In June the AES audio forensics community took advantage of another excellent opportunity to share information on research and practice in forensic science. The AES 39th International Conference, *Audio Forensics—Practices and Challenges*, was the latest meeting in the successful series that began in 2005 with the 26th Conference and continued with the 33rd Conference in 2008. While the prior AES audio forensics events were held in Denver, Colorado, this year’s event was held at the Pharmakon Conference Center in Hillerød, Denmark. The venue supported a truly international meeting, with paper presenters representing nearly a dozen different nations and an attendee list reflecting participants from more than 20 countries. The 39th International Conference reasserted AES as a key player in the audio forensics field.

Planning for the 39th Conference involved members of the AES Technical Committee on Audio Forensics and the AES headquarters staff. Work on the latest conference began in 2008 immediately following the successful 33rd Conference. Eddy Bøgh Brixen, conference chair, coordinated an outstanding committee consisting of papers cochairs Alan Cooper and Durand Begault, workshops cochairs Gordon Reid and Catalin Grigoras, treasurer S.K. Pramanik, facilities chair Katrine Bøgh Brixen, and webmaster Preben Kvist. The Com-
The late Roy Pritts and Rich Saunders (pictured on the screen below) did so much to engender interest in and support for audio forensics within the AES.

committee’s efforts culminated in an exceptional technical program and lively workshop atmosphere.

The Pharmakon Conference Center, located about an hour’s journey from Copenhagen Kastrup International Airport, has hosted several previous AES conferences. The venue was comfortable and particularly well-suited to the intimate, small group interaction that is a hallmark of AES international conferences. The attendees enjoyed the friendly and collegial quality of the Pharmakon Center—particularly the morning and afternoon coffee breaks and the dining menus.

The conference opened on Thursday morning, 17 June, with introductory remarks by Eddy Brixen. He highlighted the important roles played by the prior AES forensics conference organizers in Denver, the late Roy Pritts and the late Richard Sanders (see photos above), who had done so much to engender interest and support for audio forensics within the AES. Brixen called upon the attendees to remember Roy and Rich for their pioneering work. Following the welcome and conference overview, the morning session was devoted to presentations by three exhibitors: Cedar Audio, UK; National Center of Media Forensics, University of Colorado, Denver, USA; and DPA Microphones, Denmark. In addition to their opening presentations, the exhibitors each had space in a special room located adjacent to the main auditorium, and provided information and hands-on demonstrations throughout the conference.

The papers cochairs, Alan Cooper of the Metropolitan Police, London, and Durand Begault of Charles M. Salter Associates, San Francisco, assembled a wide variety of topical sessions including papers on audio authentication, voice identification, enhancement, speech quality and intelligibility, and general acoustical forensics.

**AUTHENTICATION**

Following a fine opening lunch on Thursday, the technical portion of the conference began with three papers on the topic of authentication. Forensic examiners determine
By examining the audio material for signs of inadvertent or deliberate alteration, verifying that the recording system operated properly and investigating the circumstances of the recording and its chain of custody, Durand Begault introduced the session’s authors and their topics.

The first paper of the session, presented by Alan Cooper, described preliminary work in automated detection of “butt-spliced” edits in a digital audio file. A butt-spliced edit is a simple deletion or insertion in a digital file. Cooper explained that while butt splicing may often cause an audible defect, not every splicing discontinuity is detectable aurally. He described a promising set of experiments to detect butt-spliced edits using sample-to-sample difference and signal correlation techniques. He acknowledged that this preliminary work is not guaranteed to detect every splice, and a variety of other techniques are under development.

The second paper on authentication techniques explained the use of magneto-optical crystals for visualizing the latent magnetic patterns present on audio recording tapes. Dagmar Boss of the Bavarian State Criminal Police (Bayerisches Landeskriminalamt) explained the use of special crystalline materials that change their optical properties in response to external magnetic fields, thereby allowing the magnetic domains on the tape to be imaged for analysis and interpretation. Although digital recording devices are rapidly supplanting the use of analog magnetic tape in many forensic circumstances, Boss emphasized the continued need for high-quality visualization techniques whenever analog tape is involved.

Rounding out the opening session, Catalin Grigoras, the newly-named director of the National Center for Media Forensics at the University of Colorado-Denver, discussed the prospects for reliable authentication of digital recordings. The challenge for the forensic audio examiner is that a skilled and clever individual may be able to alter a recording in a manner that is not detectable based on simple aural or waveform analysis. Grigoras focused his attention on recordings made with lossy perceptual audio coding, such as MP3 or WMA, and the possibility that artifacts due to successive coding and decoding, altered background reverberation, or a change from one codec to another might be detectable.

It was clear to all attendees that a key challenge for audio forensics will be methods for authenticity verification in the age of digital recording and lossy audio codecs.

**VOICE IDENTIFICATION**

After a pleasant discussion break complete with coffee, tea, and pastries, Alan Cooper introduced the next paper session covering research issues in forensic voice identification. A common task in forensic investigations is to determine the likelihood that the words in a recording of speech were uttered by a particular individual.

Eddy Brixen presented the first paper in the session. He described his investigation into the use of digital signal processing techniques to disguise the identity of the talker. An individual might deliberately choose to disguise his or her voice to avoid subsequent identification. In some cases the processed disguise is very obvious, while in other cases the alterations may be deliberately subtle and sophisticated and therefore potentially difficult to detect via forensic means. Brixen conducted a series of tests using commercial music and voice processing software to observe the spectral and temporal alterations present in several processed speech examples. He also examined the effect upon the tell-tale hum of electrical network frequency (ENF) information following processing. His conclusion is that identity concealment by signal processing voice disguise is quite feasible, and that it is difficult for even a skilled engineer to “undo” the processing to try to reveal the original, undisguised voice.

Next, Ewald Enzinger of the Austrian Academy of Sciences spoke on prospects for reliable authentication of digital recordings.
in Vienna presented his work on capturing the time-variant behavior of speech formants for diphthongs in Viennese German speech. The procedure follows the previously published work of Geoffrey Stewart Morrison in which the temporal change in formant frequencies during the diphthong are modeled with a polynomial or some other fitting function. Comparing the fitting function from an unknown talker to the functions obtained from exemplar recordings of a particular suspect may allow results suitable for reporting the likelihood ratio for the comparison. Enzinger noted that the results appear to be highly dependent on the particular speech context of the recordings, such as telephone conversations versus free speech, and also dependent upon natural phonetic and prosodic variation.

Andrey Barinov of the Speech Technology Center, St. Petersburg, Russia, explained his group’s work on the effects of mobile phone GSM (Global System for Mobile communications) source and channel coding in the context of forensic speaker identification. The GSM channel involves nonlinear and time-variant perceptual audio coding, which can alter the spectral balance and formant characteristics of the coded speech, making it difficult to form a forensic comparison between speech from a recording of a GSM phone call and the speech of an exemplar recording. Barinov’s group is working on a means for inverse processing to achieve channel compensation for GSM recordings.

Concluding the Thursday afternoon session, Anibal Ferreira of the University of Porto, Portugal, gave an overview of his work in robust speaker identification based on the relative phase relationship between harmonics of the vocal production system. Ferreira derives the “normalized relative delay” (NRD) for the periodic acoustic speech signal (or electroglotograph), and then compares the relative delay (phase difference) between the harmonic partials of two different speech examples to assess whether or not the speech of two different speakers can be discriminated. He explained that the initial experiments and test results were found to be promising for ongoing research.

At the conclusion of the Voice Identification session, the attendees enjoyed a wonderful evening dinner in the Pharmakon Conference Center dining room, a fitting finale to the day’s technical sessions and a pleasant break before the start of the evening workshop.

FRONTIERS OF FORENSIC AUDIO INVESTIGATION

Gordon Reid chaired a special evening workshop session featuring presentations by three noted experts from the United Kingdom: Alan French, an audio consultant with CEDAR Ltd; Anil Alexander, R&D Director with GriffComm Ltd; and Anna Czajkowski, an accreditation expert with Control Risks Forensics.

French led off the workshop with a presentation entitled “Time, Tide, and Technological Changes Wait for No Person”. The presentation traced the early history of audio forensic investigations by the Metropolitan Police in London, beginning in the 1970s with the investigation of audio tapes pertaining to Jeremy Thorpe, Member of Parliament, for his alleged involvement in the attempted murder of Norman Scott, a man who had claimed to be Thorpe’s former homosexual lover. The Metropolitan Police lab was asked to evaluate the authenticity of a tape presented as evidence in that high-profile case. That initial forensic audio work led steadily to 35 years of increasingly sophisticated and comprehensive lab facilities and procedures. French concluded his historical retrospective by suggesting that the future of media forensics will more and more involve audio, video, and general digital data investigations.

The next speaker, Anil Alexander, gave a fascinating presentation on the problems and prospects for dealing with the data explosion facing the audio forensics field. He noted that the ease with which modern recording equipment can obtain hours, days, or even weeks of continuous surveillance information causes the “curse of data.” How can the audio forensic examiner find the desired information from the vast data repository presented to him? He then elaborated on one of the ultimate desires of audio forensics: the ability to define an acoustic event of interest, search automatically for that event in a recording of arbitrary length, and incorporate an automatic learning algorithm to locate, discriminate, and document the desired events.

Anna Czajkowski, the last speaker of the evening’s workshop, gave a timely presentation on the issue of formal accreditation of audio forensic laboratories. ISO 17025, the international standard for assessing the competence of testing and calibration laboratories, is expected to affect audio forensic experts. Czajkowski described how audio forensic processing can be validated in a meaningful way. According to the ISO standard’s definition, validation means “confirmation, through the provision of objective evidence, that the requirements for a specific intended use have been fulfilled.” She explained that the accreditation process helps ensure that suitable equipment is used by a competent operator employing reliable methodology, thereby improving the quality of audio forensic practice while reducing the risk of error.

With the conclusion of the first day of the conference, the attendees enjoyed snacks and beverages at an informal reception in the lounge area. The friendly and collegial tradition of AES conferences was very much in evidence as everyone engaged in lively discussions of the day’s topics and presentations.
EVALUATION OF FORENSIC COMPARISON EVIDENCE AND THE LIKELIHOOD RATIO

Fresh from a fine breakfast buffet to start Friday, the second day of the conference, the attendees assembled for a special tutorial session regarding the use of the likelihood-ratio framework when evaluating audio forensic evidence. The tutorial presenter was Geoffrey Stewart Morrison, affiliated with the Australian National University, in Canberra, and the University of New South Wales, in Sydney. Morrison introduced the terminology of Bayes’ Theorem, which in the context of a forensic speaker comparison would be the posterior odds, defined by the ratio of the probability of the same speaker given the acoustical evidence, divided by the probability of a different speaker given the same acoustical evidence. In practice the forensic audio examiner would need to assess the Bayes likelihood ratio, which is the ratio of the probability of observing the provided acoustical evidence given that it was the same speaker, divided by the probability of observing the same acoustical evidence given that it was a different speaker. The remaining portion of the Bayes’ Theorem expression involves the prior odds, which treats the probability of the same speaker and the different speaker hypotheses. The posterior odds are the product of the likelihood ratio and the prior odds. Morrison consistently emphasized his point that a forensic scientist must only use the likelihood ratio, and not the prior odds, because the prior odds deal with subjective attributes like motive, opportunity, and human biases that are the province of the trier of fact (for example, a judge or jury), while the likelihood ratio should contain only objective information that is the province of the forensic examiner. He explained that for this reason it is inappropriate for a forensic scientist to report the posterior odds, and encouraged all forensic reports to stick to the likelihood ratio framework.

ENHANCEMENT OF NOISY RECORDINGS

The second half of the Friday morning schedule was devoted to a session on enhancement methods for noisy audio recordings. The first presenter, Damian Ellwart of the Gdansk University of Technology, Poland, described an adaptive filter algorithm for speech intelligibility improvement he developed with his coauthor, Andrzej Czyzewski. The promising system was tested with a mixture of speech and music produced under various environmental conditions to examine the algorithm’s noise-suppression characteristics. Future work will be done to assess the intelligibility improvement with human subjects.

Next, Gaston Hilghusen of University College London presented an interesting and important paper coauthored with Mark Huckvale entitled “Adjusting a Commercial Speech Enhancement System to Optimize Intelligibility”. The investigators reported on several interactive experiments conducted with a panel of human listeners to observe the listeners’ chosen parameter settings when listening to noisy speech through a commercially available noise-reduction system. When members of the panel were asked to adjust the noise-reduction parameters to achieve the best intelligibility, the results actually revealed a decrease in performance-based intelligibility. The study also found a substantial difference between the “optimal” settings selected by different members of the panel, indicating that opinion-based intelligibility is not an invariant standard from listener to listener.

The final paper of the morning session returned to the topic of adaptive filtering for speech enhancement. Joerg Bitzer of the Jade University of Applied Sciences and the Fraunhofer Institute for Digital Media Technology, Oldenburg, Germany, gave an overview of adaptive noise cancellation techniques for forensic applications. Adaptive systems require at least two signals, the input signal and the reference signal, having some level of correlation with each other. Bitzer explained that the recording circumstances, system nonlinearities, and complexity of the interfering noise point toward different realizations of the adaptive noise-cancellation approach, thus one size does not fit all.

The interesting and informative morning tutorial and paper sessions encouraged many questions, comments, and discussion, which continued seamlessly as the attendees made their way to the luncheon buffet.

ACOUSTICAL FORENSICS

Following lunch, the technical sessions reconvened for two papers dealing with acoustical principles. In the first presentation, “Closed-Form Spatial Decomposition of an Acoustic scene for enhancement and localization of audio objects in Forensic Analysis,” Banu Günel of the University of Surrey, UK, described a multichannel directional recording system that allows a closed-form decomposition of individual, spatially separated sound sources. Günel described the theoretical and practical foundations of the decomposition, taking a B-format microphone system (four signals) to produce very good separation of an ensemble of sound sources. The system gave very good intelligibility improvement for speech by separating the sound of the desired talker from the interfering sound and noise coming from other directions.
The second acoustical paper dealt with gunshot acoustics. Rob Maher of Montana State University, USA, presented results of an experiment involving controlled gunshot recordings of ten different firearms from ten different azimuthal directions. Maher showed that the mean sound pressure level of the muzzle blast was typically 15–20 dB lower when observed at the rear of the firearm (180° azimuth) compared to the on-axis position (0° azimuth). He also presented the time waveforms for each gun as a function of azimuth, revealing an interesting variety of distinguishing features for each firearm type. He concluded his presentation with a reminder that most gunshot forensic audio evidence includes acoustic reflections, reverberation, and clipping/distortion that will likely be more challenging for interpretation than the pristine and reflection-free waveforms obtained in his controlled experiment.

LABORATORY PROCEDURES
The second half of the afternoon session turned to the tools and techniques of contemporary professional forensic audio practice. The first presenter, Robin How of the Metropolitan Police Digital and Electronics Forensic Service, London, described the history, personnel qualifications, training, equipment, management, and research roles of the Metropolitan Police Forensic Audio Laboratory. How explained both the traditional and the emerging responsibilities of the Forensic Audio Laboratory, which is among the most experienced forensic audio facilities in Europe. He pointed out that one of the increasingly common requests in recent years has been for voice disguise: a request by the authorities to conceal the identity of a witness by rendering recorded testimony to be unrecognizable as the utterances of the protected witness, while still maintaining intelligibility for use in court or other official proceedings.

Jeff Smith of the National Center for Media Forensics at the University of Colorado-Denver, described the broader issues of digital and multimedia evidence that extend beyond the common scope of forensic digital audio examination. Audio forensic examiners are often asked to assist with investigations involving video content, computer file storage, data encryption, and many, many other facets of forensic interest in the computer age. Smith encouraged the attendees to embrace the complicated and multifaceted nature of modern “cyber” forensics, and to be ready for continuous self-study and formal education to keep pace with the dynamic nature of the profession.

AUTHENTICATION WORKSHOP
The final technical session on Friday was a special workshop on the challenges of authentication with digital audio recordings. The workshop speaker, Catalin Grigoras, presented a flow diagram summarizing the model for forensic audio data collection. He noted that the physical environment of the recording, the characteristics of the microphone, potential electrical network frequency (ENF) hum, etc., can each leave a tell-tale signature in the recording that may aid in the authentication task. Grigoras pointed out many possible indications of questionable authenticity, but also emphasized the potential difficulties in detecting surreptitious modifications of digital audio recordings.

AN EVENING OF TREATS
Friday evening offered a special treat for the conference attendees—a private tour of Frederiksborg Castle. Located on three small islets surrounded by Castle Lake in Hillerød, a short bus ride from the Pharmakon Conference Center, the castle site dates from the mid 1500s during the reign of its namesake, King Frederik II (b.1534 d.1588). The principal structures were constructed by Frederik’s son, King Christian IV (b.1577 d.1648). Several sections of the Renaissance-period castle were destroyed in a major fire in 1859, but through the good graces of J.C. Jacobsen, the founder of the Carlsberg Breweries, the castle was refurbished as a museum in 1878.
Denmark’s Museum of National History has been open to the public in Frederiksborg Castle since 1882.

The attendees enjoyed a second special treat that evening with a gourmet banquet served back at the Pharmakon Center, where the chefs had prepared a fine combination of appetizers, wines, roasted red fish, roasted Grambogård pork, and lemon Mazarin dessert, providing a delicious and memorable finale to a productive and enjoyable day.

FORENSIC AUDIO STANDARDS
Saturday, the final day of the conference, opened with a presentation and discussion of the emerging area of forensic audio accreditation and standards, hosted by Michael Piper of the U.S. Secret Service, Washington, D.C., and David Hallimore of the Houston Police Department, Houston, Texas. Both presenters are leaders of the Audio Committee within the Scientific Working Group on Digital Evidence (SWGDE, pronounced “swig-dee”), which is a cooperative organization with members from local, state, federal, and international law enforcement and investigative agencies who share information and education in the field of digital forensics. Piper and Hallimore led a discussion dealing with the implications of the U.S. National Research Council’s 2009 report entitled “Strengthening Forensic Science in the United States: A Path Forward.” The NRC report was highly critical of many forensic practices, citing the need to establish statistical reliability measures for forensic comparisons, and the need for bona fide standards for training and experience among forensic practitioners. Although the NRC report was not specifically focused on audio forensics, the implications of the report on the future admissibility of audio forensic evidence and testimony spurred the SWGDE Audio Committee to seek input on how best to address the NRC criticisms and recommendations. Piper and Hallimore invited additional comments and suggestions for future action.

AUTOMATED SPEECH PROCESSING
After the morning coffee break, Anil Alexander of GriffComm, Oxford, UK, returned to the stage to describe a technique for semiautomatic speaker segmentation for processing recorded testimony. The technique was developed for the Metropolitan Police, London, to assist with preparing recordings in which the identity of specific talkers is protected, such as the voices of undercover officers or vulnerable witnesses who have been granted anonymity by the court. Such a recording can be processed manually by a technician who identifies segments of speech by the protected talker and selectively disguises or deletes those segments. Alexander explained that this manual process is tedious and time-consuming and therefore an automatic segmentation system would be of great value. The proposed technique involving speech recognition and transition identification was found to be effective and promising, although practical issues, such as determining the appropriate speech model order and discerning simultaneous speakers (over-talking), will require further study.
The second paper on automated speech processing, “Automatic Forensic Voice Comparison Using Recording Adapted Background Models,” was presented by Timo Becker of the Federal Criminal Police Office (Bundeskriminalamt), Germany. The automatic forensic voice-comparison system uses a standard Gaussian Mixture Model (GMM) approach for text-independent speaker recognition, but with a novel adaptation to help account for the widely varying acoustical environments encountered in forensic audio recordings. The authors developed a recording adapted background model (RABM) for use instead of the classical universal background model (UBM). Test results indicate that voice-comparison performance is improved with the RABM method, but there remain several challenges having to do with database mismatch and separating the channel characteristics from the speaker’s speech characteristics.

**SPEECH QUALITY AND INTELLIGIBILITY**

Rounding out the Saturday morning sessions was the first of five papers addressing speech intelligibility. Nikolay Gaubitch of Imperial College, London, spoke about his research team’s results in estimating speech intelligibility using rapid subjective testing. The work is important because noise-reduction and quality-enhancement processing of forensic speech recordings generally results in lower intelligibility, but quantifying the tradeoff between quality and intelligibility has been hard to pin down. Gaubitch presented the Bayesian Adaptive Speech Intelligibility Estimation (BASIE) method, and described several simulations and experiments used for its characterization and validation. The pilot tests showed the ability to estimate the Speech Reception Threshold (SRT) within ±1 dB in under 30 trials. Gaubitch indicated that the BASIE method should allow a relatively straightforward approach to find appropriate levels of enhancement and noise reduction that do not compromise intelligibility.

Following the final tasty lunch of the conference, the group reconvened for the closing paper session. Dushyant Sharma of Imperial College, London, presented a paper about using the Perceptual Evaluation of Speech Quality (PESQ) algorithm to assess the effects of forensic audio processing and enhancement systems. PESQ was developed as an objective means to rate the degradations associated with telecommunications channels and codecs, eliminating the time and cost of subjective testing with human subjects. Until now, PESQ has not been applied to forensic audio processing, so Sharma’s group performed an experiment to compare PESQ with a panel of test subjects. The result was that the correlation between PESQ and the subjective ratings was poor for the substantial degradations in quality typically encountered in forensic material, indicating that some other approach will be needed for this purpose.

Next, Andrea Paoloni of the Ugo Bordoni Foundation, Rome, presented a paper describing the use of the Speech Transmission Index (STI) as a possible objective measure of signal intelligibility. The STI was developed 40 years ago as a means to assess the impact of a degraded channel on the intelligibility of speech. The STI determines the degree to which the normal spectral and temporal envelopes of a speech-like test signal are maintained through the channel. The results of several experiments indicate that the STI-based measurement is useful for classifying forensic audio systems.

Returning to the quality versus intelligibility tradeoff, Mark

---

The 39th conference committee: from left, Zoe Asta, Alan Cooper, Preben Kvist, Durand Begault, Gordon Reid, Subir Pramanik, Anna Lawaetz, Katrine Beigh Brixen, Roger Furness, Eddy Beigh Brixen, and Catalin Grigoras.
Huckvale of University College London, presented a paper entitled “Measuring the Effect of Noise Reduction on Listening Effort.” A set of intriguing experiments were conducted to see if listener effort was reduced when listening to noise-reduced speech compared to listening to the original noisy speech without processing. The major result of the experiments was that there did not appear to be any improvement in productivity when the subjects listened to speech processed for noise reduction compared to the unprocessed speech, and in certain cases the performance was actually worse with the “enhanced” speech. Huckvale suggested that enhancement systems must be designed neither to make the residual noise more speech-like, nor the residual speech more noise-like, since these attributes appear to interfere with intelligibility at the phonetic level.

The final paper of the conference was “Practical and Affordable Intelligibility Testing for Engineers and Algorithm Developers,” by Ken Worrall and Rob Fellows of Her Majesty’s Government Communications Center (HMGCC), UK. The paper was presented by their HMGCC colleague, Louise Baddeley. The project involved development of a procedure for rapid intelligibility determinations that would provide fast and useful information for an algorithm or system designer. While their proposed procedure, “Technique for Automated Comparative Intelligibility Testing” (TACIT), is not a statistically rigorous technique, Baddeley explained that it is sufficient to allow design and engineering decisions during a rapid development cycle.

AES AUDIO FORENSICS: STILL RISING AND SHINING

The AES 39th Conference was judged a great success, and AES clearly remains the leader in the field of forensic audio analysis and interpretation. The conference concluded with a sincere thank you from Eddy Brixen to the organizing committee and all of the participants. Roger Furness, AES executive director, added his words of thanks and praise to the committee, and invited all nonmembers to join the AES and to make plans to attend future Society events.

Many of the attendees took advantage of a special charter bus for transportation from the Pharmakon Center back to Copenhagen Kastrup Airport, allowing everyone time for a few extra minutes of continued discussion and conversations. Some attendees were heading home, some were embarking on vacations or other business, but all traveled with keen anticipation of another AES conference on forensic audio in the near future.

Editor’s note: The conference papers are available for purchase as a book or as a downloadable PDF at www.aes.org/publications/conferences. Individual conference papers can also be obtained from the AES E-Library at www.aes.org/e-lib. The National Center for Media Forensics at the University of Colorado Denver will be holding an Audio Forensics Workshop December 13-15. For information contact Leah Haloin at leah.haloin@ucdenver.edu or +1 303 315 5852.