THE QUEEN'S AWARDS FOR ENTERPRISE 2001

BBC

R&D White Paper

WHP 101

December 2004

EBU tests of commercial audio watermarking systems

A.J. Mason

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 101

EBU tests of commercial audio watermarking systems

Abstract

Audio watermarking has recently had a resurgence of interest, spurred on by the desire for copyright protection of digital audio recordings. Several audio watermarking techniques, some dating back more than 30 years, are described briefly here. The uses to which watermarking might be put are also summarised. Attention is then focussed on the requirements identified by the EBU applicable to distribution over the Eurovision and Euroradio networks.

The EBU issued a call for systems to meet its requirements. Subjective and objective tests were done on the systems supplied for testing. Audibility and robustness of the watermarks were measured. The results are encouraging for those considering using audio watermarking in broadcast applications.

This document was originally presented at the 116th AES convention, May 2004. Convention paper 6011.

White Papers are distributed freely on request.

Authorisation of the Chief Scientist is required for publication.

EBU tests of commercial audio watermarking systems

A.J.Mason

BBC Research and Development, Tadworth, Surrey, KT20 6NP, UK
andrew.mason@rd.bbc.co.uk

ABSTRACT

Audio watermarking has recently had a resurgence of interest, spurred on by the desire for copyright protection of digital audio recordings. Several audio watermarking techniques, some dating back more than 30 years, are described briefly here. The uses to which watermarking might be put are also summarised. Attention is then focussed on the requirements identified by the EBU applicable to distribution over the Eurovision and Euroradio networks.

The EBU issued a call for systems to meet its requirements. Subjective and objective tests were done on the systems supplied for testing. Audibility and robustness of the watermarks were measured. The results are encouraging for those considering using audio watermarking in broadcast applications.

1. INTRODUCTION

Watermarking has its origins as a means of applying distinguishing marks to paper, visible in transmitted light. When the term is applied to audio it is generally used to mean the embedding of some information in the audio signal that does not alter the perceived sound but can be detected on demand using appropriate equipment. This is shown very simply in Figure 1.

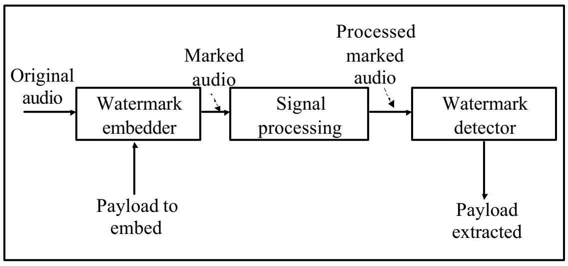

Figure 1 Simplified audio watermarking system

This paper describes briefly how this might be done and some of uses to which watermarking might be put. References are given to documents that provide more detail on the "how" and "why" of watermarking for audio (and video).

Although watermarking can be applied for a wide variety of purposes, the task of copyright protection was the one that triggered the work reported in this paper. Some of the studies that preceded this work are listed to show how watermarking technology, and the understanding of its application in this field, is maturing.

Tests on video watermarking had already been done and so the task of doing a similar job for audio was started. This involved the definition of the technical requirements, calling for systems to be submitted, deciding on test methods, testing the systems submitted, and interpreting the results. These steps are described in detail.

The work was a joint effort by several European Broadcasting Union members participating in two EBU technical project groups. The group N/WTM (watermarking), which became N/DRM-T (digital rights management - technology) initiated the tests, delegating the detailed audio aspects of the work to project group B/AIM (Audio In Multimedia).

2. AUDIO WATERMARKING: HOW

There are numerous techniques available to embed data in an audio signal. As far back as 1971 several had already been developed and evaluated by Hill[1]. Variants of those are still being used today.

Technology limited what Hill could do at that time - echo hiding was possible, but echo finding was altogether more difficult!

Technological advances have led to the enormous increase in the amount of work currently being expended in this field. Microprocessors with a clock speed of 2GHz are readily available on the home computer. This kind of signal processing power has opened up huge possibilities.

Many watermarking schemes are additive in nature. Very low-level tones might be added, their presence, absence, or modulation, being used to convey information. The level at which the tones are added is best governed by their audibility. A psycho-acoustic model can be employed to calculate what will and will not be heard by the human ear. The tones can be added at a level that will be just inaudible, with a margin for errors in the modelling.

One advancement is to add a wide band-width, noise-like signal rather than a number of discrete tones. This leads on to what is known as spread spectrum modulation. Again, a psycho-acoustic model can be used to ensure that the maximum amount of watermark energy can be added with the minimum amount of audible disturbance.

In this kind of scheme the watermark detector needs to know the noise signal that was used in the embedder. This information can be stored as the actual sequence or, more likely, as the seed to be used to start a random number generator. Detection of the watermark is performed by correlation of the watermarked signal with the noise signal. A sufficiently large peak in the correlation function is interpreted as a sign that the watermark was actually there.

In this kind of scheme the seed for the random number generator is often used as a "secret key". The watermark can only be detected by those in possession of the key. The protection of that key is crucial in applications where there are those with malicious intent. This can be quite difficult.

Echo hiding, mentioned earlier, involves the addition of small echoes to the signal. If the echoes are constrained in their amplitude and delay they are inaudible. The echoes are detected by cepstral analysis to extract the data conveyed.

Other techniques operate by controlling quantisation error. The direction of rounding in quantisation can be selected so that the quantisation error is biased in a desired direction. This then is used to add a tell-tale signal to the audio. The added signal can be made to resemble a spread-spectrum sequence as before by suitable alteration of the rounding direction.

As has been said already, there is considerable activity in this field. Several AES convention papers [2][3][4][5][6] have described recent advances. Several authors have also provided very informative works of a more general nature, covering many aspects of watermarking for both audio and video[7][8][9].

3. AUDIO WATERMARKING: WHY

So much for the techniques, but what are the reasons for watermarking? The primary one for many people has been mentioned already: copyright protection. However, there are lots of others. In the field of broadcast there are times when the association of some kind of identifier with an audio clip or programme would be very valuable.

Several places in the broadcast production process where watermarking could play a part have been described in detail elsewhere[10]. Something akin to a bar-code, embedded in the audio means that one can always know what it is, where it came from, where it's going, and so on. Watermarking at acquisition means that every contribution to a programme can be traced to its origin. This might not be relevant to the archivist, who might need to be able to identify the programme as a whole. Other requirements, nearer the distribution end of the chain, might be to identify the network carrying the programme, or, as is the case under discussion here, to be able to identify segments of the content so that the rights applicable to their use can be established and steps taken to correct any mismatch that might be found.

People have been working towards this for several years. Projects sponsored in part by the European

Commission's ACTS programme, such as OCTALIS, and TALISMAN, have produced technology and generic models within which that technology might be used. One such model provided for the use of three watermarks, referred to as W1, W2, and W3. See Figure 2.

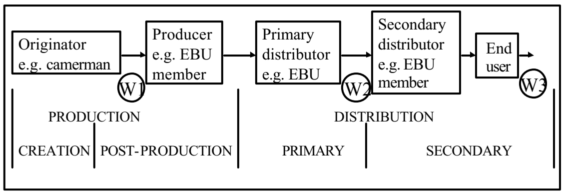

Figure 2 Generic model of watermarking in the broadcast chain

In this model the watermark W1 is used during production to identify content. W2 is applied on the distribution network (possibly at the point of reception in a suitably equipped conditional access receiver). The third watermark, W3, is applied to identify the content as received by the end user. Depending on the way any of W1, W2, and W3 are used they can link the content to its origin or its destination. The latter case has sometimes been called "fingerprinting" because it identifies individual copies of the same content distributed to different places. Unfortunately, the term "fingerprinting" is also used to refer to content recognition by feature extraction, and so is perhaps best avoided.

An EBU technical project group, N/WTM, taking this model into account, focussed its attention on W1 and W2[11][12]. The main area of interest was in the protection of copyright of television and radio programmes being distributed over the Eurovision and Euroradio networks. The intent was not so much to enforce copy-control functions at a device level, but rather to enable monitoring and tracing of the use of programme material. Technical requirements were specified and video watermarking systems were tested in due course[13].

The logical extension of the video work was into audio. Some might argue that the order was wrong: radio and music thrive whilst silent movies have had their day!

4. THE AUDIO WATERMARK TESTING PROCESS

The process started with the drafting of the technical requirements of an audio watermarking system intended for the application already outlined. The other potential applications were borne in mind, so that the requirements might be kept somewhat generic, and so the results might be seen as relevant to a wider audience.

A call for systems was made, incorporating the technical requirements, originally in June 2001. The call was revised and made again in February 2002 after problems were found with the submissions received in response to the first call.

The systems that were received were subjected to two tests in parallel. One test measured the audibility of the watermarks, the other their robustness (that is, their resistance to being removed or rendered undetectable by signal processing). The audibility was assessed by subjective tests. The robustness was measured objectively by playing processed watermarked audio into the detectors.

4.1. The EBU's technical requirements for audio watermarking

Watermarking systems can be characterised by three parameters: data capacity (D), imperceptibility (I), and robustness (R). In general, these are inter-dependent and fixing any two of the three fixes the third. Conversely, if one desires to change one of the parameters one must also change (at least) one other.

In mathematical form this can be expressed succinctly as:

$$f(D, I, R) = 0 .$$

The main technical requirements specified by the EBU for data capacity, imperceptibility, and robustness are summarised in Table 1.

A few other robustness requirements were listed in the call for systems, but were not tested due to the finite resources available.

A misunderstanding became apparent in the interpretation of the watermark minimum segment. Two different interpretations were made. One interpretation leads to embedding each watermark in two successive 2.5s segments of audio enabling detection from an arbitrarily chosen 5s of audio. The other leads to embedding a watermark once in a 5s segment enabling detection from an arbitrarily chosen 10s segment.

Audibility	Inaudible in studio and domestic conditions
Payload: Watermark minimum segment (WMS) length Data Capacity	 5s 48 bit/WMS
Detection probability	95%
False positive probability/WMS	10^{-8}
Purpose	Identification
Implementation	Real-time hardware
Maximum embedding delay	80ms
Robustness to low bit rate coding: MPEG Layer II MPEG Layer III MPEG AAC Minidisc Dolby AC-3 NICAM	 128kbit/s stereo 96kbit/s, 64kbit/s 32 kbit/s stereo 128 kbit/s stereo 728kbit/s
Robustness to production processes: Multi-band dynamic range compression Addition of voice-over Pitch-corrected time-scaling Linear speed change without pitch correction Addition of white noise Digital to analogue to digital conversion	 ±6dB +15dB ±5% ±10% -30dB
Watermark specific properties: Addition of second watermark Detection of first watermark after addition of second	

Table 1: Summary of EBU technical requirements of an audio watermarking system

4.2. Systems submitted

Three companies submitted systems for test. One of these failed to clear the "first hurdle" of informal (but expert) listening and a simple robustness test. It was not included in the exhaustive testing. The two systems that were thoroughly tested were supplied by Fraunhofer IIS and Philips.

Fraunhofer IIS supplied a system based on a Linux PC with a digital sound card. The embedder read payloads from its serial port for embedding and wrote detected payloads to a data file. Operation of the system was controlled by a simple, text-based, user interface. Two variants of the watermark detector software were provided, referred to as FhG System A and FhG System B. The user interface provided the means to run either of the detectors or the embedder.

Philips supplied a system based on Windows 2000 PCs. Two PCs were provided so that embedding and detection could be done simultaneously. Either PC could run as embedder or detector. Separate Windows programs were provided for embedding and detecting. The embedder read payloads from the serial port and the detector wrote detected payloads to the serial port as well as to a data file.

The user interfaces for both systems provided functions to select things like sample rate, mono or stereo input, and input from file.

4.3. Robustness testing method

The means by which robustness was tested is relatively simple. In essence it has already been depicted in Figure 1. A selection of test material (listed in Table 2) had a series of pseudo-random 48 bit payloads embedded into it at intervals of the watermark minimum segment. Copies of the watermarked test material were then subjected to a variety of common signal processes. Some processes were single operations, such as a simple MPEG Layer II encode and decode. Others were compound operations, involving, for example, a change of sample rate, tape recording, then transmission by one means or another.

No	Description of item	Length
1	Stereo channel ident followed by 10ms square-wave burst time-alignment	30s
2	English male speech, BBC Radio Acoustics Test CD, G2CD 01, track 22	3m 40s
3	Bass synthesizer pop music, BBC Radio Acoustics Test CD, G2CD 01, track 41	1m
4	Jazz, BBC Radio Acoustics Test CD, G2CD 01, track 31);	3m 40s
5	Classical music concert with spoken introduction minutes, recorded at BBC Broadcasting House, London	11m
6	Square-wave burst	10ms

Table 2: Test items used for robustness measurement

The watermark embedding was done at BBC R&D, the output of the embedders being recorded on Tascam DA-88 then transferred to a PC. Watermarked audio files were then transferred over the internet to the Insitut fur Rundfunktechnik (IRT) and Telewizja Polska (TVP) for processing. Processed files were transferred back for replay into the watermark detectors at BBC R&D.

In addition to a continuous replay of the processed watermarked into the detectors a segmented replay was also performed. This was used to verify the watermark minimum segment requirement. Noise signals (with known watermark payloads embedded) were added and interspersed with the processed test item, as shown in Figure 3.

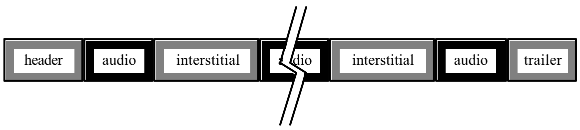

Figure 3: Segmentation and insertion of interstitial noise in audio sequence

This had the useful benefit of verifying that a signal had indeed been correctly played into a detector in the event that the processing to which the test item had been subjected had removed all watermarks.

The segmentation was done by automated generation of edit decision lists (known as "session files") for use by the popular PC editing software Cool Edit Pro.

It was the segmentation that revealed that different interpretations had been put on the watermark minimum segment requirement by the two system proponents. This was something of a disappointment because testing was well underway before this became apparent. It was eventually decided to ask the two system proponents to provide another version of their software with the alternative watermark minimum segment length. This necessitated two passes of robustness testing. To do otherwise would have resulted in a tendency for unfair comparisons to be made between results corresponding to systems with payload capacities differing by a factor of two.

The payloads detected were collected either by data logging software running on a computer connected via RS-232 to the Philips detector, or by floppy disc transfer of the Fraunhofer IIS log files. Software was written to analyse the logs of payloads, comparing them with the known sequence of pseudo-random payloads originally embedded. The robustness of a watermarking system to a particular process was expressed as the number of payloads correctly recovered as a percentage of the maximum number possible using the watermark minimum segment length.

4.4. Robustness results

As has already been stated, circumstances required that two phases of robustness testing be conducted. These correspond to watermark minimum segment length of 5s and 10s. Due to limited effort being available to conduct further tests, not all processes were repeated with the alternative watermark minimum segment systems. In the following tables entries have been left blank where results are not available for this reason.

One of the detectors was found to be unreliable in some circumstances. The entry "-" is made in the tables where it was not possible to derive a meaningful measurement from the detections it made.

Table 3 shows the results of the systems using a 5s watermark minimum segment length. These are the Philips system originally supplied and the modified Fraunhofer system. Table 4 shows the results for the systems using a 10s watermark minimum segment. These are the Fraunhofer system originally supplied and the modified Philips system.

In each table three columns of results are from continuous replay of the test material into the detectors and three are from the segmented replay. The two Fraunhofer detectors are referred to in the tables as “FhG A” and “FhG B”.

The general trend of the results is largely as was expected. Processes that significantly degrade the audio quality, coding at 32 kbit/s using MPEG AAC for example, remove most of the watermarks.

Processes that do not appear to degrade the audio quality can be unpredictable. Linear time stretching (changing the sampling rate of a recording but replaying at the original rate) can sound perfect but in some cases removed all watermarks.

It is apparent that the increase in watermark minimum segment length from 5s to 10s makes the systems more robust.

	Continuous replay			Segmented replay		
Process	FhG A bis	FhG B bis	Philips	FhG A bis 5s	FhG B bis 5s	Philips 5s
No attack	98.0	98.0	73.4	-	90.7	89.1
minidisc (BBC)	95.6	97.6	7.7	-	85.9	14.5
minidisc (TVP)			0.0			0.0
Dolby AC-3, 128kbit/s, stereo			0.0			0.0
MPEG Layer II, 128kbit/s, joint stereo	97.2	96.4	0.0	-	86.3	0.0
MP3, 96kbit/s, stereo	90.3	91.9	0.0	-	68.1	0.0
MP3, 64 kbit/s, stereo	56.5	42.7	0.0	-	15.3	0.0
AAC 32kbit/s, stereo	4.8	0.0	0.0	0.0	0.0	0.0
MPEG Layer II, 32 kbit/s, mono			0.0			0.0
linear time stretch, (10%)	0.0	97.2	0.4	0.0	0.0	2.6
pitch-corrected time-stretch, 5%	4.0	1.6	18.1	2.8	2.0	1.2
voice-over, +15dB			0.0			0.0
added white noise, -30dB			0.0			0.0
dynamic range compression	98.0	98.0	75.0	-	92.7	60.1
analogue conversion			0.0			0.0
combined audio processing			0.0			0.0
Broadcast chain 1, FM	79.4	63.7	0.0	32.7	52.0	0.0
Broadcast chain 2, NICAM			0.0			0.0
Broadcast chain 3, NICAM + MPEG	96.0	96.0	0.0	64.1	82.3	0.0
Broadcast chain 4, Dolby E+AC3			7.3			19.4
first watermark detection after application of second watermark	98.0	98.0	39.9	-	90.3	63.3
second mark detection	98.0	98.0	41.1	-	93.5	39.5

Table 3: Percentage of marks recovered with 5s watermark segments:
FhG system “bis” from second phase of tests, Philips system from first phase of tests

	Continuous replay			Segmented replay		
Process	FhG A	FhG B	Philips bis	FhG A 10s	FhG B 10s	Philips bis 10s
No attack	95.2	94.8	97.6	-	91.2	96.0
minidisc (BBC)	93.5	94.4	96.0	84.6	91.2	94.4
minidisc (TVP)	93.5	94.0		-	88.8	
Dolby AC-3, 128kbit/s, stereo	92.7	93.7		-	88.0	
MPEG Layer II, 128kbit/s, joint stereo	94.4	93.1	96.8	-	89.6	91.9
MP3, 96kbit/s, stereo	91.5	93.5	62.1	58.8	89.6	57.3
MP3, 64 kbit/s, stereo	73.8	75.0	6.5	-	63.8	8.1
AAC 32kbit/s, stereo	7.7	0.8	0.0	-	0.0	0.0
MPEG Layer II, 32 kbit/s, mono	0.0	0.4		-	0.0	
linear time stretch, (10%)	0.0	84.6	95.2	0.0	0.0	43.5
pitch-corrected time-stretch, 5%	0.4	0.8	0.0	1.6	0.0	0.0
voice-over, +15dB	0.8	0.0		-	0.8	
added white noise, -30dB	0.8	0.0		0.8	0.8	
dynamic range compression	95.2	95.2	96.0	60.4	92.0	94.4
analogue conversion	92.3	91.5		91.2	88.8	
combined audio processing	94.8	94.8		-	92.8	
Broadcast chain 1, FM	87.5	86.3	57.3	83.0	80.6	58.9
Broadcast chain 2, NICAM	91.5	90.3		78.2	90.4	
Broadcast chain 3, NICAM + MPEG	90.7	90.7	91.1	61.2	69.4	71.8
Broadcast chain 4, Dolby E+AC3	93.1	89.5		-	86.2	
first watermark detection after application of second watermark	95.2	94.4	93.5	-	88.8	96.8
second mark detection	94.0	95.6	96.8	-	95.2	95.2

Table 4: Percentage of marks recovered with 10s watermark segments:
FhG system from first phase of tests, Philips system “bis” from second phase of tests

The figures in the tables above relate to the failure of the detectors to find a watermark that was known to be embedded. This is a "false negative" detection. A "false positive" detection is one where are watermark is reported as present when one was not in fact embedded. The probability of this latter case happening should be very low. This is because of all the trouble that can be caused by false accusations of misuse!

The Philips systems (both watermark minimum segment variants) did not suffer from any false positive detections. Bearing in mind the vanishingly small probability listed in the requirements this is to be expected.

The FhG-IIS systems with a 5s watermark minimum segment did suffer from some false positive detections. Although watermarks had been embedded, the payloads reported by the detector in a small number of instances did not match any of those that had been embedded. On closer inspection the errors appeared to be due to some kind of combining of successive payloads. In 8 of the 52 tests performed, the FhG-IIS detector reported one erroneous payload (0.4% of the number embedded), in one test there were two, and in another three, erroneous payloads.

The FhG-IIS systems with a 10s watermark minimum segment did not report any false positive detections.

4.5. Subjective testing method

The assessment of audibility of the watermarking systems was done using subjective tests. Because it was not known in advance how audible the systems would be it was decided to add a forced choice to a normal ITU-R BS.1116 test method[14]. In a BS.1116 test the subject is presented with three versions of each test item. One of the versions is known to be the original. The other two versions are randomly assigned to the processed (in this case watermarked) item and the original again. The subject is required to give a grade indicating an opinion of subjective quality to the second and third versions. The forced choice was that each subject had to nominate one of the versions to be graded as being watermarked. This had to be done even if they would otherwise have given a grade of 5.0 (indicating no perceptible difference from the known original) to both.

Test items were selected by expert listeners at IRT and BBC, using the normal method described in the BS.1116 recommendation. A wide variety of audio items were watermarked by the systems provided and a subset was chosen by listening to these. To make the selection process easier, and to provide a means for training listeners, the system proponents had been asked to provide a "boosted" mode for their embedder. This boosted mode would cause watermarks to be embedded at an artificially high level so that they would be intentionally audible. Where this facility was not provided by the system proponent boosted signals were created artificially by calculation of difference signals and adding amplified versions of those to the test items.

Ten items were used with normal level watermarks in the tests. These are listed in Table 5. A small number of other items with boosted level watermarks were used for training the subjects and to provide something in the tests that they should have been able to detect. This latter point can become significant, not because of the need to verify listener reliability, but to help prevent listeners from becoming demoralized by being unable to hear any impairments in any stimuli.

No.	Item type (and source)
4	xylophone (EBU SQAM, track 41)
8	flute (EBU SQAM, track 13)
9	glockenspiel (EBU SQAM, track 35)
13	triangle (EBU SQAM, track 32)
16	violin (EBU SQAM, track 8)
27	harpsichord (EBU SQAM, track 40)
31	English male speech (EBU SQAM, track 49)
34	German male speech (EBU SQAM, track 54)
36	wind ensemble (EBU SQAM, track 36)
39	Tennis (BBC)

Table 5: Test items used in test sessions, with watermark at normal level

Subjective test sessions were conducted at IRT, France Telecom and BBC R&D. Test items were distributed to the test sites over the internet. All sites used the same subjective test software (CRC's SEAQ software). Control files for the SEAQ subjective test software were shared between the sites to ensure consistency.

Listening was done on Stax headphones (SR-404, SR-303, or SR Lambda Pro). Quiet listening rooms were used at all sites. A variety of high quality digital to analogue converters and digital sound cards were used.

The same instructions were given to all subjects at all three sites. These followed the pattern for BS.1116 tests of "basic audio quality", with an addition relating to the forced choice. The subjects were told to give a grade no higher than "4.9" to one of the stimuli to be graded. The grades of "4.9" were converted to "5.0" for the calculation of diff-grades but were used as an indication of the best guess of the subject as to which was the watermarked stimulus.

Every subject was given a training session where they could become familiar with the operation of the test equipment and with the nature of the impairments from the boosted watermark items.

Results from 25 subjects were analysed. Three analyses were performed. The first was the usual calculation of mean grade and 95% confidence interval given by the subjects to each test item. The second analysis was a Wilcoxon rank sum test on the grades for each item to establish whether the distribution of grades given by the listeners differed significantly from what would be expected if no watermark were audible. The final analysis was of the forced choice data to find out whether subjects could reliably identify the watermarked items or whether they were simply guessing.

These results are presented in the next section.

4.6. Subjective test results

The mean diff-grades and 95% confidence intervals for the two systems tested are shown in Figure 4 and Figure 5.

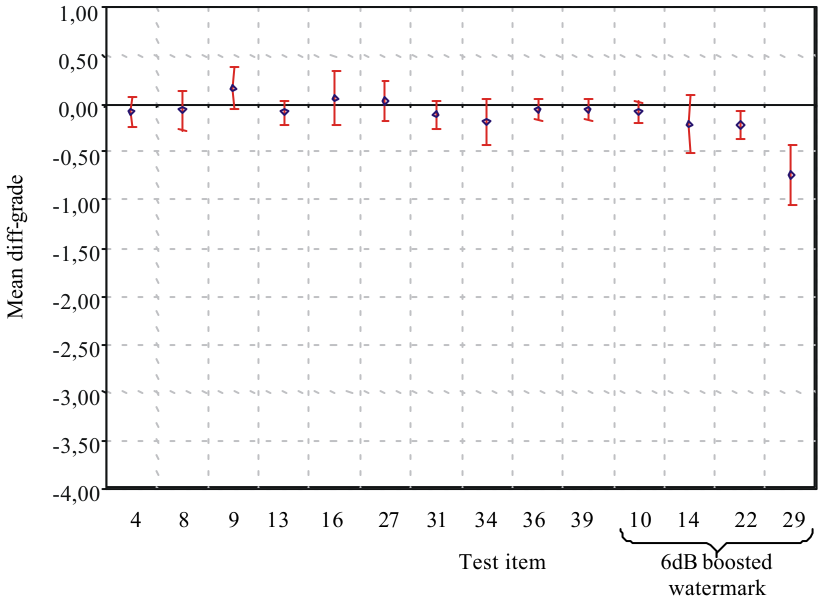

Figure 4: Mean subjective diff-grades and 95% confidence intervals for Fraunhofer IIS watermarking system

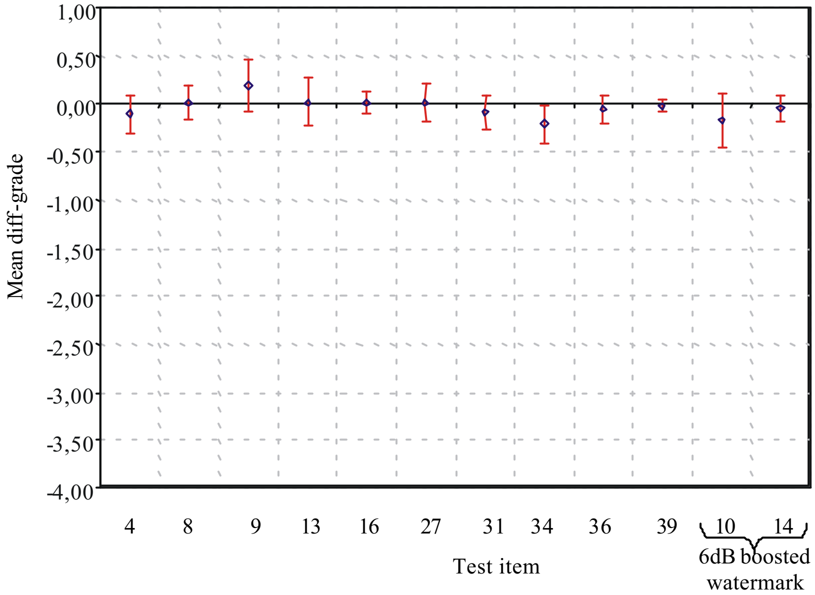

Figure 5: Mean subjective diff-grades and 95% confidence intervals for Philips watermarking system

For the Fraunhofer IIS system the mean diff-grades are close to zero with 95% confidence intervals crossing the zero axis with the exception of the boosted level items.

For the Philips system the mean diff-grades are close to zero and only one 95% confidence interval does not cross the zero axis (test item 34, normal level watermark).

The Wilcoxon rank sum test on the distributions of diff-grades gives very good agreement with the 95% confidence intervals: only one distribution, that of the grades given to item 34 with the Philips normal level watermark, show a statistically significant difference from what would be expected from an unwatermarked signal.

The watermark recognition rates calculated from the forced choice data are shown as a function of subject and of item. They incorporate 25 subjects and 10 test items. Figure 6 and Figure 7 show the results for the Fraunhofer IIS system. Figure 8 and Figure 9 show the results for the Philips system.

As is evident there are wide variations in the recognition rates of individuals. However, what is not clear is whether some people are naturally lucky or unlucky. The recognition rates as a function of test item show rather less variation.

The overall recognition rates were 46.8% for the Philips system, and 56.0% for the Fraunhofer IIS system.

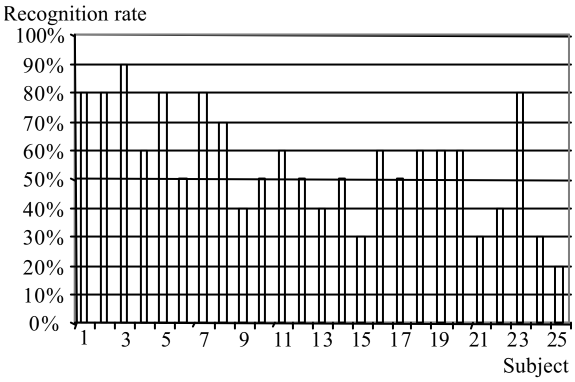

Figure 6: Recognition rate of Fraunhofer IIS watermarking system by subject

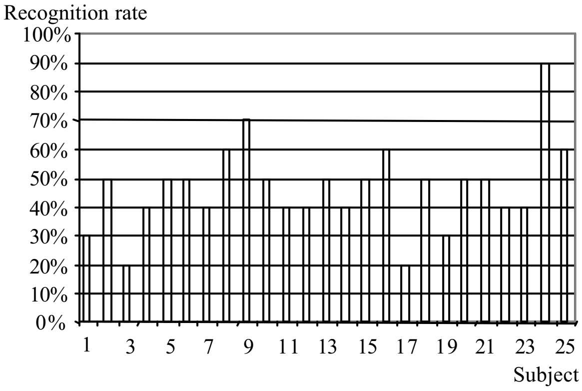

Figure 8: Recognition rate of Philips watermarking system by subject

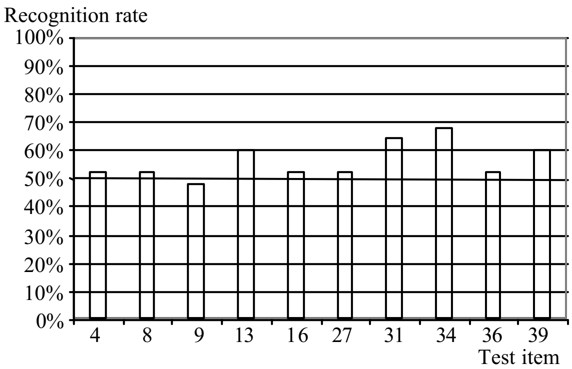

Figure 7: Recognition rate of Fraunhofer IIS watermarking system by item

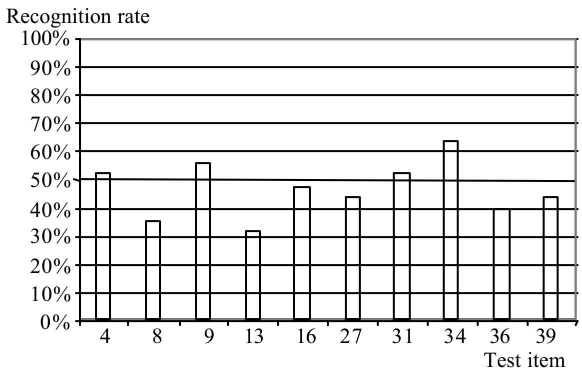

Figure 9: Recognition rate of Philips watermarking system by item

5. CONCLUSIONS

Based on several years of study of the evolution of techniques of watermarking systems, the EBU identified a set of technical requirements for audio watermarking systems that would be usable in several applications, notably copyright protection on the Eurovision and Euroradio distribution networks. The requirements specified audibility, robustness, and data capacity.

Calls for systems were issued so that manufacturers of audio watermarking technology could offer systems for testing, according to the requirements. Few companies felt able to submit systems for testing. In the event, only two of the three systems submitted were deemed sufficiently good to merit inclusion in the full set of tests. These systems were submitted by Fraunhofer IIS and Philips. Fraunhofer IIS submitted two variants of their detector, referred to as FhG-IIS System A and FhG-IIS System B.

Due to a misunderstanding, two phases of robustness tests had to be conducted. Watermarks with minimum segment lengths of 5s and 10s were tried. Fraunhofer IIS and Philips submitted alternative configurations of their systems for this.

5.1. Audibility

Subjective performance was measured using BS.1116 listening tests, into which was incorporated a forced choice. These tests were conducted on the Fraunhofer IIS 10s system and the Philips 5s system. The clear audibility of the third company's system on several test items was such that it was thought not suitable for broadcast applications and was therefore not included in the subjective tests.

The items watermarked by the Fraunhofer IIS system had mean grades with 95% confidence intervals that all crossed the zero diff-grade axis. As such it cannot be said that it was possible for the subjects to distinguish these from the originals. Only one of the items watermarked by the Philips system (test item 34, German male speech, EBU SQAM disc) had a confidence interval that did not quite cross the zero diff-grade axis. This item can be said to show a difference from the original.

The analysis of the distribution of the grades using the Wilcoxon rank sum test correlated well with this. The distribution of the grades given to item 34 watermarked with the Philips watermark showed a statistically significant difference from the original.

It must be said that none of the mean diff-grades for either the Fraunhofer IIS or the Philips system were lower than -0.25 and none of the 95% confidence intervals extended lower than a diff-grade of -0.5.

The analysis of the data obtained from the forced choice that subjects were required to make could benefit from more study. The percentage of recognitions of the watermarked items were 56.0% for Fraunhofer IIS system and 46.8% for the Philips system. These appear quite close to the 50% that would be expected by guessing. However, some individuals scored very high recognition rates (90%) and others scored very low recognition rates (20%).

5.2. Robustness

Robustness of the watermarks when subjected to a wide variety of processes was measured. Watermarks were embedded using 5s and 10s segments settings into a selection of test material totalling about 20 minutes. Copies of the watermarked material were then subjected to several different processes before being played into the detectors. The test sequence was played in two ways: uninterrupted, and then broken into isolated 5s and 10s segments.

5.2.1. Tests using 5s watermark minimum segment

The 5s watermark segment length tests revealed substantial differences between the systems. Both of the FhG-IIS systems showed good robustness to mild attacks when unsegmented signals were replayed. Some attacks, for example AAC at 32kbit/s, removed most of the watermarks. The Philips system, with a 5s watermark minimum segment length, was only robust to a very few attacks.

When the test item was replayed as isolated 5s segments the two FhG-IIS Systems behaved quite differently. FhG-IIS System A was unreliable with these segmented signals. FhG-IIS System B was still robust to mild attacks, but the detection rate fell significantly as the severity of the attack increased. For example, MPEG Layer III at 96 kbit/s left 68.1% of the watermarks, but at 64kbit/s only 15.3% remained. The segmentation did not affect the results for the Philips system: it was still not very robust to most attacks.

5.2.2. Tests using 10s watermark minimum segment

Tests conducted using a 10s watermark minimum segment again showed significant differences between the systems. As an example, the FhG-IIS Systems had a detection rate of 0.0% for segmented time-stretched material while the Philips system detected 43.5% of its watermarks. On the other hand, segmented signals that had been coded with MPEG Layer III at 64 kbit/s left 63.8% of watermarks for the FhG-IIS System B detector, but only 8.1% for the Philips detector. FhG-IIS System A was again found to be unreliable with segmented signals.

The Philips and Fraunhofer IIS systems performed similarly given mild attacks on unsegmented signals using a 10s watermark minimum segment. More severe attacks were resisted better by the Fraunhofer IIS systems.

The Fraunhofer IIS 10s and 5s systems and the Philips 10s system showed useful robustness. They were quite robust to processes that resulted in audio of high quality. Processes that significantly degraded the audio quality by low bit rate coding were found to reduce the watermark detection rate significantly. The Philips 5s system was found to be not very robust. The FhG-IIS System A detector was found to be unreliable.

Overall, the conclusions that we can draw from these tests are that it is possible to embed watermarks in audio without significantly affecting the audio quality. The watermarks can carry a data payload that could be useful for a variety of applications. The watermarks can be recovered after signal processing that might be encountered during broadcast production and transmission, provided that the audio quality is not significantly degraded.

6. ACKNOWLEDGEMENTS

The audio watermark testing presented here was done by members of the EBU project group B/AIM. Particular thanks are due to IRT, TVP, and France Télécom for the parts they played in signal preparation and subjective testing.

Thanks are also due to the system proponents for providing hardware for the tests, and for their patience and cooperation during the extended testing procedure.

I would also like to thank the BBC for permission to publish this paper.

7. REFERENCES

[1] Hill, P.C.J., "Simultaneous subliminal signalling in conventional sound circuits: a feasibility study," BBC Research Department Report, Number 1971/1

[2] Kalker, Oomen, Lemma, Haitsma, Bruekers, and van der Veen, "Robust, multi-functional, and high quality audio watermarking technology," presented at the 110th AES convention, May 2001. Preprint 5345.

[3] Neubauer and Herre, "Digital Watermarking and Its Influence on Audio Quality," presented at the 105th AES convention, September 1998. Preprint 4823.

[4] Garcia, "Digital Watermarking of Audio Signals Using a Psychoacoustic Auditory Model and Spread Spectrum Theory," presented at the 107th AES convention, September 1999. Preprint 5073

[5] Neubauer and Herre, "Advanced Watermarking and Its Applications," presented at the 109th AES convention, September 2000. Preprint 5176

[6] Neubauer, Siebenhaar, Herre, and Bauml, "New High Data Rate Audio Watermarking Based on SCS (Scalar Costa Scheme)," presented at the 113th AES convention, October 2002. Preprint 5645.

[7] Cox, I., Kilian, J., Leighton, T., Shamoon, T, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Trans. Image Proc. V6 No. 12, December 1997.

[8] Peticolas, F., Anderson, R., Kuhn, M., "Information Hiding - A survey," Proceedings of the IEEE, special issue on protection of multimedia content 87(7), July 1999.

[9] Cox, Miller, Bloom, "Digital Watermarking," Morgan Kaufmann, 2002

[10] Mason, A.J., Salmon, R.A., Devlin, J.E. Werrner, O.H., "User requirements for watermarking in broadcasting applications," presented at the International Broadcasting Convention, Amsterdam 2000.

[11] Barda and Cheveau, "Eurovision network security through access control and watermarking," EBU Technical Review No. 281, Autumn 1999.

[12] Barda and Cheveau, "Access control and watermarking," EBU Technical Review No. 282, March 2000.

[13] Cheveau, Goray and Salmon, "Watermarking - summary results of EBU tests," EBU Technical Review No. 286, March 2001.

[14] ITU-R BS.1116-1, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,” International Telecommunications Union, Geneva, Switzerland.