Intelligent Audio

Big tech loves sound that makes sense — **NIGEL JOPSON** talks to CEDAR Audio MD Gordon Reid

/ Gordon Reid

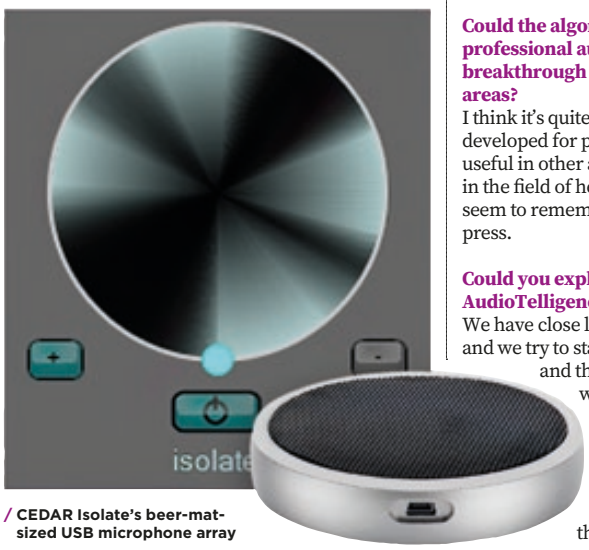
Audio was popular — but only if intelligent — at the January 2022 Consumer Electronics Show in Las Vegas. From Bosch using SoundSee microphones to record the sound of machinery and equipment on the International Space Station as a way to monitor safety, to voice-activated Sound Mirrors integrated with Amazon's Alexa. Music tech start-up Wisear were touting a technology that allows thoughts to control a smartphone's basic music playback controls — EarEEG combines an 'AI algorithm' with data captured by brain sensors. Noveto were promoting 'invisible headphones' using smart beaming technology. Companies had raised millions of dollars to investigate sound recognition for the Smart Home. Audio Anomaly Detection algorithms would learn the normative pattern of sounds within an individual home and send an alert whenever deviations occur, such as aggressive shouting or slamming doors — not in my home, please!

This big-tech enthusiasm for intelligent audio led our thoughts to the somewhat more self-effacing pro audio stalwart of restoration, dialogue noise suppression, speech enhancement and mastering — CEDAR Audio. Unbeknown to many production pros, a few years ago CEDAR spun-off a tech start-up of its own — AudioTelligence. Also based in Cambridge, the company has developed data-driven audio Blind Source Separation (BSS) technology — the separation of a set of source signals from mixed audio, without the aid of information about the sources or mixing process.

Most audio pros are aware that CEDAR is also at the forefront of speech enhancement for live surveillance and forensic audio investigation and, in July 2021, the company introduced a product aimed at these markets called Isolate. Available initially as a module in its CEDAR Trinity 5 long-term recording, monitoring and enhancement system, Isolate makes it possible to separate individual sound sources if the audio is captured using the system's beer-mat-sized USB microphone array. Its software can then identify and output the voices of multiple speakers simultaneously, allowing listeners to hear whole conversations even if they are taking place in noisy environments. This is clearly a breakthrough with possible future pro-audio applications, so we sat down with CEDAR's managing director Gordon Reid to discover more.

When you see all the fuss being made about AI audio, does it make you think, "CEDAR did that first"?

We mentioned the concept of artificial intelligence in 1992 when we launched the first hardware-based digital audio restoration product, the DC-1 Declicker. But, looking back, I think that it would be better to date its introduction to 1994 when we launched the DH-1 Dehisser. This was the first digital broadband noise reduction system that didn't require a spectral fingerprint. Up to that point, the usual method of performing broadband noise reduction in the digital realm was to take a fingerprint from a quiet bit of the audio and then, depending on how good your algorithm was, the noise could be reduced. The limiting factor was usually the point at which artefacts introduced by the process appeared — the now infamous glugging and twittering noises. At CEDAR, we felt that there had to be a way to dispense with the fingerprint by determining the noise content of the signal even when the wanted signal was present. This would also allow the algorithm to track the changing noise content which, of course, a fingerprint is unable to do. In addition to the obvious benefits for the user, this would also make the noise reduction more appropriate from moment to moment and reduce the unwanted artefacts.



/ CEDAR Isolate's beer-mat-sized USB microphone array

Was the DH-1 Dehisser an early example of AI audio?

The DH-1 calculated the spectral density of the noise using what we would nowadays call machine learning or — because people frequently misuse the term — AI. I described these aspects of the DH-1 in a draft press release, but the feedback I received before its launch was that this sounded a bit like techno-babble and that it would be better to concentrate on how the product benefitted the user. So I re-wrote the documentation, we launched the DH-1 at the European AES Convention in Amsterdam, and it went on to inspire a whole series of products that didn't require spectral fingerprints.

If you now fast-forward 15 years or so, you'll find other manufacturers had started to talk about applying machine learning and artificial intelligence to audio processing, and I think that this was a success because the time was now right. Consequently, audio pros started to become accustomed to hearing about 'AI' and the terminology also started drifting into the consumer market. Nowadays, 'AI' has become a buzzword: if your TV and your toaster have AI they must be better than those that don't!

Could the algorithms and technology in professional audio tools be used for breakthrough consumer products in other areas?

I think it's quite possible that the AI/ML techniques developed for pro-audio will prove to be very useful in other applications. One example of this is in the field of health-related diagnostics, which I seem to remember was discussed recently in the press.

Could you explain how CEDAR's sister company, AudioTelligence, came into being?

We have close links with Cambridge University, and we try to stay abreast of current audio research and then make informed decisions about what might be possible in the commercial arenas over the next few years. If we see something interesting, we discuss whether it could benefit our users and how we could work at the forefront of the new technology. Of course, we're

/ CEDAR Trinity 5

not unique in this, but it has been one of the driving forces behind CEDAR's evolution. We started out in the '80s working with sound archives and libraries, then moved quickly into pre-mastering, followed by post, audio forensics, live broadcast, and most recently location sound. At almost every stage, a technical advance made this possible.

Back in 2008, we became aware that Blind Source Separation was going to have huge potential when it matured so we allocated considerable resources to developing a practical BSS technology. We built and tested our first prototype in late 2013 and it soon became apparent that it wasn't really a pro audio product; it seemed to us that its greatest potential lay in communications and hearing assistance. CEDAR isn't set up to address the consumer and healthcare markets, so we spun-off AudioTelligence to do this. They've done some amazing work developing the technology for those users while, at CEDAR, we've continued to look at its applications in the seemingly unrelated areas of security and broadcasting.

Does Isolate have any links to previous CEDAR technologies or to technologies such as beamforming?

Similar development skills are involved, but the resulting technology is quite different. Isolate combines the underlying source separation with our 'direction of arrival' system, and it's the combination of the two that makes it possible for users to choose which sounds they want to hear. We can identify and isolate multiple speakers simultaneously, outputting a new signal containing the wanted people in a conversation without the bulk of the background noise and babble. If we also apply noise reduction to clean up the output, the results can be startling.

/ AudioTelligence, founded in 2017 as a spin-off from CEDAR

In contrast, beamforming is a technology for steering a virtual microphone in a known direction. Unfortunately, beamformers require large arrays of matched microphones so they can be unsuitable for use in some common situations; you certainly can't place one on a table in a restaurant or club to hear or interview the people around you. What you need in this situation is something that can deconvolve the entire 360° sound field into its constituent parts, rather than point a microphone at a particular speaker. So

/ AudioTelligence 'Aiso' focuses audio capture on sound sources located in a camera's field-of view

that was the task we created for ourselves back in the late-noughties: take a sound field with a mixture of wanted and unwanted sources and develop a practical technology that separates them so that listeners can choose which they want to hear.

Has there been interest in Isolate from the broadcast market, and could the technology be incorporated into a video camera?

We've had considerable interest from broadcasters, but whether BSS will take off in their field will depend upon whether they are willing to work with microphone arrays. But even today, I think that you could get some great results in noisy environments using an array sitting on a desk or something like a stable tripod. Whether camera manufacturers will want to integrate arrays into their cameras is another question and, even if they do, there will undoubtedly be additional difficulties to overcome.

Are there any plans to expand the Isolate product, to bring it into the postproduction market?

I would like to but, again, it means that people will have to become accustomed to recording using arrays with a known geometry. It's no good hanging eight microphones from the corners of the room and saying, "we'll use the audio from these to get rid of all of the unwanted sources" because it doesn't work like that.

Can we look forward to new CEDAR products for the pro-audio market this year?

We have on-going developments in all of our existing areas, and I am sure that some — if not all — of these are within a year of completion. At least one of them is based on some new research that we're quite excited about, some are evolutions of existing products, and a couple sit somewhere between.

/ CEDAR DH-1 Dehisser (1994) — with machine learning