



## AES 39th International Conference Audio Forensics - Practices and Challenges

### Session: Authentication

#### Detecting butt-spliced edits in forensic digital audio recordings

*Alan Cooper, Metropolitan Police, United Kingdom*

Digital audio recording forgeries are often created using simple editing techniques such as 'butt splicing'. Butt-splicing may leave discontinuities in the audio waveform that may or may not be audible. Detection of butt-spliced edits will be discussed and a new method to detect edits made in this manner will be proposed. The process operates in the time domain and is based on high pass filtering the audio data and modeling a discontinuity at higher frequencies where the ratio of discontinuity energy to acoustic signal energy level is improved. The model is then used as a template to search for potential edits in the filtered audio data. The technique is optimized for non-compressed audio and is capable of detecting discontinuity points within a recording that are not discernable by auditory analysis.

#### Visualisation of magnetic features on analogue audiotapes is still an important task

*Dagmar Boss, Bayerisches Landeskriminalamt, Germany*

Most of the recordings handed in for analysis are digital today but there are still cases where analogue tape material has to be investigated.

Evidence audio seized from or handed in by private persons has often been recorded in an analogue format, especially on analogue Dictaphones.

In cases where an authentication is required, visualization is most effective. Today it is carried out mainly with the help of special, artificially produced crystals that change their optical properties when brought in contact with external magnetic fields. Different devices can be used for the work with these crystals. In order to obtain good pictures of the visualization, one has to pay great attention to maintaining the right working conditions.

#### Statistical Tools for Multimedia Forensics: Compression Effects Analysis

*Catalin Grigoras, Forensic Science Center, Romania*

This paper reports an approach of using statistical tools for forensic authentication of multimedia recordings, with a specific focus on digital audio compression effects analysis. The gradual shift from analogue to digital recording systems, rapid developments in the media world and the increasingly widespread availability of digital signal processing equipments and software, as well as their ease of operation, generated a significant increase in the number of attempts to use digital multimedia evidence in every sector of litigation and criminal justice. A brief description is given to different compression algorithm effects, possibilities to detect them and authors' results on this topic. It is possible to determine a previous compression and estimate the authenticity of a digital recording.

Conference sponsor:





## Session: Speech and Forensics - Voice Identification

### Digitally Disguised Voices

**Eddy B. Brixen, EBB-consult, Denmark**

One of the newer challenges in audio forensics is the investigation of voice recordings that are suspect of being manipulated by utilizing digital signal processing. This paper presents a number of voice manipulating tools available. Several tools are developed for musical purposes. However, they are eventually applied to talking voices as well. This implies both real time processing and post processing. The voice alterations are demonstrated and the influence on acoustic parameters normally examined in connection with forensic voice analysis and authentication, are presented. This includes the effect of pitch shifting, pitch modeling, harmonizing, formant relocation, removal of voiced sounds, and its' influence on background sounds.

### Characterising Formant Tracks in Viennese Diphthongs for Forensic Speaker Comparison

**Ewald Enzinger, Acoustics Research Institute, Austrian Academy of Sciences, Austria**

This study evaluates methods that capture time-dynamic properties of diphthongs produced by speakers of Viennese German for application in a forensic setting. Polynomials, discrete cosine transform and B-splines along with experimental features based on bent-cable regression models were used to characterise the first three formant tracks of two /aE/ diphthong segments. The resulting coefficients were in turn used as parameters in a speaker discrimination procedure based on likelihood ratios which were calculated using a multi-variate kernel density formula (MVKD). A comparison of the achieved performance based on cross-validation is presented in terms of equal error rate (EER) and the log-likelihood ratio cost metric as well as DET and Tippett plots.

### Channel compensation for forensic speaker identification using inverse processing

**Andrey Barinov, Speech Technology Center Ltd., Russian Federation**

**Sergey Koval, Speech Technology Center Ltd., Russian Federation**

**Michael Stolbov, Speech Technology Center Ltd., Russian Federation**

**Pavel Ignatov, Speech Technology Center Ltd., Russian Federation**

Typically, speaker identification examination requires two audio recordings: a voice sample and a questionable recording. The questionable one is in most of cases the intercepted or recorded phone call. As far as mobile phones became the most popular way of communication the biggest part of questionable recordings is from GSM channel. Quite often these audio materials are really poor quality that is why speech enhancement and noise cancellation procedures are usually needed. The subjects of this paper are the following issues:

- influence of the GSM coding on the voice biometric traits identification parameters of speech signal;
- GSM channel compensation for voice identification.

Conference sponsor:





## **Importance of the relative delay of glottal source harmonics in speaker identification and voice register classification**

*Aníbal Ferreira, University of Porto, Portugal*

*Ricardo Sousa, University of Porto, Portugal*

In this paper proposal we focus on the real-time frequency domain analysis of speech signals, and on the extraction of suitable and meaningful features from the glottal source paving the way for robust speaker identification and voice register classification. We take advantage of an analysis-synthesis framework derived from an audio coder algorithm in order to estimate and model the relative delay between the different harmonics of the glottal source, in order to separate the influence of the source from the group delay of the vocal tract filter, and in order to monitor and extract in real-time important features correlating well with the speaker unique sound signature and the breathy, modal and pressed voice registers.

## **Session: Enhancement of Noisy Recordings**

### **Corrupted speech intelligibility improvement using an adaptive filtering based algorithm**

*Damian Ellwart, Gdansk University of Technology, Poland*

*Andrzej Czyzewski, Gdansk University of Technology, Poland*

This paper describes a technique of improving the quality of speech signals recorded under interference (adaptive filter based algorithm). Proposed algorithm is described and additional possibilities of speech intelligibility improvement are discussed. Results of the tests are presented. A way of integrating the elaborated method with an agglomeration acoustic monitoring system is proposed. The research is subsidized by the Polish Ministry of Science and Higher Education within Grant No. R00-O0005/3.

### **Adjusting a commercial speech enhancement system to optimize intelligibility**

*Gaston Hilkhuisen, University College London, United Kingdom*

*Mark Huckvale, University College London, United Kingdom*

To improve the quality of noisy recordings, sound engineers have at their disposal a variety of speech enhancement systems. It is difficult to set the parameters of these systems to maximum intelligibility, which is the topic of this paper. In a first experiment, expert listeners adjusted the settings of a commercial system while attempting to improve intelligibility. Their settings were then evaluated in a listening experiment - showing that processing actually reduced intelligibility compared to the original signal. In a second experiment a range of parameter settings for the same system were evaluated using both listeners and an intelligibility model based on a speech envelope distortion measure. The results suggest that automatic methods for optimizing parameter settings may be possible.



### **Speech Enhancement by Adaptive Noise Cancellation: Problems, Algorithms and Limits**

*Joerg Bitzer, Fachhochschule Wilhelmshaven/Oldenburg/Elsfleth, Germany*

*Matthias Brandt, Fachhochschule Wilhelmshaven/Oldenburg/Elsfleth, Germany*

In this paper, a summary about adaptive noise reduction techniques for enhancing degraded audio signals in the context of forensic application is given. The surrounding conditions for the use of adaptive systems are examined at first to understand the functionality of the algorithms that are commonly employed. Built on this analyses we will give examples of self-adaptive single-channel solutions, monophone and stereo noise-cancellers with reference signal and solutions for GSM burst removal. Furthermore, we investigate the weaknesses of the standard approaches and highlight problems that are resulting. Several difficulties generally have to be taken care of in the field of adaptive noise cancellation – approaches to those are presented, but also the limits of feasibility.

## **Session: Acoustical Forensics**

### **Performance of closed-form acoustic scene decomposition for forensic analysis**

*Banu Gunel, University of Surrey, United Kingdom*

*Grigorios Nikolopoulos, University of Surrey, United Kingdom*

*Huseyin Hacıhabiboglu, University of Surrey, United Kingdom*

*Ahmet Kondo, University of Surrey, United Kingdom*

In-situ audio recordings for forensic analysis are generally made by purpose-installed microphones in order to capture the speech and any other relevant sounds in an environment. Often, post-processing efforts are focused on cleaning these recordings from interferences. However, an acoustic scene is made up of audio objects, whose not only content, but also positions carry evidential importance. In this paper, a closed-form acoustic scene decomposition technique is presented which blindly decomposes the sound field using a compact microphone array. This decomposition both localizes and enhances the audio objects, improving the intelligibility of speech. The speech separation, recognition and localisation performance of the system will be presented for typical forensic analysis scenarios both by simulations and experiments.

### **Directional aspects of forensic gunshot recordings**

*Robert Maher, Montana State University, United States*

Crime scene forensic evidence may include audio gunshot recordings obtained from some known or assumed location with respect to the shooting position. The acoustic evidence of the gunshot depends significantly upon the relative orientation of the firearm's barrel and the recording microphone due to the inherent directional characteristics of the gun, the microphone, and the acoustical characteristics of the surrounding environment. This paper explains several important directional considerations when interpreting gunshot recordings for forensic purposes.



## Session: Laboratory Procedures

### **The Modern Forensic Audio Laboratory - a public sector perspective**

***Robin How, Metropolitan Police, United Kingdom***

The Metropolitan Police Forensic Audio Laboratory was founded in the 1960's. Its processes and policies have adapted to serve modern audio technologies and up to date Policing methods and priorities.

The paper discusses the motivations, organisation and practices of the laboratory in the light of it being government funded and working exclusively for the prosecution side of the UK adversarial judicial system. It discusses staff qualifications, training, experience, laboratory equipment, software and research programs. Through analysis and discussion of the issues and demands, it will give the reader an insight into how the UK's largest Forensic Audio Laboratory is run.

### **Computer Forensics for the Forensic Audio Professional**

***Jeff Smith, National Center for Media Forensics, University of Colorado Denver, United States***

***Marcus Rogers, Purdue University Cyber Forensics Lab Dept. of Computer & Information Technology / CERIAS, United States***

With the proliferation of digital technology and its interweaving into all facets of modern life, the landscape of crime and surveillance has evolved placing new demands on the forensic analyst working with digital audio. In order to address these new demands, practitioners will need to embrace technology dealing with the acquisition and recovery of digital data as well as advanced file searching and information gathering. This paper presents an overview of computer forensics for the forensic audio professional in an effort to help define their role in a digital world revolving around personal computers and mobile devices. While focusing on the overlap and synergy between audio and computer forensics, suggestions are made for taking on challenges presented in the acquisition and analysis of digital evidence. Topics include: history, overview of hardware and software, existing and developing standards and best practices, and limitations.

## Session: Speech and Forensics - Automated Systems

### **Voice carving in Police Dialogue: Forensic application of Automatic Speaker Segmentation**

***Anil Alexander, Griff Comm Ltd, United Kingdom***

***Oscar Forth, Griff Comm Ltd, United Kingdom***

***Robin How, Metropolitan Police Forensic Audio Laboratory, United Kingdom***

In cases involving vulnerable or intimidated subjects, witnesses may agree to give evidence conditional to their anonymity being maintained. Disguise or masking, which is the equivalent of pixelation in video, can be applied in an audio setting. Witness interviews of significance, with the judge's agreement, can be 'pixelated' or disguised in order to mask the identity of the speaker(s). Semi-automatic speaker segmentation or diarization can be applied to disguising voices of vulnerable and intimidated victims, as well as undercover officers. We examine the application of semi-automatic speaker recognition to aid the quick segmentation of audio intended for disguise or masking and significantly reduce processing times.

Conference sponsor:





### **Automatic Forensic Voice Comparison Using Recording Adapted Background Models (#40)**

***Timo Becker, Federal Criminal Police Office, Germany***

***Michael Jessen, Federal Criminal Police Office, Germany***

***Sebastian Alsbach, University of Applied Science Koblenz, Germany***

***Franz Broß, University of Applied Science Koblenz, Germany***

We present the automatic forensic voice comparison system SPES which is developed in cooperation with the Federal Criminal Police Office of Germany (Bundeskriminalamt - BKA), the University of Applied Sciences Koblenz and the Department of Phonetics, University of Trier, Germany. The system is based on the classical GMM-UBM framework based on MAP adaption as described by Reynolds et al. This GMM-UBM framework is extended by creating specific background models for each suspect recording based on the similarity of speakers. The less general and more specific background models lead to better discrimination ability and more reliable results. We show how the usage of recording adapted background models outperforms the standard GMM-UBM approach in different recording conditions.

## **Session: Speech quality and intelligibility assessment**

### **Bayesian Adaptive Method for Estimating Speech Intelligibility in Noise**

***Nikolay Gaubitch, Imperial College London, United Kingdom***

***Mike Brookes, Imperial College London, United Kingdom***

***Patrick Naylor, Imperial College London, United Kingdom***

We present a tool for efficient evaluation of the psychometric function for speech intelligibility in noise. The core of this tool is an adaptive Bayesian method, which estimates a threshold and the slope of the psychometric function for a given noise sample. The signal-to-noise ratio at each subsequent iteration is chosen such that the expected information to be gained by this iteration is maximized. We demonstrate the use of the tool using car noise and babble noise. We also show how the tool can be used to evaluate the effects of speech processing algorithms on intelligibility.

### **C-QUAL - A Validation of PESQ Using Degradations Encountered in Forensic and Law Enforcement Audio**

***Dushyant Sharma, Imperial College London, United Kingdom***

***Gaston Hikhuisen, University College London, United Kingdom***

***Patrick Naylor, Imperial College London, United Kingdom***

Assessment of speech quality of law-enforcement audio recordings is important as degradations introduced by non ideal recording conditions can reduce the intelligence value of such recordings. Furthermore a model that predicts speech quality could be beneficial for assessing the performance of audio collection and enhancement systems. The Perceptual Evaluation of Speech Quality (PESQ) algorithm (ITU P.862) has been validated for degradations common in telecommunications. In this paper we apply PESQ to degradations typically encountered in law-enforcement. Also we present a subjectively labeled database (C-Qual) containing distortions encountered in law enforcement scenarios. Comparing the prediction by PESQ and the observed opinions provided by the listeners shows that PESQ is less suitable for estimating speech quality in this context.

Conference sponsor:





### **Objective measures of speech intelligibility in forensic applications**

**Andrea Paoloni, Fondazione Ugo Bordoni, Italy**

**Giovanni Costantini, Università Tor Vergata, Italy**

**Massimiliano Todisco, Università Tor Vergata, Italy**

In order to decide whether the transposition of the speech originated from a lawful interception in the written text is the actual speech signal or only the views of the transcriber is required to have an objective method to measure speech intelligibility. Unfortunately, the forensic expert does not have the original signal, therefore, must make its assessment based on the only available signal. This paper addresses the issue by using three different approaches to evaluate the signal and a corpus which is measured voice intelligibility. The approaches are tested with 3 different types of noise and compared results with the speech intelligibility scores. Measures based on Speech Transmission Index have proven to be reliable for predicting the intelligibility in forensic applications.

### **Measuring the Effect of Noise Reduction on Listening Effort**

**Mark Huckvale, University College London, United Kingdom**

**Ngawang Frasi, University College London, United Kingdom**

Noise reduction (NR) has become widely applied in the forensic audio domain to “improve” the quality of noisy speech recordings. In this paper we consider how such processing affects listener productivity in everyday speech communication tasks. Two measures are presented: one based on reaction time to spoken digits, and one based on finding errors in a transcript of a spoken conversation. We explain why such tasks are a useful complement to measures based on intelligibility then present the methodology and results for two evaluations of these measures using MMSE processing on speech corrupted by babble and car-noise. Finally we discuss the implications for the use of NR techniques and for models of how signal quality affects speech communication.

### **Intelligibility Testing as an Engineering Tool**

**Robert Fellows, HMGCC, United Kingdom**

**Kenneth Worrall, HMGCC, United Kingdom**

Speech communication systems can only be effective if they are highly optimized. Where systems are intelligibility critical, intelligibility testing is the key to deciding whether a system is viable and which system components are best.

Conventional intelligibility tests can be expensive to run, and have an insatiable appetite for fresh test material. TACIT (Technique for Comparative Intelligibility Testing) was developed to provide a quick easy intelligibility test needing only a small pool of listeners, and capable of delivering, in only a few hours, a result good enough to influence design decisions.

TACIT is now used on a daily basis, but its impact has been magnified by the development of new ways of presenting the results to customers and managers.

Conference sponsor:





## Tutorials

### T1: The likelihood-ratio framework for the evaluation of forensic-comparison evidence

**Geoffrey Stewart Morrison**<sup>1,2</sup>

<sup>1</sup> School of Language Studies, Australian National University, Canberra, Australia

<sup>2</sup> School of Electrical Engineering & Telecommunications, University of New South Wales, Sydney, Australia

This tutorial provides an introduction to the forensic comparison of audio recordings in the likelihood-ratio framework. Examples are drawn from audio recordings of human voices, but the principles and techniques can be applied to audio recordings of any source. The tutorial covers the topics essential for an understanding of the likelihood-ratio framework and its application to the forensic comparison of audio recordings including: What is a forensic likelihood ratio? Why is the likelihood-ratio framework the logically correct way to evaluate forensic evidence? How is a forensic likelihood ratio calculated? How is the validity and reliability of the likelihood-ratio output of a forensic-comparison system evaluated? What factors affect the validity and reliability of a forensic-comparison system and how might validity and reliability be improved?

### T2: SWGDE AND Forensic Audio STANDARDS

**Michael Piper**<sup>1,4</sup>, **David Hallimore**<sup>2,4</sup>, and **John POWELL**<sup>3,4</sup>

<sup>1</sup> U.S. Secret Service, Washington, D.C., USA

<sup>2</sup> Houston Police Department, Houston, TX, USA

<sup>3</sup> Los Angeles Sheriff's Department, USA

<sup>4</sup> Scientific Working Group for Digital Evidence, USA

The Scientific Working Group on Digital Evidence (SWGDE) recently published a position paper on the U.S. National Research Council's February 18, 2009 report to Congress entitled "Strengthening Forensic Science in the United States: A Path Forward" in which the Council set forth a broad overview of the state of forensic science in the United States along with several recommendations for improvement. An overview of *The Report*, SWGDE's response, and its Audio Committee's current projects will be presented.

In each of the topics we intend to touch on (Forensic Audio Best Practices, minimum standards, accreditation/certification, training, ENF, etc.), there will be many points suitable for audience input and discussion. One of the key purposes of this presentation is to solicit input from the forensic audio community for the work we are undertaking at SWGDE.

Conference sponsor:





## Workshops

### W1: Frontiers of Forensic Audio Investigation

Chairman:

**Gordon Reid, Managing Director, CEDAR Audio Ltd, UK**

Presenters:

**Anna Czajkowski, Control Risks, UK**

**Anil Alexander, Head of R&D, GriffComm Ltd, Oxford, UK**

**Alan French, Audio and Video Forensics Manager, FTS, UK**

Session 1: Anna Czajkowski  
**ISO17025 Accreditation in Forensic Audio**

Session 2: Anil Alexander  
**Audio winnowing: Finding whom or what you want to hear from a recording**

Session 3: Alan French  
**Time, tide and technological changes wait for no person.**

### W2: Overview of Digital Audio Authentication

Chairman:

**Catalin Grigoras, Forensic Science Center, Romania**

This session will emphasize some of the latest developments in forensic authentication of digital recordings, like Electric Network Frequency (ENF) and lossy compression analysis. It will also present the most common source or errors, the nowadays technological limits and the need for quality assurance.

The presentations will be interactive, in a lecture-discussion format with images, case studies and relevant research results.

Conference sponsor:

