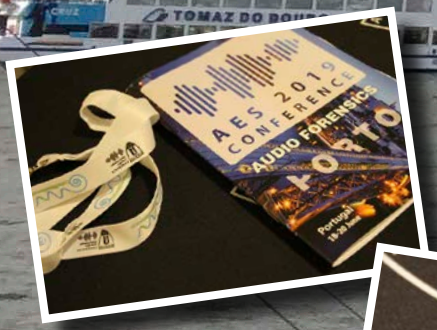
International Conference

Audio Forensics
Trust Powered by Science

Porto, Portugal
18–20 June, 2019

CONFERENCE REPORT



Conference cochairs Aníbal Ferreira, speaking, and Eddy Brixen.

For the seventh time, the Audio Engineering Society's international research experts in audio forensic analysis gathered for a fascinating series of papers, workshops, tutorials, and discussion. The 2019 AES International Conference on Audio Forensics, *Trust Powered by Science*, attracted 70 participants from 14 countries around the world to gather in Vila Nova de Gaia, Portugal, just south across the Douro River from the historic city of Porto. The first AES audio forensics conference was held in Denver, USA, in 2005, so this year marked 14 years of growing AES leadership in audio forensics research and education. In addition to the prior conferences in Denver (2005, 2008, and 2012), AES audio forensics conferences have been held in Hillerød, Denmark (2010), London, U.K. (2014), and Arlington, Virginia, USA (2017).

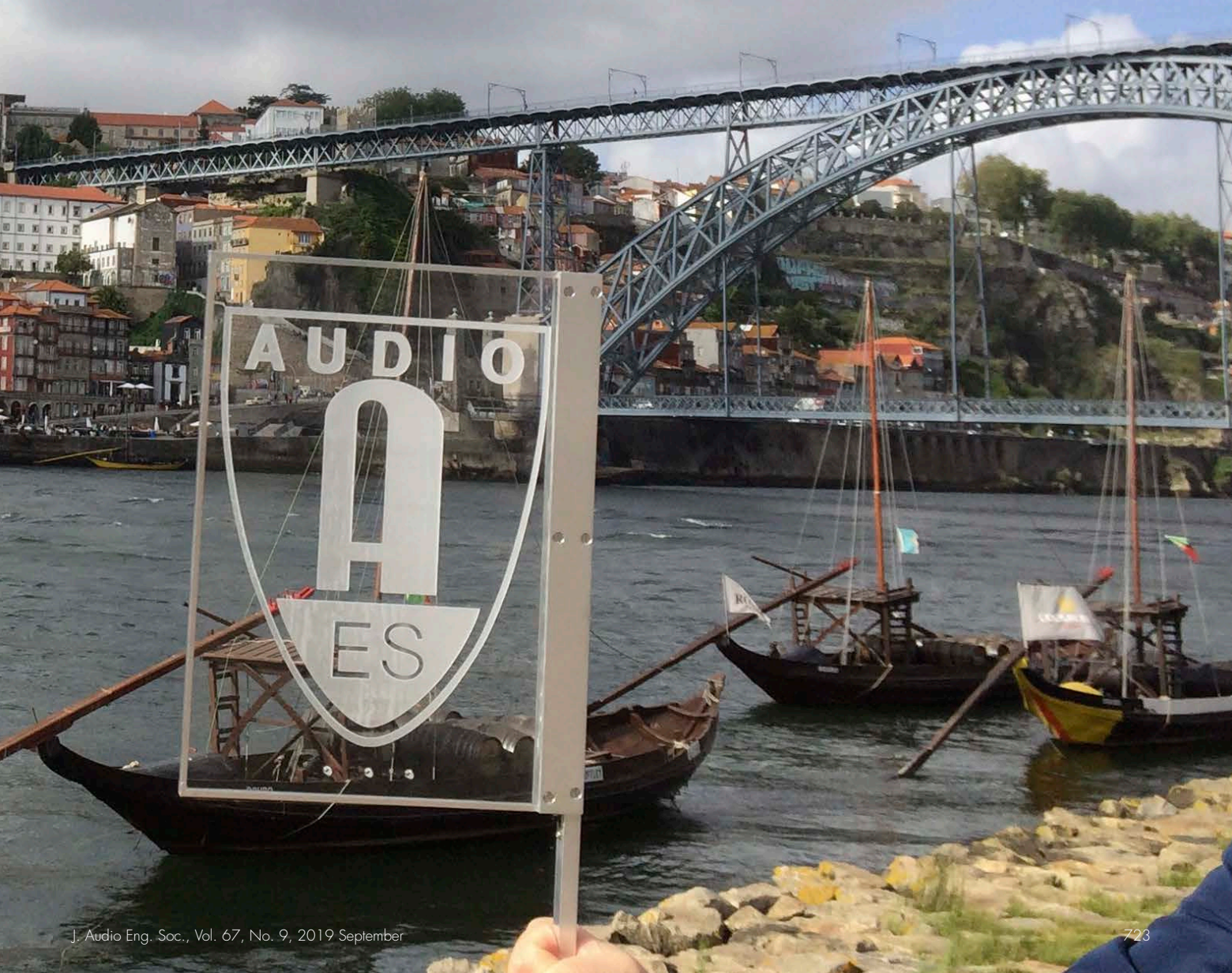
The AES international conference delegates included forensic examiners from crime labs, software developers, university researchers, educators, experienced private consultants, students, and many individuals interested in finding out more about the audio forensics field.

Conference cochairs Aníbal Ferreira and Eddy Brixen selected a fine variety of tutorial and workshop presenters. The meeting took place at the Holiday Inn Porto-Gaia, located in the Vila Nova de Gaia community about 1.5 km south of Porto's picturesque riverfront area. The hotel was about 45 minutes ride from the airport using the convenient Porto Metro light rail system.

Douglas Lacey and Durand Begault served as cochairs for the paper sessions, while Catalin Grigoras and Diamantino Freitas organized the workshop sessions. Jeff Smith performed important duties in marketing the conference and managing the website. Antonio de Oliveira chaired the sponsorship and demonstration aspects of the conference, and José M. N. Vieira was the conference treasurer. The conference received excellent support and volunteers from the University of Porto and from the Portuguese AES Section "Associação Portuguesa de Engenharia de Áudio" (APEA).

WELCOME TO PORTO

All AES international conferences include the valued feature of special opportunities for delegates to establish professional contacts, engage in face-to-face discussions, and learn about current trends in the audio research fields. Needless to say, the conference hotel and the easily accessible neighborhoods of the Porto and Gaia communities provided a great venue for the conference.



Porto, population 288,000, is the second largest city in Portugal, after Lisbon. The Porto metropolitan area is home to more than 2 million people. The central portion of the community is located on the banks of the Douro River, several kilometers inland from the Atlantic Ocean. The area was inhabited by pre-Celtic people as early as 300 BCE, making it among the most ancient established population centers in Europe. The settlement was called Portus Cale during the Roman Empire, and over the centuries the Portus Cale name is believed to have morphed into Portugal. Porto's sheltered access to the ocean made it a prime location for shipbuilding and commerce.

Since the 18th century the area has become perhaps best known for the production of *vinho do Porto*, the fortified wine known throughout the world as Port, that takes its name from the Porto region. Today, the Douro River separates the city of Porto on the north bank from the city of Gaia on the south bank, but six major bridges carry traffic and pedestrians between the cities. The Gaia riverfront area hosts numerous Port wine cellars, while on the north side Porto's Ribeira district features attractive shops and restaurants, all connected via narrow streets and delightful cobblestone walkways. It was also noted that J.K. Rowling, the author of the Harry Potter book series, lived in Porto in the early 1990s, and took some of her inspiration for details in her novels from the sights and sounds of the city, including Porto's ornate Livraria Lello Bookstore as the template for the wizard-supply shops in her fanciful Diagon Alley.

DAY 1—CONFERENCE OPENING

The conference opened promptly at 8:45AM on 18 June with welcoming remarks from Aníbal Ferreira, cochair of the conference and a professor at the nearby University of Porto. Additional greetings came from Diamantino Freitas, the current chair of the local AES Section (APEA), from Filomena Oliveira, the representative of the Porto Convention and Visitors Bureau, and from João Falcão e Cunha, the Dean of the College of Engineering of the University of Porto.

Opening welcomes from, top to bottom, Anníbal Ferreira, Eddy Brixen, Diamantino Freitas, João Falcão e Cunha, and Filomena Oliveira.

Aníbal Ferreira and cochair Eddy Brixen also thanked the sponsors and exhibitors, led by Platinum Sponsors CEDAR Audio Ltd. and Oxford Wave Research. Keith McElveen of Wave Sciences provided hands-on demonstrations of microphone array systems as part of the exhibition, too.

Karlheinz Brandenburg—opening keynote speaker

KEYNOTE LECTURE 1

The first scheduled presentation was an outstanding keynote address by Karlheinz Brandenburg of the Technical University Ilmenau, Germany, and the Fraunhofer Institute for Digital Media Technology. Brandenburg is well-known in the AES as one of the key contributors to the important field of perceptual audio coding. His keynote presentation, entitled "Lossy Compression in Digital Evidence: Easier or More Difficult?" introduced the basic principles of source coding for speech and music, and then focused on the emergence of perception-based lossy coding that exploits the strengths and weaknesses of human psychoacoustics to achieve greater compression performance. Examples of lossy compression include MPEG-1 Layer 3 ("MP3"), Dolby Digital, Microsoft Windows Media Audio ("WMA"), and MPEG-2 Advanced Audio Coding ("AAC"), among others.

Brandenburg explained the psychoacoustic principles underlying perceptual audio data compression, and emphasized the importance of both frequency and temporal masking in modern coding algorithms. Because the coding procedures are tied closely to estimates of human perception, he cautioned audio forensic examiners about relying upon precise waveform information when interpreting signals reconstructed from perceptually-coded data streams. Similarly, the quantization levels of the original uncompressed material will be absent from lossy coding output. Nevertheless, Brandenburg indicated that voice biometrics, scene classification, and probably electrical network frequency (ENF) analysis could be satisfactory with wideband perceptual audio coders as long as the original source material had the necessary low-frequency bandwidth.

Brandenburg concluded his remarks by mentioning recent work to uncover "coding footprints" that may be present in the encoded file, such as the use of hidden and unused ("don't care") bits in the standard bitstreams that may have been used to store hidden metadata. He also mentioned research into ways to estimate which perceptual audio coder might have been used (or not used) to create an uncompressed output file. With the ever-increasing use of perceptual coders in audio devices, Brandenburg's insights were greatly appreciated.

TECHNICAL PROGRAM—DAY 1

Paper Session 1: Gunshot Analysis 1

Following the intriguing keynote presentation, the first technical session of the conference included two papers. Doug Lacey, the session chair, introduced the first paper, "Overview of Forensic Gunshot Analysis Techniques," by Durand Begault, Steve Beck, and

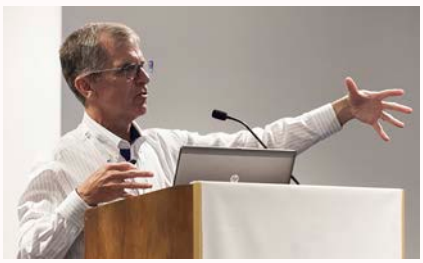
Rob Maher, which was presented by Durand Begault. Begault explained that the purpose of the paper was to collect and itemize the best-practices for analyzing gunshot recordings based on a review of techniques used in actual forensic cases presented in court. He noted that methods for analyzing the sounds of firearms date back to the detection of artillery in World War I, and that recent examinations generally involve a combination of critical listening, waveform analysis, spectral analysis, and amplitude envelope analysis. Begault also mentioned the challenges of working with an active investigation in which other evidence may be available, such as the location of bullet holes or the count of spent cartridges collected at the crime scene. Such associative/corroborative information can potentially influence the interpretation of audio forensic information in an improper manner.


Durand Begault explains gunshot analysis.

The second paper on the session was authored by Steve Beck. Beck described the use of cross correlation in the paper "A Short-Time Cross Correlation with Application to Forensic Gunshot Analysis." He has found cross correlation to be useful in gunshot cases, generally for finding the best time lag when determining relative time delays. Although cross correlation is sometimes proposed for gunshot similarity determination, Beck has found that the procedure is difficult to apply confidently in many cross-range, cross-angle, and cross-firearm comparisons due to the numerous sources of signal variation that may be present.

Paper Session 2: Gunshot Analysis 2

Following a morning coffee break featuring cookies and other treats, the next session began with a paper presented by Rob Maher of Montana State University. In his paper entitled "Shot-to-Shot Variation in Gunshot Acoustics Experiments," Maher showed examples from a controlled experiment involving successive shots from the same firearm. The experimental question was to determine if there were measurable shot-to-shot variations in the recordings for the successive "identical" shots, and if so, to assess the possible explanations for the differences. Maher showed examples of ten shots from a Colt 1911 pistol, a Glock 19 pistol, a Ruger SP101 revolver, a .308 Winchester rifle, and a Stag AR-15 rifle, using commercial off-the-shelf ammunition for each gun. The results showed clearly measurable differences among the successive shots, leading to the conclusion that manufacturing variability in parameters such as bullet size, cartridge crimping, propellant load, propellant consistency, etc., can have a discernible effect upon the gunshot acoustical recordings. Maher summed up his recommendations for audio forensic examiners by advising great care before assuming that the recorded sound of a particular firearm is precisely repeatable in an investigation involving multiple gunshot sounds.


Rob Maher details shot-to-shot variation.

Next, Steve Beck returned to the podium for his paper "Who Fired When: Associating Multiple Audio Events from Uncalibrated Receivers." Beck described a systematic procedure for analyzing multiple audio recordings of the same gunshot incident from spatially distributed positions. The Time Difference of Arrival (TDOA) method of estimating source position works well if the receiver positions are known, the source and receivers are not moving, and the receivers are not in a degenerate relative position, such as all on a radial line from the source. Beck presented an example using unsynchronized receivers and two separate simulated gunshots, demonstrating that the timing between shots from the simulated source positions is maintained at the receivers, even though the absolute time difference will generally vary due to the different acoustical path lengths. Beck showed how the relative time differences can help identify which sound event (gunshot) went with which source. He also showed an example using the multiple recordings made of the assassination attempt of U.S. President Ronald Reagan on March 30, 1981, as the president left a speaking engagement at the Washington Hilton Hotel.

The final paper before lunch, "Improved Gunshot Classification by Using Artificial Data," by authors Christian Busse, Thomas Krause, Jörn Ostermann, and Jörg Bitzer, of Oldenburg and Hannover, Germany, described their work developing automated systems for gunshot classification based on audio recordings. Jörg Bitzer presented the paper. He explained that one of the difficulties in training a machine-learning algorithm occurs when the signals of interest are rare in the available database. Most learning algorithms work best if there is a very large number of training examples available, and unusual sounds such as gunshots are often rare in the available training sets. The approach used by the team was to generate a set of artificial shot sounds based on a simple model of a gunshot and augment the classification training set with the synthesized sounds. The results were sufficiently encouraging to continue further research using this strategy. A key requirement will be to determine the required variability in the training set to account for the acoustical differences of the recording, such as reflections, reverberation, firearm orientation, clipping, and so forth.

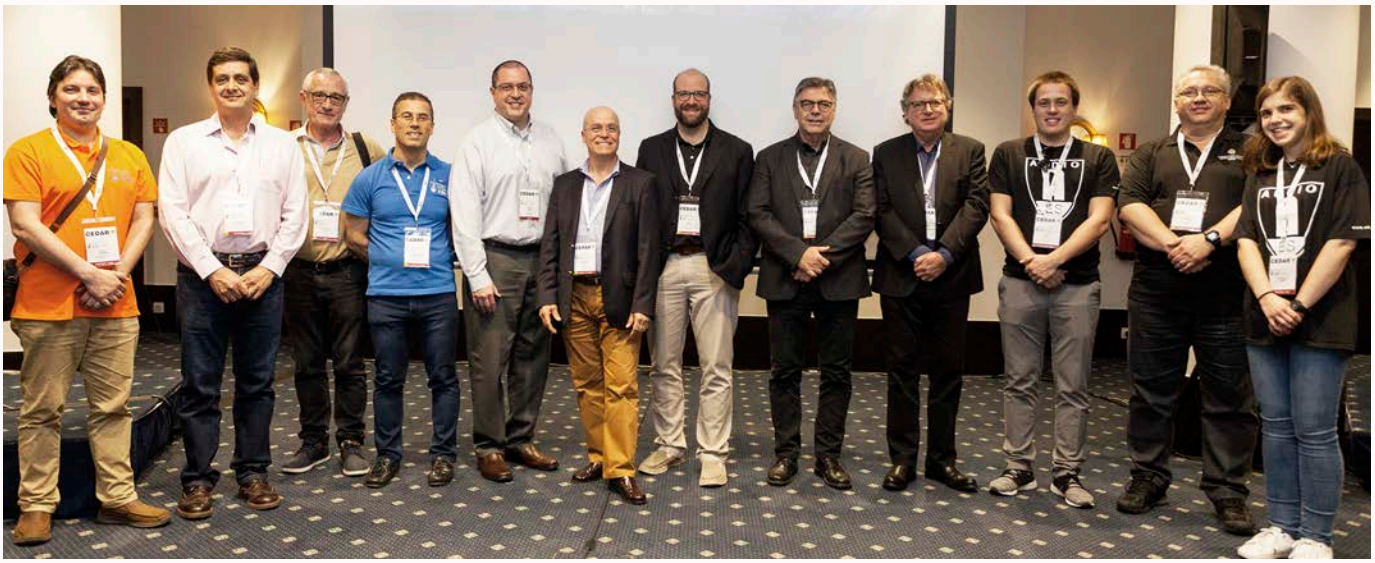
Poster Paper

During the conference, the organizers scheduled a poster paper presentation during the lunchtime exposition. One poster was presented by Jeff Smith, Catalin Grigoras, and Cole Whitecotton of the National Center for Media Forensics, University of Colorado Denver. The poster, entitled "An Updated Triage Approach for Voice Memos Recordings, iOS 12.x," gave new results to supplement a paper from the 2017 AES Audio Forensics Conference held in Arlington, VA. The work provided an empirical decision tree that an examiner could use when assessing the integrity of a Voice Memos recording in Apple iPhone brand mobile phones.


Jeff Smith

Tutorial 1: Forensic Audio Authentication

Following a pleasant and relaxing break for lunch, Catalin Grigoras presented a tutorial providing a summary review of the latest developments in examinations of digital audio authen-



The conference committee meets for a group photograph: from left, Marco Oliveira, José Vieira, António Oliveira, Aníbal Ferreira, Douglas Lacey, Diamantino Freitas, Jeff Smith, Eddy Brixen, Durand Begault, João Silva, Catalin Grigoras, and Francisca Brito.

ticity. Grigoras explained that determining authenticity can be difficult in the digital age, because many high-quality editing packages are now available. His approach has been to study carefully the metadata accompanying a recording, since the details of the metadata may reveal a file that has been opened and edited with a software package that leaves behind telltale indications. Most audio recorders and processors have distinctive metadata fields. However, it is always possible that a particularly clever adversary could edit the binary metadata to conceal changes to the file. Grigoras also noted that some types of cut-and-paste edits use interpolation and cross-fades that use higher numerical precision than the native format of the original recording, and this may leave detectable artifacts. Similarly, some audio editor packages may up-sample the original recording, revealed by the presence of out-of-band spectral artifacts.

The topic of this tutorial was of great interest due to recent reports of "deep fakes" and "cloned voices" appearing in the news. The delegates engaged in a good discussion about authenticity in the context of perceptually-coded material, and Karlheinz Brandenburg and others weighed in on the possibility of artifacts due to tandem coding (i.e., compressed material that is decoded, edited, and then re-encoded).

Paper Session 3: Authentication 1

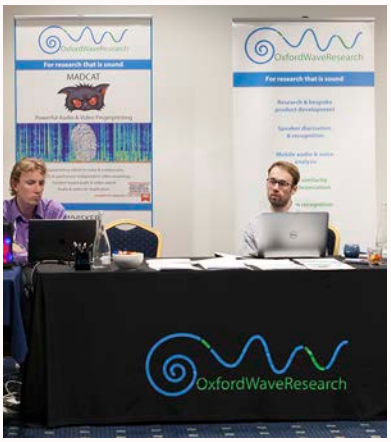
The Tuesday afternoon technical paper sessions began with two presentations dealing with forensic authentication topics. The first paper, "Inverse Decoding of PCM A-law and μ-law," by Luca Cuccovillo and Patrick Aichroth of the Fraunhofer Institute for Digital Media Technology, dealt with the examination of audio files for traces of A-law or μ-law companding. Luca Cuccovillo presented the paper. A-law and μ-law PCM encoding was developed originally for the public switched telephone network ("landline" phones) to convert a 13-bit (A-law) or a 14-bit (μ-law) linear PCM sample into a logarithmic 8-bit sample, thereby reducing the number of bits required for storage or transmission while maintaining a satisfactory signal-to-quantization-noise ratio. Because these methods of companding map multiple input sample values to a single code value, the statistical distribution of output sample values will be non-uniform, and this can be detected in the output data stream. Cuccovillo described an experiment with an audio file originally coded with A-law or μ-law and then subsequently tampered with by inserting uncoded audio samples from another source. He stated that the detection of signal segments without traces of companding can be useful in evaluating the consistency and trustworthiness of an audio file reportedly containing A-law or μ-law material.

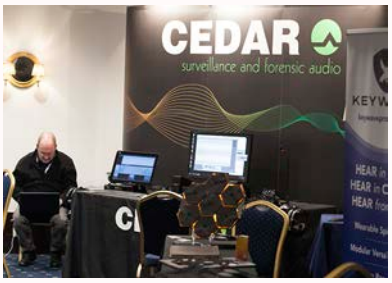
The second paper in the session was also presented by Luca Cuccovillo. The paper entitled "Copy-Move Forgery Detection and Localization Via Partial Audio Matching," by Milica Maksimovic, Luca Cuccovillo, and Patrick Aichroth, of the Fraunhofer Institute for Digital Media Technology, described experiments to detect segmental sonic "fingerprints" in an audio recording and seek duplicates of the recorded material that would suggest that a forger had deliberately copied and re-inserted duplicate audio material elsewhere in the file. For example, a forger might try to take a recorded utterance of "yes" at one point in the file and copy-paste it over an utterance of "no" somewhere else in the file, thereby deliberately changing the meaning and context while still using the talker's authentic samples. Cuccovillo explained the research team's approach, which involved adapting an audio "fingerprint" method with feature vectors representing segments of 100–150 ms. He concluded that there is more work to be done, but the initial results are useful and encouraging.

Paper Session 4: Authentication 2

An afternoon break provided the delegates the opportunity to review the technical exhibition while enjoying some snacks and beverages. Next, the authentication papers continued with three papers considering audio material recorded by the Voice Memos software on Apple brand smartphones.

The first paper, "Forensic Authenticity Analyses of the Metadata in Re-encoded M4A iPhone iOS 12.1.2 Voice Memos Files," by Bruce E. Koenig and Douglas S. Lacey of BEK TEK LLC, continued a line of work the authors have described in previous papers. Doug Lacey explained some of the new features present in the Apple iOS 12 release of the Voice Memos application, and reported on the research involving downloading audio files from the phone, opening the files with audio editor software on a personal computer, then simply saving the files unaltered using the "Save As…" feature of the software. When Koenig and Lacey compare the original files to





Platinum sponsors Oxford Wave Research (left) and CEDAR (above).

the re-encoded files, they note obvious metadata differences in all cases. Thus, Lacey recommended the importance of understanding the native metadata structure when attempting to assess the likelihood that an iPhone audio file purported to be original might actually have been edited or otherwise manipulated.

Next, James Zjalic of Verden Forensics, Birmingham, England, presented his paper "Detection of In-App Apple Voice Memos Edits through Extrapolation Artifacts." Zjalic explained his investigation of the postrecording features of the Voice Memos app, such as Replace, Edit, and Trim. He identified several artifacts of the in-app edits, including what appears to be out-of-band energy in the waveform, possibly indicating up-sampling during the cross-fade at an edit point. For future work, Zjalic suggests additional examination of the artifacts to help rule in or rule out the presence of an edit in the file.

For the final paper in the afternoon session, Catalin Grigoras of the National Center for Media Forensics, University of Colorado Denver presented his paper, coauthored with Jeff M. Smith, entitled "Forensic Analysis of AAC Encoding on Apple iPhone Voice Memos Recordings." The research involved applying various methods to detect decoded WAV file traces that might indicate that the original recording was stored with lossy compression. The results indicated that some particular combinations of iPhone operating system and Voice Memos versions resulted in distinctive features that could be detected, but there was no clear indication why the otherwise similar system versions would have different audio features.

Workshop 1: Forensic Audio

The Tuesday afternoon sessions of the conference concluded with a fascinating workshop presentation by Chris Smith of the London Metropolitan Police Service audio forensics lab. Smith explained the intriguing scope and variety of work in one of the largest police organizations in the world: more than 30,000 sworn officers and 10,000 civilian support staff serving the nearly 9 million people living and working in greater London. The audio forensics lab comprises six specialists who perform casework, court presentations, special projects, and some related research.

The lab's work comes from police investigations, internal affairs inquiries, and some work in support of smaller police forces elsewhere in the United Kingdom. Smith explained that cases involving sexual assault, physical assault, and murder are the most common. Although digital forensics is increasingly important, the Metropolitan Police audio lab still must handle a wide range of audio source formats, still including some analog audio cassettes and VHS videotapes. The lab sees a growing number of recordings from body-worn recorders, mobile phones, and home surveillance systems.

The lab's workload includes format conversion, editing to conceal/protect privacy, deliberate voice disguise for courtroom use, and general needs for audio enhancement and authenticity determination. The lab generally refers requests for speaker/talker comparisons to outside experts.

Although the audio forensics lab works within the Metropolitan Police system, it is the duty of the lab to serve the court, not the police. In 2008, the United Kingdom established an independent office of the Forensic Science Regulator to ensure "that the provision of forensic science services across the criminal justice system is subject to an appropriate regime of scientific quality standards." Among other recommendations, the Forensic Science Regulator has mandated that forensic laboratories such as the Metropolitan Police audio lab seek International Organization for Standardization / International Electrotechnical Commission (ISO/IEC) 17025 accreditation, covering general requirements for the competence of testing and calibration laboratories. The ISO/IEC 17025 certification process is complicated and time-consuming, and Smith reported that no U.K. forensics lab is currently accredited. Nevertheless, he expects the Metropolitan Police audio lab to complete the accreditation requirements within the next year.

As the first day technical sessions of the conference came to an end, the delegates were invited to attend a delightful and relaxing cocktail party in the spacious lobby of the hotel. Conversations ranged from highly technical questions about the day's presentations, to recommendations for evening dining and entertainment in the nearby neighborhoods of Vila Nova de Gaia.

DAY 2—KEYNOTE LECTURE 2

Wednesday morning dawned with a fine breakfast at the hotel preceding a special keynote lecture entitled "Source-Filter Processing for Audio Forensics," by Prof. Udo Zölzer of Helmut Schmidt University in Hamburg, Germany. Zölzer explained his extensive prior work in audio signal processing using filterbank techniques, both in the low-latency time domain and in the higher-efficiency block-based frequency domain. He showed very helpful examples of time-frequency processing to separate the spectral envelope from the source excitation signal, and the corresponding applications of these principles to de-noising and signal quality enhancement. Among his recommendations was the use of spatialization to avoid listener fatigue during speech transcription: getting the sound "outside the head" can help provide a less artificial aural impression.

TECHNICAL PROGRAM—DAY 2

Paper Session 5: Speaker Recognition/Analysis 1

The Day 2 technical sessions began with two papers on forensic speaker recognition. The first paper, "Premature Overspecialization in Emotion Recognition Systems," was presented by Gustavo Assunção on behalf of his co-authors Fernando Perdigão and Paulo Menezes, of the University of Coimbra, Portugal. The University of Coimbra was established during the 13th century and is one of the oldest in Europe. Coimbra is located approximately 130 km south of Porto, on the banks of the Mondego River. Assunção described their research to determine the likely emotional state of a talker based upon a recording of the talker's speech. Speech emotion recognition (SER) systems that postpone the decision on emotional state, thereby allowing time to adapt to the particular characteristics of a given talker, can show better performance. The research group's results using a speaker-adapted method applied to a database of emo-

tional speech utterances have been good. Assunção indicated that the group's future work will explore ways to include more context information to gain additional accuracy in emotion estimates.

The second paper was by Hafiz Malik and Raghavendar Changalvala of the University of Michigan-Dearborn. In "Fighting AI with AI: Fake Speech Detection Using Deep Learning," Malik explained the growing concern about so-called "cloned speech" and "deep fakes," which refer to computational methods able to modify or synthesize recorded speech in such a way that it is virtually identical perceptually to a real speech utterance by a particular individual. Cloned speech impersonating the voice of a real person could potentially defeat a voice access system or even frame an innocent person with a threatening or embarrassing fake recording. Malik described some preliminary experiments that have worked well to distinguish between cloned speech and bona fide speech using a particular database. The conference delegates posed many questions about the strengths and weaknesses of various approaches to assess speech authenticity.

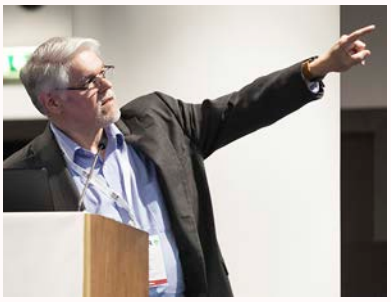
Platinum Sponsors: Special Talks

The morning coffee break lead into special presentations by individuals from the conference's Platinum Sponsors. The first presentation was by Gordon Reid, Managing Director of CEDAR Audio Ltd. Reid has been a longtime supporter of AES Audio Forensics sessions and conferences. His presentation, entitled "Narrowing the Gap Between Live Surveillance and the Forensic Laboratory," described a newly-developed "satellite" recording device intended for use by field agents who are not forensic audio signal processing experts. The device includes a set of simple controls, some basic real-time enhancement features, and additional access controls providing encryption and secure download support to a server system.

The second special presentation was "The Who, the When and the What—Challenges in the Development of Real-World Solutions for Forensic Audio Processing," by Anil Alexander, CEO and co-founder of Oxford Wave Research. Alexander explained that Oxford Wave Research develops products and solutions based on law enforcement requirements and special requests. He often finds that simple and effective techniques are very important in practice, and the most effective techniques are not always the most advanced or novel. One example is Oxford Wave Research's software product VOCALISE, an acronym for Voice Comparison and Analysis of the Likelihood of Speech Evidence. The software provides automatic speaker recognition tools that allow comparison of recorded speech from an unknown talker to a database of speech from known talkers using both traditional phonetic parameters and automatic detection of spectral features. Another useful feature is diarization: automatic separation of speech intervals by separate talkers in a conversation.

Tutorial 2: Automatic Speaker Recognition

Following the lunch break, the delegates reconvened in the presentation hall for a tutorial on forensic automatic speaker recognition by Eliud Bonilla of the Johns Hopkins University Applied Physics Laboratory. Bonilla provided a fine introduction to speaker recognition principles and concepts from the point of view of a practitioner. He showed several examples of speech samples compared to a training database, and emphasized the importance of having features that can lead to a strong hypothesis test regarding the talker's identity. Among the challenges cited were the known factors that cause dynamic changes in an individual's speech, such as mood, illness, age, fatigue, intoxication, and other deliberate and nondeliberate changes that can vary over time.



Eliud Bonilla introduces speaker recognition.

Paper Session 6: Speaker Recognition/Analysis 2

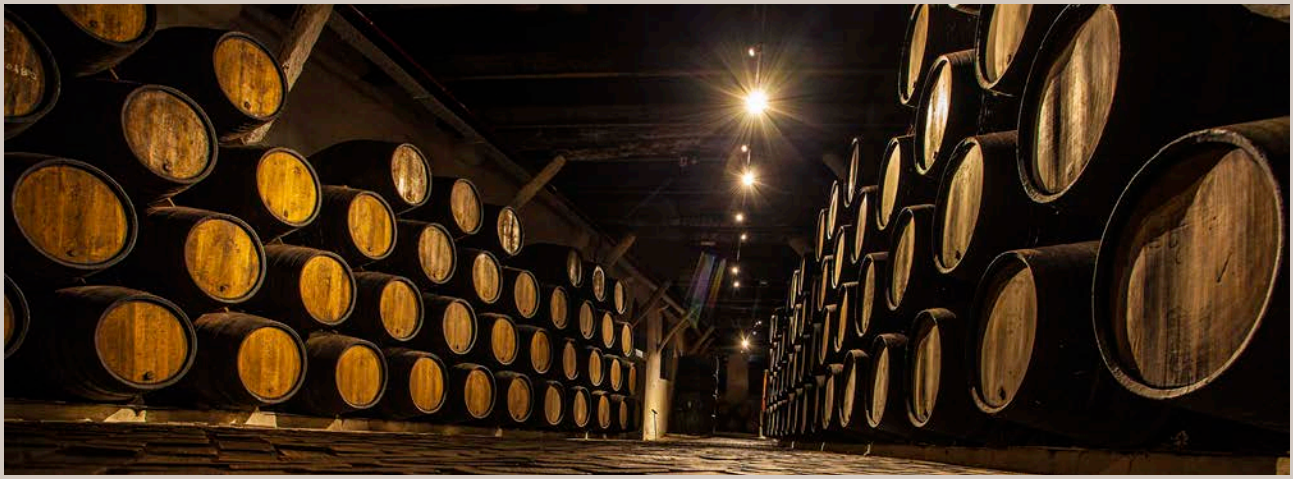
Next on the afternoon's agenda were technical papers focused upon speaker recognition and analysis. First, Finnian Kelly of Oxford Wave Research presented the paper "Deep Neural Network Based Forensic Automatic Speaker Recognition in VOCALISE Using x-vectors," which he co-authored with Oscar Forth, Samuel Kent, and Anil Alexander of Oxford Wave Research, and Linda Gerlach of Philipps-Universität Marburg, Germany. Kelly presented the background principles of so-called deep neural networks, which take a complex set of features and produce a low-dimension representation known as an x-vector. The x-vector method allows a larger number of training samples, including modifications such as the addition of reverberation to some of the training samples. He explained that the performance of the VOCALISE software has improved with the use of the x-vector framework.

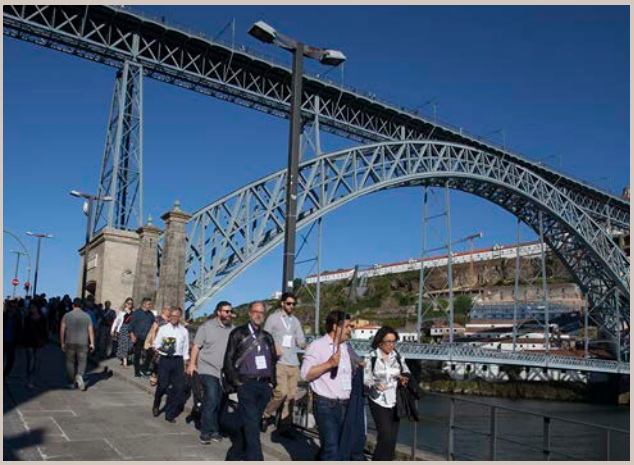
The second paper in the session was "Phonetic-Oriented Identification of Twin Speakers Using 4-second Vowel Sounds and a Combination of a Shift-Invariant Phase Feature (NRD), MFCCs and F0 information," by conference co-chair Aníbal J. Ferreira of the University of Porto. Ferreira described a challenging project to compare and contrast the speech of twins, triplets, and close relatives: individuals who would likely have nearly identical physical vocal tract dimensions. The research focuses on better feature selection for the speech models. His work involves the use of traditional mel-frequency cepstral coefficients (MFCCs) and fundamental frequency analysis (f_0), as well as the Normalized Relative Delay (NRD), which is based on prior work to understand the phase information in audio coding and compression. The work has shown some favorable results, even with relatively short speech segments of only several seconds in duration.

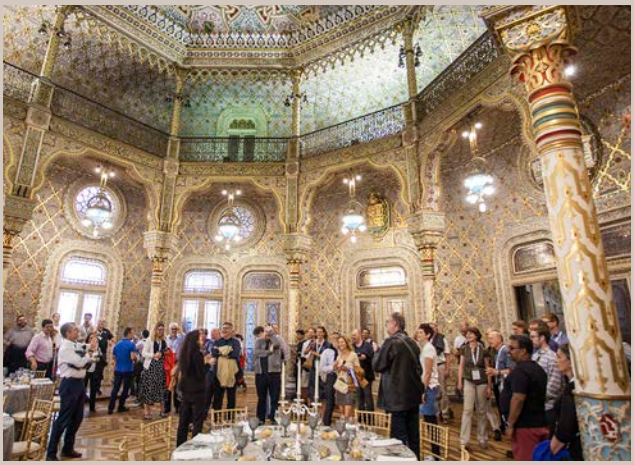
Workshop 2: Gunshot Analysis

Following an afternoon break, the technical activities of Day 2 concluded with a workshop on gunshot analysis. Durand Begault led off the workshop with several observations concerning the application of urban gunshot detection systems in audio forensics cases. The scientific and mathematical principles of multilateration are well known, but Begault shared his concern that uninitiated individuals might be "wowed" by the use of advanced technology rather than applying a critical eye to the important differences between theoretical principles and the three-dimensional, non-line-of-sight conditions actually found in an urban scene. He advised caution in applying findings from urban gunshot detection that lack a strong statistical basis for the "hit" rate and "miss" rate of the detection and localization algorithms.

The second workshop presenter, Steve Beck, discussed the complexity of dealing with human earwitnesses. He explained that while humans are actually quite good at recognizing sounds in general, humans are quite unreliable at reporting specific details of what they heard. Different listeners are likely to give different reports about the timing, loudness, and sequencing of audible sound events. The audio forensic examiner has the ability to use physical principles to estimate sound propagation,







Venues for the evening social event: clockwise from the top, the Sandeman Port wine cellar, the Bolsa Palace, and the Dom Luís I Bridge.

spreading loss, refraction, attenuation, etc., but the forensic question can often involve the likelihood that a particular sound would be detectable by an average listener under some specified geometrical distances and meteorological conditions. This sort of investigation will have to include models of psychoacoustics and human perception.

Next, Rob Maher presented several practical examples of gunshot sounds recorded by multiple receivers, emphasizing the significant differences between the audio signals. He explained that multiple recordings can offer the forensic examiner options for multilateration, timing analysis, and firearm characterization, but even with multiple simultaneous recordings obtained concurrently, the different recording positions, different device settings, and unsynchronized start/stop times may provide unexpectedly different information. Maher cautioned developers working to create automatic firearm recognition systems that attempting to "train" a classification system using only a small number of gunshot recordings, or recordings obtained under a limited range of circumstances, is not advised.

EVENING SOCIAL EVENT

Upon the conclusion of the successful second day of the conference, many attendees joined a special opportunity for a walking tour and dinner in the riverfront area of Porto. The delegates boarded a bus at the hotel for a short ride to the south bank of the Douro River, and being guided by a custom-built and cleverly lighted "AES" sign carried by the tour leader, walked along the promenade to the historic Port wine aging cellars of Sandeman in Vila Nova de Gaia. Established in 1790 by Londoner George Sandeman, the vast cave containing hundreds of barrels and casks of Port wine provided a dark, cool, and redolent stroll for the group, culminating in a delightful visit to the wine tasting rooms to sample the famous beverage.

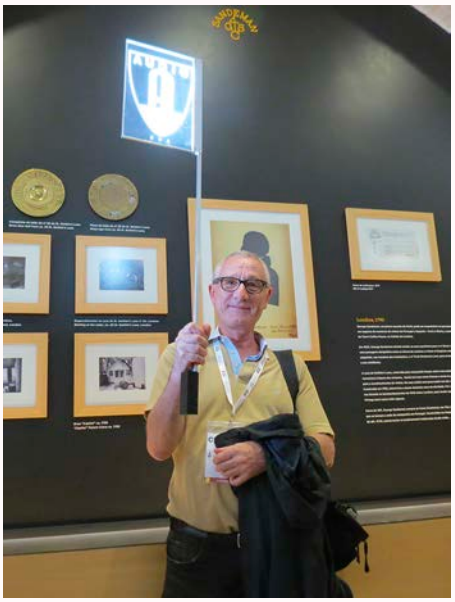
Following the Sandeman tour and refreshments, the group walked along the river to the base of the distinctive and elegant Dom Luís I Bridge that spans the deep Douro River valley. Completed in 1886, the arch-style bridge features a lower deck at the level of the river to accommodate pedestrians and cars, while an upper deck soars 85 meters overhead at the top of the arch, carrying the light rail line and pedestrian traffic. After following the leader's lighted AES sign across the bridge, the AES group proceeded through the lively Ribeira Square district on the Porto side of the river, and then up the narrow streets to the Palácio da Bolsa, or Stock Exchange Palace. The Bolsa Palace was completed in 1850, with interior decorations continuing into the 20th century. Filled with beautiful paintings, frescoes, and other unique furnishings, the site is now both an active business venue and a tourist attraction. After a guided tour of the building, the delegates were invited into

the Palace's remarkable restaurant O Comercial for a lovely and relaxing multicourse Portuguese dinner.

TECHNICAL PROGRAM—DAY 3

Tutorial 3: Deep Learning for Audio Synthesis, Separation, and Enhancement

The final day of the conference began with many delegates discussing the enjoyable and memorable social time on Wednesday evening. The morning's scheduled tutorial presentation was adjusted to accommodate the unavailability of the originally scheduled speaker. Fortunately, an excellent substitute presenter was available via a Skype hookup. Shahan Nercessian, senior research engineer with audio signal processing company iZotope, Inc., of Cambridge, Massachusetts, spoke about recent developments in machine learning applications for audio. Nercessian began by providing a summary of the terminology and history of machine learning, neural networks, and deep learning. He then focused on applications involving convolutional neural networks and recurrent neural networks as a means to classify complicated input sources, such as audio signals. He concluded his presentation with contemporary applications in which a neural network is used to synthesize new signals, such as performing text-to-speech synthesis with the "voice" speaking in a way to mimic the speech of a desired talker.



António Oliveira, committee person responsible for sponsorship and demonstrations, with his illuminated AES sign.

Paper Session 7: Scene/Microphone Classification & Steganography 1

Following the tutorial, the final paper sessions of the conference focused on telltale signal features that identify a particular recording device or recording space. The first paper involved research to identify a recording location using pre-trained convolutional neural networks (CNNs) and a post-trained deep neural networks (DNNs). "Bag-of-Features Models Based on C-DNN Network for Acoustic Scene Classification," was presented by Lam Pham on behalf of his coauthors Ian McLoughlin, Huy Phan, and Ramaswamy Palaniappan of the University of Kent, UK, and Yue Lang of Huawei Technologies Co. Ltd., Shenzhen, China. The research system has been applied to several acoustic scene classification tasks, and has showed good performance relative to other systems. The conference audience had many questions about the implementation choices and details.

Next, Lam Pham presented a second paper, "Beyond Equal-Length Snippets: How Long is Sufficient to Recognize an Audio Scene," on behalf of his co-authors Huy Phan, Oliver Y. Chén, and Maarten De Vos of the University of Oxford, Ian McLoughlin of the University of Kent, and Philipp Koch and Alfred Mertins of the University of Lübeck, Germany. This paper described research to apply neural network systems to acoustic scene classification with "fusion" of neural network outputs. The researchers found that model fusion was advantageous if the test signal lengths were shorter than 20 seconds. With longer signal segments, there was a smaller performance gain from the fusion procedure. The audience members asked questions about the test conditions, and how the accuracy of the results was determined.

Paper Session 8: Scene/Microphone Classification & Steganography 2

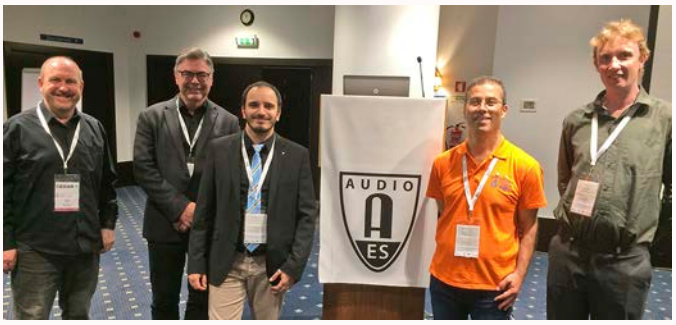
For the final session, Khalid Mahmood Malik of Oakland University, Michigan, USA presented the paper "Exploiting Frequency Response for the Identification of Microphone Using Artificial Neural Networks," co-authored by Azeem Hafeez and Hafiz Malik of the University of Michigan-Dearborn. The research employed a database of microphone frequency-response measurements. The researchers recorded 50-second sinusoidal tones at 80 frequencies between 100 Hz and 8 kHz, and then attempted to match the observed frequency content of actual recordings to the microphone characteristics. The promising results were questioned by several audience members due to the non-anechoic response measurement technique, and the difficulty separating the influence of the microphone from all of the other acoustical frequency response factors of the recording room and spoken source.

Conference cochair Eddy Brixen of EBB-consult, Smorum, Denmark, presented his paper "Directivity and Sensitivity of Cell Phones: iPhone 7," describing his experiments to understand the directional characteristics of the microphones in a handheld mobile phone that are active during a telephone call. He tested an Apple brand iPhone 7 model for its directional pickup behavior when recording in the Voice Memos application, as well as when performing a voice telephone call. Brixen found that the phone's automatic gain control and noise reduction features had a big impact upon the recorded signal when in telephone mode. The phone suppressed background sounds below approximately 70 dB(A) when no foreground sounds were present, and this raises a concern about attempts to interpret background sounds from mobile phone recordings. In summary, he found that the phone worked well for its designed purpose of conveying intelligible foreground speech transmission and suppressing competing background "noise."

Concluding the conference paper sessions, author João Moutinho presented "Steganography for Indoor Location," coauthored by Diamantino Freitas, and Rui E. Araújo, of the University of Porto. The research dealt with techniques for indoor location that would work in areas where global positioning system (GPS) radio signals cannot be detected, such as inside private buildings, transportation stations, and museums. The concept was to insert inaudible location data into the audio material presented through existing public address loudspeakers, and then to have a special smartphone application able to detect and decode the acoustic information. The steganography (data-hiding) can be achieved using spread spectrum (minimally audible broadband noise), echo hiding (data concealed by reverb tails), or some other method. The results of various live tests have been good, and the authors are looking forward to wider deployment of the indoor location system in Portugal.

CONFERENCE BEST PAPER AWARD

As the conference came to a close, the organizing committee and the Platinum Sponsors selected a winning paper based upon their review of the conference proceedings. The best paper recipient was "Inverse Decoding of PCM A-law and μ-law," by Luca Cuccovillo and Patrick Aichroth of the Fraunhofer Institute for Digital Media Technology. Luca Cuccovillo and Patrick Aichroth had previously

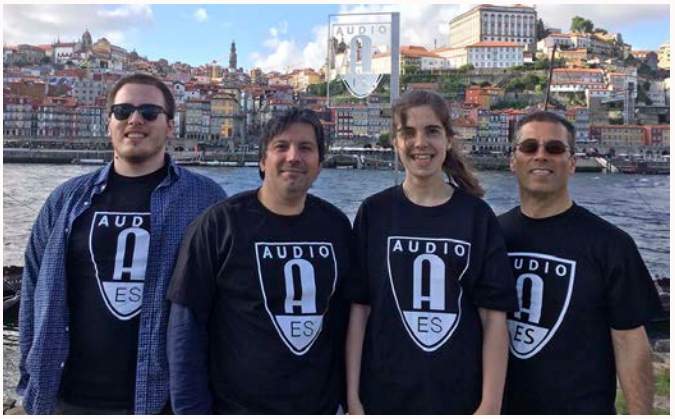


Best paper award presentation: from left, Gordon Reid, Eddy Brixen, author Luca Cuccovillo, Aníbal Ferreira, and author Oscar Forth.

received the best paper award at the 2017 AES Audio Forensics Conference, so the authors extended their fine record of providing key technical contributions to the field of audio forensics.

AES AUDIO FORENSICS: TRUST POWERED BY SCIENCE

The AES 2019 Audio Forensics Conference cochairs Aníbal Ferreira and Eddy Brixen closed the successful conference with thanks to the organizing committee and to the local volunteers who worked to ensure the comfort and focus of the delegates. "Obrigado, adeus e viagens seguras," goodbye and safe travels, from Porto, Portugal. The meeting continued the strong tradition established by the six prior AES forensics conferences, with the participants eager to hear of plans for future AES events in the field of audio forensics.



Student volunteers, from left, João Silva, Marco Oliveira and Francisca Brito, with Aníbal Ferreira, conference cochair.

Editor's note: AES Members can access the conference papers free of charge via the AES E-library at http://www.aes.org/e-lib/ or http://www.aes.org/publications/conferences/?confNum=ID-192

Delegates and their entourage gather for a group photograph at the Sandeman wine cellars.