WIDE-AREA NETWORKING & REMOTE PRODUCTION

PATRICK WARRINGTON
TECHNICAL DIRECTOR, CALREC AUDIO LTD, HEBDEN BRIDGE, UNITED KINGDOM,
EMAIL: PATRICK.WARRINGTON@CALREC.COM

The use of wide-area audio networking makes possible the remote production of live TV events, where large geographical distances separate the production facilities and the events that are being televised. There is growing interest in this area from broadcasters, both in the potential for significant reductions in production costs as well as improvements in production quality. But there are difficulties to overcome if this is to become practical in a wide variety of circumstances. This paper explores some of the technical challenges, in particular, the transportation of audio, and the issues of reliability, redundancy and synchronisation, and briefly looks at a variety of relevant technologies and standards.

Introduction

The world of broadcast has come a long way since the first state-controlled networks flickered into life in the 1930s for a couple of hours of didactic, tightly planned live-to-air content, to be transmitted once and then lost forever. Today's broadcast consumers are well on the way to being able to watch what they like, whenever they like, via a wealth of top-up, pay-to-view and free-to-air services.

And in this era of lucrative DVD boxed set releases and consumers downloading and/or watching whole TV series at times of their own choosing, the non-commercial attitudes that allowed reams of recorded programming to be erased in the 1960s and 70s (including, famously, all of the original 1960s master video tapes of the BBC's Doctor Who) on the grounds that no-one would ever want to see programmes again after their initial transmission — these attitudes truly seem to belong to the Paleozoic Era. But never mind the 1960s — already, there can be no-one under the age of 15 who can recall the pre-YouTube era before 2005, a time when one couldn't simply dial up video of practically anything from a handheld device. These young people may be working in broadcast and beginning to shape the future of the industry within the next five years — an interesting thought.

Despite the instant accessibility enjoyed by the modern consumer of so many forms of visual and audio content, it's still widely accepted that there are certain events whose excitement is best enjoyed live — or if not live, then captured live in detail and watched in near-real time. Big sports fixtures, news stories or state occasions all fall into this category. Here, broadcast has as important a role to play as ever, because not everyone can squeeze into a baseball stadium to watch the World Series Final, occupy the front pew at Westminster Abbey at a Royal Wedding (or Coronation…), or see over-salaried football players literally tearing chunks out of each other from a few feet away.

Of course, remote broadcasts are not new. In the UK, the first official outside broadcast was the coronation of King George VI, which was handled by just three cameras. Needless to say, since the 1930s, things have become a lot more complicated; at the wedding of Prince William and Kate Middleton, the BBC deployed over 100 HD cameras. And with increasing complexity comes increasing cost. Even small outside broadcasts are expensive, and this puts a limit on the number of events that can be covered.

This paper will explore the concept of remote production, and how close it is to a practical reality. By remote production, what is meant is the idea of live remote broadcasting, where the creative part of audio production, with a skilled audio operator sitting at a mixing console, does not take place at the event site, but at a studio facility some distance away — perhaps in a different city, country or continent.

This is quite a new concept that is only now becoming possible with the evolution of the necessary technology. We are at the point where manufacturers can start to make products for remote production that are more than laboratory experiments that don't constantly require expensive, specialised IT support. Perhaps most importantly, such products are now not only within reach but also reliable, and broadcasters stand a chance of being comfortable using them. After all, a product based on entirely sound theory that nobody wants to use in practice is unlikely to gain wide acceptance.

This paper discusses some of the challenges of remote production, and some possible solutions. But not everyone reading this will have a broadcasting background, so we'll begin by offering a brief overview of the kinds of audio connectivity that you find in broadcast environments.

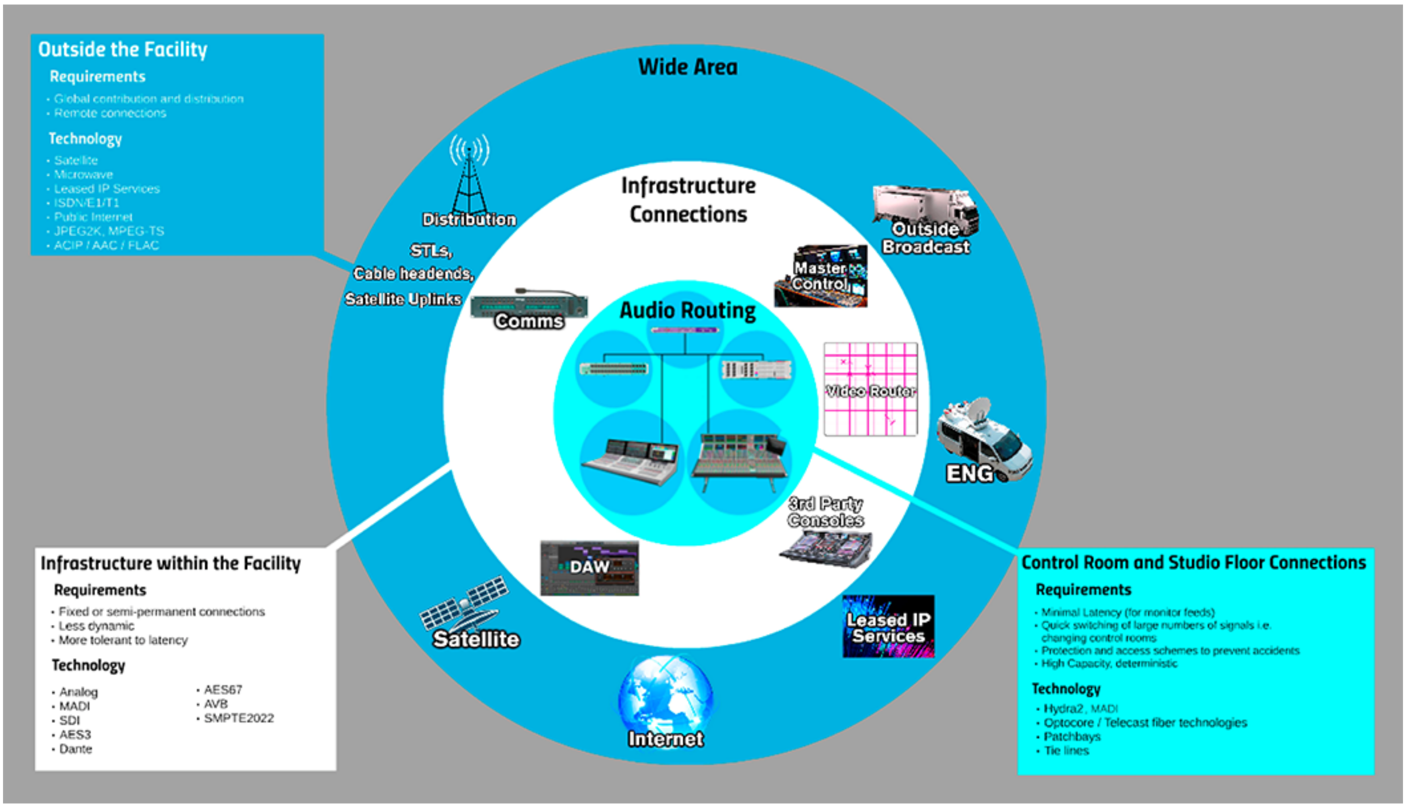
Knowing Your Onions

Native English speakers will be familiar with the above phrase, which means 'knowing your chosen specialist subject in detail'. Metaphors are popular in communications technology anyway, but in this case, an edible alium is a particularly apposite reference, and thus the diagrammatic representation of the world of broadcast audio above will be referred to throughout the rest of this paper as the onion model. Not because it makes people cry, but because it's layered.

What this diagram is trying to show is that there are different types of audio interconnections within a typical broadcast plant, all of which have different characteristics and map to different applications.

The centre is where much of the audio content is generated and routed, with microphones in studios and on people's lapels, connecting to a mixing console along low-latency paths. In other words, this layer is about signal acquisition and transporting signals from the studio floor to control rooms and mixing consoles. In the middle is the audio operator

THE BROADCAST AUDIO 'ONION' MODEL



who creates programme feeds, as well as monitor feeds for presenters and performers, often destined for in-ear devices. Signal latency is critical — if a monitor feed containing some element of the performers' own voices is delayed more than a few milliseconds, it becomes a distraction and may lead to talent-related tantrums.

In analogue days, all of this was very simple, consisting of mics plugged into consoles directly or through tie-lines and patch bays. Nowadays, other interconnection technologies are used. It is more common to find stage boxes located inside studios which contain the mic amps and converters needed to digitise the audio, close to the source, so that it can be transported with no further loss of quality to a control room that may be several hundred metres away.

All broadcast console manufacturers have products to do these jobs, operating via audio networking protocols designed to allow control rooms and studios to connect and share resources seamlessly, and making it easy for any given programme to be made in a different studio or control room without any manual repatching of signals, and thus giving clients a lot of flexibility in their management of studio resources. The protocol developed at Calrec carries thousands of low-latency audio channels, and also control information that allows port parameters to be controlled, such as microphone amp gain. The Calrec solution also monitors the network, and offers other services, for example the protection of routes to prevent accidental interruption or over-patching of critical signals, and integration with other broadcast control systems.

Transports used in this portion of the 'onion' are characterized by low-latency, totally deterministic performance, redundancy, and ideally, instantaneous on-demand routing (via push button or salvo) that is easily controlled from the mixing console surface by the audio operator, or by the facilities manager, from a control panel.

Moving to the second ring of the onion in the above diagram, we find that this does not directly concern itself with signal acquisition; rather, it is concerned with the communication of signals between different broadcast equipment, for example from the audio mixing console (in the inner ring), to communications systems like those from Riedel and RTS, to video routers like those offered by Evertz and Grass Valley. We may also see connections to other mixing consoles, perhaps for backup or sub-mixing, and to DAWs (such as Pro Tools or Pyramix) for recording. We also see connections to master control for programme transmission. In this ring, audio connections are likely to involve lots of channels, and they are likely to be far less dynamic — they may be set up and left for years. They are also more tolerant of latency.

Because of the large numbers of signals, MADI is a very popular interconnection protocol in this ring. We also see SDI, as well as good old AES3. In the future, we can expect to see more AoIP, or Audio over IP — Dante, Ravenna, and most importantly, AES67 and possibly AVB. In fact, this second layer should provide the perfect conditions for use of AES67 and AVB, because of its capacity and (hopefully) its interoperability, and because its higher latency is not a problem.

It is very important to understand that despite much talk and hype about them, AES67 and AVB are essentially cable-replacement technologies, not dynamic routing technologies. Perhaps they can be in the future, but there are no control mechanisms agreed on that would allow equipment to be controlled, for example to set up and tear down streams, or control port parameters. Also, IP streams do not lend themselves to dynamic, rapid configuration. A stream is not well suited to routing a single signal from here to there, or worse, a salvo of routes. Streams need to be large in order to be efficient, but small in order to be flexible. If each signal had its own stream, this flexibility would be attained, but audio over IP does not work like that. For example, if a stage box provides a 64-channel stream for a console, but then a piece of QA gear wants a stereo pair from the stream, difficulties will ensue, especially if the QA gear can't handle the entire 64-channel stream. It's notionally possible to plan a set of stream configurations that will strike the right balance for connections that remain static, but it is a hugely complex, and perhaps impractical undertaking to reconfigure IP streams and manage stream widths in order to meet signal routing demands, which change dramatically and rapidly. In short, it seems that push-button routing and salvos do not fit easily with AoIP or AVB.

Nonetheless, both AoIP and AVB do have an application sweet spot, namely the interconnection of third-party equipment. Where we now see thousands of cables under floors and in racks, with AoIP and/or AVB, in the future we might see hundreds or even tens of cables, as IT speeds increase. Very elegant solutions will be possible, cost-effective and hopefully future-proofed, but the benefits won't come for free. It will take some design effort to create a network that will work under all circumstances. IT staff may have to be involved in making sure the bandwidth is there to meet the traffic profile, setting up the right QoS and so on. But once the routes are in place, then as long as there are no surprises, AoIP- and AVB-based solutions should work consistently and reliably.

There is still a healthy debate between the advocates of AES67 and AVB — AVB supporters point to their chosen protocol's superior performance and immunity to network traffic volatility. AES67 fans will drily observe that switch capacity is cheap and getting cheaper, and will wish you luck in your quest to obtain AVB-compatible switch products.

We can also expect to see SMPTE2022 replacing SDI as video over IP in the future, as IP video routing becomes prevalent. Experiments are being conducted today, but it will probably be three to five years before there is anything usable.

Turning to the onion's outer ring, this is where communication beyond the campus, known in the business as Wide Area Communication, takes place. It falls into two types — contribution and distribution. Contribution includes back haul from remotes, which range from major sports productions to single-camera news reports and live studio hook-ups, such as interviews between people in different regions. Distribution includes STLs, feeds to cable head-ends and satellite uplinks. And, of course, directly to consumers through IP-TV.

Technologies used for long-haul or Wide Area audio connections have traditionally included satellite, a variety of synchronous data services like E1/T1, ISDN, and X.21, but more recently have included IP. In the case of big remote vehicles, the backhaul audio may be piggy-backed into a video feed that is data compressed into ASI, JPEG2K or an MPEG-TS stream and returned over a leased IP line, via satellite or a microwave link. In radio we see the use of IP codecs and IP audio streaming using ACIP and other proprietary mechanisms, usually employing data compression with AAC or FLAC.

Dante, Ravenna and AES67 are conspicuous by their absence from the above list, and indeed, their use in Wide Area networking applications is at best very limited. The transmission conditions need to be perfect such that no data packets are lost (high-quality MPLS has uptime in the 'five 9s', ie. 99.999% of the time). The network being used also has to support all the standards required for Dante, Ravenna or AES67, such as PTP — and put simply, it may not.

That concludes a description of the broadcast audio 'onion' as it is today. But how might it look in the future? And what about Remote Production — how could that work?

Remotely Interesting

Today, for anything other than the simplest single-camera news remote, it is necessary for all the broadcast equipment to be taken to the venue. This means using one or more enormous mobile production facility, usually situated in a big truck, with CCUs, a vision switcher, a comms system, an audio console, and lots more, including a great many staff to operate it.

But what if all the raw video and audio could be transported back to a studio facility so that the production could take place there? It's a simple idea in principle. It would mean that the amount of equipment and, crucially, staff required at the venue could be reduced. A production team would still be required, but they would not need to travel. Imagine the cost savings if all the production staff were not required to spend two days at a premiership or Bundesliga football match:

no airfares, no hotel rooms, no per diem expenses. And now imagine the savings if production teams didn't need to travel halfway around the world for weeks on end to cover the Olympics or the World Cup?

More importantly to anyone reading this who is not an accountant, the production quality can be improved, thanks to the better listening environment afforded by a studio (surround monitoring in OB vehicles is always an unsatisfactory compromise), and the equipment that a studio can offer over a remote production vehicle as a result of being less constrained by space requirements. So the solution is to transport the raw audio and video back from the venue to a production facility, rather than taking the production facility to the venue. It's content contribution on an entirely new scale. If this sounds like a tricky proposition, it is. There are currently lots of problems to overcome, but we can expect it to be a practical reality soon.

The question is: how? Firstly, the audio signals have to be acquired. As we've discussed, there are already ways to transport audio channels over long distances, but the impact on the production workflow needs to be considered. Ideally, audio operators should be seamlessly in control of signal acquisition, as they are now, their workflow unaltered; introducing new equipment or controls can only be counter-productive. In other words, we don't want to force a new operational paradigm onto audio operators who already have extremely demanding jobs. Ideally, the workflow should be the same for a remote production as it is in a studio, with mic gain and phantom power switching governed from the control surface, and port patching that can be stored into memories, all under the immediate control of the audio operators, and not on some other piece of equipment.

One way to achieve this would be to use the same stage boxes employed in a studio environment, albeit perhaps slightly ruggedied for use in the field. However, they should connect to the console and be controllable in the same way. The sole difference would be their location, which could be at any distance.

Immediately, there is a connectivity issue. How can multiple microphone feeds, control signals, and camera feeds be transported between Old Trafford in Manchester and Sky's Production facility in West London, or from the Dodger Stadium in LA to the NBC Sports facility on the East Coast? Or from the Olympics in Rio to a production studio in Berlin? The answer certainly involves IP, which is the only game in town for these distances.

IP and packet switching is a fine technology, although it was not designed for real-time use. We know that IP packets, once dispatched into a network, can take different paths, may have inconsistent travel times, or may go missing. None of this affects your email messages, as higher-level protocols deal with those, but it spells big trouble for a real-time media stream. Of all commodities on the planet, I can't think of one that has a shorter shelf life than an audio sample in a live stream. By the time a missing packet is detected, it's too late — there is already a hole in your picture, or in your audio.

For small, local, private IP networks, all of these problems can be managed, and reliable IP connections established. This is why AES67 or Dante can work well in that environment. But for wide-area networks, it's a different story.

IP connections themselves also vary in quality — MPLS can give 'private network' bandwidth performance, but at a cost. Leased services can be rented by the hour, but coverage is variable. BBC technicians complain of patchy service even in London, and have resorted to using aggregated 3G connections — several phone connections — for some radio outside broadcasts.

Nevertheless, IP networks are expanding everywhere, and they are growing in capacity and value for money year by year. They are already extensively used in some broadcast applications, in particular for radio, to connect remote studios. Sometimes this is to allow discussions and interviews with people who cannot be physically present, and sometimes it is to distribute live programme feeds to remote transmitters. So although IP networks remain patchy at the moment, they will provide a solution in the long run. Time is on IP's side. Nielson's Law of IP bandwidth predicts that user bandwidth grows by 50% per year (10% less than Moore's Law). So far, data from 1983 to 2014 fits the prediction.

Our ability to process is also increasing more rapidly than our ability to move data. This has always been the case. It is what leads to ideas like object audio, where the components of an audio production can be sent to consumers' homes so that it can be processed locally specifically to their taste.

Given time, there will be enough IP bandwidth, at sufficiently low cost for our requirements. And in fact, these requirements are themselves pretty small. Take the example of football; there will be several mics and some commentators' feeds; maybe 24 channels, with a few in the reverse direction (such as comms and monitor feeds). The areas of concern are more to do with access to connectivity; namely, there aren't always high-bandwidth data points where you need them. Increasingly, sports venues have good connectivity, but of course news happens anywhere, often in highly inconvenient places.

Back to the stage box, capturing PCM audio. How exactly will this data be transported? Firstly, some kind of codec will be needed to convert the PCM into a format that can be transported over IP. And there will need to be a transport protocol.

This is where some say we could use AES67. It is possible, but not ideal, since AES67, Ravenna and AVB rely on PTP clock distribution, as mentioned earlier, and the latter may or may not work well over WANs, depending on the IP provider. It might be possible to use a reduced form of AES67, with synchronisation provided via GPS clocks at either end, but ACIP is probably a more practical solution. This does not use PTP, but relies on a receiver reconstructing a clock by looking at the rate of packet reception.

Of course, all of the above is rendered useless without an IP service. As mentioned above, at sports venues it may be possible to lease substantial bandwidth for as long as is needed, if there is a fibre access point nearby, as is done for back haul right now. Some operators may have access to an MPLS service which can provide guaranteed, uncontended bandwidth, control of IP addresses and the aforementioned reliability in the 'five nines'. There's also the option of aggregated 3G and 4G mobile network bandwidth. And finally, there's the biggest IP network on the planet — the public Internet — where anything can happen, but quite often nothing happens. Like the Wild West, it is certainly no place for live audio streaming, although some fearless types do use it for live audio, occasionally for radio location reporting. This requires codecs that are carefully designed to adapt to the Internet's unpredictable behavior, by companies such as Tieline Technology.

Whatever IP connection is available, it may be desirable to data-compress the audio stream, perhaps to make better use of available bandwidth, reduce costs, or both. Bandwidth may well be at a premium if video feeds from six or more cameras have to fit onto the same IP service. Various options are available, including MPEG, AAC, and even FLAC (free lossless audio compression). These offer a variety of data compression ratios that trade bandwidth for audio quality, latency and processor cycles.

Data-compressed audio can also be more greatly affected by packet loss. In some frame-based schemes, such as MPEG layer 2, if a single packet is lost, an entire 24-millisecond block of audio goes with it, which is much more noticeable and harder to conceal than a single packet containing just a handful of samples.

If it is a requirement that no audio packets are lost over IP, then some kind of error correction scheme must be considered; packet loss even occurs over high-quality MPLS connections. FEC is one such scheme, working like CD's Reed-Solomon. Extra data is sent which allows for a certain density of errors to be fully corrected. Unfortunately, the nature of IP errors is that they often come in bursts, which will quickly exhaust FEC's correction capabilities.

Another alternative is to use a dual-stream mechanism. This is expensive, as it involves sending a second complete stream, identical to the original. At the receiver, the best connection is selected on a packet-by-packet basis, entirely transparently. So if one stream is disrupted, then you have the second to keep you going. If you are able to arrange it such that the two streams travel though alternate network paths, the protection can be very good.

However, the penalty for either of these schemes, and for the data compression mentioned earlier, is latency. Such protection requires more buffering, so the delay that the audio suffers travelling from microphone to mixing console is longer.

Why is latency an issue? It may not be in some circumstances. However, in most remote productions, where commentators are present in the venue, they will need in-ear monitor feeds. This allows them to hear a mixture of the programme audio, the producer, their co-hosts and themselves. If the monitor mix is created in a mixing console back at the facility, many miles away, then the audio that ends up in the ear of the commentator will have suffered a round trip's worth of delay, which might easily add up to several hundred milliseconds. Even at best, it will be tens of milliseconds, which will not make for smooth operations. Trying to talk while listening to a delayed version of your voice is next to impossible, even for a seasoned pro.

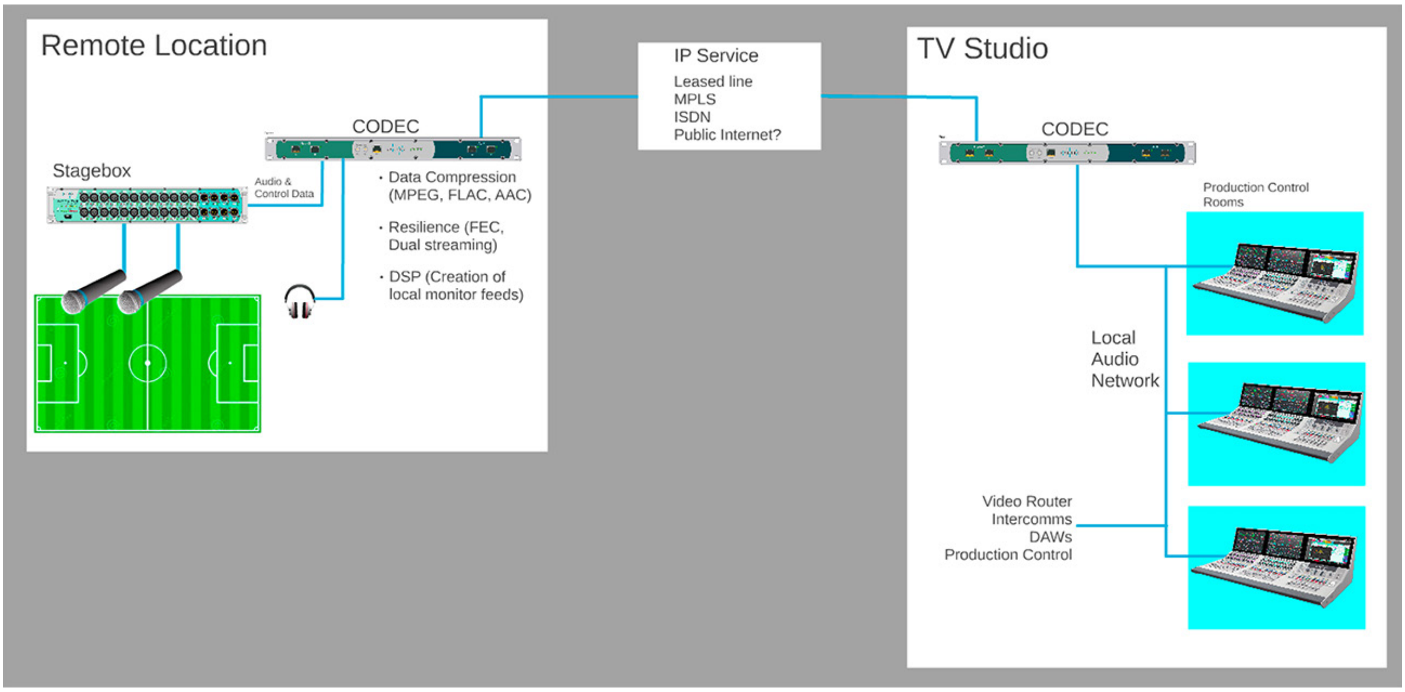
A solution to this problem can be provided by adding some DSP on the far end codec that allows monitor feeds to be created locally at the live event, by combining the local commentators' microphone feeds with programme feeds and producer talkback from the production facility. This way, there is no round-trip penalty imposed on the commentators' own voices, and audible delays remain within tolerable limits.

Of course, there are further issues, such as the problem of synchronizing audio back to picture. Lip sync is always an issue, but if audio and video are subject to different data compression schemes, the problem may be much worse. Clever automatic mechanisms can be derived to compensate for differential delays, but these are beyond the scope of this paper.

Security can be another bugbear. If the IP stream is to pass through an IP network that is public, or at least accessible to others, it must be encrypted to prevent unauthorized access, with more potential scope for increased latency and/or corruption.

These are the biggest of the technical challenges to remote production. But there is one still greater for equipment manufacturers. All of the requirements and challenges detailed above must be surmounted in such a way that the operators' workflows are not changed, or made any more complex; their role is already challenging enough. It is the belief of this author that the many technical complexities of remote production must be absorbed and managed by the mixing console equipment, such that the audio operator remains free to concentrate on the creative aspects of sound design and production quality.

A REPRESENTATION OF THE REMOTE PRODUCTION MODEL DESCRIBED IN THIS PAPER



Summary

In summary, remote production will come. The temptations of ever-cheaper IP bandwidth, and the prospects of large-scale cost savings, will combine to make it irresistible.

By way of a final example, a large sports broadcaster in the USA is aiming, in the future, to broadcast far more college sport than is currently possible and even high-school sports events. High schools have neither the skills nor budget to produce live sports, but they probably can lay out microphones and point cameras and connect them to a codec of some sort, allowing the production process to be handled by staff at a broadcast facility a long distance away.

It is often said that a production team will always have to be physically present at the largest, most prestigious events. However, even if this is true, by reducing the costs of live production for medium and smaller-scale events, it may be possible to cover more events, and make more localized coverage possible — an exciting prospect.