MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Combining video, audio and noise data in noise monitoring

Manvell, Douglas[1]
EMS Brüel & Kjaer
Skodsborgvej 307, Nærum, Denmark

Currie, David[2]
EMS Brüel & Kjaer
Skodsborgvej 307, Nærum, Denmark

ABSTRACT

One of the primary traits of noise monitoring is that, unlike noise measurements, there is no operator on site to take notes and photos or to observe (look and listen) the environment. Recording the sound signal from the measurement microphone in noise monitoring terminals enables source identification and categorisation into specific or residual sound to better determine noise limit compliance. Adding a camera to the noise monitoring system enables videos or a series of still pictures to correlate with noise levels and sound recordings to further enhance source identification and categorisation into specific or residual sound. This paper describes the techniques involved and discusses best practice in combining video, audio and noise data in state-of-the-art noise monitoring solutions to not only enable source identification and categorisation but also to do so efficiently. The paper will also highlight areas of future development and research.

Keywords: Noise, Environment, Instrumentation
I-INCE Classification of Subject Number: 71

1. INTRODUCTION

1.1 The Human Expert

The acoustic specialist uses their eyes and ears to supplement environmental noise measurements. Their observations of the situation are an important resource in environmental assessment. In many cases, they take pictures and record the sound to help document their assessment and help identify and document specific noise levels [1]. With this, environmental noise assessments are more robust.

Although several sources of data are available [2], noise levels, video and audio replicate the most important senses and data available to an on-site expert. Recording the sound signal from the measurement microphone in noise monitoring terminals enables source identification and categorisation into specific or residual sound to better determine noise limit compliance.

---

[1] douglas.manvell@emsbk.com
[2] david.currie@emsbk.com

Adding a camera to the noise monitoring system [3] enables videos or a series of still pictures to correlate with noise levels and sound recordings to further enhance source identification and categorisation into specific or residual sound.

1.2 The Remote Human Expert

One of the primary traits of noise monitoring is that, unlike noise measurements, there is no operator on site to take notes and photos or to observe (look and listen to) the environment. On the other hand, remote monitoring enables the acoustic specialist to effectively be in more than one place at a time, and enables them to look back in time so that they can evaluate a situation after the fact. Rather than trying to clone acoustic experts, it is more feasible to deploy multiple monitoring hardware to record salient information.

Attempts to utilise sound and video in noise monitoring are not new. This paper looks at some of the attempts made and the lessons learned, and it supplements this this with a discussion of techniques that can be used, and emerging technology that can be correctly applied to benefit the expert as well as the client and decision makers.

1.3 From Limited Data to Big Data

Historically, data storage and transfer was expensive and time consuming. Hence, early remote monitoring applications tended to report only a few salient parameters, operate in alert/event modes where additional details are recorded only at times of interest, and minimize file size by utilising compression techniques. This was used in audio, video and in noise with eg periodic reports and, for statistics, specific levels and distribution data as opposed to storing all the samples. Although data was limited, data gathering intelligence was built into these remote monitoring systems to help enable compliance decisions and troubleshooting to be made.

Today, data storage is economically viable and transfer of large amounts of data is, although potentially costly, possible. There is no longer the same need to strictly determine suitable data reduction strategies. Instead, intelligence needs to be built in on the data presentation through big data analytics which examines large and varied data sets to uncover information that can help organizations make informed business decisions [4].

Nevertheless, industry-defined best practices are often based on solutions from previous generations of remote monitoring systems. Efficient systems need to combine meeting industry standards, exploiting the capabilities of big data, and providing efficient application and business decision workflows.

2. RECORDING THE SOUND SIGNAL

Recording the sound signal from the measurement microphone in noise monitoring terminals enables source identification and categorisation into specific or residual sound to better determine noise limit compliance. This can be done manually or automatically using AI, both after a training period.

Digital signal recording can be done with different inputs, ranges, quality & format (ie with or without compression).

Typically, the recording is of the frequency weighted signal from the measurement transducer. Normally the Z-weighted signal is used for listening, although using the C-weighting can reduce wind noise – this reduces the maximum levels in the recording, reducing the risk of overload when looking at quiet noise levels. In addition, sound recordings are often linked to noise events (where the noise level exceeds trigger conditions such as a broadband level over a minimum duration). Here, the recording duration can be set to a fixed limit or the end of the event. For long events, a maximum recording duration is recommended to reduce analysis time and data transmission costs. With noise events, it is useful to include recording of some seconds of the signal in advance of the event (so called pre-recording). This is useful as the time leading up to the trigger often contains useful audio information. To enable this, in some noise monitors, this functionality requires a sound recording buffer to enable this period to be recorded as event trigger conditions can take some time to be exceeded.

Early sound recording implementations used 16-bit or less resulting in a recording range of over 80 dB (theoretically up to 96 dB but typically some bits are used for metadata, etc). Although this sounds good, in reality it is insufficient for some monitoring applications, for example for monitoring jet aircraft fly-overs in quiet rural areas where the noise levels can vary from under 30 dB to over 120 dB. 24-bit enables over 120 dB which is sufficient for covering the entire measurement range of the monitor from its noise floor to the maximum measurement level of its microphone. The larger the recording range, the more it becomes necessary to be able to adjust the playback gain to enable signals to be heard.

The quality of the recording is defined primarily by the sampling rate. This is effectively 2 times the maximum upper frequency in the recording due to anti-aliasing. Ranges from 8 to 48 kHz are often used. 48 kHz enables recording over the entire human listening range and is thus optimal for signal analysis of audio. However, it uses a lot of memory and bandwidth in transmitting the signal to the server for listening or processing. Thus, smaller sampling rates are often used. For a lot of environmental noise where the main content is between 50 to 1000 Hz, 8 kHz sampling is ample – increasing recording capacity and reducing transmission time (and thus cost) by a factor of 6. The sound recordings are a significant percentage of the data required to be transferred (see below).

Table 1 Memory required as a function of sampling rate

Sampling Rate (kHz)		8	16	24	48
Memory (KB/s)	16-bit	16	32	48	96
	24-bit	24	48	72	144

With the invention of MP3 in 1993 [5] and its widespread use in the music industry, it was quickly taken up in the mid- to late-1990s in noise monitors, when bandwidth and storage were still at a premium as a way to optimise recording quality within the limitations of data transfer, even in batch transfer mode. Even more so if real-time listening was required. Recordings are similar in size to 8 kHz 16-bit uncompressed recordings.

MP3 decoding and encoding technology was patented and, up to 2017, fees for their use had to be paid by system integrators. In addition, the compression complicated subsequent data processing and analysis. Thus, solutions avoiding compression but utilising to 8 kHz 16-bit uncompressed recordings also thrived. Over the last decade, the bandwidth and cost of remote communications has significantly increased and decreased, respectively, thus reducing the drive to compress the recordings.

Another technique that is used is automatic gain control (AGC) [6] which takes a signal with a large dynamic range and compresses it (eg from 120 dB to 40 dB) to ease manual listening and recognition of the signal on playback. This is ideal for listening to the measured signal during attended measurements to ensure that the charecteristics of the specific sound are audible in the sound recording. It was introduced when high dynamic range recordings became available and soon became a naterual part of noise monitoring systems. In some systems, the user can set the Peak Recording Level of the recording to provide enough signal for the associated audio playback function to be more easily audible. Settings may be lower than this if the monitor records lower noise levels in quiet surroundings. With a recording range of 90dB, it could be set to eg 110dB to enables coverage from a typical NMT's noise floor.

However, in practice, the AGC method has a couple of drawbacks. The technique can result in what sounds like clicks or beeps in the audio signal when it is used in combination with lower sampling rates such as 8 kHz. This is caused by the AGC adjustments. It is only related to the audio clip used for recognising the noise and has no impact on the measurement. So, to paraphrase Shakespeare, "To AGC or not to AGC, that is the question". This is resolved in the EMS Brüel & Kjaer AU-3639 NMT [7] by giving the user the option to set up the audio recording with or without AGC. We discovered that not having AGC enhances source recognition by listening by enabling the increase in noise levels to be reflected in increased volume of the playback. Auto Gain Control (AGC) is particularly noticeable in road, rail or aircraft vehicle pass-bys where the gain is reset several times during the passby, this giving a different psychoacoustic impression compared to the sound in real-life.

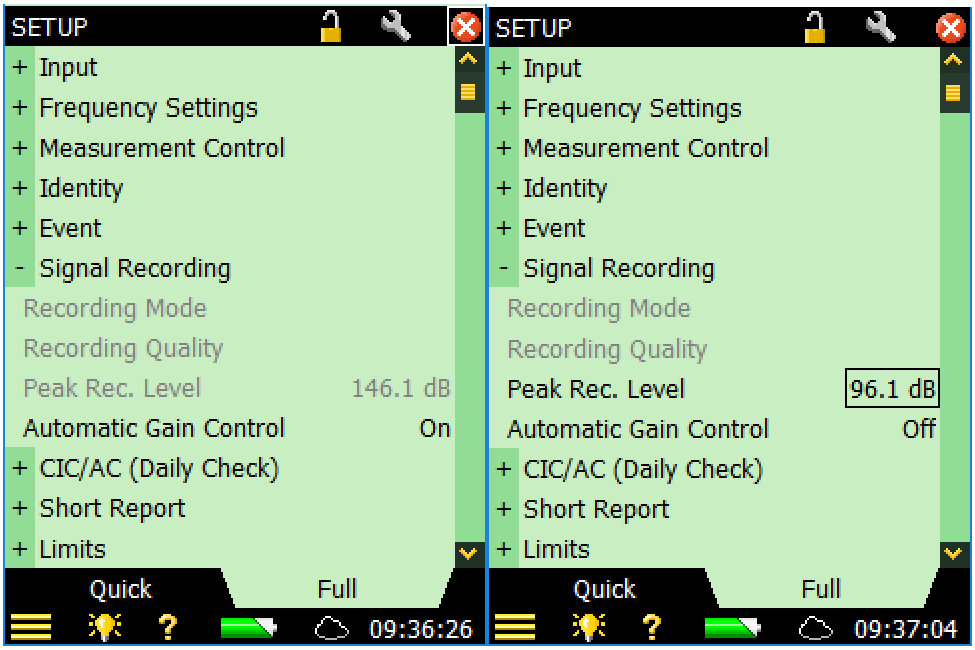

Figure 1 A noise monitor display showing set up of signal recording with and without AGC, in this case the EMS Brüel & Kjaer AU-3639 NMT

However, without AGC, the user is required to set the Peak Recording Level of the recording. In our solution, the recommended and default setting is 90dB (nominal value, adjusted by calibration) which provides enough signal for the audio playback function on a PC to be more easily audible. Settings may be lower than this if the NMT records lower noise levels in quiet surroundings.

Utilising AGC enables the user to listen to both quiet and loud noises without having to manually set up the NMT recording but sacrifices the dynamic psychoacoustic aspects that sometimes ease recognition. This may prove useful when measuring relatively close to impulsive sources such as around construction sites or close to mines in cases where the user is interested in hearing the residual sound as well as the sound causing the alert. AGC has the side effect of making wind noise in relatively quiet areas more audible.

An important part of audio is the increased operational costs and bandwidth required. Typical noise data including 1/3-octave data is approximately 1GB / month. Adding audio costs ca 1 MB / minute for a 16-bit 8 kHz sampling WAV file so even with only 5 minutes of sound recording per hour results in another 2 GB of data, tripling the data transfer needs. This has to be taken account of in the communication bandwidth and the associated subscription costs. It is one of the reasons why 24/7 audio is not widely used. However, in the future, assuming improvements in bandwidth and extrapolating historical reductions in communication costs, 24/7 audio recordings could become viable. Nevertheless, the main task when monitoring remains ensuring data intelligence rather than simply making detailed data available.

Stereo recordings are not relevant for noise monitors as only a single microphone is used. Thus, listening perception is somewhat limited compared to binaural recordings. However, it is, in most cases, sufficient for source recognition.

3. ADDING A CAMERA

Adding a camera to the noise monitoring system enables videos or a series of still pictures to correlate with noise levels and sound recordings to further enhance source identification and categorisation into specific or residual sound.

While surveillance video cameras have been used for many years to monitor public spaces and businesses for security purposes [8], these systems were typically not suitable until recent times for remote cloud-based storage systems due to the bandwidth required to transmit the data. Traditionally, surveillance cameras communicated live data directly to a central server, where a user could find video clips or images of interest by looking at the time in question. For example, in New York City surveillance cameras were used in conjunction with noise monitors to monitor areas of the city where overnight deliveries were made. At the time of a noise alert or complaint, the videos were looked at manually by the users to establish who or what the sound source was.

In more recent times, advances in both camera technology, compression technology and the previously mentioned communications improvements have meant that new approaches can be taken to further enhance source identification. It is the combination of these advances that have reduced the limitations that made the use of cameras costly and less-effective then they could have been.

Digital camera technology improvements have been driven by technology changes that have seen a decrease in the size of the pixels possible in camera digital sensors. This decrease in size has led to an increase in the resolution which now allows digital surveillance cameras the capability of much higher resolutions such as High definition, 4K or even 8K, giving the viewer a much more capable image in terms of video and image fidelity. These improvements have also resulted in higher quality sensors with greater dynamic range, giving greater contrast and better low light capabilities.

For night vision these high contrast low light cameras are further enhanced with the addition of infrared LEDs build into the camera, or when necessary, a separate infrared illuminator can be combined with the camera to enhance the video image.

The IOT evolution has pushed surveillance camera technology into the age of smart devices. Surveillance cameras now have microprocessors on board and are capable of being completely controlled remotely, allowing control of camera resolution, focus and sensitivity as well as giving the user the ability to pan, tilt and zoom the camera on request. Also possible with this newer generation of IOT cameras is that developers can now program the camera to perform such functions as video compression, facial recognition and other security-conscious operations. Cameras now have on-board storage, which can be used to store data without the need to constantly transfer data to a central storage device, thus potentially minimising bandwidth costs. This storage can be periodically polled, and the videos or images stored can be compressed prior to transmission to the server. The IOT also provides the possibility for all of this to occur automatically without human intervention.

Other artificial intelligence such as facial recognition or licence plate recognition is also capable to be programmed directly onto the camera itself. It is very common for a camera located on the gate of a carpark to recognise and print the licence plate of a car upon entry to the carpark. This recognition can happen within seconds and the printed on the parking ticket. Once a user has paid for the parking using the pay stations, the system will automatically notify the computer controlling the gate on the exit that the car with the licence plate originally recognized is now clear to leave, and it is not necessary to use the parking ticket to exit the carpark anymore. The camera on the exit again recognises the licence plate and automatically opens the gate.

Video compression codec progression has given significant savings to the amount of data needed to store for any given video clip. In “Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)” [9], the comparison of several generations by means of peak signal-to-noise ratio (PSNR). The findings were that HEVC encoders can achieve equivalent subjective reproduction quality as encoders that conform to H.264/MPEG-4 AVC when using approximately 50% less bit rate on average.

Figure 2 A typical outdoor camera

4. COMBINING IT ALL - TECHNIQUES AND BEST PRACTICES

4.1 Previous Experiences

The earliest system combining audio, video and noise levels I can remember being involved in was for a project in 2001 around a construction site in Los Angeles. The system was built up around a Brüel & Kjær Type 7802 Noise Montoring Software which connected to the NMTs over modems and a video server running the cameras. Here, 12 NMTs were used with 2 digital video cameras.

The NMTs recorded the event noise levels as well as MP3 (compressed) format sound files. This data was downloaded from the NMTs to the noise server either daily or on request from the user. Each NMT had 3 months of backup data.

Each camera was a digital standalone unit with separate IP address and capable of simultaneously transmitting 1 frame/sec digital data to the video server. It was possible to remotely see 30 frame/sec real-time video data over the internet. The two servers are bridged together.

- Separate video server with 3 months storage of 24X7, 1-frame/sec rate data
- Main server for noise levels, sound recordings, and playback and control of the entire event playback functionality

Once an event was selected it was possible to see the noise level time profile with the time updated, listen to the sound file and play the selected (muted) video together.

Figure 3 Screen shot from the 2001 Los Angeles project

The standard mode of operation for the sound recording download was for the system to call up the NMT and request the MP3 sound file. For large numbers of events (eg 100 per day per NMT), it was recommended that a higher speed communication (DSL) be made available at each location. It was estimated that downloading this would take approximately 5 minutes for each NMT – ie 1 hours for the all 12 NMTs in the system. However, on demand downloading resulted in minimal playback delays.

In 2008, a we delivered another deployed system around a port, this time based on Brüel & Kjær Environmental Noise Management System Type 7843. Here, all data was stored on the same server, making the system architecture simpler and more robust. In addition, the cameras were integrated with the NMTs so that data was only transferred to the server when there were noise events, reducing the operating cost of the system.

However, this had a significant drawback in that the noise monitors and cameras were located at noise sensitive properties some 2-300 m away from the main noise source of the port area. As the main complaints and issues occurred typically during the evening, the camera had difficulty in helping to identify up the cause of the alert due to the distance and, importantly, the fact that the area was floodlit, resulting in poor picture resolution of the suspected noise sources, no matter how the light setting of the cameras were done.

4.2 Current Best Practice

In 2018, EMS Brüel & Kjaer developed a modern noise management solution including cameras on the basis of research into best practice. This new solution exploits that the camera can record 24/7 video to its on-board memory, and does not utilise any communication bandwidth, unless a noise exceedance has been detected. When that occurs, the Sentinel system requests appropriate video clips from the camera, to show what occurred during the noise exceedance, and then present the user with the clip in question. Thus, large amounts of bandwidth are not required to send all video to the server. In addition, the user does not need to scroll through large amounts of video to find an action that caused an exceedance - instead, it is provided to them automatically.

Multiple cameras can also be used and provide videos of the period from different angles or locations in the sound scape.

This video clip is also (where appropriate) synchronised with the audio from the IEC 61672 Class 1 microphone, so that, during playback, the operator is able to not only hear the audio from the noise monitor but is able to video the video of the same period sync'd together.

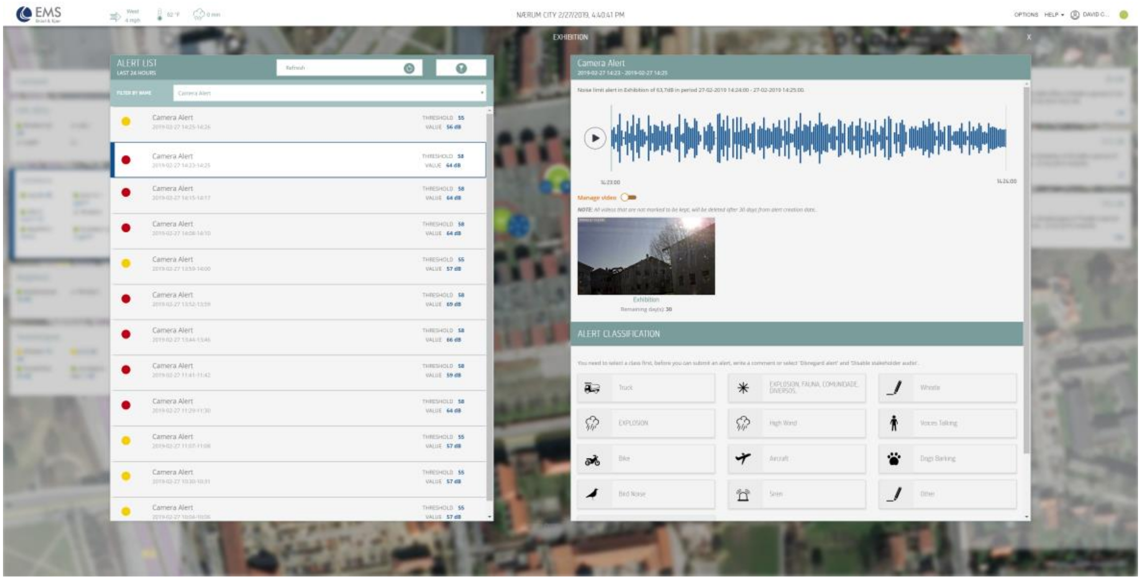

Figure 4 Sentinel alert handling window with video [3]

A good example of this configuration is in Edmonton, where the system has been used to record the licence plates of vehicles that have made a noise exceedance. In this case, the camera provides the video sync'd with the audio from the measurement microphone to allow the Edmonton council to evaluate the sound source.

For longer periods of time (for example a 1 hour LAeq), the system provides a time-lapse of the activities in front of the camera, rather than a real time feed. In this case, the user only needs to review a time-lapse of a couple of minutes to establish the pattern of behaviour that led to the exceedance (time-lapses are not sync'd with audio).

5. FUTURE DEVELOPMENT AND RESEARCH

On the basis that communications bandwidth continues to improve while its cost continues to decrease, data quality and quantity could increase until such time as further enhancement does not give additional benefit.

24/7 recordings as opposed to selected data sets could become viable and thus more widespread. Nevertheless, the main task when monitoring remains ensuring data intelligence rather than simply making detailed data available. Thus, development in this direction has perhaps limited interest, particularly in the long term, even though data storage costs may not be extortionate, and even if AI were utilized to sift through data to help determine outcomes.

What is more promising is increased data quality, in particular in combination with AI. Additional data could help reduce the number of false positive limit exceedences and provide better insight into the causes of non-compliance.

How much binaural recordings could benefit remote monitoring and source identification in operational systems needs to be investigated.

An interesting area of research is into how monitors could be developed that emulate human ability to localise and identify sound sources. This could involve stereo cameras and microphones on a moveable "head". And, if the monitor's "brain" was trained and calibrated to thus quantify and identify specific sources and their sound levels. Research into this could lead to a new generation of monitors that one could call a "Weatherproof Acoustic Consultant Robot". Following on from this, it would be beneficial that these be standardised to defined quality standards, and independently verified to meet these.

6. CONCLUSIONS

With no operator on site to take notes and photos or to observe, audio and visual recordings greatly enhance noise monitoring. Recording the sound signal from the measurement microphone in noise monitoring terminals or adding images of the source area enables enhanced source identification and categorisation into specific or residual sound to better determine noise limit compliance. Based on an analysis of monitoring needs and available technology, this paper has discussed best practice in combining video, audio and noise data to efficiently enable source identification and categorisation. The authors look forward to the results of future development and research.

7. REFERENCES

1. Acoustics — Description, measurement and assessment of environmental noise — Part 1: Basic quantities and assessment procedures. *ISO 1996-1:2016*
2. D. Manvell & P. Stollery, "New techniques to determine specific noise for increasing the effectiveness of continuous unattended noise monitoring systems", *Proceedings of Internoise 2014*, Australia (2014)
3. https://www.emsbk.com/sentinel/
4. A. De Mauro, M. Greco, M. Grimaldi "*A Formal definition of Big Data based on its essential Features*". *Library Review*. **65**: 122–135. doi:10.1108/LR-06-2015-0061 (2016)
5. https://en.wikipedia.org/wiki/MP3 "*A More Loss-Tolerant RTP Payload Format for MP3 Audio — RFC 5219*". IETF. February 2008. Retrieved 4 December 2014.
6. https://en.wikipedia.org/wiki/Automatic_gain_control
7. Noise Monitoring Terminals Types 3639 and 3655 Product Data sheet BP 2379, EMS Brüel & Kjaer (2018)
8. http://en.wikipedia/org/wiki/Closed-circuit_television
9. J. Ohm, G. Sullivan, H. Schwarz, T. Tan and T. Wiegand "*Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)*", http://iphome.hhi.de/wiegand/assets/pdfs/2012_12_IEEE-HEVC-Performance.pdf (2012)