THE QUEEN'S AWARDS FOR ENTERPRISE 2001

BBC

R&D White Paper

WHP 057

August 2004

Audio watermarking

R.Walker

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 057

Audio watermarking.

R. Walker.

Abstract

The paper contains a brief overview of modern methods for embedding additional data in audio signals for the purposes of access control or forensic tracking. The paper also contains a summary of work carried out at the BBC Research Department in the late 1960s and early 1970s on the same topics. It concludes that the earlier work anticipated almost all modern audio watermarking methods. In some cases, even the exact numerical values of the system parameters proposed are essentially the same as those used today in some systems. At that time, the limitations of technology and the absence of today's more serious motivation meant that the work was not continued at that time. The work also anticipated elements of today's audio bit rate reduction coding systems though, again, the future need was not anticipated at the time.

Key words: audio, watermark, coding, bit rate reduction.

White Papers are distributed freely on request.

Authorisation of the Chief Scientist is required for publication.

BBC Research & Development
White Paper WHP 057

Audio watermarking.

R. Walker.

1. Hiding messages - an introduction and a brief history.

The motivations for embedding hidden markers or messages in any kind of media content are essentially self-evident. Putting a 'secret' mark in an object can enhance security by making counterfeiting more difficult, as in the case of bank notes, or enable more reliable identification for forensic purposes. The process is commonly known as "watermarking". It is related to steganography[1].

Historically, steganography has a long history. Amongst Greco-Roman civilisations, one of the earliest documents describing the transmission of a hidden message comes from the Histories of Herodotus, *c.* 440 BCE. At that time in Greece, text was written on wax covered wooden tablets. In one story, Demaratus wanted to notify Sparta that Xerxes intended to invade Greece. To avoid detection, he scraped the wax of the tablets and wrote a message on the underlying wood. He then covered the tablets with wax again. The tablets appeared to be unused so they passed inspection by sentries without question. Unfortunately, the key to this message was not known to the recipients, who took some time to discover it – an early illustration of the importance of key management!

A common form of hidden writing is by the use of 'invisible' inks. Such inks were used with great success as recently as World War II. At that time, steganographic technology consisted almost exclusively of invisible inks. Because of the ease of activating those simple inks and with improvements in chemistry, more sophisticated inks were developed. The messages had to be processed with a prescribed and secret sequence of chemicals (the key), in the same sort of way as photographic images.

Null ciphers (unencrypted messages) were also used, and still are. Because of the recognisable patterns produced by many coding schemes, suspect communications can be detected by automatic filters. By camouflaging the real message in an innocent sounding one, the apparently innocent messages are allowed to pass. An example of a message containing such a null cipher is:

> *News Eight Weather: Tonight increasing snow. Unexpected precipitation smothers eastern towns. Be extremely cautious and use snowtires especially heading east. The highway is not knowingly slippery. Highway evacuation is suspected. Police report emergency situations in downtown ending near Tuesday.*

Taking the first letter in each word, the following message can be derived:

> "*Newt is upset because he thinks he is President.*"

The following message was (allegedly) actually sent by a German spy in WWII:

---

[1] Steganography is the art of hiding messages. When the message content is related to the host carrier then the hidden message is a watermark. The watermark set also includes messages that are not hidden. The sort of hidden watermarks discussed in most of this paper are from the intersection of the sets of steganography and watermarking.

Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by-products, ejecting suets and vegetable oils.

Taking the second letter in each word the following message emerges (though it is interesting to note that almost all references to this famous example omit the hyphen and thus get the actual decoding incorrect):

"Pershing sails from NY June 1."

All steganograpic systems are limited by a fundamental three-way conflict between visibility, robustness and data rate. The complexity and sophistication of the system increases with the degradation that the marked material has to withstand, the amount of embedded data it has to carry and with reductions in the required perceptibility of the embedded mark.

2. Audio watermarking.

In the audio industry today there is much interest in copyright management and protection. Embedding some form of 'hidden signal' or watermark in the audio stream is seen as a potential method for managing the use of the material. To ensure that only those with the right to access it can do so, methods have been proposed to include electronic 'gatekeepers' in audio equipment. That way, unauthorised reproduction, and especially unauthorised copying, could be prevented. However, there are many serious problems with this concept. It is not at all clear that they can be overcome sufficiently to provide a reliable and effective control system.

An alternative use is to mark content solely to enable its provenance to be traced. This usually needs lower watermark amplitudes, mainly because detection can be spread over a rather longer interval than for 'real time' control. That significantly eases the problem of finding a compromise between the three principal system properties. It is known that some commercial recording companies are now using such systems to trace 'leaks' that occur during the production process.

Some of the problems can be listed as:-

a) Audio quality.

It is obvious that the system must not excessively impair the audio quality. The acceptability of the impairments depends on the intended audience. Some people argue that any alteration of the original sound is unacceptable. They probably argue from an idealistic position, that the best current audio quality is only barely acceptable and still needs further improvement before it could be described as 'perfect'.

Others, working from a more pragmatic position, recognise that most of the potential audience is either not that critical or is not in a position to judge audio quality to that extent, because of their limited hardware facilities or aural capabilities. To those, a degree of impairment that might actually be finite but still imperceptible would, by definition, be acceptable.

Yet others, listening under very poor conditions, as in a car or via a low bit rate channel, might even tolerate some clearly perceptible impairments, especially if they could not distinguish them from the other impairments.

The wide range of these different conditions presents a very difficult problem for watermarking system application. It is unlikely that a single watermark would survive translation between those different environments.

b) Reliability

Any system intended to control access by a legitimate customer to legally acquired material will have to be reliable. It takes only a tiny amount of reported difficulty to give a system a bad reputation commercially. One of the serious potential problems would be the (probable) lack of acceptance of liability by vendors. In both the pre-recorded audio and the broadcasting industries, the hardware and the content are generally provided by different suppliers. The potential for each supplier to blame the other's product, leaving the customer without recourse, is self-evident. In fact, at the time of writing, just such a conflict is taking place over other sorts of copy-restriction techniques. It has become an exceptionally contentious issue, one that may be responsible for some of the recent decline in retail CD sales, at least according to some industry opinions.

Any watermarking system used for access control would have to detect the watermark quickly and reliably and default to not preventing access when it could not do so. It is somewhat less important that access by non-authorised persons is properly barred. It is doubtful whether any protection system with adequate performance in that respect could be developed. Whether it could also be done with acceptable audio impairment is seriously doubtful.

Identification applications need not be so reliable, at least in the short term. A longer detection period provides extra detection performance, allowing a reduction in watermark amplitude and audibility. It means that the required degree of reliability can be attained over longer extracts, perhaps most of a 'single' record track of, say, three to five minutes.

c) Robustness

To prevent unauthorised access, a watermarking system has to be resistant to significant degradation of the material. Many of the potential customers for illegally copied audio (or video) are quite tolerant of impairments to the material. Any included watermark would also be impaired. The watermarking system has to be resistant to those impairments. The general view is that the watermark should survive until the content becomes "of no commercial value".

d) Audio coding for bit-rate reduction.

The inherent conflicts between bit-rate reduction systems that try to remove inaudible components and watermarking systems that try to hide additional data inaudibly are self-evident.

The rest of this paper is concerned with the technology of audio watermarking. Questions about its commercial application are for others to decide.

3. Spread-spectrum modulation.

The problem of conveying information at a relatively low rate through a channel with perhaps a wide bandwidth but a very poor signal-noise ratio is a common one. In many such applications, security is also important. A modulation system called 'spread spectrum' was developed for such purposes. The first known patent application was in Germany in 1940/41 for secure military communications over radio links. The first US patent on spread spectrum came from Hedy Lamarr, the Hollywood movie actress, and George Antheil, an avant gard composer. That patent was granted in 1942, but the details were a military secret for many years.

The principles of spread spectrum modulation are well known and much used today. The technology has evolved into commercial systems. It is the foundation of modern portable communications system (mobile phone), wireless LAN standards and the Global Positioning System (GPS).

One implementation uses a frequency-agile transmitter/receiver combination. The carrier is switched rapidly amongst a large set of potential frequencies. The switching sequence is created by some form of pseudo-random pattern generator. If the receiver can be synchronised initially and remains in step (a non-trivial matter), the transient channel frequencies are re-combined to give a useful, nominally non time-varying, channel. The effect is to spread the radio frequency energy over a wide band, with corresponding reductions in the amplitudes of individual components. The power spectral density can be reduced to around or below the ambient noise level and are thus made undetectable, except by a properly synchronised receiver. Another implementation, usually called the 'direct' method, modulates the pseudo-random pattern with the data to produce a signal with a much wider bandwidth. That in turn is used to modulate the radio-frequency carrier. The sidebands of the resulting signal extend over a bandwidth of the same order as the pseudo-random signal rather than that of the message.

The message data bit rate is always much less than the pseudo-random pattern rate, which is generally known as the 'chip' rate (the notion being that the message is divided into small chips and each one is sent by a separate 'route'). Recombining all of the little 'chips' recreates the original message. The recombination process involves correlation and averaging, which increases the signal/noise ratio of the hidden message by 3dB for every doubling of the number of chips included. A system with 1000 chips per bit of data would result in an improvement of about 30 dB. That would be sufficient to raise a signal from around the background noise level to a level suitable for binary (two-level) detection.

For video or audio watermarking, a similar process can be used. The message, in this case the watermark data, is used to modulate a much finer pseudo-random pattern. For each bit of data, the pattern may have hundreds, thousands or even tens of thousands of chips. The modulated pattern can be added to the audio or video signal at a very low level, in principle imperceptibly. For audio, the pattern can be added to the signal in the time domain as a noise-like component or added to (or used to modulate) the components in some convenient transform domain (Fourier, DCT, Wavelet, Hadamard, etc.). For video, the pattern can be added to the signal in the 2D spatial domain, the 3D space-time domain or, again, in some convenient transform domain.

The data can be recovered by appropriate filtering of the combined signal. In practice, the combined signal is usually cross-correlated with the known pseudo-random pattern. The cross-correlation function gives information about the phase, amplitude and polarity of the original watermark data (assuming that it is actually detectable). The synchronisation problem is easier than for real-time communications because the whole signal can be stored (though there can little doubt that modern communication systems now exist where the whole spectrum can be stored and processed retrospectively). The cross-correlation process provides the decoding of all possible phases simultaneously. The only secret part required for decoding is the pseudo-random sequence or the seed number and algorithm required to create it. The method itself may be made public.

Many (if not all) video and audio watermarking systems use some form of spread-spectrum modulation. However, it is used simply as a means of embedding the low-rate data in the host carrier. The purpose is to optimise the embedding and detection of the watermark data by matching the bandwidth and noise level characteristics of the two signals. Spread-spectrum modulation is not an intrinsic component of watermarking systems. In all of the detailed discussion of methods that follows in the next section, the embedded watermark data is usually, but not necessarily, the result of some form of spread-spectrum modulation process.

4. Modern watermarking methods for audio.

The principles of commercial watermarking systems are secret and can only be deduced from the promotional material, user parameters and the claimed resistance to attacks. Much more openness is encountered amongst research workers and in the patent literature. However, whether any of that newer research work or the patent claims will survive to commercial exploitation is yet to be determined.

What follows is derived from published descriptions and, especially, the claimed performances of systems. In some cases, the published results of tests on specific commercial systems have put into the public domain significant insights into the operating principles of those systems. It is a very brief summary of a large amount of publicly accessible information, some parts of it more reliable than others. None of it has been obtained from any privileged access to systems or information.

The principles of watermarking can be divided into a small number of categories, based (loosely) on their mode of operation :-

Time-domain methods.

i. Adding noise.
ii. Adding echoes.
iii. Modifying phase.
iv. Amplitude modulation.

Frequency (or, more generally, transform) domain methods.

v. Adding modulated carriers.
vi. Adding noise in a transform domain.
vii. Subtracting frequency bands (adding notches).
viii. Combination of adding notches and then adding carriers.
ix. 'Out-of-band' signals.
x. Modification of the spectral distribution.

Coded domain (MPEG, etc.) methods.

xi. Modifying coding coefficients by biased error distribution.
xii. Adding noise-like signal to coding coefficients.

In all of these methods, account may be taken of the human auditory system (HAS) sensitivities to embed the watermark signal selectively, improving the balance between added watermark energy and perceptibility. It may well be that any practicable watermarking system has to do that to achieve a workable compromise, but it is not a fundamental requirement. Also, some particular transform domain may be considered better for discriminating between perceptible and imperceptible features of the two signals. Transformation to that domain allows that discrimination to be better made. It may then be (and usually is) convenient to embed the watermark data in that domain also. Alternatively, the perceptually-shaped watermark may be transformed back to the time domain and added there as an amplitude component.

Similar considerations are fundamental to the principles of the low-bit rate coding schemes (e.g. MPEG) that are now so widespread and so successful. Watermarking is not very different in that respect. Indeed, some of the most successful watermarking systems work in the same transform domain as the bit rate reduction system. In one sense that is inevitable and unavoidable. One of the substantial conflicts between watermarking and bit rate reduction coding is that the watermark must be at or near the boundary of imperceptibility and, at the same time, the coding system is designed to remove inaudible or just audible features in order to be able to discard data. It is clearly sensible to make the watermarking system aware of the potential effects of coding systems.

As an example of perceptual masking for video watermarking, some consider the wavelet domain to be the best, at least of the readily calculable domains, for modelling the human visual system properties. They argue that working in that domain allows the best discrimination to be made between those visual features that are most important and usually largest in amplitude, where the watermark information can best be hidden.

It is practically axiomatic that steganographic data must be hidden in the parts of the signal that are important to the signal's integrity. Otherwise, it is too easily removed, for example by an appropriately matched filter, without seriously affecting the signal quality. Examples from history are the various kinds of early ‘spoilers’ intended to disrupt unlicensed analogue copying. Most, if not all, were easily removed by simple filters because they worked in parts of the spectrum that could be removed without serious effect on the basic audio quality. Similar examples of copy protection from the video domain involve modification of the synchronising pulses. Because the protection is not applied to the content, it is easily removed without affecting the quality of the content at all. It might be argued that modern digital methods of signal storage and transmission have eliminated the need for watermarking because digital data can be marked using additional data or protected by encryption. Actually, they are the primary reason for the present resurgence in interest. The substantial problem with those digital methods is that they are entirely dissociated from the content. Once removed from the protected environment, as it must be to be used, the mark or protection is lost. The principle of watermarking is that the mark, at least, stays as an integral part of the content.

5. The methods in detail.

This section is intended to give an outline of the various schemes proposed for adding detectable and useful data in an inaudible way to an audio signal. No implication is intended by their inclusion that the schemes actually do or potentially can achieve their objectives. In practice, almost all watermarking schemes so far proposed, tested or evaluated throughout the world fail to be simultaneously inaudible, secure, robust to a large range of attacks and provide a significantly useful hidden channel data rate. There may of course be others, less public.

5.1. Time-domain.

i. **Adding noise.**

The watermark may be added to the audio stream simply as a low-level signal. If the watermark is subjectively not very different from random noise, then it merely increases the effective background noise level of the audio signal. The most common way of implementing such a system is by using spread spectrum modulation. However, other modulation systems are possible.

One scheme, as used in both audio and video MOLE systems, involves least significant bit modulation. On the basis that it is better to sacrifice the signal information carried by the least significant bit in order to avoid the significantly worse effects of further uncontrolled encoding operations, it was used entirely to carry auxiliary data. Such a system can support a high data rate, equal in bits per second to the sampling frequency, but is not at all robust. Such a watermark signal is likely to be seriously compromised by any processing that modified the signal envelope. In practice, it can only be used in digitally coherent systems (those that actually retain the original bits).

ii. **Adding ‘echoes’.**

It has been long established that the human auditory system is insensitive to ‘echoes’, provided they meet certain criteria. This was first established formally by Haas in the early 1950’s [1]. These ‘echoes’ are not, of course, what is commonly understood as real echoes, on a time scale of significant fractions of a second or longer because those are, by definition, audible. Here, ‘echo’ is put in single quote marks (‘’) to indicate this distinction. On a much shorter time-scale, around 20ms, a repetition of the signal is not perceptible provided the amplitude is somewhat lower than the original signal. For the purposes of watermarking, the amplitude of the ‘echo’ amplitude need not be very low, which gives some scope for embedding inaudible watermark data having sufficient energy to be subsequently detectable. Typically, data may be embedded by using an ‘echo’ at, say, 10ms to indicate a binary ‘0’ and at, say, 30 ms to indicate a binary ‘1’. The amplitudes might be between -20 and -40 dB relative to the original signal. The sequence of binary data may be used to carry the watermark data directly at a relatively slow rate or as part of a spread spectrum modulation system at a higher rate.

Detection is usually based on transformation to the Fourier domain and extraction of the impulse response or, more likely, through transformation to the signal cepstrum. The added ‘echoes’ can then be identified and decoded.

iii. **Modifying the phase.**
Audio is said to be phase insensitive, in that the HAS is not aware of either absolute or relative phase of the signal as a whole, or of parts of the signal relative to other parts. This is highly contentious, particularly amongst ‘audiophiles’, but is the basis of at least one current watermarking system.

This method has been included under “Time domain methods” because phase is akin to time, even though the method itself requires the use of Fourier transformation for its implementation. Some may argue with this classification – it is not an important issue.

In the method, the signal is transformed to the Fourier domain in blocks. The relative phases of the Fourier components are modulated to carry the watermark information. Usually, the phase of a reference block is set to a reference value and the code information carried by phase offsets from that reference. The signal is then transformed back to the time domain. Alternatively, the phase differences between components in a single block may be set to specific values. In all cases, the imposed changes in phase must be moderated to reduce audibility.

Detection is by the same sequence of transformations and calculation of the differences between the phases of the Fourier components. The modulation again may be either the raw watermark data or spread spectrum encoding.

iv. **Amplitude modulation.**
In principle, the time domain envelope of the signal may be modified to carry watermark data. One modern method relies on removing very brief segments of the audio signal – on the basis that short breaks are not (very) perceptible.

Alternatively, the overall envelope may be modulated on a longer time-scale. Although no audio system is definitely known to work in this way, a commercial video system intended for the cinema may work by introducing an artificial ‘flicker’ or frame-rate variation in the overall brightness.

It is conceptually possible that an audio system (with necessarily a very low watermark data rate) could work in a similar fashion.

5.2. Frequency (or other transform) domain.

Instead of adding the watermark signal in the original signal domain, it may be added in some transform domain. Almost any transform domain could be used, but most work has concentrated on the common Fourier, DCT, Hadamard or Wavelet transforms, which are variously thought to be either readily computable or offer beneficial mappings to human perception systems or of course both.

It is usually convenient to use a domain in which the properties of the human perceptual system can best be expressed. However, this is not a necessary limitation of the watermark system, though the choice may have significant impact on the instrumental complexity.

The following example methods are presented as modification of the common frequency domain, i.e. the Fourier transform domain, but the descriptions are perfectly general. In some cases, the methods could be implemented by linear filtering, or even as analogue systems. Then the complexities of the digital numerical transformations are avoided. The operational principle might be the same, whether the filtering was carried out with assemblies of copper, iron and polycarbonate or with dsp chips.

v. **Adding modulated carriers.**
At least one modern watermarking system uses the principle of adding modulated carrier frequencies to an audio signal. To be successful, such a system needs to have a system for calculating and making use of psycho-acoustic audibility thresholds. Almost any added tone will

be audible sometimes, if its amplitude is not controlled by some masking threshold function. The tone or tones are additionally modulated to convey the watermark information.

Such a system is simple in concept and capable of being made imperceptible (if the masking calculations can be done with sufficient accuracy) but probably not very secure. It wouldn't take long for an attacker to determine the frequencies and distort them sufficiently to cause a detector to fail.

vi. **Adding noise in a transform domain.**
Instead of a noise-like signal being added in the original signal domain (Method (i)), the watermark can be added in a suitable transform domain. For efficiency, it is likely that the same domain would be used to calculate the perceptibility and masking functions. The method is the same except for the required transformations.

vii. **Subtracting frequency bands (adding notches).**
The idea of removing some narrow frequency bands is a simple and quite old idea. The subjective effects on most types of audio signal are small. However, the potential data rate is relatively low, because audio normally contains periods when some frequency bands are not used.

Such a system would also be quite easy to detect and overcome.

viii. **Combination of adding notches and then adding carriers.**
Methods (vi) and (vii) can be combined. At first, this seems to be a trivial combination. The BBC used the idea at one time to carry the compression data from an audio compressor to a matching expander ("pilot tone compandor system") [2]. In that case, the data carrier was removed afterwards, so there was no requirement for inaudibility, except for the secondary effects of intermodulation distortion in the distribution system. The idea was also used (and perhaps still is being used) in the 'Chipmunk' system, developed in the former USSR.

However, there may be some potential for a modern version. In principle, if the added carriers had similar characteristics to the audio that had originally been in each band then the system would be less perceptible. It would also be very robust, especially to digital coding where the watermark signal would be more likely to be seen as part of the audio signal. In the past that might have been difficult to implement. Today, the method of Spectral Band Replication (SBR) first proposed for Digital Radio Mondiale and now also being incorporated into some MPEG audio standards [3] might form a basis for calculating such apparently similar but computationally distinct signals[2].

ix. **'Out-of-band' signals.**
The addition of signals outside the normal audio spectrum is an old idea. It has been used in the past as a 'spoiler' to interfere with copying using analogue tape recording. The intended principle was that it would cause audible intermodulation components with the bias signal in the record amplifier/head. It never worked in practice and is trivial to overcome. It is completely obsolete.

Any watermarking system based on out-of-band carriers would be similarly trivial to overcome, even if it would pass through modern digital processing systems with their sharp band-limiting filters.

x. **Modification of the spectral distribution.**
Some systems have been proposed that modify the spectral weighting. If carried out in a high resolution Fourier domain, that would be the same as adding noise in a transform domain similar

---

2 Patent applied for, No. 0415750.9, 14th July 2004.

to (vi) above but by multiplication (modulation) rather than addition. If carried out over a broad spectrum, it would be the frequency-domain equivalent of (iv) above, with similar limitations.

5.3. Coded domain (MPEG, AC3, DTS, APT X100, ATRAC, PASC, etc.).

xi. **Modifying coding coefficients by biased error distribution.**
In an audio coder intended for bit-rate reduction the signal is coarsely quantized. That is how the bit-rate reduction is achieved. The decision thresholds for that quantization are based on subjective modelling of the perception thresholds in a number of relatively narrow frequency bands. However, as in any system based on subjectively assigned decision thresholds, the thresholds themselves are not immutable. They are set during system development as compromises between subjective quality and output bit rate targets. It may well be, and usually is, the case that the thresholds can be altered slightly without serious effect on the output quality.

If those decision thresholds were altered in a small but pre-determined way by the watermark signal then the effect on the quality would be small, but those alterations could be detectable in a receiver. At least one current watermarking scheme is based on this approach.

This modulation scheme has the advantages of (a) having the perception model already calculated for the coding function and (b) in principle at least having information about what effect subsequent coding systems might have on the watermarked signal.

xii. **Adding noise-like signal to coding coefficients.**
As an alternative to modifying the coding coefficients by biased error distribution, the watermark data can simply be added to the coefficients, as in any other additive system working in a transform domain.

6. Security

One of the key issues in any kind of covert communication is that of security. The trend in audio and video watermarking is towards a very few, commercial systems. That trend is heavily encouraged by the patent pool arrangements and the modern aggressive approach to patent exploitation. Many companies now see their main source of income being from intellectual property licenses rather than from manufacture. At least two of the most active companies in watermarking are typified by that approach. Indeed, at least one of those has essentially no other source of income. This means that the entire world might ultimately be limited to using perhaps just one or two systems.

That has serious implications for security. It is especially serious if the watermarking system is used to control access to content by means of the hardware in the customer's home. There will be many millions of watermark detectors available to people trying to avoid the control systems. They will be able to experiment with concealment or reverse engineering ('hacking') at a rate limited only by their local hardware.

The SDMI challenge [4] illustrated how systems could be overcome, even without effective access to a detector and even when there was no significant commercial motivation. With available detectors and any degree of profit motive, the situation will only be worse. It seems likely that, for access control at least, watermarking would become ineffective in matter of days from its first release.

The users of such watermarking systems will have to rely entirely on the new, and quite draconian, legal controls over methods of bypassing security systems, as in the "Digital Millennium Copyright Act" (DMCA) in the USA and in the EU Copyright Directive .

For forensic tracing, the situation for watermarking is very much better. Because of the longer integration time, the watermark signal will probably be at a much lower level and the detection process will also be more reliable. The detectors themselves will be few and securely confined to trusted locations/personnel. Overall, it is very easy to envisage widespread use of watermarking for tracing leaks during the production process. In those cases, it may be that more different types of watermarking system will be used, with perhaps some proprietary developments. That will result in greatly enhanced security.

7. Early BBC work.

In the period up to 1971, work was done at BBC Research Department on techniques for hiding audio control signals in conventional sound circuits [5]. The motivation was to try save the cost of separate control circuits for programme distribution. At that time, there was no concept of electronic (or any other sort of) content protection. Indeed, very few throughout the entire BBC thought that content had much value once it had been transmitted (as witnessed by the wholesale destruction of what would now be valuable recordings).

Those ideas are now far outside any patent potential, because of the time lapse and the fact that they have been openly published. However, their publication does represent a robust record of prior art in the field of what is now called audio watermarking. It could give the BBC complete freedom in any attempt to develop, and even license for production, a modern watermarking system based on any of those ideas.

The report came to light recently in response to an enquiry from a developer of modern audio watermarking systems and a current holder of patents in the field, evidently carrying out research into prior art. As the information about this prior BBC art has become more widely disseminated, other 'patent holders' have also been showing interest.

Ref. 5 describes 10 "aural limitations" that might be exploited in embedding hidden signals. It goes on to list 16 "Theoretical Methods" for signalling and explores in detail, including subjective testing of audibility, four likely candidates, for data rates around 50 baud. Some of the ideas were described then as 'not practicable'. However, modern technology, in particular the ability to carry out real-time orthogonal transformations, has made all of them feasible now. The fact that they were described then as impractical does not diminish the claim to 'prior art'.

The following table lists the ideas expressed then, in the words of that time, with some modern comments. The original material is quoted more or less fully to save readers having to cross-reference to the old report (now available from the R&D Web site, http://www.bbc.co.uk/rd/pubs/).

Ref. 5 code letter	**"Aural limitations" from Ref. 5. - monophonic signals assumed.**	**Comments.**
a	**Frequency weighting characteristics of human hearing;** (Fletcher-Munsen audibility curves); the subjective loudness of individual sound components depends jointly upon their intensity and frequency. For example, at very low sound intensities near to the audible threshold, the ear's sensitivity at 50 Hz is about 50 dB less than at the most sensitive frequency (about 2.5 kHz).	By implication, an exploitation of masking, perhaps by the inherent human audibility threshold. However, it is difficult to imagine that as a practicable limitation because it would be dependant on final reproduced loudness.
b	**Insensitivity of the ear to phase information;** (Ohms law of hearing); this is normally understood to imply that the ear can perceive amplitude and frequency but not phase. This formulation must break down if phase changes significantly modify the actual envelope of the signal. For complex sounds, dispersion should probably not be allowed to exceed about 8 ms between any two frequency components; this figure is the presently quoted upper limit in group-delay difference between the maximum usable frequency and the band-centre for high-quality music lines.	Used exactly as described by some modern watermarking systems.

Ref. 5 code letter	"Aural limitations" from Ref. 5. - monophonic signals assumed.	Comments.
c	**Sound masking phenomena;** in its simplest form, this effect shows itself as the suppression of quiet sounds by relatively loud sounds of comparable [similar] frequency. The degree of masking is generally reduced if the frequency separation between quiet and loud components increases.	A direct description of masking, as still understood today.
d	**Subjective tone generation;** the non-linearity of the ear can generate extra subjective sound components which are not present in the input waveform. If the input signal comprises simple tones, the extra components are harmonic tones and also sum and difference frequencies.	Harmonic and intermodulation products created by HAS non-linearity. By implication, these created tones might be used to mask added watermark components.
e	**Theory of missing fundamentals;** this effect is a type of dual of (d) above. If the fundamental frequency component is omitted from a complex sound, the presence of the related harmonics often allows it to be heard subjectively.	Subjective regeneration of missing fundamentals. By implication, these created tones might be used to mask added watermark components.
f	**Time discrimination (Haas effect);** this relates to the 'echo' tolerance of the ear and, although under special experimental conditions, time intervals of about 10 ms can be noticed, it is normally only possible to discriminate the echo if the interval is more than about 25 ms.	A perfect, though very brief, description of systems based on adding 'echoes'. Though the idea was correct, the expression is incomplete. Modern expressions of the effect would also take into account the influence of amplitude, producing a function of audibility versus both time delay and amplitude.
g	**Frequency response tolerance;** the ear can apparently tolerate small perturbations in the amplitude-frequency response of the sound channel but there is little published information on the subject. Intelligibility measurements on speech have revealed that spectral humps reduce intelligibility more than corresponding spectral depressions. Results obtained from work on loudspeakers have shown that variations of about ±2 dB over the audio band can be tolerated; the ear is more sensitive than this, however, in respect of smooth amplitude slopes across the spectrum.	Tolerance to minor alterations to frequency domain response. By implication, this insensitivity to frequency response alteration might be used to carry watermark data.
h	Programme drop-out tolerance; investigations into the audibility of tape drop-outs in magnetically recorded sound signals have shown that regular low-frequency (<2 Hz) attenuation notches of reasonable depth (>20 dB) can be inaudible if their duration time is less than about 1-2 ms. The ear is rather more tolerant than this to isolated random drop-outs.	Tolerance to short breaks in the signal. It is likely that modern quality standards would require rather less disruption for it to be inaudible. At least one modern audio watermarking system works on exactly these principles.
i	**Frequency tolerance;** this is a very small effect because the ear is astonishingly sensitive to frequency changes. The minimum perceptible frequency change varies with sound level and is between 2 and 4 Hz over the lower part of the audio spectrum but beyond about 2 kHz the minimum perceptible shift is an approximately constant fraction of the nominal frequency. (A feature of f.d.m. carrier circuits is that the entire sound spectrum can be subject to small frequency shifts; recent subjective work on this topic has shown that a maximum shift of ±2 Hz is normally acceptable).	HAS acuity in frequency domain. The report acknowledges that HAS is very sensitive to pitch shifts and by implication this is not a usable property, though the necessary accuracy of filtering seen then as a substantial practical difficulty is now achievable using transform methods. Modern digital systems might also distort the spectrum, though more likely by proportional scaling rather than a uniform shift.
j	**Intelligibility of frequency-compressed speech;** work on the analysis and synthesis of speech signals in connection with Vocoders has revealed some interesting possibilities which may be applicable to the present problem. The fundamental principle behind the operation of most Vocoder channels is	A partial description of modern psycho-acoustic coding systems. The terminology is obsolescent and written in terms of speech vocoders, but most of the aspects are there – frequency band division, coarse quantisation, but masking only by

Ref. 5 code letter	"Aural limitations" from Ref. 5. - monophonic signals assumed.	Comments.
	the assumption that the ear behaves as a short-term frequency analyser so the information may be transmitted as packets of 'elementary' time-frequency signals. The characteristics of speech are such that, for acceptable intelligibility, only a limited number of these elementary signals need to be sent; moreover, speech signals may be further compressed by coarse quantisation and, in the limit, infinitely clipped (i.e. unity-bit) signals can be sent without complete loss of intelligibility. These results, however, do not necessarily apply to the other types of programme information.	implication. With the benefit of hindsight, it is a great loss that the significance of this work was not recognised at the time. However, the technology of the time would certainly have limited the potential development . The need for bit-rate reduction was also many years in the future. Otherwise, the BBC could have been the original holder of the core patents on modern compression systems.

The report then goes on to list 16 "Theoretical Methods for Simultaneous Subliminal Signalling". They are presented in Ref. 5 as Table 1. The table identifies the basis of the system (additive or multiplicative), the "Signalling System", the "Modulation Method" and the "Subliminal Basis" (which of the previously identified 10 "Aural limitations" is being exploited). That table is replicated here, with an additional column showing which of the modern methods described above comes closest to the same thing and some comments.

System, Type and number	Signalling system	Modulation	Subliminal basis	Modern systems and comments.
Additive				
1	Low level l.f. tone	pulse or f.s.k.	(a)	(v) The system description appears to assume masking by the auditory threshold. No modern system is likely to work like this because of the sensitivity to gain changes. However, it was reported even in 1971 that such a system was in use (1971, Ref. 15).
2	Low level h.f. tone	pulse or f.s.k.	(a), (d)	(v) As above. Both of these systems are specific examples of the more general class of masking employed in System 3.
3	Masked signals	Analogue or digital	(c)	(All) All watermarking systems have to rely on masking of some sort if they are to be inaudible (that is effectively the definition inaudibility). The accompanying text contains an exact description of the use of masking for both "tone and pulse signals" and modulation of the watermark for perceptual masking.
4	Noise signals	pseudo-random binary	(c)	(i) The accompanying text contains an exact description of spread-spectrum modulation, though not described as such because the term was not then in common use (the first commercial use was more than ten years in the future). The example methodology is, however, exact.
5	Spectrum perturbation of added noise	binary switching	(c)	No exact parallel. It has elements of Methods (v), (vi) (vii), (viii), but the description emphasises the use of modulation of the spectral distribution of the added noise to carry the watermark.
Multiplicative				
6	Signal phase switching	phase modulation	(b)	An essentially exact description of Method (iii). Of course, the concept of real-time Fourier Transforms was then in its infancy (though a reference is given to an internal BBC Research Department Technical Note on the subject).
7	Constant envelope	phase or binary p.c.m.	(j), (b)	(iii) A variation of Method 6, restricted to high frequencies and described differently

System, Type and number	Signalling system	Modulation	Subliminal basis	Modern systems and comments.
				because of implementation practicalities at that time. Today, both would probably be implemented through transform methods and would amount to the same thing.
8	Amplitude drop outs	p.c.m.	(h)	(iv) The final sentence describes the possible (probable) need for error-correcting codes. It also includes the idea of inserting "pulses" in the drop-out period to aid detection and, most importantly, to reduce impairment – a description of Method (viii).
9	Frequency notches	multilevel p.c.m.	(e), (g)	An exact description of Method (vii). Another system based on modification of the Fourier domain.
10	Quantisation switching	binary p.c.m.	(j)	No modern parallel. The idea of switching the quantisation of the signal intermittently to convey information is the same as adding bursts of modulated noise (Method (i)). Of course, switching the quantisation accuracy in narrow frequency bands is the very essence of most modern bit-rate reduction coding systems.
11	Programme frequency perturbation	f.s.k.	(b), (i)	No modern system works by shifting the centre frequencies of the signal components (as far as is known). At the time it was thought to require exceptionally precise analysis. Today, the analysis is trivial but the ear still has very acute sensitivity to frequency shifts, especially inharmonic ones.
12	Artificial reverberation	Binary p.c.m.	(f)	(ii) A precise and detailed description of a modern (and, it is believed, widely used) audio watermarking system. Even some of the numerical values are close to those currently understood to be employed.
Hybrid				
13	Tone-burst synchronisation	Binary	(a), (d)	The idea behind these "hybrid" systems was to combine "additive" and "multiplicative" systems so that one could provide synchronisation for the other. The problem was seen as the interference of the programme material with the detection because of its variability. By using one of the modulation methods to provide 'slots' into which tones or pulses could be inserted, that effect could be overcome. All of these methods are combinations of methods already described, in some cases even the combination has also been described. That does not, of course, detract from their potential viability.
14	Amplitude drop out insert	Binary p.c.m.	(h), (j)	
15	Frequency notch insert	Multilevel p.c.m.	(e), (g), (a)	
16	Quadrature phase signals	a.m., p.m. or p.c.m.	(a), (b)	This system is a variation on the phase modulation concept, (iii). In this proposal, the phase components of the audio signal are converted to "all odd" or "all even" and the resulting 'empty' set used for carrying the watermark data. It belongs under the "hybrid" category because it alters the audio signal to 'make space for' the watermark.

Ref. 5 goes on to evaluate some of the proposals subjectively, using of models of the systems. No work was carried out on actually sending data over the hidden channel. The technology of the day would have made some of the more promising proposals very difficult to implement as functional systems. The objective at the time was stated as "… not to design an actual system, but estimate the maximum subliminal signalling rates of the methods investigated."

8. Conclusions

The two main objectives of this paper were to provide a general background to audio watermarking methods and to summarise the work carried out in 1969/70 at BBC Research Department on sending hidden signals on audio circuits.

The reasons for using audio watermarking, the potential applications and the problems faced by a watermarking system have been described. The paper has also described, in general terms, the weaknesses of the human auditory system that can be exploited for embedding hidden messages.

It concludes that watermarking for access control is likely to face severe difficulties, both technical audio ones and security issues. Watermarking for content tracking is likely to be more successful because it doesn't require the same speed of detection and is less susceptible to damage or decryption.

The earlier BBC work has been analysed and compared to modern methods. It has been shown that the earlier work, though expressed in terms that are at least partly obsolescent now, pre-dated much of the modern approach, some of it close to exactly. At least one modern watermarking method was described then to the extent of some of the appropriate numerical values. In general, the work would probably stand as valid prior art in any application of audio watermarking today.

As an aside, the same earlier work came close to describing modern audio coding systems. However, at the time, the need for that kind of digital audio coding was not present and the available technology would have been a severe limitation.

9. References

1. 1951. Haas, H. "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache;" PhD Thesis, University of Gottingen, Gottingen, Germany. Also published in Acustica **1**, 49-58 (1951) and as "The Influence of a Single Echo on the Audibility of Speech," *J. Audio Eng. Soc.*, Vol. 20 (Mar. 1972), pp. 145-159.

2. 1969. Manson, W.I. and Shorter, D.E.L. "A pcm sound in syncs distribution system: syllabic rate companding. BBC RD Report 1969/48.

3. See, for example "http://www.codingtechnologies.com/products/assets/SBR_White_Paper_v1.pdf"

4. See, for example http://www.cs.princeton.edu/sip/sdmi/.

5. 1971. Hill. P.C.J. "Simultaneous subliminal signalling in conventional sound circuits". BBC Research Department Report No. RD1971/1 (http://www.bbc.co.uk/rd/pubs/reports/1971-01.pdf)