as being what they believed the words to be raised from 1 to 32 (of 96), and then after hearing mock expert testimony to the contrary that dropped to 26. For listeners in a control group who heard mock expert testimony that the words spoken were “he shot them all” the number reporting that they believed the words to be “I shot the prick” rose from 2 to 3 (of 94). Finally, both groups were told that the words spoken were definitely not “I shot the prick” at which point the number of listeners reporting this as being what they believed the words to be dropped to 17 for the first group, but rose to 11 for the control group. This demonstrated that although very few listeners heard the words “I shot the prick” without being prompted, a substantial proportion could be induced to hear these words if they were suggested to them, and, more disturbingly, even if the suggestion came in the form of being told that these were not the words.

One expert (Rose, whose evidence the defense held back for potential presentation in the trial proper rather than in the Supreme Court hearing) opined in his report that what anybody heard was irrelevant, what mattered was what Bain said, that the best way to assess this was via acoustic analysis rather than auditory perception, and that the proper way to evaluate the strength of the evidence was via a likelihood ratio, i.e., what are the relative likelihoods of getting the acoustic properties of the disputed utterance had the speaker said “I shot the prick” versus had he said “I can’t breathe”. As a research project Morrison & Hoy [101] conducted a preliminary version of such an analysis using telephone recordings of a speaker mimicking the

speaking style of the disputed analysis. The speaker produced about 40 tokens of each phrase. A form of cepstral analysis was conducted to extract acoustic information from the first speech sound in the known tokens of “shot” [ʃ] and “can’t” [ç] and from the speech sound in the equivalent position in the disputed utterance. Statistical models were trained and tested. The measured acoustic properties of the disputed utterance were found to be approximately 31 000 times more likely under the “can’t” hypothesis than under the “shot” hypothesis.

### ***3.7 Voice-based lie detectors***

Not mentioned in previous Interpol Forensic Science Symposium reviews, there was some controversy around a paper on voice-based lie detectors (formally voice stress analyzers) published by Eriksson & Lacerda [102]. The paper included criticism of a particular commercial product, and the manufacturer of that product threatened to sue the journal publisher. The publisher withdrew the paper from their website. Other recent papers published on the topic include Hollien et al [103], Harnsberger et al [104], Harnsberger [105], Horvath et al [106], and Lacerda [107]. These papers reported on theoretical and empirical assessments of commercial systems whose explicit or implied function is lie detection via acoustic analysis of voice signals. There may be a placebo effect whereby speakers who believe an effective lie-detection system is in use are less likely to lie, and human listeners may be able to perceive that speakers are lying at levels slightly above chance, but beyond that none of the studies found any substantial evidence in support of the hypothesis that any of the systems performed at levels above chance.

## **4. Audio Enhancement (Jeff M. Smith)**

### ***4.1 Introduction***

The enhancement, or clarification, of forensic audio is a common task related to the processing and analysis of audio evidence. This is because recordings made by law enforcement, intelligence, or the general public, which end up as forensic evidence are commonly made in non-ideal environments with non-ideal equipment leading to degraded quality and a poor ratio of signal to noise (SNR). The general goals for the enhancement of forensic audio include: to increase intelligibility of speech present in a recording which may increase the accuracy of transcription and number of words present in a transcript, to decrease listener fatigue due to recorded interferences, and to decrease the SNR in the preprocessing of recorded material for automatic speech and speaker recognition systems.

Early innovations in this area still impact the set of current solutions including spectral subtractive algorithms [108] and statistical model based algorithms [109] that are applicable to monaural recordings. Where multiple microphone sources are available, spatial filtering by means of beamforming [110] and

Independent Component Analysis (ICA) [111] can be effectively applied. Some more recent advances in this area will be described below including a discussion of research into new algorithms for speech enhancement. Additionally, special attention will be given to recent developments in the evaluation of speech intelligibility, which has recently and naturally evolved within this mature field.

Since speech enhancement research is well established and research contributions in this area are very frequent, the impact of innovative research is hard to evaluate soon after initial publication. This literature review therefore will focus on a few novel and relevant publications in the main areas related to forensic audio enhancement: monaural and binaural approaches, deconvolution, speech intelligibility evaluation, and the new areas of Compressive Sensing (CS) and Computational Auditory Scene Analysis (CASA).

#### **4.2 Reference works**

There are two recent reference publications related to this field. The 2nd Edition of the Encyclopedia of Forensic Sciences featured a chapter on Forensic Audio Enhancement and Authentication [112] by Grigoras & Smith. In this chapter the authors present a basic procedure for the handling and processing of forensic audio for both enhancement and authentication. Additionally, references to best practices are provided.

Loizou's 2nd Edition of Speech Enhancement: Theory and Practice [113] was published which continues to be a valuable reference in the area of speech enhancement. The new addition pays special attention of speech intelligibility including two new chapters on the subject.

The Scientific Working Group on Digital Evidence (SWGDE) publishes guidelines and best practices related to computer and mobile phone forensics as well as forensic audio. The Audio Committee made up of law enforcement and academia released the "Core Competencies for Forensic Audio v1.0" in September of 2011, which complements the previously drafted "Best Practices for Forensic Audio v1.0" from 2008. These documents are valuable resources for the drafting of laboratory practices and Standard Operating Procedures (SOPs) respecting consensus driven best practices for forensic audio processing and enhancement.

#### **4.3 Enhancement of Monaural and Binaural Recordings and Future Areas of Research**

Two interesting papers related to tone removal from recordings were presented at the AES 46th International Conference on Audio Forensics. Haddad & Noga [114] present a novel method for removing tone interferences by utilizing a super resolution spectrum analysis technique to remove the poles of the unwanted signal. In testing, this method showed better results than the traditional notch filter when preprocessing material for speaker

recognition tasks. Nordlund & McElveen [115] present a solution for removing non-stationary tonal noises by whitening the signal's noise floor and identifying tonal peaks for subtraction. This achieves higher-resolution subtraction reducing error and distortion.

Another useful approach in forensic audio enhancement is the separation of signals in a monaural recording by using commercially available material present in the recording (music, TV broadcast, etc.) to synthesize binaural reference cancellation. The problem with application of this method in forensics is in time domain alignment and drift of the often low-quality source to the commercial reference material. Ding & Havelock [116] propose a drift-compensated adaptive filter (DCAF) to achieve better cancellation while Alexander et al [117] apply landmark-based acoustic fingerprinting, similar to what is used in Shazam and other music identification services, to automatically align material.

Recent research by Paliwal et al [118] into processing noisy audio signals in the modulation domain has shown an improvement over traditional acoustic spectral subtraction. Another exploration of processing in the modulation domain by Zhang & Zhao [119] achieves binaural blind source separation.

In another growing area of research, computational auditory scene analysis (CASA), researchers seek to emulate with a machine the human ability to overcome the so-called "cocktail party effect". It has been shown by Wang [120] that the main concern of CASA is use of the ideal time-frequency binary mask (IBM). Recent papers on IBM estimation in speech enhancement include May et al [121] and Jensen & Hendriks [122].

Another new area of research that has had profound effect in many areas is compressed sensing or CS introduced by Candès et al and Donoho [123] [124]. This technique can help acquire and reconstruct a signal from a sparse or underrepresented dataset allowing the entire signal to be determined from relatively few measurements; fewer than those set forth by the Nyquist theorem (which requires twice the highest sampled frequency). D. Wu et al [125] have explored application of CS based speech enhancement finding that compressed speech and noise via discrete cosine transform (DCT) achieves proper signal sparsity for compressed sensing. Low et al [126] provide a good overview of compressive sensing and speech enhancement. P. Wu et al have also used CS in multichannel dereverberation, or deconvolution, of audio signals [127].

#### ***4.4 Speech Quality vs. Speech Intelligibility***

As discussed earlier, one crucial aim in the enhancement of forensic audio recordings is to increase intelligibility of speech material in order to increase accuracy and words present in a transcript. Recently, it has been found that the classical methods of enhancement are effective at increasing the signal quality (increase in SNR) but do not increase intelligibility AND may actually make speech less intelligible. Subjective listening tests by Hu and Loizou



[128] using the NOIZUS database shed light on this. Hilkhuisen et al [129] recently found congruent results in testing three algorithms (spectral subtraction, MMSE, and subspace) with difficult noise types (car and talker babble).

Thus, recent changes have taken place in the research and development of speech enhancement algorithms focused on speech intelligibility. Loizou & Kim [130] discuss this further and add additional findings of interest like that fact that in testing subspace algorithms perform worst in overall quality but perform well in terms of preserving speech intelligibility. They urge researchers focusing on intelligibility to maximize greater than 0 dB the segmental SNR in the frequency domain. A predictive measure for determining speech intelligibility has been proposed Taal et al [131] with a short-time objective intelligibility measure (STOI) as a reliable means for obtaining evaluation data while avoiding costly listening experiments.

Researchers at the Center for Law Enforcement Audio Research (clear-labs.com) in the UK investigate this area with special attention to processing forensic audio by examiners. Hilkhuisen et al [132] investigate improvement of intelligibility based on parameter settings of commercial equipment chosen by experts attempting to increase intelligibility. Findings were that while parameter settings varied greatly, experts attempting to enhance noisy speech propose parameter settings which generally deteriorate intelligibility. In another interesting paper, Hilkhuisen et al [133] investigate whether repeated listening to audio material (replay) improved intelligibility or understanding of utterances. This is important because experts and those preparing transcriptions commonly replay audio material. The study found that after replaying 5 times, listener performance saturated while listeners themselves underestimated their performance believing it improved after replaying 5 times. The authors conclude that replay can improve intelligibility performance but may lead to overconfidence.

## 5. Organizations

Forensic audio analysis is a growing community that has members in several international working groups:

- AAFS - American Academy of Forensic Science: Within the American Academy of Forensic Science is the newly formed Digital and Multimedia Sciences section that includes forensic audio and speech analysis. <http://aafs.org/digital-multimedia-sciences>
- AES - Audio Engineering Society: The Audio Engineering Society is devoted exclusively to audio technology. Founded in the United States in 1948, the AES has grown to become an international organization that unites audio engineers, creative artists, scientists and students worldwide by promoting advances in audio and disseminating new knowledge and research. <http://www.aes.org/>
- ENFSI FSAAWG – European Network of Forensic Science Institutes Forensic Speech and Audio Analysis Working Group: a European group that

is focused on all aspects of forensic audio and speech analysis, including linguistics. "Membership of FSAAWG is open to representatives from all ENFSI member institutes. Members have to be active in the areas of forensic speech and audio analysis." "Representatives from non-ENFSI members who are active in the field of forensic speech and audio analysis examinations can apply for associate membership." <http://www.enfsi.eu/page.php?uid=63>

- The Forensic Acoustics Subcommittee (FAS) of the Acoustical Society of America (ASA) was established in 2010 and organizes a special session at an ASA meeting approximately once per year. "Membership of the ASA Forensic Acoustics Subcommittee is open to current members of the ASA." Website: <http://asa.forensic-acoustics.net/>

- The Forensic Speech Science Committee (FSSC) of the Australasian Speech Science and Technology Association (ASSTA) was established in 1996. Membership of the committee is by invitation. Website: <http://www.assta.org/?q=assta-forensic-speech-science-committee>

- IAFPA - The International Association for Forensic Acoustics and Phonetics was established in 1991 and holds an annual conference. "Full membership is available to established phoneticians and acousticians with operational and/or academic interests in forensic applications of phonetics or acoustics." Website: <http://www.iafpa.net/>

- NCMF - National Center for Media Forensics: an American center that is part of the University of Colorado which has a strong forensic audio program in addition to research and education in video and image forensics. <http://www.ucdenver.edu/academics/colleges/CAM/Centers/ncmf/Pages/ncmf.aspx>

- SWGDE - Scientific Working Group on Digital Evidence: an American group that includes a forensic audio committee that has produced best practices manuals and is promoting research on forensic audio. <http://www.swgde.org/>

- SWG-Speaker - Scientific Working Group on Speaker recognition: a new created American group to support and promote the scientific foundations and practice of speaker recognition, voice data collection, measurement, transmission, and retrieval. <http://swg-speaker.org/>

## 6. Acknowledgements

We thank Michael Jessen of the Bundeskriminalamt Germany, Bruce Koenig and Doug Lacey of BEK TEK LLC, Mark Huckvale and Gaston Hilkhuisen of the University College London, James D Harnsberger of the University of Florida, and Francisco Lacerda of Stockholm University for assistance in preparation this manuscript.

## 7. References

- [1] Koenig BE, Lacey DS. "Forensic Authenticity Analyses of the Header Data in Transcoded WMA Files From Small Olympus Audio Recorders." JAES Volume 60 Issue 4 pp. 255-265; April 2012

- [2] Shen Y, Jia J, Cai L. "Detecting Compressed AMR-format Audio Recordings". PCC2012
- [3] Luo D, Luo W, Yang R, Huang J. Compression History Identification for Digital Audio Signal. ICASSP 2012
- [4] Koenig BE, Lacey DS, Grigoras C, Price SG, Smith JM. "Evaluation of the Average DC Offset Values for Nine Small Digital Audio Recorders". Journal of the Audio Engineering Society, Volume 61 Issue 6 pp. 439-448; June 2013
- [5] Malik H, Miller JW. "Microphone Identification Using Higher-Order Statistics". 46th AES International Conference: Audio Forensics; June 2012
- [6] Malik H, Zhao H. "Recording Environment Identification Using Acoustic Reverberation". IEEE ICASSP 2012
- [7] Pan X, Zhang X, Lyu S. "Detecting Splicing in Digital Audios Using Local Noise Level Estimation". IEEE ICASSP 2012
- [8] Chen J, Xiang S, Liu W, Huang H. "Exposing Digital Audio Forgeries in Time Domain by Using Singularity Analysis with Wavelets". The 1st ACM Workshop on Information Hiding and Multimedia Security; June 2013
- [9] Boss D. "Using Ripple Signals for the Authentication of Audio Material". 46th AES International Conference: Audio Forensics; June 2012
- [10] Bao Y. "Electric network frequency estimation based on fractional Fourier Transform for audio authenticity". Technical Acoustics, Vol.30, No.4, Pt.2, August 2011
- [11] Su H, Garg R, Hajj-Ahmad A, Wu M. ENF "Analysis on Recaptured Audio Recordings". IEEE ICASSP 2013
- [12] Nicolalde DPR, Apolinario JA, Biscainho LWP. "Audio Authenticity: Detecting ENF Discontinuity With High Precision Phase Analysis". IEEE Transactions on Information Forensics and Security, Vol. 5, No. 3, September 2010
- [13] Nicolalde DPR, Apolinario JA, Biscainho LWP. "Audio Authenticity Based on the Discontinuity of ENF Higher Harmonics". EUSIPCO 2013
- [14] Coetzee S. "Phase and Amplitude Analysis of the ENF for Digital Audio Authentication". 46th AES International Conference: Audio Forensics; June 2012
- [15] Yuan Z, Liu Y, Conners R, Liu Y. "Effects of Oscillator Errors on Electric Network Frequency Analysis". 46th AES International Conference: Audio Forensics; June 2012
- [16] Liu Y, Chai J, Conners R, Liu Y. "A Study of the Accuracy and Precision of Quadratic Frequency Interpolation for ENF Estimation". 46th AES International Conference: Audio Forensics; June 2012
- [17] Yuan Z, Liu Y, Conners R, Liu Y. "Using Simple Monte Carlo Methods and a Grid Database to Determine the Operational parameters for the ENF Matching Process". 46th AES International Conference: Audio Forensics; June 2012
- [18] Archer H. "Quantifying Effects of Lossy Compression on Electric Network Frequency Signals". 46th AES International Conference: Audio Forensics; June 2012
- [19] Grigoras C, Jenkins C, Smith JM. "Advances in ENF Database Configuration for Forensic Authentication of Digital Media". Proceedings of the Audio Engineering Society 131st Convention; October 2011

- [20] Jenkins C, Steinhour J. "Advances in Electric Network Frequency Acquisition Systems and Stand Alone Probe Applications for the Authentication of Digital Media". 46th AES International Conference: Audio Forensics; June 2012
- [21] Grigoras C, Smith J. "Advances in ENF Analysis for Digital Media Authentication". 46th AES International Conference: Audio Forensics; June 2012
- [22] Korycki R. "Methods of Time-Frequency Analysis in Authentication of Digital Audio Recordings. *International Journal of Electronics and Telecommunications*", Vol.56, No.3, pp. 257-262; September 2010
- [23] Korycki R. "Time and spectral analysis methods with machine learning for the authentication of digital audio recordings". *Forensic Science International*, Vol. 230, pp. 117-126; March 2013
- [24] Gupta S, Cho S, Kuo CJ. "Current Developments and Future Trends in Audio Authentication. *IEEE Multimedia*", Vol.19 Issue 1; January 2012
- [25] Grigoras C, Rappaport D, Smith JM. "Analytical Framework for Digital Audio Authentication". 46th AES International Conference: Audio Forensics; June 2012
- [26] Koenig BE, Lacey DS. "An Inconclusive Digital Audio Authenticity Examination: A Unique Case". *Journal of Forensic Sciences*, 2011
- [27] Jessen M. "The Forensic Phonetician: Forensic Speaker Identification by Experts", in *The Routledge Handbook of Forensic Linguistics*, Coulthard M, Johnson A, Eds. Abingdon, UK: Routledge, 2010, pp. 378–394.
- [28] Morrison GS. "Forensic Voice Comparison", in *Expert Evidence*, Freckelton I, Selby H, Eds, Sydney, Australia: Thomson Reuters, 2010, ch. 99.
- [29] Amino K, Osanai T, Makinae H, Arai T. "Historical and Procedural Overview of Forensic Speaker Recognition as a Science", in Neustein A, Patil HA, Eds, *Forensic Speaker Recognition*, New York: Springer, 2012, pp. 3–20. doi: 10.1007/978-1-4614-0263-3\_1
- [30] Perrot P, Chollet G. "Helping the Forensic Research Institute of the French Gendarmerie to Identify a Suspect in the Presence of Voice Disguise or Voice Forgery", in *Forensic Speaker Recognition*, Neustein A, Patil HA, Eds, New York: Springer, 2012, pp. 469–503. doi:10.1007/978-1-4614-0263-3\_16
- [31] Gold E, French P. "International Practices in Forensic Speaker Comparison", *International Journal of Speech, Language, and the Law*, 2011, vol. 18, pp. 293–307. doi:10.1558/ijssl.v18i2.293
- [32] Morrison GS, "Distinguishing between Forensic Science and Forensic Pseudoscience: Testing of Validity and Reliability, and Approaches to Forensic Voice Comparison", *Science & Justice*, 2013. doi:10.1016/j.scijus.2013.07.004
- [33] French JP, Harrison P. "Position Statement Concerning Use of Impressionistic Likelihood Terms in Forensic Speaker Comparison Cases", *International Journal of Speech, Language, and the Law*, 2007, vol. 14, pp. 137–144. doi: 10.1558/ijssl.v14i1.137
- [34] Balding DJ. *Weight-Of-Evidence for Forensic DNA Profiles*. Chichester, UK: Wiley, 2005.

- [35] Buckleton J. "A framework for interpreting evidence", in *Forensic DNA Evidence Interpretation*, Buckleton J, Triggs CM, Walsh SJ, Eds, Boca Raton, FL: CRC, 2005, pp. 27–63.
- [36] Evett IW. "Interpretation: a personal odyssey", in *The Use of Statistics in Forensic Science*, Aitken CGG, Stoney DA, Eds, Chichester, UK: Ellis Horwood, 1991, pp. 9–22.
- [37] Robertson B, Vignaux GA. *Interpreting Evidence*, Chichester, UK: Wiley, 1995.
- [38] Rose P, Morrison GS. "A Response to the UK Position Statement on Forensic Speaker Comparison", *International Journal of Speech, Language and the Law*, 2009, vol. 16, pp. 139–163. doi:10.1558/ijssl.v16i1.139
- [39] Morrison GS. "Forensic Voice Comparison and the Paradigm Shift", *Science & Justice*, 2009, vol. 49, pp. 298–308. doi:10.1016/j.scijus.2009.09.002
- [40] French P, Nolan F, Foulkes P, Harrison P, McDougall K. "The UK Position Statement on Forensic Speaker Comparison: A Rejoinder to Rose and Morrison", *International Journal of Speech, Language, and the Law*, 2012, vol. 17, pp. 143–152. doi:10.1558/ijssl.v17i1.143
- [41] Evett IW, Aitken CGG, Berger CEH, Buckleton JS, Champod C, Curran JM, Dawid AP, Gill P, González-Rodríguez J, Jackson G, Kloosterman A, Lovelock T, Lucy D, Margot P, McKenna L, Meuwly D, Neumann C, Nic Daeid N, Nordgaard A, Puch-Solis R, Rasmusson B, Radmayne M, Roberts P, Robertson B, Roux C, Sjerps MJ, Taroni F, Tjin-A-Tsoi T, Vignaux GA, Willis SM, Zadora G. "Expressing Evaluative Opinions: A Position Statement", *Science & Justice*, 2011, vol. 51, pp. 1–2. doi:10.1016/j.scijus.2011.01.002 doi:10.1016/S1355-0306(98)72117-3
- [42] Berger CEH, Buckleton J, Champod C, Evett IW, Jackson G. "Evidence Evaluation: A Response to the Court of Appeal Judgment in *R v T*", *Science & Justice*, 2011, vol. 51, pp. 43–49. doi:10.1016/j.scijus.2011.03.005
- [43] Redmayne M, Roberts P, Aitken CGG, Jackson G. "Forensic Science Evidence in Question", *Criminal Law Review*, 2011, vol. 5, pp. 347–356.
- [44] Robertson B, Vignaux GA, Berger CEH. "Extending the Confusion about Bayes", *Modern Law Review*, 2011, vol. 74, pp. 444–455. doi:10.1111/j.1468-2230.2011.00857.x
- [45] Morrison GS. "The Likelihood-Ratio Framework and Forensic Evidence in Court: A Response to *R v T*", *International Journal of Evidence and Proof*, 2012, vol. 16, pp. 1–29. doi:10.1350/ijep.2012.16.1.390
- [46] Aitken C, Roberts P, Jackson G. *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses. Practitioner Guide No 1*, Royal Statistical Society's Working Group on Statistics and the Law, 2010. <http://www.rss.org.uk/site/cms/contentviewarticle.asp?article=1132>
- [47] Expert Working Group on Human Factors in Latent Print Analysis. *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. Gaithersburg, MD: US Department of Commerce, National Institute of Standards and Technology, 2012. [http://www.nist.gov/manuscript-publication-search.cfm?pub\\_id=910745](http://www.nist.gov/manuscript-publication-search.cfm?pub_id=910745)

- [48] Broeders APA. "Forensic Speech and Audio Analysis Forensic Linguistics – 1998 to 2001 – A Review", Proceedings of the 13th International Forensic Science Symposium, Lyon, 2011, pp. D2-53–D2-84.
- [49] Broeders APA. "Forensic Speech and Audio Analysis Forensic Linguistics – A Review: 2001 to 2004", Proceedings of the 14th International Forensic Science Symposium, Lyon, 2004, pp. 171–188.
- [50] Bijhold J, Ruifrok A, Jessen M, Geradts Z, Ehrhardt S, Alberink I. "Forensic Audio and Visual Evidence – A Review: 2004 to 2007", Proceedings of the 15th Interpol Forensic Science Symposium, Lyon, 2007, pp. 372–413.
- [51] Kriigel CR, Smith G, Graves M. "Audio Analysis – Review: 2007-2010", Proceedings of the 16th International Forensic Science Symposium, Lyon, 2010, pp. 379–396.
- [52] National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press, 2009.
- [53] Morrison GS, Ochoa F, Thiruvaran T. "Database Selection for forensic Voice Comparison", Proceedings of Odyssey 2012: The Language and Speaker Recognition Workshop, Singapore, 2012, pp. 62–77.
- [54] Morrison GS, Rose P, Zhang C. "Protocol for the Collection of Databases of Recordings for Forensic-voice-comparison Research and Practice", Australian Journal of Forensic Sciences, 2012, vol. 44, pp. 155–167. doi:10.1080/00450618.2011.630412
- [55] Morrison GS, Zhang C, Rose P. (2011). "An Empirical Estimate of the Precision of Likelihood Ratios from a Forensic-Voice-Comparison System", Forensic Science International, 2011, vol. 208, pp. 59–65. doi:10.1016/j.forsciint.2010.11.001
- [56] Morrison GS, Thiruvaran T, Epps J. "Estimating the Precision of the Likelihood-Ratio Output of a Forensic-Voice-Comparison System", in Proceedings of Odyssey 2010: The Language and Speaker Recognition Workshop, Brno, Cernocký H, Burget L, Eds, International Speech Communication Association, 2010, pp. 63–70.
- [57] Morrison GS. "Measuring the Validity and Reliability of Forensic Likelihood-Ratio Systems", Science & Justice, 2011, vol. 51, pp. 91–98. doi:10.1016/j.scijus.2011.03.002
- [58] Ramos D, González-Rodríguez J. "Reliable Support: Measuring Calibration of Likelihood Ratios", Science & Justice, 2013. doi: 10.1016/j.forsciint.2013.04.014
- [59] Becker T, Solewicz Y, Jardine G, Gfrörer S. "Comparing Automatic Forensic Voice Comparison Systems Under Forensic Conditions", Proceedings of the 46th Audio Engineering Society Conference on Audio Forensics: Recording, Recovery, Analysis, and Interpretation, Denver, 2012, pp. 197–202.
- [60] Solewicz, Y., Becker, T., Jardine, G., & Gfrörer, S. (2012). Comparison of speaker recognition systems on a real forensic benchmark. Proceedings of Odyssey 2012: The Speaker and Language Recognition Workshop, Singapore, pp. 86–91.

- [61] Becker T, Jessen M, Alsbach S, Broß F, Meier T. "SPES: The BKA Forensic Automatic Voice Comparison System", in Proceedings of Odyssey: The Speaker and Language Recognition Workshop, Brno, Cernocký H, Burget L, Eds, International Speech Communication Association, 2010, pp. 58–62.
- [62] Becker T, Jessen M, Alsbach S, Broß F, Meier T. "Automatic Forensic Voice Comparison Using Recording Adapted Background Models", Proceedings of the Audio Engineering Society 39th International Conference on Audio Forensics, Hillerød, 2010, pp. 162–166.
- [63] Becker T. Automatischer forensischer Stimmenvergleich (Automatic forensic voice comparison). PhD dissertation, Universität Trier, 2012.
- [64] Rose P. "Yes, Not Too Bad - Likelihood Ratio-Based Forensic Voice Comparison in a \$150 Million Telephone Fraud", Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney, 2012, pp. 161–164.
- [65] Enzinger E. "Mismatched distances from speakers to telephone in a forensic-voice-comparison case", Proceedings of the 21st International Congress on Acoustics, Montréal, Proceedings of Meetings Online, 2013, vol. 19, paper 060039. doi:10.1121/1.4805425
- [66] Chen A, Rose P. "Likelihood Ratio-Based Forensic Voice Comparison with the Cantonese Triphthong /iau/", Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney, 2012, pp. 197–200.
- [67] Enzinger E. "Characterizing Formant Tracks in Viennese Diphthongs for Forensic Speaker Comparison", Proceedings of the Audio Engineering Society 39th International Conference on Audio Forensics, Hillerød, 2010, pp. 47–52.
- [68] Hughes V. "Establishing Typicality: A Closer Look at Individual Formants", Proceedings of the 21st International Congress on Acoustics, Montréal, Proceedings of Meetings Online, 2013, vol. 19, paper 060042. doi:10.1121/1.4798775
- [69] Jialin P, Rose P. "Likelihood Ratio-Based Forensic Voice Comparison with the Cantonese Diphthong /ei/ F-Pattern", Proc Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney, 2012, pp. 205–208.
- [70] Li J, Rose P. "Likelihood Ratio-Based Forensic Voice Comparison with F-Pattern and Tonal F0 from the Cantonese /ɔy/ Diphthong", Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney, 2012, pp. 201–204.
- [71] Morrison GS. "A Comparison of Procedures for the Calculation of Forensic Likelihood Ratios From Acoustic-Phonetic Data: Multivariate Kernel Density (MVKD) Versus Gaussian Mixture Model – Universal Background Model (GMM-UBM)", Speech Communication, 2011, vol. 53, pp. 242–256. doi:10.1016/j.specom.2010.09.005
- [72] Morrison GS. "Vowel Inherent Spectral Change in Forensic Voice Comparison", in Vowel Inherent Spectral Change, Morrison GS, Assmann PF, Eds, Heidelberg, Germany: Springer-Verlag, 2013, pp. 263–283. doi:10.1007/978-3-642-14209-3\_11

- [73] Pingjai S, Ishihara S, Sidwell PJ. "A Likelihood Ratio-based Forensic Voice Comparison Using Formant Trajectories of Thai Diphthongs", Proceedings of the 21st International Congress on Acoustics, Montréal, Proceedings of Meetings Online, 2013, vol. 19, paper 060043. doi:10.1121/1.4799433
- [74] Rhodes RW. Assessing the Strength of Non-Contemporaneous Forensic Speech Evidence. PhD dissertation, University of York, 2012.
- [75] Zhang C, Morrison GS, Thiruvananthapuram T. "Forensic Voice Comparison Using Chinese /iau/", Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, 2011, pp. 2280–2283.
- [76] Zhang C, Morrison GS, Ochoa F, Enzinger E. "Reliability of Human-Supervised Formant-Trajectory Measurement for Forensic Voice Comparison", Journal of the Acoustical Society of America, 2013, vol. 133, pp. EL54–EL60. doi:10.1121/1.4773223.
- [77] Zhang C, Morrison GS, Enzinger E, Ochoa F. "Effects of Telephone Transmission on the Performance of Formant-Trajectory-Based Forensic Voice Comparison – Female Voices", Speech Communication, 2013, vol. 55, pp. 796–813. doi:10.1016/j.specom.2013.01.011.
- [78] Zhang C, Enzinger E. "Fusion of Multiple Formant-Trajectory- and Fundamental-Frequency-Based Forensic-Voice-Comparison Systems: Chinese /ei/, /ai/, and /iau1/", Proceedings of the 21st International Congress on Acoustics, Montréal, Proceedings of Meetings Online, 2013, vol. 19, paper 060044. doi:10.1121/1.4798793.
- [79] Kinoshita Y, Ishihara S. "F0 Can Tell Us More: Speaker Classification Using the Long Term Distribution", Proceedings of the 13th Australasian International Conference on Speech Science and Technology, Melbourne, 2010, pp. 50–53.
- [80] Zheng R, Rose P. "Likelihood Ratio-Based Forensic Voice Comparison with Cantonese Short-Term Fundamental Frequency Distribution Parameters", Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney, 2012, pp. 153–156.
- [81] Wang CY, Rose P. "Likelihood Ratio-Based Forensic Voice Comparison with Cantonese /i/ F-Pattern and Tonal F0", Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney, 2012, pp. 209–212.
- [82] Enzinger E, Zhang C, Morrison GS. "Voice Source Features for Forensic Voice Comparison – an Evaluation of the GLOTTEX Software Package", Proceedings of Odyssey 2012: The Speaker and Language Recognition Workshop, Singapore, 2012, pp. 78–85.
- [83] Kavanagh C. New Consonantal Acoustic Parameters for Forensic Speaker Comparison. PhD dissertation, University of York, 2012.
- [84] Kavanagh C. "Exploring Duration and Spectral Parameters of English /m/ for Forensic Speaker Comparison", Proceedings of the 21st International Congress on Acoustics, Montréal, Proceedings of Meetings Online, 2013, vol. 19, paper 060040. doi:10.1121/1.4798992.



- [85] Yim ACS, Rose P. "Are Nasals Better? Likelihood Ratio-Based Forensic Voice Comparison with Segmental Cepstra from Cantonese and Japanese Syllabic/Mora Nasals", Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney, 2012, pp. 5–8.
- [86] Rose P. "Forensic Voice Comparison with Japanese Vowel Acoustics - A Likelihood Ratio-Based Approach Using Segmental Cepstra", Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, 2011, pp. 1718–1721.
- [87] Rose P. "Forensic Voice Comparison with Secular Shibboleths - A Hybrid Fused GMM-Multivariate Likelihood Ratio-Based Approach Using Alveolo-Palatal Fricative Cepstral Spectra", Proceedings of the International Conference on Audio, Speech and Signal Processing, Prague, 2011, pp. 5900–5903.
- [88] Rose P. "More Is Better: Likelihood Ratio-Based Forensic Voice Comparison with Vocalic Segmental Cepstra Frontends", International Journal of Speech, Language and the Law, 2013, vol. 20, pp. 77–116. doi:10.1558/ijssl.v20i1.77
- [89] Gold E, French P, Harrison P. "Examining Long-term Formant Distributions as a Discriminant in Forensic Speaker Comparisons under a Likelihood Ratio Framework", Proceedings of the 21st International Congress on Acoustics, Montréal, Proceedings of Meetings Online, 2013, vol. 19, paper 060041. doi:10.1121/1.4800285.
- [90] Rose P, Winter E. "Traditional Forensic Voice Comparison with Female Formants: Gaussian mixture model and multivariate likelihood ratio analyses", Proceedings of the 13th Australasian International Conference on Speech Science and Technology, Melbourne, 2010, pp. 205–208.
- [91] Reynolds DA, Quatieri TF, Dunn RB. "Speaker Verification Using Adapted Gaussian Mixture Models", Digital Signal Processing, 2000, vol. 10, pp. 19–41. doi:10.1006/dspr.1999.0361
- [92] Aitken CGG, Lucy D. "Evaluation of Trace Evidence in the Form of Multivariate Data", Applied Statistics, 2004, vol. 53, pp. 109–122. doi:10.1046/j.0035-9254.2003.05271.x
- [93] Morrison GS. "Tutorial on Logistic-regression Calibration and Fusion: Converting a Score to a Likelihood Ratio", Australian Journal of Forensic Sciences, 2013, vol. 45, pp. 173–197. doi:10.1080/00450618.2012.733025
- [94] Enzinger E, Morrison GS. "The Importance of Using Between-Session Test Data in Evaluating the Performance of Forensic-Voice-Comparison Systems", Proceedings of the 14th Australasian International Conference on Speech Science and Technology, Sydney, 2012, pp. 137–140.
- [95] Schwartz R, Campbell JP, Shen W, Sturim DE, Campbell WM, Richardson FS, Dunn RB, Granville R. "USSS-MITLL 2010 human assisted speaker recognition", Proceedings of the International Conference on Audio, Speech and Signal Processing, Prague, 2011, pp. 5904–5907.

- [96] Mendes D, Ferreira A. "Speaker Identification Using Phonetic Segmentation and Normalized Relative Delays of Source Harmonics", Proceedings of the 46th Audio Engineering Society Conference on Audio Forensics: Recording, Recovery, Analysis, and Interpretation, Denver, 2012, pp. 215–222.
- [97] Thaitechawat S, Foulkes P. "Discrimination of Speakers Using Tone and Formant Dynamics in Thai", Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, 2011, pp. 1978–1981.
- [98] Künzel HJ. "Automatic Speaker Recognition with Crosslanguage Speech Material", International Journal of Speech, Language and the Law, 2013, vol. 20, pp. 21–44. doi:10.1558/ijssl.v20i1.21
- [99] Innes B. "R v David Bain – a Unique Case in New Zealand Legal and Linguistic History", International Journal of Speech, Language and the Law, 2011, vol. 18, pp. 145–155. doi:10.1558/ijssl.v18i1.145
- [100] Fraser H, Stevenson B, Marks T. "Interpretation of a Crisis Call: Persistence of a Primed Perception of a Disputed Utterance", International Journal of Speech, Language and the Law, 2011, vol. 18, pp. 261–292. doi:10.1558/ijssl.v18i1.145
- [101] Morrison GS, Hoy MC. "What Did Bain Really Say? A Preliminary Forensic Analysis of the Disputed Utterance Based on Data, Acoustic Analysis, Statistical Models, Calculation of Likelihood Ratios, and Testing of Validity", Proceedings of the 46th Audio Engineering Society Conference on Audio Forensics: Recording, Recovery, Analysis, and Interpretation, Denver, 2012, pp. 203–207.
- [102] Eriksson A, Lacerda F. "Charlatantry in Forensic Speech Science: A Problem to be Taken Seriously", International Journal of Speech, Language and the Law, 2007, vol. 14, pp. 169–193. doi:10.1558/ijssl.2007.14.2.169
- [103] Hollien H, Harnsberger JD, Martin CA, Hollien KA. "Evaluation of the NITV CVSA", Journal of Forensic Sciences, 2008, vol. 53, pp. 183–193. doi:10.1111/j.1556-4029.2007.00596.x
- [104] Harnsberger JD, Hollien H, Martin CA, Hollien KA. "Stress and Deception in Speech: Evaluating Layered Voice Analysis", Journal of Forensic Sciences, 2009, vol. 54, no. 3, pp. 642–650. doi:10.1111/j.1556-4029.2009.01026.x
- [105] Harnsberger JD. "Best Practices in the Evaluation of Speech Technology: The Case of Voice Stress Analyzers", paper presented at 21st International Congress on Acoustics, Montréal, 2–7 June 2013. Available: <http://montreal2013.forensic-acoustics.net/>
- [106] Horvath F, McCloughan J, Weatherman D, Slowik S. "The Accuracy of Auditors' and Layered Voice Analysis (LVA) Operators' Judgments of Truth and Deception During Police Questioning", Journal of Forensic Sciences, 2013, vol. 58, pp. 385–392. doi:10.1111/1556-4029.12066
- [107] Lacerda F. "Voice Stress Analyses: Science and Pseudoscience", Proceedings of the 21st International Congress on Acoustics, Montréal, Proceedings of Meetings Online, 2013, vol. 19, paper 060003. doi:10.1121/1.4799435
- [108] Boll, S. "Suppression of acoustic noise in speech using spectral subtraction". IEEE Trans. Acoust. Speech Signal Process., 1979, ASSP-27(2), 113-120.

- [109] Ephraim, Y., & Malah, D. "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator". *IEEE Trans. Acoust. Speech Signal Process.*, 1984, ASSP-28, 137-145.
- [110] van Veen, B., & Buckley, K. "Beamforming: A versatile approach to spatial filtering". *IEEE ASSP Magazine*, 1988, April, 4-24
- [111] Jutten, C., & Herault, J. "Blind separation of sources, part 1: An adaptive algorithm based on neuromimetic architecture". *Signal Processing*, 1991, 24, 1-10.
- [112] Grigoras, C., & Smith, J. M. "Audio Enhancement and Authentication". In J. Siegel, & P. Saukko (Eds.), *Encyclopedia of Forensic Sciences*, 2nd ed., 2013, pp. 315-526. Academic Press.
- [113] Loizou, P. *Speech Enhancement: Theory and Practice*, 2nd ed., 2013, Boca Raton, FL, USA: CRC Press.
- [114] Haddad, D. M., & Noga, A. J. "Tone removal using a band focus speech reconstruction algorithm". *Audio Engineering Society 46th International Conf. Audio Forensics*, 2012, Denver.
- [115] Nordlund, S., & McElveen, J. K. "Enhancing low SNR speech corrupted by non-stationary tonal noises". *Audio Engineering Society 46th International Conf. Audio Forensics*, 2012, Denver.
- [116] Ding, H., & Havelock, D. "Drift-compensated adaptive filtering for improving speech intelligibility in cases with asynchronous inputs". *EURASIP Journal on Advances in Signal Processing*, 2010
- [117] Alexander, A., Forth, O., & Tunstall, D. "Music and noise fingerprinting and reference cancellation applied to forensic audio enhancement". *Audio Engineering Society 46th International Conf. Audio Forensics*, 2012, Denver.
- [118] Paliwal, K., Wojcicki, K., & Schwerin, B. "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain". *Speech Communication*, 2010, 52, 450-475.
- [119] Zhang, Y., & Zhao, Y. "Modulation domain blind speech separation in noisy environments". *Speech Communication*, 2012, <http://dx.doi.org/10.1016/j.specom.2013.06.014>.
- [120] Wang, D. "On ideal binary mask as the computational goal of auditory scene analysis". In P. Divenyi (Ed.), *Speech Separation of Humans and Machines*, 2005, (pp. 181-197). Norwell, MA: Kluwer Academic.
- [121] May, T., van de Par, S., & Kohlrausch, A. "Noise robust speaker recognition combining missing data techniques and universal background modeling". *IEEE Trans. on Audio, Speech, and Lang. Process.*, 2012, 20 (1), 108-121.
- [122] Jensen, J., & Hendriks, R. "Spectral magnitude minimum mean-square error estimation using binary and continuous gain functions". *IEEE Trans. on Audio, Speech, and Lang. Process.*, 2012, 20 (1), 92-102.
- [123] Candès, E.J., Romberg, J.K., Tao, T. "Stable signal recovery from incomplete and inaccurate measurements". *Communications on Pure and Applied Mathematics*, 2006, 59(8): 1207. doi:10.1002/cpa.20124
- [124] Donoho, D. "Compressed Sensing". *IEEE Trans. on Information Theory*, 2006, 52 (4), 1289-1306.
- [125] Wu, D., Zhu, W., & Swamy, M. "On sparsity issues in compressive sensing based speech enhancement". *IEEE International Symposium on Circuits and Systems*, 2012 (pp. 285-288). Seoul.

- [126] Low, S. Y., Pham, D. S., & Venkatesh, S. "Compressive speech enhancement". *Speech Communication*, 2013, 55, 757-768.
- [127] Wu, P., Epain, N., & Jin, C. "A dereverberation algorithm for spherical microphone arrays using compressed sensing techniques". *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2012, (pp. 4053-4056). Kyoto.
- [128] Hu, Y., & Loizou, P. C. "Subjective comparison and evaluation of speech enhancement algorithms". *Speech Communication*, 2007, 49, 588-601.
- [129] Hilkhuisen, G., Gaubitch, N., Brookes, M., & Huckvale, M. "Effects of noise suppression on intelligibility: Dependency on signal-to-noise ratios". *Journal of the Acoustical Society of America*, 2012, 131 (1), 531-539.
- [130] Loizou, P., & Kim, G. "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions". *IEEE Transactions on Audio, Speech, and Lang. Process.*, 2011, 19 (1).
- [131] Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. "An algorithm for intelligibility prediction of time-frequency weighted noisy speech". *IEEE Transactions on Audio, Speech, and Lang Process.*, 2011, 19 (7), 2125-2136.
- [132] Hilkhuisen, G., Gaubitch, N., & Huckvale, M. "Effects of noise suppression on intelligibility: Experts' opinions and naive normal-hearing listeners performance". *Journal of Speech, Language, and Hearing Research*, 2013, 56, 404-415.
- [133] Hilkhuisen, G., Lloyd, J., & Huckvale, M. "Effects of replay on the intelligibility of noisy speech". *Audio Engineering Society 46th International Conf. Audio Forensics*, 2012, Denver.