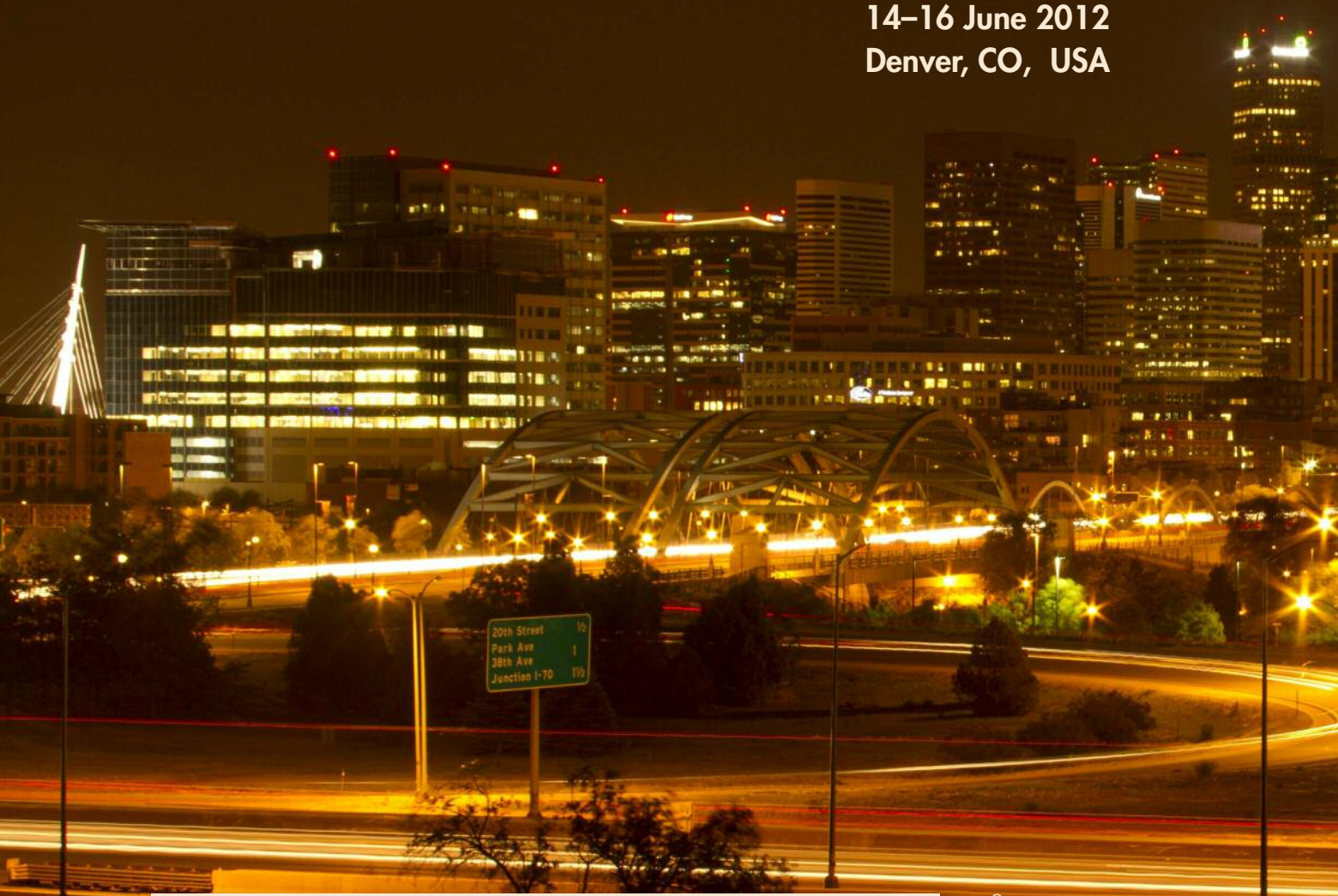


AES 46th International Conference

Audio Forensics—Recording, Recovery, Analysis, and Interpretation

14–16 June 2012

Denver, CO, USA



In June 2012 the city of Denver, Colorado, welcomed the AES audio forensics community back to the Mile-High City for another outstanding conference. Participants from all over the world gathered to share information on research and practice in forensic science. The AES 46th International Conference, *Audio Forensics—Recording, Recovery, Analysis, and Interpretation*, was the most recent AES event focusing on the field of audio forensic analysis and interpretation. The sequence of AES audio forensic conferences began in 2005 with the 26th AES Conference held in Denver. The 33rd Conference returned to Denver in 2008, followed by the 39th Conference that convened in 2010 in Hillerød, Denmark.

The 85 registered participants included representatives from more than 15 countries. The conference brought together an outstanding combination of practitioners, researchers, law-enforcement professionals, attorneys, and many other experts and students all sharing an interest in the latest developments and contributions to audio forensic science made by AES members.

Sponsors

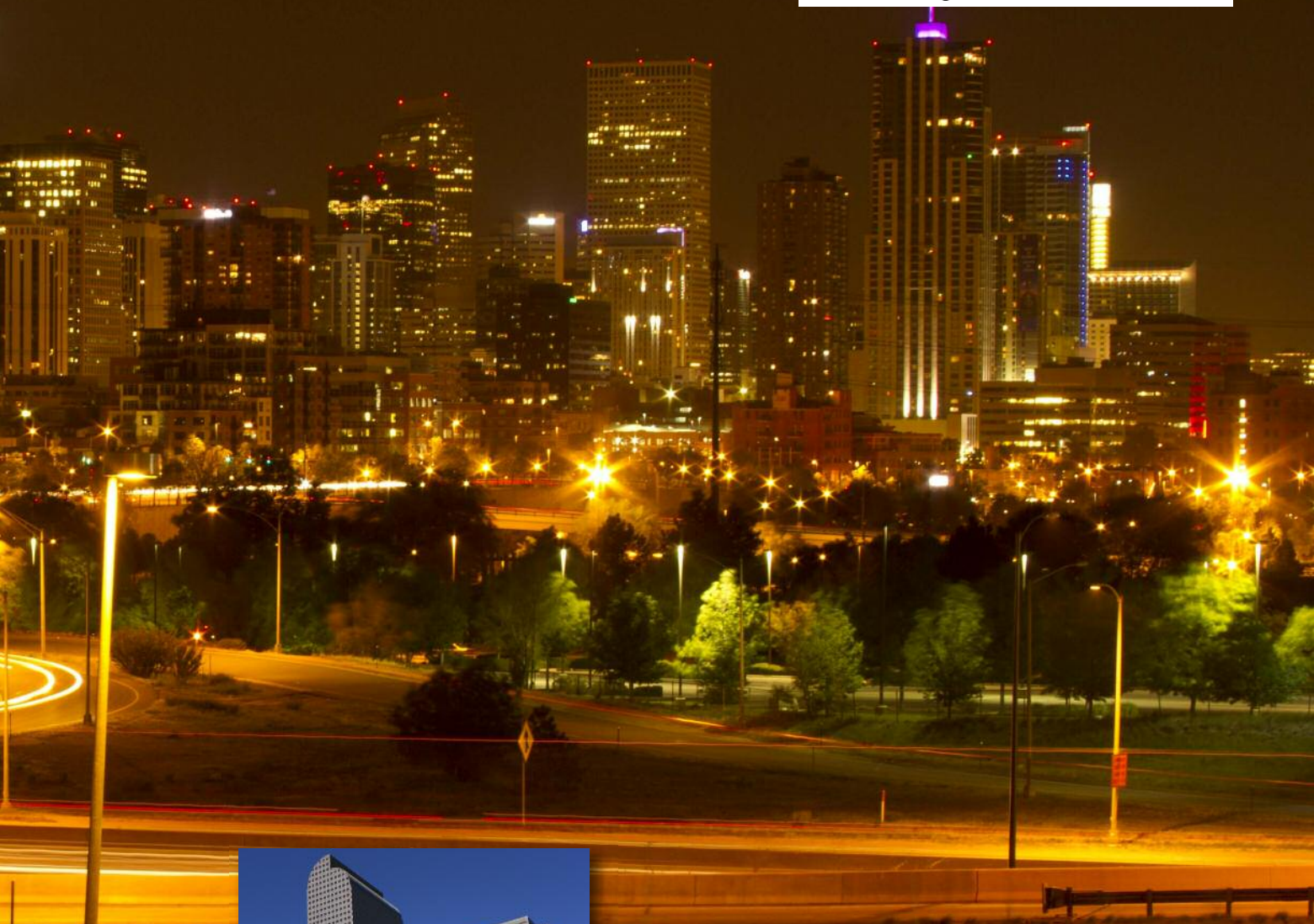


Hosted by



University of Colorado
Denver

National Center for Media Forensics
College of Arts & Media



The Warwick Hotel, adjacent to the
National Center for Media Forensics



Statue of a cowboy riding a bucking
bronco portrays Denver's pioneering
spirit.

The many months of planning prior to the 46th Conference involved members of the AES Technical Committee on Audio Forensics, the AES Colorado Section, the AES University of Colorado Denver Student Section, and the AES headquarters staff. The 46th Conference emulated the success of the prior audio forensics meetings by recruiting an outstanding organizing committee. Jeff M. Smith, conference chair, was joined by papers co-chairs Catalin Grigoras and Durand Begault, workshops co-chairs Eddy Bøgh Brixen and Christopher Peltier, treasurer Joe Erickson, facilities and registration coordinator Leah Haloin, Social Coordinator Kellyn Smith, and local committee chairs Wanda Newman and Ann Sanders. The committee's superlative efforts created a top-notch technical program and a fun and collegial conference atmosphere.

The conference venue was the Warwick Denver Hotel, located in bustling downtown Denver next door to the National Center for Media Forensics and only about three blocks from the Colorado State Capitol building. The accommodations and meeting facilities were well-appointed and comfortable. The attendees took advantage of the traditional AES international conference format, which is deliberately designed to provide opportunities for small-group interaction and easy face-to-face discussion among the participants. It was clear that the attendees made full use of the conference venue's fine features.

CONFERENCE OPENING

The conference opened on Thursday morning, 14 June, with a very pleasant mid-summer day. Despite the visibility of smoke from the High Park wildfire in the mountains 120 km (75 miles) northwest of Denver, the conditions downtown in the Mile High City were clear and very comfortable. The formal program began with introductory remarks by Jeff Smith, conference chair. Smith, associate director of the National Center for Media Forensics of the University of Colorado Denver (NCMF UCD), gave a warm welcome and introduction to the participants, sponsors, and special guests and provided a compelling overview of the conference. Smith thanked the local organizing committee and volunteers from the AES

Colorado Section and the UCD AES Student Section. Smith introduced Roderick Nairn, UCD provost and vice chancellor for academic and student affairs, who added his official words of welcome on behalf of the University of Colorado Denver campus.

Key sponsors and exhibitors for the AES 46th Conference included Agnitio, Blue Collar Audio, Cognitech, Digital Audio Corporation, and AES Sustaining Member iZotope, Inc. In addition to brief opening presentations, the exhibitors each provided information and hands-on demonstrations throughout the conference in the poster presentation area immediately adjacent to the main conference room.



Phil Mellinger during his keynote address on the Watergate tapes.

Jeff Smith acknowledged Geoffrey Stewart Morrison, who had provided a special pre-conference evening event entitled "Workshop on Validity and Reliability in Forensic Voice Comparison." He also acknowledged the NCMF staff for providing special pre- and postconference training courses entitled "Forensic Authentication of Digital Audio" and "Forensic Audio Enhancement."

KEYNOTE LECTURE

Phil Mellinger, chief scientist at Trusted Knight Corporation, provided a special opening keynote presentation based upon his published article from *Forensic Magazine* (vol. 8, no. 1, Feb/Mar 2011, pp. 19–24) about his recent work investigating the infamous "18½ minute gap" in one of the key Watergate tapes recorded June 20, 1972, from the Nixon White House. In 2004, Mellinger became interested in some of the unresolved mysteries of the Watergate era, such as the unknown (at that time) identity of the individual known as "Deep Throat" who had provided information to *Washington Post* reporters Bob Woodward and Carl Bernstein. Even after Deep Throat's identity was revealed in 2005 to be W. Mark Felt, FBI associate director (1971–1973), the Watergate topics remained of interest to Mellinger, and he explained how he became particularly curious about the mystery of how the 18½ minute gap occurred. Various theories have been proposed over the years about who might have been involved in the erasure besides President Nixon's secretary, Rose Mary Woods. Mellinger presented his evidence that the "gap" tape was erased in three separate stages: first a 4-minute-35-second portion using a recorder in Rose Mary Woods' office, a second 12-minute-46-second erasure that took place somewhere other than Woods' office, and a final 1-minute-9-second erasure while again in Woods' office.

TECHNICAL PROGRAM

The conference papers co-chairs, Catalin Grigoras of the National Center for Media Forensics, Denver, CO, and Durand Begault of Charles M. Salter Associates, San Francisco, CA, put together a wide range of paper sessions covering audio forensic research and practice. The conference workshops co-chairs, Eddy Brixen of EBB-Consult, Denmark, and Christopher Peltier of Charles M. Salter Associates, San Francisco, did an equally fine job developing a fascinating slate of topics and workshop presenters.



Jeff Smith, conference chair, opens the event.



Roger Furness, AES deputy director, welcomes delegates to Denver.

Workshop 1: Forensic Audio Enhancement

The technical portion of the conference began with a special workshop session on audio enhancement, moderated by Christopher Peltier. The first presenter on the workshop was Mark Huckvale of University College London, London, UK, on the topic of enhancement of speech in noise. Huckvale described several research projects conducted by the Centre for Law Enforcement Audio Research (CLEAR), a joint research center of University College London and Imperial College London. Among their findings was a way to describe the effect of audio enhancement processing as a “noise” shift in the psychometric function, which helps describe the common situation in which enhanced audio is judged to be of better perceptual quality than the original signal, but the intelligibility of the speech is actually degraded. Researchers at the CLEAR center performed a study with twelve enhancement modules from five commercial audio enhancement systems over the full range of each systems’ control parameters and found that all had at least one set of parameters that slightly increased intelligibility (up to 2 dB improvement), but the modules also had settings that degraded intelligibility by an even greater amount (up to 3.8 dB degradation). Based on these findings, Huckvale’s group recommends that forensic examiners be very careful about choosing the appropriate product and parameter settings to suit the signal and the processing scenario.

The second workshop presenter was Eddy Brixen, who provided a very interesting set of recommendations regarding the preparation of audio material for presentation to untrained listening panels, such as juries in court proceedings. Listeners in these situations are generally not screened for hearing impairment, and the playback circumstances and acoustical surroundings are typically uncontrolled. Brixen’s recommendations include the need to explain carefully what is to be heard, to make sure the playback level is appropriate, and to review the examples using the same equipment that will be used in court.

Paper Session 1: Enhancement of Forensic Audio I

After the opening lunch break on Thursday, the conference paper presentations began with the topic of audio enhancement. The first paper on the session concerned error concealment in audio signals. The work by Stephan Preihs, Fabian-Robert Stöter, and Jörn Ostermann of Leibniz Universität Han-



Authors, from top: Mark Huckvale, Richard Conners, Daniel Rappaport, Geoffrey Stuart Morrison, Keith McElveen

nover, was entitled “Low Delay Error Concealment for Audio Signals.” The presentation described two model-based methods suitable for real-time extrapolation of missing audio samples in a stream using Kalman filtering or variable-order linear prediction. Although not specifically tied to audio forensics, the error concealment concepts could be applied to recover degraded audio from forensic sources.

The second enhancement paper was entitled “Music and Noise Fingerprinting and Reference Cancellation Applied to Forensic Audio Enhancement,” by Anil Alexander and Oscar Forth of Oxford Wave Research, and Donald Tunstall of Digital Audio Corporation. The authors described an interesting method to cancel interfering sound in a forensic audio recording if the unwanted sound is from a known source, such as a commercial music recording or an archived broadcast. Their approach is to identify the interfering sound material, time-align the known material with the forensic recording, and then use a least-mean-square (LMS) algorithm to model adaptively the acoustical effects of the room in which the recording was made. The approach can also be applied if simultaneous recordings are made in the same room using spatially separated microphones or even using completely separate recording systems (e.g., two smartphones). As long as the time-varying time alignment of the recordings can be estimated, the cancellation quality can be very good.

Poster Session 1: Miscellaneous Techniques

A scheduled break in the middle of the afternoon provided an opportunity for the participants to enjoy some snacks and beverages while visiting with the conference exhibitors and the first poster session authors.

A poster entitled “Tone Removal Using a Band Focus Speech Reconstruction Algorithm” by Darren M. Haddad and Andrew J. Noga of the Air Force Research Laboratory, Rome, NY, presented a method to remove very narrowband interference from audio recordings using a super-resolution spectrum-analysis technique.

Two of the posters dealt with the observation that some digital audio recorders and coding algorithms might introduce a systematic DC offset that might be useful for audio authentication. “The Effects of Audio Compression Algorithms on DC Offset,” by Daniel Fuller, and “The Gain Effect: How Does it Affect the DC?” by Sean Jacobson, both of the National Center for Media Forensics, Denver, CO, reported on experiments to measure and classify the offset characteristics.

The fourth poster in session 1 was “Designing an Automated Gunshot Detection and Image Response System,” by Jordan R. Graves, also of NCMF, Denver. Graves was unfortunately unable to attend the session, but his poster described plans for a camera system to be steered by acoustic detectors and triangulation.

Paper Session 2: Enhancement of Forensic Audio II

Following the enjoyable break and poster session, the late afternoon technical session convened with two paper presentations on the topic of audio enhancement. The first paper, “Enhancing Low SNR Speech Corrupted by Non-Stationary Tonal Noises,” by Scott Nordlund and J. Keith McElveen, covered a preprocessing technique to model the tonal peaks in the noise spectrum under the assumption that the noise characteristics vary more slowly than the desired speech spectrum. The process reduces (whitens) the noise spectrum by reducing the spectral peaks. The resulting signal is then more suitable for subsequent processing with spectral subtraction or some other noise reduction technique.

The next paper entitled “Effects of Replay on the Intelligibility of Noisy Speech” was presented by Mark Huckvale of University College London. The paper described a very interesting experiment to test the typical intuition that repeated listening to a noisy speech recording will improve its intelligibility. The results indicate that listeners do get improvement by listening more than once, but the improvement does not continue to increase after four or five repetitions, even though the listeners think that their performance continues to increase upon hearing even more repetitions. Huckvale explained that the intelligibility improvement attributable to the multiple repetitions is approximately the same as a 1.5 dB improvement in signal-to-noise ratio (SNR).

Upon the conclusion of the first day of the conference, the attendees gathered for conversation, snacks, and beverages at a cocktail reception in the 15th floor ballroom. Amid views of the east-central Denver neighborhoods, the AES conference participants enjoyed the opportunity to relax and unwind while engaging one another in lively discussions of the day’s topics and

presentations. The combination of first-time attendees and AES members who have attended numerous AES conferences provided an excellent setting for collegial discussion.

Paper Session 3: ENF Analysis I

Friday, June 15, 2012, opened with a fine informal buffet breakfast and coffee service. The technical sessions began with a set of technical papers covering various aspects of Electrical Network Frequency (ENF) analysis. ENF refers to the tell-tale presence of “hum” in an audio recording due to leakage or coupling of the alternating current (AC) signal from the electrical power grid into the audio circuits. If an ENF signature is detected in the audio, it is possible to compare the subtle frequency fluctuations in the ENF signal to a database of electrical grid frequency history information and thereby estimate the date and time at which the audio recording was made. The electrical grid frequency tends to vary around its nominal value (e.g., 60 Hz in the U.S., 50 Hz in Europe) due to the normal, random variation in generator capacity and electrical load changes on the grid.

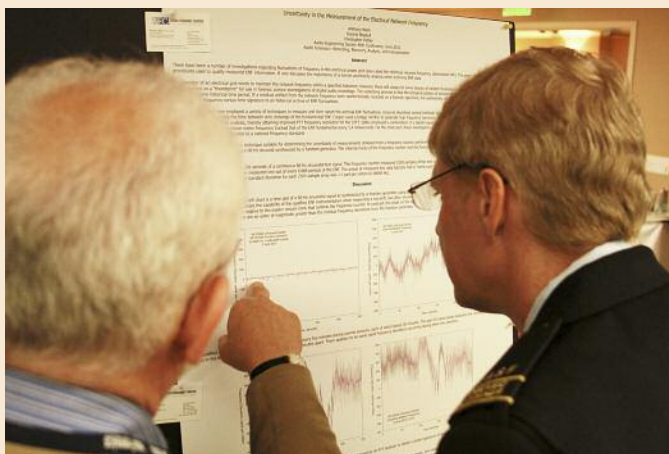
The first ENF paper “Phase & Amplitude Analysis of the ENF for Digital Audio Authentication,” was by Sean Coetzee of Prism Forensics, Los Angeles, CA. Coetzee described his work in examining the phase of the extracted ENF signal to detect discontinuities or phase “jumps” that might indicate tampering. Next, Richard Connors of Virginia Tech, Blacksburg, VA, presented a paper entitled “Effects of Oscillator Errors on ENF Analysis,” which investigated the effect of crystal oscillator discrepancies between the recording device and the ENF reference database. If the recorder’s clock crystal is inaccurate, a systematic frequency offset will be present between the ENF extracted from the audio signal and the separately determined grid reference. Connors’ research team has developed a reliable procedure to detect and compensate for the discrepancies attributable to the crystal oscillator frequency errors.

The third paper on the ENF session was by Harrison Archer of the National Center for Media Forensics, Denver. Archer’s paper, “Quantifying Effects of Lossy Compression on ENF Signals,” reported on his MS thesis work that examined the effects of ten compression algorithms on 100 different hours of recorded ENF signals sampled over a five-month period. Archer found that the ENF information appeared satisfactory for automated matching with eight of the ten tested coding algorithms. The test was done on ENF signals alone, and therefore future work will be needed to determine the effect on the low-frequency ENF band when regular audio material is present, especially in the case of perceptual audio codecs.

The final paper on the Friday morning session was “A Study of the Accuracy and Precision of Quadratic Frequency Interpolation for ENF Estimation,” presented by Richard Connors of Virginia Tech. The research involved the use of parabolic (second-order) interpolation of discrete Fourier transform (DFT) magnitude spectra to refine the frequency estimate for underlying signal components presumed to be quasi-sinusoidal, such as the ENF component in an audio recording. The work largely duplicated and confirmed the results from prior studies in sinusoidal analysis dating back to the 1980s.



Delegates concentrate hard during a presentation on Electrical Network Frequency analysis.



Interested delegates study the poster on ENF by Nash, Begault, and Peltier.

Workshop 2: WinHex for Forensic Audio Analysis

The Friday morning program concluded with a special workshop presentation by Doug Lacey of BEK TEK LLC, Clifton, VA. Lacey provided an overview and simple tutorial of digital forensics features of the WinHex software package. WinHex is a product of X-Ways Software Technology AG of Cologne, Germany. WinHex provides a wide variety of software features for displaying binary digital file contents in hexadecimal form, giving a visual representation of the file contents. Lacey explained several forensic analysis situations in which the interpretation of digital file contents can be helpful, such as interpretation of unknown file formats and proprietary coding methods.

Paper Session 4: ENF Analysis II

After an enjoyable lunch break, the attendees reconvened for the Friday afternoon paper presentations to hear about additional results on ENF analysis. The first paper, "Advances in Electric Network Frequency Acquisition Systems and Stand Alone Probe Applications for the Authentication of Digital Media," was presented by Chris Jenkins of Blue Collar Audio, Denver, CO. Jenkins described several important considerations necessary to ensure the quality and integrity of ENF data acquisition and storage systems for use as reference databases. In particular, Jenkins pointed out the issue of time base and sampling synchronization, since the digital sample clock timing in most analog to digital conversion systems may not be sufficiently accurate for long-term timing reliability.

The second paper in the session was another presentation by Richard Connors of Virginia Tech. His presentation, "Using Simple Monte Carlo Methods and a Grid Database to Determine the Operational Parameters for the ENF Matching Process," described an investigation of how reliable the matching process may be when comparing an ENF record extracted from an audio recording with the reference database. Connors pointed out that the extracted ENF will have noise and distortion that could reduce the confidence of finding a match between the evidentiary recording and the database. Additional work will be needed to quantify the typical noise characteristics of extracted ENF sequences, and to determine the degree to which the noise affects a forensic examiner's interpretation of the ENF results.

The ENF session concluded with a paper authored by Catalin Grigoras and Jeff Smith of the National Center for Media

Forensics entitled "Advances in ENF Analysis for Digital Media Authentication." Grigoras explained a series of investigations into the quality and reliability of ENF extraction and analysis. A potentially large range of candidate matches occur when comparing an extracted ENF segment with a reference database using crosscorrelation or using mean quadratic difference. The length of the evidentiary recording and the criteria used to establish a match can have a large effect on the reliability of the process.

Poster Session 2: ENF Analysis

The mid-afternoon break for refreshments, exhibits, and posters included three presentations on ENF analysis. A poster by Nicholas Ng of the National Center for Media Forensics described an assessment of the integrity and consistency of the two ENF databases currently being recorded in Denver and in Las Vegas. Alireza Sanaei of Anglia Ruskin University, Cambridge, UK, presented his work entitled "Tuning and Optimization of an ENF Extraction Algorithm," which involved optimizing ENF extraction by using the network frequency harmonics, not just the fundamental frequency. Given a certain frequency measurement precision, Sanaei showed that the fundamental frequency estimate is improved by looking at the harmonic frequencies and then dividing by the harmonic number. The third poster on the session was presented by Anthony Nash of Charles M. Salter Associates, Inc., of San Francisco, CA. Nash's presentation, coauthored by Durand Begault and Christopher Peltier, considered how best to quantify the accuracy and precision of ENF measurements, particularly the ability to trace the frequency measurement to a standard time reference. Nash expressed his recommendation that the ultimate usefulness of ENF databases will require quantifying the measurement uncertainty of frequency counters and frequency measurement algorithms that are calibrated and traceable back to a national frequency standard.

Paper Session 5: Audio Authentication

As the conference momentum continued to build, the final Friday paper session comprised a collection of four papers in the area of audio interpretation and authentication. The first paper, "Analytical Framework for Digital Audio Authentication," was presented by Daniel Rappaport of the National Center for Media Forensics. Rappaport gave an interesting overview of the authentication opportunities and challenges in the era of digital audio recording, editing, and processing. The second session paper was by Hafiz Malik of the University of Michigan-Dearborn, Dearborn, MI, who presented "Microphone Identification Using Higher-Order Statistics." Malik's work involved a set of experiments to determine if there are sufficient unique artifacts to be found in an audio recording that are attributable to the microphone used to make the recording. The goal is to have a method to detect forgeries or fabricated evidence by noting inconsistencies between the recorded signal and the microphone characteristics. Malik explained that this is a difficult problem because it is not easy to separate the distortion of the microphone from the spectral characteristics of the signal itself. Further work will be needed to deal with the many variables involved in the signal chain of even a seemingly simple audio recording, and to handle the nonlinear aspects that may influence the higher-order statistical methodology.

The third presentation was entitled "Using Ripple Signals for the Authentication of Audio," by Dagmar Boss of the Bayerisches

Landeskriminalamt (Bavarian State Criminal Police Office), München (Munich), Germany. Boss explained that special signaling tones are used in the electrical power system in some countries such as Germany, Australia, New Zealand, South Africa, and the United Kingdom. The special signaling tones, known as ripple signals, are in the 100 Hz to 2 kHz range, and are used to provide load management of the electrical grid. Devices attached to the power grid can detect the ripple signals and adjust their load behavior, for example, by shutting down nonessential systems during peak load conditions. Just like the ENF signals, the ripple signals can be coupled into an audio recording system and may be detectable in the recorded audio data. Boss suggested that the ripple signals could potentially provide an additional piece of evidence for assessing digital audio authenticity.

The final paper on the Friday session was “Evaluation of the Average DC Offset Values for Nine Small Digital Audio Recorders,” by Bruce Koenig of BEK TEK LLC, Clifton, VA. Koenig described the results of an experiment to see if recordings made with nine different off-the-shelf digital audio recorders might exhibit DC offset characteristics that would be meaningful for evaluating forensic audio authenticity. Unfortunately, the preliminary results were found to have inconsistencies and a sufficiently large variance that would likely rule out this approach for authenticity assessment.

Friday Social Event: Banjo Billy and Katie Mullen

Eager to wind down after a couple days of intense learning and discussion, the conference attendees welcomed the opportunity to enjoy an informal social event on Friday evening. The conference organizers chartered an excursion through downtown Denver with Banjo Billy, a local Denver personality who guides historic tours on a bus modified to look like the ramshackle home of a hillbilly. The bus picked everyone up at the hotel and then drove around the downtown area while Banjo Billy shared Denver history, jokes, ghost stories, and other local lore while the riders enjoyed a selection of beverages obtained from several Colorado breweries and wineries.

Fortunately, the all-too-soon conclusion of the bus tour did not signal the end of the festivities, as the group next enjoyed a fine reception and dinner at Katie Mullen’s Irish Restaurant and Pub, located on the fashionable 16th Street Mall in the Sheraton Denver Downtown Hotel. Following dinner, many conference participants headed for a stroll along the 16th Street Mall to enjoy the sights and sounds of a comfortable Friday evening, before walking a few blocks back to the conference venue at the Warwick Hotel.

Paper Session 6: Miscellaneous Techniques

Saturday, June 16, 2012, marked the third and final day of the conference. With the Friday evening social activities still fresh in everyone’s memory, the breakfast buffet and a large mug of coffee or tea provided a nice way to gear-up for the technical sessions.

The opening session on Saturday included four presentations on miscellaneous audio forensics techniques. The first presenter was Mark Huckvale of University College London, who gave a lively and informative talk entitled “Effectiveness of Electronic Voice Disguise Between Friends.” The paper dealt with the audio forensic situation in which the identity of a witness is to be disguised to protect him or her from potential retribution. The challenge is to maintain intelligibility while concealing the witness’ identity even if the defendant is actually well-acquainted with the witness. Huckvale described an experiment in which a group of students who all knew each other very well were asked to identify classmates based on audio recordings of each student reading a short script. The students had a 90% success rate in identifying one another with undisguised playback of the recordings.

The researchers found that even with rather extreme pitch and simulated vocal tract length changes, such as a shift of 8 semitones (frequency shift factor of 1.6) and a tract length change of 120%, identification performance was still better than chance, and certain distinctive talkers were found to be very recognizable. Based on these results, Huckvale suggested that there is a significant risk of identifying a familiar talker even with extensive vocal disguise, and this risk should be considered very carefully by the prosecuting authority.

Next, Keith McElveen of Wave Sciences Corporation, Charleston, SC, presented a paper entitled “GMM-Based Efficient Language Identification,” that described the use of Gaussian Mixture Models (GMMs) for statistical classification and identification from a recording of speech uttered in an unknown language. McElveen explained the use of a perceptual minimum variance distortionless response (PMVDR) algorithm for the processing front-end. The system successively compares the input speech segment to multiple GMMs, with each GMM trained for a particular language, then selects the best match. The results were very good using the 1994 Oregon Graduate Institute (OGI) Multilanguage Telephone Speech database, with somewhat lower performance using the African Speech Technology (AST) telephone speech database. McElveen suggested that the lower AST performance was due to a mixture of dialects present in each of the AST language groups.



Exhibitors iZotope (top) and Cognitech (bottom) show their analysis and capture tools to delegates.

The third paper, “Automatic Search and Classification of Sound Sources in Long-Term Surveillance Recordings,” was authored by Robert C. Maher and Joseph Studniarz of Montana State University, Bozeman, MT. Maher explained the challenges involved in forensic interpretation of audio surveillance recordings that are days, weeks, or even months in duration. Maher’s research has employed automatic search methods based on spectral templates and two-dimensional (time–frequency) filtering to identify changes in the sonic texture. Maher showed an example experiment that found a probable gunshot within a 30-hour recording.

The final paper on the Saturday morning session was “Carving and Reorganizing Fragmented MP3 Files Using Syntactic and Spectral Information,” presented by Sascha Zmudzinski of the Fraunhofer Institute for Secure Information Technology, Darmstadt, Germany. File carving refers to the recovery and reconnection of deleted or damaged files in a digital storage system, such as a hard disk drive. Modern computer operating systems typically are able to utilize noncontiguous sectors on the disk for storing large files, so a single file may have sections stored in many different physical locations on the disk, which is known as file fragmentation. If the file table is deleted or damaged, it may be difficult to find and piece together the fragmented information. Zmudzinski’s team has developed several techniques to perform file carving for MP3 files, including locating the MP3 header frame codes, applying file structure and frame length rules, and then using spectral matching from the MP3 data to infer the file, frame, and sector continuity.

Workshop 3: Expert Testimony

The third workshop for the conference commenced after a brief morning break for refreshments and conversation. Jeff Smith moderated the workshop, entitled “Forensic Science and Expert Testimony: Insights and Strategies from the Advocate’s Perspective,” and introduced the three workshop presenters.

The first presentation was in the form of a pre-recorded video produced by Peter Weinberg, a trial and litigation consultant with Litigare Litigation & Trial Design Consulting, LLC, of Denver, CO. In the video, Weinberg described the role of the expert witness in legal proceedings and emphasized the difference between advocacy for one side or the other in the legal arguments versus being an advocate for the truth. He recommended that the expert witness serve in the role of educator: working to be an effective teacher for the court and for the jury. He expressed his opinion that a good educator must be credible, trusted, and authoritative on technical matters, while also appearing friendly, interested, and engaged. An expert witness who is aloof, argumentative, or seems to be acting as an advocate for the client can often be discounted by the jury.

The second presenter for the expert witness workshop was Joseph Mathews, director of government relations for Milliman Care Guidelines, Seattle, WA. Mathews was unable to be in Denver, but he provided his presentation via an interactive Skype hookup. Mathews spoke about the importance of expert witnesses in civil litigation. Attorneys very quickly think about securing

expert testimony when they begin working on a new case. Depositions are very important in civil litigation and Mathews suggested that expert witnesses be carefully prepared and rehearsed prior to a tough deposition, including being of consistent tone and demeanor during both direct and cross examination.

The final presenter for the Expert Testimony workshop was Henry R. Reeve, Denver District Attorney’s Office, who spoke about the role of expert testimony in criminal cases. Reeve emphasized the high stakes involved in any criminal proceeding: the court must determine whether or not the government will be allowed to deprive an individual of his or her basic rights and freedom. Reeve recommended developing a clear understanding of each person’s role in the legal process. An expert witness must not try to “out-lawyer” the attorneys, as the roles of the expert and the attorney are different in the court proceeding and judges and juries are not interested in having expert witnesses spar with the legal counsel. He also explained that in most criminal legal venues in the United States the notes prepared by an expert working on a criminal case are generally discoverable, meaning that the expert’s notes must be provided to the prosecuting and defense attorneys.

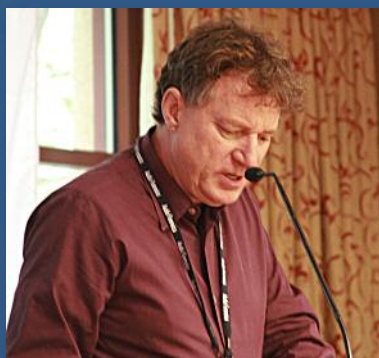
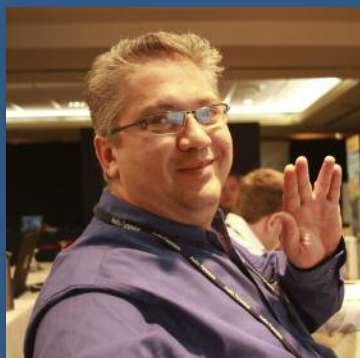
Paper Session 7: Laboratory Practices

After a relaxing Saturday lunch break, the attendees reconvened for the final afternoon of the conference. David Hallimore of the Houston Police Department and Michael Piper of the U.S. Secret Service provided information and remarks about the Scientific Working Group on Digital Evidence (SWGDE) and its current effort to define a core-competencies document for forensic audio examiners. Hallimore and Piper explained that the core-competencies document is being developed in response to the National Academy of Sciences’ 2009 report “Strengthening



Banjo Billy’s Bus Tour as enjoyed by many delegates to the 46th conference.

SOCIAL EVENTS



Top row, from left, Catalin Grigoras, Durand Begault, and Daniel Fuller (volunteer)

Middle row, from left, Wanda Newman and Ann Sanders, Jeff Smith, and Leah Haloin

Bottom row, Eddy Brixen (left) and Christopher Peltier (right), flanking author Mark Huckvale

CONFERENCE COMMITTEE MEMBERS

Forensic Science in the United States: A Path Forward,” that strongly criticizes the forensic sciences in general for a lack of enforceable standards and accredited education and training opportunities. The SWGDE group is seeking additional feedback from the audio forensics community as the core-competencies document is revised and finalized.

Paper Session 8: Speaker Recognition and Comparison I

The afternoon presentations continued with a paper session on speaker recognition. The first presentation was “Comparing Auto-

matic Forensic Voice Comparison Systems Under Forensic Conditions,” by Timo Becker of the Bundeskriminalamt (Federal Criminal Police Office), Germany. The experiment compared seven custom and commercial voice-comparison systems using audio material from the German Speech Database produced by the Bavarian State Police. The results showed that the seven systems all performed about equally, but each system reacted differently to changes in channel characteristics, spoken dialect, and speaker accent. Becker recommends that human forensic examiners must be involved as much as possible in assessing and evaluating the results of the automatic systems.

Next, Geoffrey Morrison of the University of New South Wales,

Australia, presented a paper under the intriguing title “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.” The audio material was from a notorious 1995 murder case and 2009 retrial in New Zealand. The dispute centered on a marginally intelligible utterance in an emergency-call-center recording that had highly conflicting interpretations by the prosecution and the defense. Morrison described his work to develop a likelihood ratio of the probability that the word “shot” was uttered (prosecution theory) versus the probability that the word “can’t” was uttered (defense theory), using acoustical analysis of that portion of the audio recording. Morrison concluded that the likelihood that the defendant had uttered “can’t” was 31,000 times more likely than the likelihood that the word “shot” was spoken. He concluded his presentation with several recommendations for future investigations of this type and he shared some concerns about ensuring validity of the testing process.

Poster Session 3: Speaker Recognition and Comparison

Following the Session 8 paper presentations, the afternoon break and poster session continued the same theme, with four fascinating posters covering topics in voice comparison and voice recognition.

“Development of System for Forensic Applications Using Spanish Words,” presented by Jose Benito Trángol, Universidad Nacional Autónoma de México, México City.

“Case Studies from MPRJ (District Attorney’s Office of Rio de Janeiro, Brazil): Voice Disguise and Automatic Speaker Recognition (ASR),” presented by Eline Portela, Maria Gargaglione, and Mônica Azzariti, CSI/Ministério Público do Estado, Rio de Janeiro, Brazil.

“Influence of Recording Distance on Voice Quality for Use in Speaker Recognition,” by Brian Prendergast, National Center for Media Forensics, Denver.

“Synthetic Voice Forgery and Voice Comparison,” by Guillaume Galou, Gendarmerie Nationale Criminal Research Institute, France.

The presenters provided many key insights into the voice comparison field, and the attendees enjoyed many interesting and informative conversations in the poster area and overflowing into the lobby corridor.

Paper Session 9: Speaker Recognition and Comparison II

Wrapping up the afternoon break and posters, Jeff Smith and the organizing committee invited everyone to bring their beverages and snacks back to the main conference room and reconvened the final paper session of the conference.

Anibal Ferreira, University of Porto, Portugal, explained his work analyzing speech recordings for tell-tale phase changes that may be useful in speaker identification. His presentation, “Speaker Identification Using Phonetic Segmentation and Normalized Relative Delays of Source Harmonics,” reported on the use of phonetic segmentation and phase extraction for speaker ID purposes, particularly using the vowel portions of the speech segments. Ferreira has found that including phase information can result in better classification performance.



Brian Prendergast of the National Center for Media Forensics expounds the ideas on his poster.

The final paper, “Securing Speaker Verification System Against Replay Attack,” was presented by Hafiz Malik of the University of Michigan-Dearborn. Malik explained that some types of speaker verification systems are used to grant access to restricted facilities or online data transactions, but a potential weakness of such systems is a replay attack in which a prior recording of the authorized talker is played back in an attempt to fool the verification system. Malik’s preliminary work has been focused on modeling microphone nonlinearities and higher-order statistics to detect differences between a live-spoken phrase and a phrase played back from a prior recording.

AES AUDIO FORENSICS: SETTING THE STANDARD

The AES 46th Conference continued the tradition from the three prior AES forensics conferences with superior audio forensics papers and workshops. AES reaffirmed its place as the leading professional group in the field of forensic audio analysis and interpretation. Jeff Smith, conference chair, and Roger Furness, AES deputy director, expressed thanks and praise to the conference committee, and encouraged all participants to be active AES members and to continue presenting their work at future AES events.

It is a great time to be involved in the emerging field of audio forensics. The 46th Conference attendees headed home from Denver invigorated with new and innovative ideas and in ardent anticipation of the next AES Conference on Forensic Audio.

Editor’s note: You can purchase a copy of the conference proceedings at www.aes.org/publications/conferences. Copies of individual papers are available at www.aes.org/e-lib.