Cole Whitecotton

Deepfakes and media forensics — **NIGEL JOPSON** discovers CSI: Audio

> The detection of fakes is going to get more and more difficult

A report from Deeptrace, the Netherlands-based cyber security group, identified 7,964 deepfake online videos at the start of 2019. After nine months, the figure nearly doubled to 14,678, and has been growing since. In March 2019 it emerged criminals had used artificial intelligence-based software to impersonate a chief executive's voice, ordering a fraudulent transfer of €220,000. The CEO of a UK-based energy firm thought he was speaking on the phone with the boss of his firm's German parent company, who asked him to send the funds to a Hungarian supplier (according to the company's insurance firm, Euler Hermes Group SA). In April 2020, Extinction Rebellion (XR) activists released a deepfake video of the Belgian Prime Minister Shophie Wilmès making a speech linking Covid-19 to the climate crisis.

"There is a broad attack surface here — not just military and political but also insurance, law enforcement and commerce," says Matt Turek, programme manager for MediFor, a media forensics research program led by the Defense Advanced Research Projects Agency (DARPA), part of the US defence department. DARPA is also working on a program named SemaFor, whose point is to identify semantic inadequacies in deepfakes.

While other areas of professional audio are in a state of flux, one field is experiencing rapid growth with the production of more and more source material every day — whether 'deepfake', surveillance, reportage or law enforcement. This is audio forensics, the use of advanced audio analysis and filtering to try to extract voices and meaning from badly-captured and noisy recordings, as well as methods of validating and presenting the resulting evidence to courts and other bodies.

Strangely, there's currently no formal qualification in this field. If your aim is to become 'CSI: Audio', what can you do? The University of Colorado in Denver is trying to rectify this missing training: sign up for a Master of Science in Recording Arts and you can focus on Media Forensics in the MSRA-MF degree programme offered by their National Center for Media Forensics (NCMF). The intention is to take students from a wide range of disciplines and prepare them for careers in the fields of audio and video forensics, as well as other areas of hi-tech crime fighting. We caught up with one of the NCMF team, Cole Whitecotton, to discover what it takes to be and audio sleuth.

How did you become involved with the NCMF University of Denver, Colorado?
I'm an alumnus of the programme, having graduated from it with a Master's degree three years ago. At the time, the Center was doing work for DARPA — the US Defense Advanced Research Projects Agency — and I became involved right at the beginning of the project. DARPA is the organisation responsible for the development of emerging technologies for use by the US military and, when I completed my Masters, the Center opened up a position that combined working on the project together with classroom support and online teaching.

Is this part of the initiative to identify 'deep fake' audio?
We were one team out of many working on detecting deep fakes. Our job was to generate fake audio and video materials that some of the other teams then tested to try to detect them. One of these was using straight-up machine learning AI, and others were working with hybrid versions that were a mix of learned and trained algorithms along with traditional

> **“It’s an arms race and I don’t think it’s ever going to go away**

methods like green screen, rotoscoping, and so on. We had meetings at least twice a year and met a lot of people ranging from big name universities, to the FBI, to companies like Honeywell, that are involved in AI. The project is moving to the next level now, to what they’re calling Semafor, or semantic forensics. I think that there’s going to be even more research into deep audio fakes during this stage.

Is finding fake audio harder than detecting a fake image?
Right now, not so much, but I think it will be. Companies like Lyrebird have made a lot of progress towards imitation and generating new, fake voices. Adobe also had something that they released as a beta version a while back [Adobe VoCo, Resolution V16.1] but after a couple of months they decided to pull it, because it seemed dangerous. All of these companies have been working to remove the robotic sound that for decades has been the tell-tale of a fake voice and, just like there’s thispersondoesnotexist.com I think that there’s going to be thisvoicedoesnotexist.com or something similar soon. The detection of fakes is going to get more and more difficult.

Moving away from the DARPA project, can you tell us about NCMF’s courses?
There are two areas of teaching audio forensics at NCMF; there’s the Masters programme and there are the training courses for law enforcement. In our new classroom, we have racks of equipment at the back of the room, and there’s a station back there that allows you to use the various systems and see them working. In my day, we had a sort of offshoot — a closet really — where we had all of the audio equipment, so we didn’t get to spend a lot of time with it. But now that it’s in the room it’s easy to fire it up to show it to the whole class. Unfortunately, we can have anywhere from eight to ten students in a class, so not everyone has the chance to use the systems but, after the demonstration, the students and trainees can start to experiment and learn what they can do.

How do you demonstrate audio forensic equipment to a class of students?
For example, our former Associate Director Jeff Smith has a good example of CEDAR processing that he loves to demonstrate during our three-day course on forensic audio analysis and enhancement that’s geared toward people working in forensic labs. He talks with loud music in the foreground to simulate a bar, and then uses reference audio from a CD to remove the music in real-time using the system’s Time Align and Cross Channel Adaptive Filter modules. A lot of times that we do this, no-one even hears that there is somebody talking in the original… and then we remove the music. It’s a great example of ‘here is something that sounds like nothing’ and then finding the meaning in it. The biggest reaction for people who don’t really have a lot of audio background is: “I didn’t know that was possible… I see stuff like this on TV and movies but I always assumed it was Hollywood magic!” We run into that a lot.

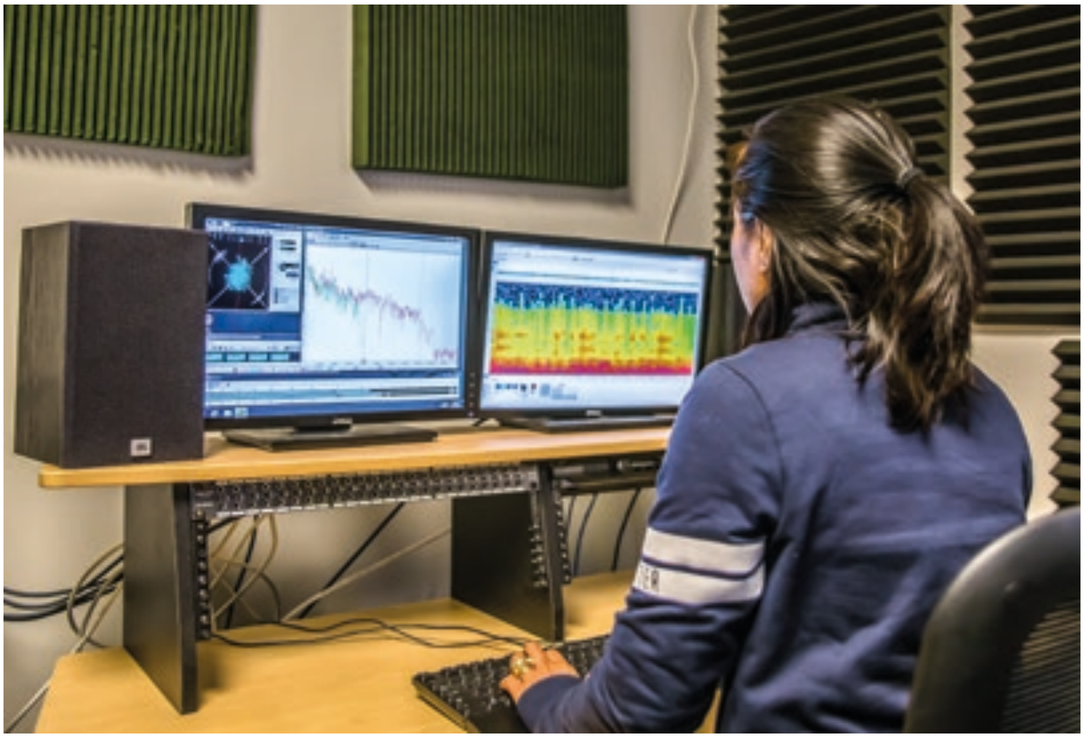
National Center for Media Forensics
COLLEGE OF ARTS & MEDIA
UNIVERSITY OF COLORADO DENVER

/ **Jeff M. Smith Associate Director NCMF (and Chair of the AES Technical Committee on Audio Forensics)**

You've mentioned the CEDAR Cambridge Forensic System. Can you tell us more about your use of that?
The aspect I like most about CEDAR is the real-time stuff. It's almost like the difference between a node-based system and a non-linear editor. I love how you can connect this process to that process, and how the UI shows you a visual representation of what audio is coming in, what's going out, and what each process is doing. Another feature that Jeff really likes is how you can create markers in modules like the spectrum analyser and then transfer them to other processes like Debuzz, or the EQs, to create the right filters. We also love to use CEDAR for background images. It looks so cool that anytime we do any classroom pictures we have its UI up in the background. Because you can have a bunch of different windows open for a single project, it's also how CSI looks. But that brings up another point. I'm sure that a lot of people get into this field because of shows like CSI, and their expectations can be way too high because of that!

Is operating audio equipment part of the course?
We're not here to teach our students specifically how to use CEDAR. We have many different audio and video systems and this is important, because each law enforcement agency has different tools from the next. So when the students come to us they can see those they know right next to others that they don't. Every tool does something slightly different, and everybody has its own way of working. The Masters programme is designed to make sure the students have a broad foundation that they can then build on. Enhancement is one part of this, audio authentication is another, understanding how audio is generated is another, and so on. Then we add the things you need to know if you're going to become a forensic expert — things like report writing, testimony and how to present what you did.

“Huge amounts of forensic data are being generated



So the focus is more on understanding the theory and workflow?
Anybody can push buttons, but you have to understand what the buttons do, when you would want them to do it, and why. That's all part of the course, preparing the attendees for working in different fields such as an FBI lab, or in local law enforcement, or in private companies like Target or Walmart. The first semester concentrates on legal stuff - the federal rules of evidence and what it means to be an expert witness. Then we have semesters that concentrate on audio, or on video, or on MatLab or Python scripting, and so on. We've also had people who have come here to help them develop their own forensic tools, but they don't necessarily have the computer science and coding skills necessary. If they did, they would also need a certain amount of luck... having the right idea at the right time. I can't remember the number of times when people have said to me: "oh, that idea was obvious, I could have done that" — but they didn't. So we ask them: "The person who did that probably had no more information than you, so why didn't you think of that idea first?" It's something that's really worth thinking about.

NCMF occasionally carries out fieldwork, and we noticed that you put out a press release about a counter-terrorism case that the Center worked on a while back. Can you tell us anything about that?
Jeff and Catalin (NCMF Director, Catalin Grigoras) are rock stars of the US forensic world, really well known and well regarded, so people come to us to ask for help and we sometimes work pro-bono with local police and other agencies. But I can't tell you anything more than that. What I can tell you is this — huge amounts of forensic data are being generated. The amount keeps growing and growing and growing, and that's the biggest problem: dealing with the immense amount of digital data by cataloguing and categorising, as well as making sure that you're not missing relevant information and not handing over irrelevant information. It's just going to get worse as time goes on. It's an arms race and I don't think it's ever going to go away.