DIGITAL AUDIO WATERMARKING FOR BROADCAST MONITORING AND CONTENT IDENTIFICATION

Ron Healy B.Sc.

A THESIS SUBMITTED FOR THE QUALIFICATION OF
MASTER OF SCIENCE IN COMPUTER SCIENCE

National University of Ireland, Maynooth
Ollscoil na hÉireann, Má Nuad

Department of Computer Science

OCTOBER 2009

HEAD OF DEPARTMENT: DR. ADAM WINSTANLEY, PH.D.
RESEARCH SUPERVISOR: DR. JOSEPH TIMONEY, PH.D.
MR. TOM LYSAGHT.

Declaration:

I hereby certify that this material, which I now submit for assessment on the program of study leading to the award of Master of Science in Computer Science by Research, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:____

Dated:____

Contents

List of figures

Abstract

Copyright legislation was prompted exactly 300 years ago by a desire to protect authors against exploitation of their work by others. With regard to modern content owners, Digital Rights Management (DRM) issues have become very important since the advent of the Internet. Piracy, or illegal copying, costs content owners billions of dollars every year. DRM is just one tool that can assist content owners in exercising their rights. Two categories of DRM technologies have evolved in digital signal processing recently, namely digital fingerprinting and digital watermarking. One area of Copyright that is consistently overlooked in DRM developments is 'Public Performance'.

The research described in this thesis analysed the administration of public performance rights within the music industry in general, with specific focus on the collective rights and broadcasting sectors in Ireland. Limitations in the administration of artists' rights were identified. The impact of these limitations on the careers of developing artists was evaluated.

A digital audio watermarking scheme is proposed that would meet the requirements of both the broadcast and collective rights sectors. The goal of the scheme is to embed a standard identifier within an audio signal via modification of its spectral properties in such a way that it would be robust and perceptually transparent. Modification of the audio signal spectrum was attempted in a variety of ways. A method based on a super-resolution frequency identification technique was found to be most effective. The watermarking scheme was evaluated for robustness and found to be extremely effective in recovering embedded watermarks in music signals using a semi-blind decoding process. The final digital audio watermarking algorithm proposed facilitates the development of other applications in the domain of broadcast monitoring for the purposes of equitable royalty distribution along with additional applications and extension to other domains.

Dedication

This work is dedicated to Ann Healy. It would not have been possible to even begin to contemplate a return to full time third level education without the tacit and express support of Ann and the rest of the family. In the ten years since I returned to education ***we*** have achieved many qualifications. The knowledge that Ann would be happy if I achieved the status of '*the most highly educated coach driver in the Country*' provided the support that allowed me to afford as much concentration as possible to the requirements of my B.Sc. Degree and then my M.Sc. research without fear of failure.

The biggest obstacles faced, as is the case with many students, are generally motivational and financial. In my case, with no grant aid or other funding over the three years of this M.Sc. research, I was nevertheless able to spend only half of each week working and the other half studying because the 'auditor general' in our house ensured that whatever finances were available were used to best advantage. At no time was there every any financial pressure to cease studying and work full time. This drastically reduced a whole host of financial pressures and considerations.

Acknowledgements

I would like to acknowledge the consistent and patient support of my Supervisor, Dr. Joseph Timoney Ph.D., at the Department of Computer Science in N.U.I. Maynooth. Dr. Timoney was almost solely responsible for ensuring the work did not come to a permanent halt on more than one occasion. At points during the research that led to this conclusion, his contention that the end result was not only worth the undertaking but, more importantly, also within my capabilities was encouragement enough to persuade me to continue when I might otherwise have given up.

I would also like to acknowledge the contribution of Jian Wang, postgraduate researcher at the Department of Computer Science in N.U.I. Maynooth for his assistance over the final phase of this work when the limitations in my technical ability in mathematics and signal processing were exceeded. Similarly, I wish to acknowledge the assistance of Jonathan Lambert, Dr. Conor Mc Elhinney and the numerous staff and postgraduate researchers at NUI Maynooth who assisted along the way with my understanding of technical concepts as well as the world of Academia.

I would also like to thank the Head of the Department of Computer Science at N.U.I. Maynooth, Dr. Adam Winstanley, for authorising funding that afforded me the privilege of attending various Academic Conferences through the final year of my research in order to present my work and elicit reaction to it. Any reaction I *did* receive was favourable and these reactions from individuals with no connection to my work or the Department encouraged me to believe that my work was both technically competent and, more importantly, practical and worthwhile. Without this extremely important financial support from the Department, the publication of academic papers at these Conferences would have been financially unjustifiable. The difference that these Conference attendances made to my personal motivation was dramatic and immediate.

Finally, I would like to acknowledge the support of my Employer, Sean Reid, and my Supervisor, Mary Maguire, for allowing me the flexibility in my work arrangements for the duration of this research. This enabled me to prioritise my studies, particularly at times when they demanded most of my attention and concentration.

Previous publications

Healy, R., and Timoney, J., ‘An Application of Digital Audio Watermarking to Broadcast Monitoring’, Paper accepted for publication at the *China-Ireland International Conference on Information and Communications Technologies*, Beijing, China, September 2008.

Healy, R. and Timoney, J., ‘300 years of copyright: have we gone full circle? On the use of technology to address limitations in distributing public performance broadcast royalties’, *Socio-Legal Studies Association Annual Conference*, UK, April 2009.

Healy, R. and Timoney, J. ‘Digital Audio Watermarking with Semi-Blind Detection for In-Car Music content Identification’, *Proceedings of the AES 36th International Conference: Automotive Audio—Sound in Motion*, Dearborn, Michigan, USA, June 2009.

Healy, R. and Timoney, J. ‘Limitations in the Distribution of Public Performance Royalties in Ireland’. *Conference on Intellectual Property*, New Rochelle, USA, June 2009.

Wang, J., Healy, R. and Timoney, J. ‘Digital Audio Watermarking by Magnitude Modification of Frequency Components Using the CSPE Algorithm’. *Proceedings of the China-Ireland International Conference on Information and Communications Technologies*, NUI Maynooth, Ireland, August 2009

Wang, J., Healy, R. and Timoney, J. 'Perceptually Transparent Audio Watermarking of Real Audio Signals Based On The CSPE Algorithm', Irish Signals and Systems Conference, Cork, Ireland, June 2010.

These publications are available at www.cs.nuim.ie/~rhealy/ *and* www.irishunsigned.com/papers

Chapter 1: Introduction

In recent years there has been much research conducted in the area of audio and video compression and manipulation using digital signal processing techniques. The ubiquitous MP3, more properly titled MPEG-1 Audio Layer 3, is almost 20 years old having been approved as an ISO standard in 1991 [**1**]. It was, however, only one step in a long process of the exploitation of research into how the human hearing system works. As the MP3 is but a single step in the development of audio encoding based on perceptual factors, so there has been much further development and the MP3 has since been superseded by later codecs. While research and development efforts continue to be expended in an attempt to find ways of making file sizes (and therefore transmission times) smaller without affecting perceived quality, modern research in the area of audio processing has devolved from a focus on compression to one of using such knowledge for digital rights management applications.

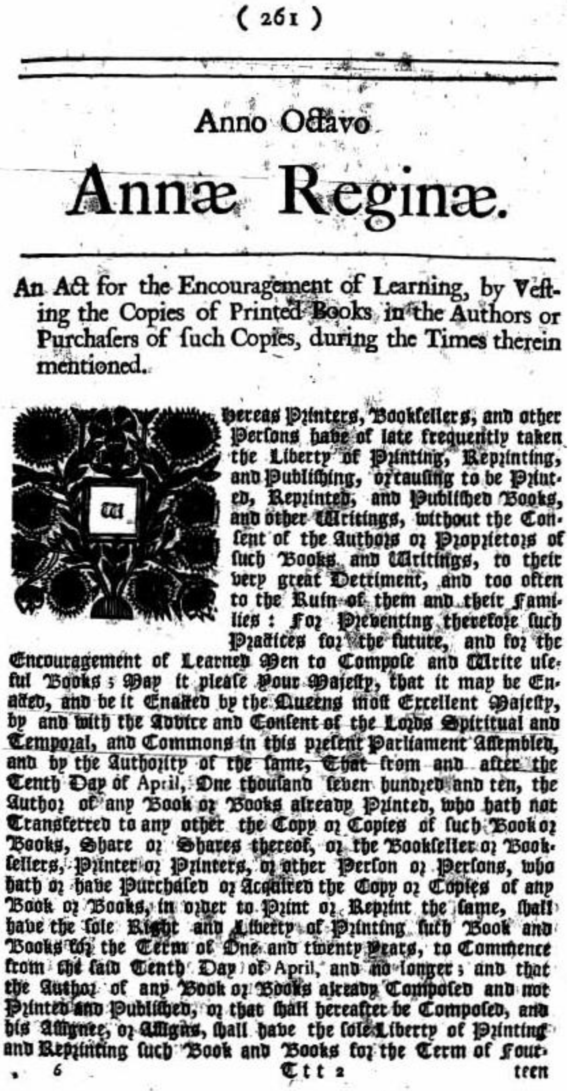
This shift in focus has been driven, in part, by the needs of the Entertainment industry to find means for protecting, tracking or identifying intellectual property such as photographs, music and movies. The SDMI (The *Secure Digital Music Initiative*) [**2**], a group consisting of more than 200 companies in the fields of I.T., Music and Entertainment, Consumer Electronics as well Security and Internet Service Providers, issued a challenge at the turn of the century, with regard to digital music, which invited investigation into the area of digital fingerprinting and watermarking as a content protection mechanism in the music industry. The challenge that was issued invited members of the public to break, hack or otherwise compromise various data-encoding techniques and technologies that the SDMI had developed as an initial step towards 'Digital Rights Management' (DRM) standardisation. The technology that the SDMI purported to recommend to the Industry was broken by outside sources almost immediately [**3**]. Crave *et al* noted that the SDMI challenge provided invited attackers with less information than would be available to an everyday 'pirate', and was apparently therefore deliberately intended to be a limited test of its technologies.

Eventually, the SDMI folded, claiming that it was awaiting further technical improvements before implementing DRM technologies. One of the reasons identified for the SDMI's failure was that the technologies then available were insufficient to achieve the aim of completely hiding an added watermark from those expert or talented listeners described as having 'golden ears'. This meant that there was no way of preventing potential removal of the watermark, since at least some of the human population could detect whether a file had been watermarked, detection being the first step before attack on a watermarking scheme.

Still, DRM-focused research continued apace. Much research has gone into copy control, copy limitation or copy prevention as these are the areas where stakeholders see most potential losses [**4**]. The ease of availability and simplicity of copying digital media (such as films and music) impacts heavily on sales. While losses might not be as high as the Music industry claims, because not everyone who downloads or copies a file is a potential customer, there is no doubt that digital copying and internet availability is damaging music sales. The 'Recording Industry Association of America' (RIAA) states that it is estimated by the 'Institute for Policy Innovation' (www.ipi.org) that illegal copying costs the industry US$12.5 billion [**4**] globally in lost music sales alone, along with more than 70,000 jobs in the US. This relates only to music piracy and does not include the potentially larger losses caused by the illegal distribution of DVD / TV / film. The obvious reaction to this, in commercial terms, is to try to limit or prevent illegal copying.

1.1 Background to this work

Parallel to the requirement of the corporate sector of the Entertainment industry are the requirements of the multitude of artists, performers and associated producers that create the material in the first place and for whom Copyright legislation was originally intended. Copyright is of course ancient in comparison to Computer Science, essentially being born in the 'Statute of Anne' in 1710 in England, but it is nevertheless a valid and topical area of active computer science research. The 'Statute of Anne' was introduced to prevent authors and their assignees being exploited by unauthorised re-printers copying, or pirating, their works and doing so '*to their very great detriment and too often to the Ruin of them and their Families'* [**5**]. The first page of the Statute is reproduced in Figure 1.1 and it clearly states the motivation of the act in the introductory paragraph as being for the '*Encouragement of Learning*' and further contends that this can be achieved by '*Vesting the copies of the Printed Books in the Authors or Purchasers of such Copies'* [**5**]. In the case of this early statute the works in question were books. However, the concept soon migrated to other creative areas and Copyright and its derivatives now extend to a wide variety of creative endeavours. Perhaps the most common would be in the Entertainment industry.



(261)

Anno Octavo

Annæ Reginæ.

An Act for the Encouragement of Learning, by Veſting the Copies of Printed Books in the Authors or Purchaſers of ſuch Copies, during the Times therein mentioned.

Whereas Printers, Bookſellers, and other Perſons have of late frequently taken the Liberty of Printing, Reprinting, and Publiſhing, or cauſing to be Printed, Reprinted, and Publiſhed Books, and other Writings, without the Conſent of the Authors or Proprietors of ſuch Books and Writings, to their very great Detriment, and too often to the Ruin of them and their Families: For Preventing therefore ſuch Practices for the future, and for the Encouragement of Learned Men to Compoſe and Write uſeful Books; May it pleaſe Your Majeſty, that it may be Enacted, and be it Enacted by the Queens moſt Excellent Majeſty, by and with the Advice and Conſent of the Lords Spiritual and Temporal, and Commons in this preſent Parliament Aſſembled, and by the Authority of the ſame, That from and after the Tenth Day of April, One thouſand ſeven hundred and ten, the Author of any Book or Books already Printed, who hath not Transferred to any other the Copy or Copies of ſuch Book or Books, Share or Shares thereof, or the Bookſeller or Bookſellers, Printer or Printers, or other Perſon or Perſons, who hath or have Purchaſed or Acquired the Copy or Copies of any Book or Books, in order to Print or Reprint the ſame, ſhall have the ſole Right and Liberty of Printing ſuch Book and Books for the Term of One and twenty Years, to Commence from the ſaid Tenth Day of April, and no longer; and that the Author of any Book or Books already Compoſed and not Printed and Publiſhed, or that ſhall hereafter be Compoſed, and his Aſſignee, or Aſſigns, ſhall have the ſole Liberty of Printing and Reprinting ſuch Book and Books for the Term of Fourteen

6 Ttt 2 teen

Figure 1.1: The first page of the 'Statute of Anne' (1710), generally considered to be the world's first copyright legislation [5].

While the Statute was intended to prevent the exploitation of authors' works by what was then the perfectly legal reprinting of paper publications, it opened the debate into wider areas. Within 65 years the scope of the legislation had widened to include, in the words of Lord Chief Justice de Grey, speaking in the House of Lords, '*composers of music, the engravers of copper-plates, the inventors of machines'* [**6**]. In that year (1774) the case of '*Donaldson v Beckett'* was debated at the House of Lords in England and the Attorney General observed that booksellers, who had previously been re-printing and reselling books without recourse to the author, had not *'ever concerned themselves about authors, but had generally confined the substance of their prayers to the legislature, to the security of their own property'* [**6**].

The corporate sector of the creative industries then, as now, would seem to have had little concern for authors while furthering their own ends. After the case was settled, it was held that the author had certain inalienable rights that he or she could choose to avail of, waive or assign. In the judgement it was further observed that '*literary works, like all others, will be undertaken and pursued with greater spirit, when, to the motives of public utility and fame, is added the inducement of private emolument*'. This is the basis for the development of modern Copyright: that an author of a work has rights that he or she can choose to either use, waive or limit, and that the potential for profiting on an ongoing basis from their work, by availing of their rights, is an incentive to further development of these and similar works.

It is almost exactly three hundred years since the Statute of Anne was enacted. Digital Rights Management (DRM) technologies in digital audio and video have received much attention in recent years with various efforts made to protect content from illegal copying, use or distribution. Some schemes were technically successful but not well received by end-users so therefore not successful in implementation. Others were not particularly successful technically, falling to the efforts of 'hackers' and other attacks. Efforts made by the corporate sector of the Entertainment industry to enable and standardise DRM technologies has so far served one purpose, namely protecting '*the security of their own property*' [**6**]. Some organisations have, of course, made attempts to ensure that artists are protected. However, comparatively small research output in the areas of DRM is targeted at schemes which do not primarily protect the corporate sector.

The world's most well-known digital music retailer, Apple's *iTunes* store, recently agreed to remove all Digital Rights Management (or *'electronic protection measures'*) restrictions from its music [**7**] and the rationale behind this decision can be illustrated by the widely reported fact that Norway's Consumer Ombudsman agreed to drop his country's legal challenges to *iTunes* use of DRM. It had been contended by the Norwegian Ombudsman, among others, that DRM technologies employed by Apple were restrictive to consumers, denying them the right to transfer purchased music to other devices. This issue was also being monitored closely by other European countries but the threat of any legal action was subsequently dropped after Apple removed DRM protections. Apple's defence of their DRM measures in the past may have been cloaked in claims of protecting the artist but they were in fact never about the artist. Opponents of DRM might claim they were more like 18th century publishers who had never '*concerned themselves about authors'* but rather were more interested in '*the security of their own property*' [**6**].

Notwithstanding the general acceptance of the need for copy protection and prevention and the obvious financial losses incurred by the recorded music industry, one area of digital rights management that receives comparatively little attention is the collation of data and subsequent distribution of royalties from public performance licensing. 'Public Performance' is an area specifically legislated for in modern Copyright and it is potentially an important source of revenue for Copyright holders (authors and their assignees). It is for this reason alone that the music *publishing* industry exists. It should be noted in this regard that 'public performance' in terms of music publishing is quite distinct from traditional sheet music publishing.

Research into potential technologies for the protection or monitoring of public performances is quite limited in scale and scope. One reason for this lack of urgency may be because breaches of public performance Copyright are not causing any tangible financial losses to the Music industry in the same way that illegal copying does. Indeed, the opposite may be the case, at least in some jurisdictions. It is apparent that, in some royalty distribution systems, not only do incorrect royalty distributions negatively affect some musicians and performers, they can actually create the reverse effect for which the concept of Copyright was invented before it evolved into an economy that is today worth more than €5Bn in Europe [**8**]. Since the concept of collective rights administration was born in Europe and the various European organisations have developed a multitude of mechanisms for the task, the amount of revenue generated from these activities is almost triple what was collected in the US and almost five times what was collected in Japan (on 2004 figures) [**8**]. It follows therefore, that artists and performers who have their collective rights administered by European organisations must be at an advantage compared to international counterparts.

Unfortunately, instead of providing accurate royalty payments to those who have had their works used in a 'public performance' capacity (including TV and Radio), thereby adding '*the inducement of private emolument*' [**6**] to an author's other potential rewards, today's royalty distribution systems often penalise developing and unrepresented artists while over-compensating well-established artists, corporate publishers and Copyright owners. This is perhaps why research in the area of audio coding for the monitoring of public performances such as radio and TV broadcasts is not as well resourced as that which deals with protection against illegal copying. Instead of costing the corporate sector and established artists, it can be shown to be rewarding them more than it should.

1.2 Introduction to the problem-domain

In order to understand the proposed technological solution to the problem of equitable monitoring of public performances for royalty distribution it is necessary to have a broad understanding of the technologies available, how they differ and to what purposes they are better suited. There are many different techniques within each sub-discipline but a broad overview is given here and expanded later.

1.2.1 Digital audio fingerprinting

Digital audio fingerprinting involves analysing a signal in some way in order to create a set of representative data that will be used as a reference at a later date in order to compare against a new ‘fingerprint’ taken in the same manner from a candidate signal. The two are compared in order to see if they are the same, thereby identifying the candidate fingerprint as being from the same source as the reference fingerprint. This concept is analogous to fingerprinting a person as it is used to provide acceptably individual identifying data (the reference print) that can later be used to identify whether the candidate print (e.g. a print taken at a crime scene) belongs to the same person from which the reference print was taken. The basic concept is illustrated in figure 1.2.

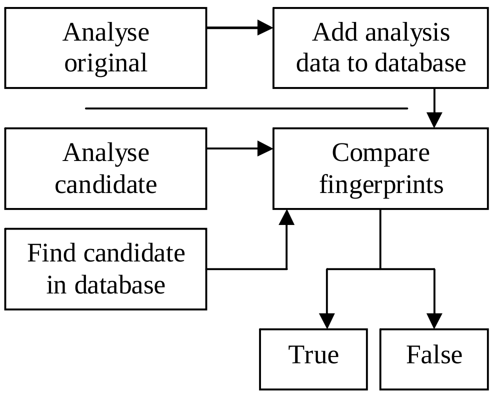


Figure 1.2: Basic fingerprinting scheme.

Fingerprinting techniques are used to some extent in broadcast monitoring to allow content owners or producers to track use of the content, distribution and/or audience reach. Implementations include ‘AudioID’, developed by the Fraunhofer Institute for the Digital Media Technology group headed by Prof. Karlheinz Brandenburg – who also headed the development of the MP3 standard, as well as the commercially available proprietary ‘Media Analytics’ technology from New Media Lab. There are many other implementations of fingerprinting techniques used for broadcast monitoring. However, they are also commonly available for other tasks related to content identification. One of the better known fingerprint-based scheme’s is the ‘Gracenote’ database (formerly the Compact Disc Data Base, or ‘CDDB’). This is an internet-based database of millions of tracks which is used to provide internet look-up for automatic recognition of Compact Discs as well as for on the fly identification of music as it is being played.

Fingerprint-based techniques can be used for many content-identification processes but they have some important limitations. For example, if a work is made publicly available and the content creator is not previously aware of the potential for fingerprint identification, then no fingerprint will be made available to the monitoring organisation before public release. Perhaps the largest broadcast monitoring provider in the UK and Ireland, Nielsen Music Control, has a database of 500,000 [**12**] pieces but this is obviously not the full range of all recorded music. If the monitoring organisation has no copy of the fingerprint of a piece of audio, then it might as well not exist for the purpose of monitoring. Another problem is that of versions: if an author remixes or otherwise alters a piece after release and initial fingerprinting, then the new piece is different so its fingerprint will be different. The monitoring organisation must have a copy of every version of a work that is made publicly available. However, perhaps the most obvious problem with fingerprinting techniques is the necessity to have a large and continuously increasing data store of fingerprints to be able to monitor for current and future releases into perpetuity.

According to Melinda Newman, West Coast Bureau chief of Billboard magazine, in 2004, *'There are about 30,000 albums released a year'* [**9**]. If the average album has only ten tracks, this amounts to 300,000 tracks per year in the US alone and only by official major record labels and subsidiaries. In the UK market, figures taken from the Music industry periodical '*Music Week*' indicate that there are approximately 11,000 albums released annually[i]. Again, taking a conservative estimate of 10 tracks per album, this equates to 110,000 album tracks in the UK. Adding in various single mixes, radio edits, collaborative remixes and sampled derivations – not to mention live or recorded-as-live performances on TV and radio - it becomes clear that the collection of digital fingerprints from which royalties are calculated is very limited. Nielsen's database accounts for little

[i] Data compiled privately by David Reid (Music industry A&R representative, Choice Music Prize founder) from 'Music Week' magazine and Official Chart Company figures for 2006, 2007.

over 12 months worth of *new* US and UK releases alone. Given that releases in different territories may be mixed and mastered differently, the limitations widen.

This takes no account of the thousands of artists worldwide who now release albums without the involvement of the record industry in any form. By way of illustration, there are only more than 200 albums and more than 500 singles released in Ireland each year[i]. Of these, a large number are by artists who have no corporate record label or publisher contracts and instead release their work independently. It is estimated by *'IrishUnsigned.com'*, a web-based promotional organisation for developing artists, that approximately 50 albums, singles or EPs are released in Ireland by independent artists each year. This equates to approximately 7-8% of all releases in the territory. It is not suggested that the same ratio exists in other domains but the advent of digital-only releases will certainly not cause a decline in the ratio of independent/corporate releases. Even if only 1% of singles and albums released in the UK and US was independent of corporate record label or publisher involvement, this would still be a substantial number, not least because it is those artists who most need equitable treatment to ensure their careers are *'undertaken and pursued with greater spirit'*, as England's Attorney General said in 1744 [**6**].

There is no way to estimate the number of individual tracks made publicly available in a given year but it is likely to be well above 500,000 internationally. Apple's *iTunes* store claims to have as many as 10 million tracks in its system [**7**] in January 2009. Providers like Nielsen, it has to be said, only monitor what they are specifically requested to monitor. Since there is a cost involved, owners usually only pay for the monitoring of singles released and likely to be broadcast on radio/TV – assuming they know of the possibility of doing so.

On considering one of the most requested tracks in the US, Led Zeppelin's *'Stairway to Heaven'*, such limitations in the opt-in systems used for broadcast monitoring

---

[i] Figures compiled from research undertaken through 'IrishUnsigned' based on confidential sales chart data supplied to the Music industry by Chart Track UK.

become more obvious. The track was not originally released as a single yet has amassed almost 1.5 million radio plays in the US alone [**10**] [**11**]. The track *did* eventually chart in many countries when digital downloads became chart-eligible but approximately 1.4 million of those US radio plays (and countless tens of thousands worldwide) would never have appeared on any airplay chart as there was no single released. While this is an extreme case, it does serve to illustrate the problem of only monitoring the airwaves for a small selection of tracks according to little known opt-in measures.

Another interesting point to note is that the Performance Rights Society (PRS), the UK's performance rights agency, agreed a deal in mid-2008 to use data supplied by 'Nielsen Music Control'. This has led to a claimed increase in the accuracy of royalty distribution to 90% [**12**]. It is also claimed that this partnership will *'double the accuracy of radio royalty payments to its members'*. While it makes sense that the majority of broadcasts from major broadcasters are of commercial releases, the problems caused by inaccurate distributions will still continue. This is because the final 10% of royalties from the major broadcasters, and the *'remaining smaller commercial radio stations [which] will continue to be paid by taking samples of the music broadcast throughout the year'* are likely to include a disproportionate number of minor and/or less commercially successful artists. It does have to be accepted that, limited as it is, this is a major improvement on previous systems.

Unfortunately, it is still the less established, less well-informed and less well resourced developing artists who will be left out of the distributions. They are, in effect, being dealt with exactly as was the case when Copyright was first legislated for. In fact, these same artists will now be even *more* likely to be impacted '*to their very great detriment and too often to the Ruin of them and their Families'* [**6**]. This is because some of the money they *should* receive from the PRS for even occasional plays on large or small broadcasters *may* be, as with all of the fingerprint-based royalty administration processes, incorrectly distributed instead to more well known artists. The service provided by 'Nielsen

Music Control' to the PRS relies on audio fingerprinting technology, as explained on their website[i].

1.2.2 Digital audio watermarking

Digital audio watermarking, as is suggested by its name, can be visualised as similar to watermarking of images by photographers or content owners, as in Figure 1.3, or watermarking of notes by banknote issuers to prevent or inhibit unauthorised copying, to prove ownership or to prove authenticity. Generally, the purpose of the watermark is not to physically or technically prevent copying but to make unauthorised copies either of little value or noticeably invalid. This same purpose generally applies to digital audio watermarking.

Figure 1.3: An example of a watermark used on a photographic image to make unauthorised copying an unattractive option [13].

Essentially, in the case of audio watermarking, the process involves adding some form of information – the watermark – to some signal – the host – in order that it can be recovered and decoded at a later date and used to prove the authenticity or identity of the candidate presented. The process is outlined in figure 1.4.

---

[i] 'Nielsen Music Control uses a unique patented electronic fingerprinting technology 'Medicor' developed for the sole use of Nielsen Music Control and the direct specific needs of the music industry'. http://www.nielsenmusiccontrol.com. Accessed October 23rd 2009.

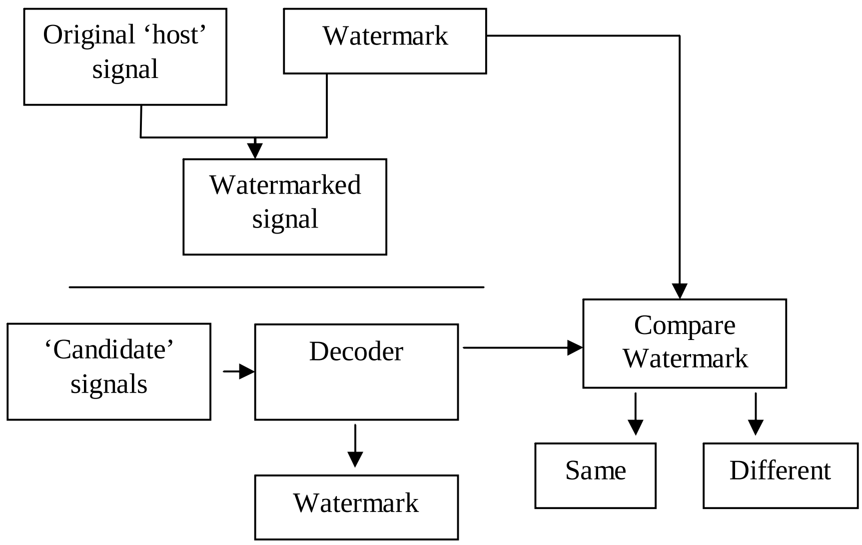


Figure 1.4: Basic watermarking scheme.

One interesting example of the use of digital audio watermarking for authenticity verification is in the real-time watermarking of witness and defendant statements at the time of interview with the police to prove they were not tampered with at a later date [**14**]. This technique is termed a 'fragile' or 'semi fragile' watermarking scheme as the host, watermark, or both would be noticeably damaged by any form of manipulation of the watermarked audio signal, in the same way as the photograph in Figure 1.3 would be noticeably damaged by removal of its watermark.

While it might appear that security of the watermark against removal should be a major consideration for all watermarking schemes it should be noted that the domain of audio or broadcast monitoring offers no realistic advantage to broadcasters, listeners or even potential pirates by the illicit removal of the watermark if that watermark serves only to identify the work in question or to provide added value to the end-user. The attempted removal of a digital watermark would in fact be likely to damage the audio in terms of quality.

One aspect of watermarking schemes that is sometimes overlooked or otherwise relegated to an incidental consideration is the detection procedure. As mentioned in Section 1.2.1, digital fingerprinting only works if there is a stored set of data to which a candidate fingerprint can be compared. This is one of its major limitations, particularly for less well-informed artists. Some watermarking schemes also specify detection or decoding processes that require access to either the original (unwatermarked) audio or to some other related data. These are called 'informed' watermark detection schemes [**15**] and they certainly have their uses in audio watermarking. However, broadcast monitoring would not benefit from such a system to any great extent as it would suffer the same limitations as fingerprint-based monitoring.

The ideal scenario in most applications would be for decoding to be possible in the complete absence of the original audio or any information related to it, except perhaps the knowledge that it actually has a watermark. This is called 'blind' or 'zero knowledge' decoding depending on its application [**16**]. A realistic compromise exists whereby the decoder might have access to some information *relating to* the host audio or the watermark. This is called 'semi-blind' decoding [**72**]. An extension to this is the case where there is no prior knowledge of the original host audio, nor the watermark itself, but there is known information relating to the watermarking process. This is also a form of 'semi-blind decoding' but it is perhaps more useful. In the case of a transparent, standardised watermarking technique, using standard pre-defined input values, the decode process could, indeed, become 'almost blind'.

1.3 Problem statement

The distribution of royalty payments for the authorised use of copyrighted works is a very important factor in the career of any artist. As proponents of early Copyright legislation clearly outlined in its promotion [**5**] [**6**], these considerations were particularly important to developing artists. Current processes used for the allocation of fees generated from licensed users to content owners, according to content use, are inaccurate, inefficient and weighted (at least incidentally) against developing artists. This is caused by ad hoc and inadequate mechanisms for the reporting of content used by licensed music users. The result of this is that, at least in some case, artists are abandoning their careers due in part to inadequate financial reward. The solution to this problem is an efficient and equitable system for monitoring the output of broadcasters and other licensed music users in order to be able, as much as possible, to ensure the correct remuneration is allocated to content owners when their content is used by licensed users.

Digital audio fingerprinting techniques are currently used in the broadcast monitoring domain in order to create these reports. However, audio fingerprinting techniques have some inherent limitations and disadvantages which, when fingerprinting alone is used as the basis for reporting, can actually lead to a worsening of the situation as far as developing artists are concerned. The solution to the problem is a transparent and accessible digital watermarking scheme. A successful digital audio watermarking scheme would address many, if not all, of the limitations of a digital fingerprinting scheme in relation to accuracy and equitability.

1.4 Structure of the remainder of this document

Chapter 2 provides an introduction and explanation of the area of Copyright as it relates to music and the creative processes, with specific reference to Public Performance copyright. Analysis of the Irish music industry is provided by way of illustrating how the rights provided for in this area are actually implemented and administered practically. Brief analysis of the global collective rights licensing sector is also provided for context. Issues and limitations that arise in administering some of the rights provided for by Copyright legislation are addressed and the effect of these limitations is outlined, specifically with reference to developing artists.

Chapter 3 investigates how the existing systems and structures in the Music industry may have a detrimental impact on artistic development, particularly for those at the threshold of their artistic career. A brief outline is presented of the various metrics that are currently employed for the furtherance of both an individual career as well as the perpetuation of the Music industry itself. The problems and limitations encountered in the industry, as a result of reliance on this information, are also discussed.

Chapter 4 provides an introduction to sound and the human auditory system. This is followed by an explanation of how humans perceive sound, how the ear deals with conflicting sounds and how the brain processes the resultant signals from the ear. This section also suggests ways in which the limitations of the human auditory system can be exploited to achieve various goals. The concept of data hiding is introduced and explained in general terms.

Later in the section an explanation is provided of how digital signal processing techniques are implemented and how the digitisation of sound is undertaken. An overview of sound file digitisation, storage, manipulation and compression is also provided in general terms. The chapter continues with an explanation of the two main audio identification techniques, namely watermarking and fingerprinting, before comparing the two techniques and identifying problems and limitations with the use of digital fingerprinting as the *de facto* standard for identification of audio in a broadcast environment.

Chapter 5 evaluates the technical domain of audio watermarking techniques and technologies. Various disparate audio watermarking techniques are outlined and examples of their implementation as published in the literature are provided. In each case, the relative advantages and disadvantages of each technique are mentioned as appropriate.

Chapter 6 details the watermarking scheme proposed as a solution to the various limitations in current implementations of audio identification and monitoring systems. A digital audio watermarking scheme is described from the initial design phase, through multiple phases to a scheme that addresses with some level of success most problems associated with broadcast monitoring for audio identification. The chapter also discusses some issues relating to watermarking schemes used for various applications. The development of the scheme is described in three distinct phases with reference to the various problems encountered and how they were solved or circumvented. Results and analysis of the output of the proposed scheme are provided.

Chapter 7 provides a conclusion to the research undertaken and outlines the advantages and disadvantages of the proposed system compared to existing schemes as it pertains to the domain of audio-monitoring for equitable and efficient royalty collection and distribution. A suggested implementation for a complete broadcast monitoring process is then outlined.

Chapter 8 suggests future work that may be undertaken to develop the proposed scheme in different domains, as well as suggesting developments within the audio domain that might take advantage of the characteristics of the scheme to provide alternative applications for the basic watermarking technique.

1.5 Summary

In this chapter, the problem domain was introduced and some analysis of the problems faced was provided. The history, background, development and motivation were briefly discussed. A description and brief explanation of the advantages and disadvantages of broadcast monitoring techniques based on audio fingerprinting and watermarking was provided. In the next chapter, the relationship between Copyright, licensing, royalties and content owners is explored with specific attention paid to the benefits of an efficient and equitable monitoring and distribution system for developing artists.

Chapter 2: Copyright, royalties and Music

This chapter introduces the fundamental concepts that underpin Copyright legislation and examines these concepts insofar as such legislation pertains to the Entertainment industry in general and the Music industry in particular. Specific attention is paid to the area of 'public performance' Copyright, which includes broadcast performances of pre-recorded works. The existing systems of royalty reporting, collection and administration are outlined and briefly evaluated before an overview of the economic value of the sector is given.

2.1 What is Copyright and what can be copyrighted?

The concept of Copyright has been around for centuries. Since the first major milestone in its legislative development, Copyright concepts have centred around the premise that the Copyright in a piece of work should belong automatically and without any special requirements – except perhaps those taken to prove ownership **-** to the person who created it (the author), unless and until he or she decides to give that right to a third party, whether for a fee or otherwise[i]. This right should extend for a fixed period and then expire so that the work is part of the 'public domain'. Initially, Copyright was a notion that was developed and legislated in each country and each had its own requirements and solutions. In 1883, the '*Paris Convention for the Protection of Industrial Property*' [**17**] extended French legislation to automatically protect content owners from one co-signatory of the Treaty against exploitation in territory controlled by another co-signatory. Later, in 1886, the Berne Convention [**18**] solidified the cross-border nature of both Copyright and protection from infringement.

[i] The 'Statute of Anne' in 1710 first recognised the inherent and automatic rights of the author where previous legislation defined Copyright as, amongst other things, a revenue-generating tool for the State. See http://www.nisus.se/archive/050902e.html for a background of Copyright, including the period prior to 1710.

According to Mr. Tom Kitt T.D, in a Dáil[i] debate on the introduction of recent legislation, the purpose of the concept of Copyright in modern terms is '*an attempt to strike a balance between the needs of creators and the interests of the wider public*' [**19**]. No mention is made of the needs of any industry segment. In the same debate, Mr. Stanton T.D. stated '*There is an educational and information deficit in our society concerning Copyright and intellectual property'*. Most people know that Copyright exists, and most people are peripherally aware that its purpose is to compensate artists and authors, including songwriters, when their works are used for anything other than the private listening/reading pleasure of the individual who holds a legitimate copy.

Notwithstanding this observation, it is likely that only a small proportion of the public at large will know the facts of Copyright. In a nutshell, Copyright is the term for the right of the creator of a work to allow or prevent certain acts being done to/with the work, such as copying, broadcasting, selling, lending etc. Allowing for the possibility that others may abuse their Copyright, legislation provides for procedures to recover fees (royalties) for such breaches and also allows for punishments in law for such breaches. In practical terms, almost anything that can be committed to a physical representation can be copyrighted. The Copyright will be held in most cases by the person or organisation that first made it available publicly or the person or organisation that can prove its existence on a date earlier than anyone else[ii]. For example, if a lyric-writer tells a writing partner the lines for a song and his partner commits the lyrics to music, transcribes the song to paper-notation, thereby recording it, and subsequently registers the piece, the lyricist has lost Copyright as he did not commit the lyrics to a physical medium. Even if he had done so, the partner would still own the Copyright as the lyricist would have no way of proving he had created the lyrics first.

---

[i] The 'Dáil' is the Irish Parliament.

[ii] There is no actual requirement to copyright a work as Copyright is automatic and vested in the creator. However, it is sometimes important to be able to prove it was copyrighted on or before a certain date. See IMRO's advice in this regard in their FAQ for music makers, online at: http://www.imro.ie/faq/music_makers.shtml, question 2. Accessed 24th October 2009.

Registering a Copyright is not strictly necessary for it to exist in Law. Copyright exists (or '*subsists*', to quote the act [**23**]) automatically. The process of registering a piece of literary or musical work only serves to prove, using a trusted third-party or other verifiably secure system, that the piece existed on the date of registration and the Copyright is/was owned by the registrant. There are, as with any Legislation, numerous exceptions, loopholes and provisos. However, in the context of this thesis, Copyright ownership and registration need not be considered in any detail. Suffice to say that, while Copyright exists automatically upon creation of a work, it can be difficult to prove prior ownership of that work at a specific date in the past if no copy of the work can be proven to have existed at that date.

Copyright legislation provides for the Copyright owner (or their representative[i]) to license third parties for any and all specified uses of the copyrighted works. In some cases, such as open-source licensing in the software arena, the licenses exist but it still costs nothing to use the works. In other cases, the author may state that licenses must be sought for any and all uses, whether physical or otherwise, of their work, by anyone other than for private use by an individual who holds a legitimate copy of the work. Essentially, the Copyright owner can make his or her requirements for legitimate users be as wide or as narrow as he or she wishes since they are the owner of the item in question.

The licensing of most of the rights associated with Copyright can be assigned to third parties to administer. Some of the rights that are routinely assigned to third-parties, specifically in the area under consideration (the music industry) are[i]:

- 'Duplication rights' (to a record label, for example, to make CDs or to a library or other institution to make photocopies).
- 'Synchronisation rights' (the right to add, or synchronise music/audio to film and video).

[i] Most collective rights organisations act on behalf of content owners to administer some, but not necessarily all, of their rights. 'Broadcast Music Inc.' is one such organisation that offers multiple licensing options, including those listed here. http://www.bmi.com/licensing/entry/533606 Accessed 23rd October 2009.

- ‘Public Performance rights’ (the right to make music or video available to others, who have not themselves purchased it or obtained a legitimate copy).
- ‘Digital Performance rights’ (permit the use of the sound recording for digital transmission, e.g. Internet streaming).

Again, there are exemptions. It is the public performance right that is dealt with in this work although the techniques examined can be adapted to address other concerns in the Digital Rights Management arena, such as ownership verification, tracking leaked pre-release content and so on. Similarly, although it is not the focus of this document, those concerned with areas as diverse as covert communications, additional broadcasting revenue streams and automated video subtitling can adapt the techniques to suit their requirements.

2.1.1 Public performance rights and royalties

The area of ‘Copyright control’ in music has been heavily researched in recent years, with a particular focus on the protection of copyrighted material from illegal copying. Almost from the beginning of recorded music, it has been copied by home users. As time and technology progressed such copying became easier. The cassette tape allowed for easy recording direct from radio. Video cassettes allowed for easy copying direct from television. Industry schemes to deal with such illegal copying have always been reactionary and – to a certain extent – predictable. The addition of a levy to all cassette sales [**20**] was one solution to the industry’s perceived loss of income but it did not solve the actual problem of illegal copying. Rather, it merely financially compensated record labels for such illegal copying to a limited degree. Schemes such as this do not make any real difference to the group for whom the concept of Copyright was designed to protect. Instead they serve to compensate a group who have been assigned *some* of the rights enshrined in Copyright (the right to duplicate or copy) for what are, at best, notional losses that are incurred when illegal copies of a work are made, on the assumption that those who use the illegal copy would otherwise always pay for it. This is, of course, patently not the case.

With the advent of widespread use of the Internet in the home and workplace, music has been illegally shared on a much wider scale and with much less difficulty. The focus of research, particularly with commercial backing or motivation, is in the area of preventing 'pirating' of music and video. This has led to an explosion of the development of various techniques to tackle the problem. However, it is obvious from the proliferation of pirate sources and the lack of any real success in restricting pirating that any attempt to prevent illegal copying is not much more than an exercise in staying one step *behind* the pirates.

Indeed, some believe that pirating should be allowed and that it should benefit the creator (as opposed to the current owner) of the copyright. In Germany, as far back as 18th century, widespread 'pirating' of published material (books) was shown to lead to increases in sales due to increased awareness and to the growth of public lending libraries [**20**]. In recent discussions in legal circles in the United States of America, some sections of the profession suggest Copyright legislation can actually work against those it is intended to protect. It is suggested that copyright legislation merely serves as a means of artificially inflating the revenues generated by industry organisations and the highest-earning bracket of the music industry at the expense of the lower-earning, developing, and more vulnerable section [**21**].

Moreover, many developing artists and performers realise that widespread copying of their works is often a way to increase awareness of them and therefore to create a potential audience and market for future works. The rise of some 'stars' in recent years has been developed on this premise, most notably Lily Allen and 'The Arctic Monkeys' in the UK. Record labels and other Copyright owners (who are not necessarily the creators of the works but are instead licensed to exploit the works) have different perspectives on the necessity of Copyright legislation. It is this sector that has financed and fuelled recent trends in protecting Copyright and punishing transgressors such as Napster, Kazaa, Limewire and Pirate Bay. Those most likely to have been tackled by industry efforts to address piracy are not actually pirates themselves. They are instead providers of services which are used to facilitate or expedite piracy.

The current situation that exists has similarities to the 15th and 16th centuries, prior to the development of copyright. In the past, the actual creator of a piece would sell it to a third party such as a book publisher for further exploitation and it was the publishers and producers (analogous to today's record labels) that benefited most from exploitation of the works. The focus of much research in the area has in recent years switched from compression to tracking and identifying the proliferation of high-capacity pirates. These high-capacity or high-profile illegal copiers are seen to be the main cause of lost earnings as well as being most likely to lead to a successful prosecution and therefore most useful as an example to the public. The impact of these measures or, in fact, the impact of any copyright-related research on the development of the artist is a distant secondary concern of related research.

Other than copying, with or without permission, the second main area of Copyright legislation - and one that is often overlooked - is that of 'public performance' royalties. Essentially, the owner of the Copyright of any piece of music made available to the public is entitled to a fee for such public performance. The Irish Courts have ruled that *'…a performance of music which takes place outside the domestic or family circle of the audience be regarded as a performance in public. It does not matter whether the audience has paid, whether they have come for the sole purpose of listening to the music or whether the music is performed by musicians or by mechanical means such as a radio, CD or a tape'* [**22**]

This definition can be extended to cover any use of music in public, even in a setting as innocuous as a group of friends sitting around a CD player while out camping. However, in reality this description is usually referred to in terms of royalties due when a song is played on the radio or TV, or when a song is played as background music in an establishment of some sort that is open to the public or where employees have access to the music in question (deliberately or otherwise).

Legally, any Copyright owner can monitor the airwaves and, when they notice a 'play' of their piece, they can then contact the relevant broadcaster and negotiate a fee for the use of that material. However, this is obviously impractical and the cost/benefit trade-off would make it a worthless exercise. Instead, Copyright owners are usually represented by an agent or organisation which will license their catalogue of material for use along with material from its other members. License fees or royalties are then collected and distributed according to the number of times a piece has been reportedly made available on radio or TV over a given period. In practical terms, in Ireland, this task falls to IMRO (*the 'Irish Music Rights Organisation'*). IMRO is a royalty collection and distribution agent representing over 3000 Irish songwriters which is authorised under legislation by the 'Controller' of the copyright act. IMRO also administers reciprocal collection agreements with international equivalent organisations. There are other such organisations in Ireland and their roles will be mentioned within this document but for clarity it will generally focus on IMRO in terms of analysing the current state of royalty collection and distribution in Ireland as the requirements of other collective rights societies are similar.

2.1.2 The need for a royalty collection and distribution process

Copyright collection is now a huge industry. Moreover, it involves an almost limitless number of potential inputs in order to accurately complete a report of royalties due. However, even before such a 'distribution report' can be prepared, the royalties must be collected. As mentioned in the previous paragraph, this task can theoretically be undertaken by the owner of a work or works but there are inherent problems with this theoretical idea:

- The owner of the work(s) must be aware that the work has been used in some manner
- The license for use of such work(s) must be negotiated, even retrospectively
- The license fee (royalty payment) agreed for such use must be collected

In the current climate, in Ireland alone, there are dozens of radio stations and a number of TV stations. There are then those broadcasters who broadcast into Ireland from another territory. In modern digital communications, there are literally thousands of potential outlets for the owner of a work to monitor. Of course, if they are the owner of more than one work, they must monitor for each. Furthermore, since there is no way of knowing when (or even if) a given broadcaster will use any given work, then they must all be monitored continuously and simultaneously. This is an unrealistic prospect.

Even if the beneficial owner of a group of works, such as a publisher or other aggregator, might want to perform this task privately to ensure compliance and maximise returns for their works, the cost of undertaking such a task would far outweigh the benefit. Monitoring Irish-originated broadcasters alone would require an outlay of more than €5 million per annum, per publisher (or artist) calculated below:

According to the current information (2008) on the website of the *Broadcasting Commission of Ireland* [i], there are more than 50 licensed radio broadcasters in the State, along with 13 television broadcasters and re-broadcasters. This equates to over 1500 hours of broadcast per day (63 x 24 hours) or over half a million hours per year. Assuming the task could be done by a 24-hour roster of monitoring staff, manpower payments at the legal minimum wage of €8.65 per hour (2008) would produce a cost of almost €5 million per annum. It is obvious that it is therefore not a viable proposition for any artist or publisher to undertake to perform the monitoring task themselves, particularly if they have either a small collection of works or are unlikely to be extensively featured on broadcast media.

When one considers the above, and then widens the perspective to take in the scope of thousands of artists in Ireland (including defunct artists), and then widens it again to take in the tens of thousands of artists worldwide and the thousands of licensed

[i] http://www.bci.ie

broadcasters that must be considered, the problem becomes almost impossible to visualise, let alone solve.

It is with the scale of this problem in mind that artists and performers have, over the years since the end of the nineteenth century, grouped together as collection societies to perform the task. This arrangement now has a legal basis in most European countries. In Ireland, for example, the '*Copyright and Related Rights Act 2000*' [**23**] permits a representative subgroup of a larger group (such as songwriters, performers etc) to represent the wider body of their peers under license in order to administer their rights. The collection society in question is conferred with certain legal rights which make the administration of these rights not only easier to manage but easier to enforce since they have the backing of legislation and the threat of punitive action against non-compliant users of the works they represent.

2.2 The existing royalty collection and distribution process

Copyright collection and distribution in the current environment relies on anachronistic, ad-hoc and unscientific amalgamation of procedures and systems. The owner of a piece of copyrighted material (a song, for example) registers it with a collection and distribution organisation (such as IMRO) and assigns permission to the organisation to collect royalty fees from plays on TV and radio or other public performances. At the other end of the process, when songs are played on TV or radio the broadcaster is required to submit a list of all songs broadcast along with time and date for referencing. Where such data is not submitted, the Copyright owner or their assignee has the Statutory right to demand it from the broadcasters or other users.

The process briefly outlined would seem like a reasonable one. In practice, however, it is useful only to a point. Some of the main flaws in the system will be examined along with the known erroneous outcomes that actually exacerbate the problems associated with the system in the first place. Limited as it is, the system is not operated to any great efficiency or accuracy. Much data remains uncollected, collected data is often unverified and a minimal account is taken of non-mainstream music broadcasting.

According to IMRO, revenues generated by collection of the blanket license fee charged for public performance rights is grouped and distributed according to various rules. In essence, '*for broadcasters, each station's revenue is distributed as a separate pool on the basis of logs submitted by the broadcasters*'[i]. This would suggest that if 'Broadcaster A' pays a blanket fee, then this fee should be distributed to Copyright owners according to all, exactly all, and only all, of the songs that it has broadcast

Of course, there are exceptions. RTE1, Lyric FM and Radio Na Gaeltachta are treated as one broadcast pool. On television, RTE 1 and RTE 2 are similarly 'pooled' into one distribution[i]. This might not seem like it should be a problem but it can cause the wrong Copyright owner to be paid or the wrong amount to be paid to a recipient. Of course, in a scheme where one recipient gains, another one (or more than one) has to lose.

When calculating how much of the royalty pool each Copyright owner should get, there are two distinct metrics: number of plays and duration of play(s). Again, according to IMRO, '*All TV stations are paid out on a duration basis, as are all RTE Radio stations plus Newstalk. All other stations are paid on a per play basis*'[i]. There will obviously be advantages and disadvantages in either, depending on the individual event (or 'play'). For example, if a track is only one minute long, and it is paid-per-play, then it will receive the same royalty as a track that is 12 minutes long. Alternatively, in a pay-per-duration distribution, a 12-minute piece would earn twelve times the amount of a one-minute piece. It is therefore advantageous for shorter pieces to be played on stations that are distributed on a per-play basis and for longer pieces to be played on stations that are paid out on a per-duration basis.

[i] From private communications with IMRO

Since there is no realistic way of identifying 'quality' as all arts are subjective, then payments have to be made on quantity. For example, John Cage's '*Four minutes thirty three seconds*' (also known as 4'33") is entitled to the same royalty payment as any other piece, despite having no actual sounds (other than auditorium background sounds). In fact, if a royalty payout is calculated on a per-play basis, then this piece of 'silence' would generate the same royalty as any other piece of work, no matter how long or complex. On stations distributed on a per-duration basis, the Cage piece (which was 'played' by the BBC Symphony Orchestra and broadcast live on BBC Radio 3 on January 16th, 2004 to mark his death [**24**]) would generate more royalties than the vast majority of other works, since modern music is generally shorter than 4.5 minutes and, even where it is longer, it is generally shortened for radio and TV broadcasts.

If the letter of the law is to be followed [**23**], royalty distributions should, in theory, be calculated on the basis that a piece played on a small local radio station with a small audience would be entitled to a royalty payment correspondingly smaller than if it was played on a National broadcaster with a correspondingly large audience. Furthermore, a piece played on any given station in a 'quiet' time should not still be paid the same royalty as a piece played during 'peak' listener hours for the same station as the 'public' to which the performance is available is correspondingly larger. Distributions are, therefore, only partially based on audience size and are, in reality, based more on *potential* audience size. A work broadcast to a few thousand people will currently often be awarded the same royalty as one played at peak time to tens of thousands.

2.2.1 Why is there a need for a new system?

The systems and process in place for the collection and distribution of royalty payments do not really address the practical problems faced by – in particular – the developing artist. As mentioned in section 2.1.2, the royalty agent (IMRO, for example) has the right to demand data on playlists of music broadcast by radio and TV stations. Rather than demanding such information from all broadcast sources, IMRO contents itself with only some data. According to IMRO, '*Stations are categorised either as full census,* *sampled or mixed. Where a station is sampled, the value paid for each play is correspondingly higher than it would be if the station were subject to a full census. If a station is on a mixed rate this means that they are delivering full census reports for all automated programming (i.e. where music is played out using an automated/playout system)'*[**25**]

While a 'full census' of a station's output would be the most accurate, and therefore fairest, data to base the payouts on, it is currently not available for every broadcaster. In fact, the definition of 'full census' can be misleading as this suggests that a station provides complete listings of all music broadcast in a given period. In reality, while most stations do try to achieve this level of accuracy they do not do so consistently as there are often tracks that are played on a specialist show that are not in the list and, ironically, these are likely to be the shows that broadcast music from less well known artists.

Perhaps the most revealing part of IMRO's description of their distribution calculations is the statement *'Where a station is sampled, the value paid for each play is correspondingly higher than it would be if the station were subject to a full census'* [**25**]. What this means is that a 'sample' of the data is taken for these broadcasters and the entire blanket fee of royalties collected is then distributed according to these samples. The result of this sample-based distribution is that pieces with very few plays are correspondingly less likely to be reported on the sample, while the full royalty fee is then distributed among those that are reported. In fact, in many cases, a piece that is played on stations that are sample-reported may not ever be part of the data submitted to IMRO and other royalty agents and may never generate any royalties from public performance, regardless of the number of times they were broadcast. A list of licensed broadcasters and their reporting categorisation is shown in Table 2.1 in Section 2.3.

If 'Song A' is played ten times and does not appear in a station's sample, while 'Song B' is played ten times and does appear in the sample, then – all other things being equal – 'Song B' will generate more royalties than it is entitled to as its share of the overall royalty pool will include some of the portion that otherwise should have been distributed for ‘Song A’. More importantly, if ‘Song B’ is only played once, and this play is included in the sample, then ‘Song B’ may well end up with rather large payouts (comparatively speaking) and some of these payouts would actually belong to the owner of ‘Song A’. In order to overcome this problem, and to make the collection and distribution of royalties fairer, in particular for works played less often, a complete and accurate list of all works broadcast at all times on all stations needs to be available.

While the problem of inaccurate reporting in the domestic market primarily affects developing artists, even major artists are adversely affected by the lack of accurate data from the international markets. There is no accurate information available to a collective rights organisation for broadcasters outside its own jurisdiction. This is a huge problem, both in logistical terms and in terms of the potential royalties that go uncollected. What generally happens is that the relevant collection society in a jurisdiction will inform a sister-organisation with whom it has a reciprocal arrangement of the amount of royalties it collected that are due to that sister organisation. There is no real scope for querying the information provided since there is no record of plays of works that should be paid to members of the society.

2.3 What is the ‘collective rights’ market worth?

In commercial terms, the collection and distribution of royalties may seem inconsequential and lacking any commercial imperative to change, especially when compared to the resources and efforts being committed to the other main Copyright area of piracy. However, in the case of IMRO’s Irish marketplace alone, license fee income is in the region of €36 million per annum, of which approximately €31 million is distributed to members for public use of their registered and copyrighted material. The remainder is absorbed by administration, including the actual generation of distribution details [**26**]. There are other royalty agents and collectors in Ireland and, of course, Ireland a comparatively small market. In European terms, the revenue generated by this sector is circa €5 billion and growing [**8**].

By way of illustration of some of the issues that arise when allocating royalties correctly, consider a well-known song. If a radio station played ‘My Way’ twice, performed once by Frank Sinatra and once by U2, then the Copyright owner of that song would be entitled to a portion of the IMRO license fees collected for two individual instances. However, IMRO only deals with royalties owing to songwriters and publishers. There is then the *performance* royalty fee which is due to the person or group actually performing the recording that was made available. In this case, this would be due to Frank Sinatra’s estate for one performance and to U2 for the other. In Ireland the recently-formed ‘Recorded Artists And Performers’ (RAAP) administers the collection and distribution of the performers’ royalties and, as such, RAAP have as much use as IMRO for fair and accurate reporting of broadcast data. RAAP distributed over €3 million to Irish performers alone in 2005 [**27**]. The public performance royalties generated by performers’ collecting agents in Ireland and the EU is approximately 10% of the value of the royalties generated by writers’ collection agents [**8**].

While the revenue generated by public performance licensing for giants of the Irish music industry like U2 would be expected to be substantial enough to warrant careful administration [**28**], it could also make a huge difference to grass-roots developing artists and songwriters. A relatively small payout for an independent artist or songwriter could easily make the difference between pursuing their art and choosing an alternative career. Some of these artists could go on to become the U2’s of the future on the basis of such income. The amount of revenue generated from high-rotation radio play for even one or two songs in a year could be enough to finance the recording and release an album.

2.3.1 Royalty payments to content owners

The amount of money paid to an artist for a single play of a single work by a single broadcaster varies significantly according to the reach and target audience of the broadcaster, as well as the type (i.e. commercial broadcaster or community broadcaster). Table 2.1 illustrates the variations, using IMRO figures from 2005 [**25**]. Table 2.1 also lists those broadcasters for whom only sampled data is available and those for whom full census data is available. Note from Table 2.1 that, while some broadcasters provide full data for automated broadcasts, they only provide sample data for non-automated broadcasts. There is no indication of the ratio of automated to manual broadcasts.

Description	Sample Days (Per Month)	Rate	Per
2FM - General	All	€4.40	Minute
96 FM - General	All	€1.83	Play
98 FM - General	All	€3.50	Play
Beat 102-103 FM - General	All	€0.31	Play
Clare FM - General	3	€3.84	Play
County Sound - General	All	€1.34	Play
East Coast Radio - General	5	€3.68	Play
FM 104 - General	All	€2.62	Play
Galway Bay FM - General	All	€0.65	Play
Highland Radio - General	2	€16.87	Play
KCLR 96fm - General	All	€0.35	Play
KFM - General	All	€0.18	Play
Limerick's Live 95FM - General Mixed -	Full Census Automated & 4 Days/Month Non Automated	€0.85	Play
LM/FM - General	2	€10.18	Play
Midland Radio 3 - General	2	€9.80	Play
MWR FM - General	4	€29.31	Play
NewsTalk 106 FM - General	All	€4.83	Minute
NWR - General	4	€10.37	Play
Ocean FM - General	All	€0.19	Play
Q102 - General	All	€0.86	Play
Radio Kerry - General	2	€23.82	Play
Red FM - General	All	€0.25	Play
RTE Radio 1/Lyric FM/RNaG	All	€2.40	Minute
RTE TV - General	All	€4.22	Minute
Shannonside/Northern Sound - General	3	€11.99	Play
South East Radio - General Mixed	Full Census Automated & 6 Days/Month Non Automated	€0.49	Play
SPIN 103.8FM - General All €0.40 Play			
Tipp FM - General Mixed	Full Census Automated & 4 Days/Month Non Automated	€0.86	Play
TnaG - General	1 in 5 days	€0.64	Minute
Today FM - General	All	€4.92	Play
TV3 - General	All	€3.88	Minute
WLR - General Mixed	Full Census Automated & 2 Days/Month Non Automated	€0.90	Play

Table 2.1: IMRO distribution run from 2005 listing rates paid to each artist for a single play of a single work. It also lists those broadcasters who supply full / limited data.

The discrepancy between the number of plays reported and the actual number of plays as discussed above is only one area of concern for developing artists. Is likely to exacerbate the problems experienced by these artists and songwriters because not only are they less likely to be played on lucrative and influential media outlets such as Radio and Television, but, even if they do manage to get such exposure, their rightful income is often not received. Additionally, it is often added to the income distributed to other artists and songwriters who have appeared extensively on reported playlists. These artists and songwriters are, of course, often the more successful and well-known performers (which is why they receive more exposure in the first place).

In fact, it is clear from the information provided in Table 2.1 that a well-informed artist can benefit greatly from information available about distribution runs. If an artist had a work played on, for example, Radio Kerry, on one of the two days that the broadcaster is sampled, this single broadcast of a single work would be paid €23.82. Another artist may have their work broadcast on each of the other 26-29 days (excluded from the sample data) and would be paid nothing. If this is extrapolated across the whole broadcast year and duplicated in other 'sampled' broadcast reports worldwide, the scale of the problem can be seen. While the variables (broadcasters, rates and metrics) may change over time, the fact remains that well-informed or well-connected artists can stand to gain dramatically higher royalties at the expense of under-informed or under-represented artists.

2.4 Summary

This chapter has introduced the fundamental concepts that underpin Copyright legislation in relation to the music industry. Existing royalty reporting, collection and administration systems were outlined and briefly evaluated before an overview of the economic value of the sector was given. The importance of royalties for the development of an artistic career has been recognised since the inception of a legal Copyright framework. However, there are other factors that may be considered to be equally important to the career development of artists. Some of these factors are connected to the same topic of broadcast monitoring and reporting. These issues will be discussed in Chapter 3.

Chapter 3: The Developing Artist.

In Chapter 3, additional factors that can impact on the relative success of the careers of developing artists are outlined, specifically in relation to the standing of these artists in the public eye and in the wider industry. Various metrics for gauging early-career successes are introduced and are related back to the importance of fair and equitable reporting of broadcast data as a means of measuring and artist's appeal and exposure. Some problems that exist in this regard are discussed with a view to solving them with techniques extended from those used to solve the problems discussed in Chapter 2.

3.1 Popularity networks and attractedness

It is perhaps a well-known paradigm in everyday life that popularity breeds popularity. The music industry, in particular, is founded on the whole concept of popularity. Individual organisations within the industry attempt to create or promote some sort of 'critical mass' of buyers or listeners of a particular act so that they may become 'mainstream' and generate a return on the initial investment as soon as possible.

In recent years, the development of 'social networks' and 'popularity networks' has been exploited as practical marketing tools and deliberately utilised to propel acts towards achieving a 'critical mass'. As mentioned in Section 2.1.1, Lily Allen and 'The Arctic Monkeys' are just two artists who can count popularity networks as being responsible to a comparatively large degree for their success in their early careers. In terms of music, fashion and myriad other taste-based products, the desired result of marketing is to make an act (or product) fashionable. While creating a 'fad' can generate enough of a return on investment, making an act become fashionable while increasing the likelihood of their long-term success is the key to long term survival and growth in the music industry. Many promotional strategies are employed in an attempt to create or promote an act, with the hope that some acts will become commercially successful for the benefit of both themselves and the industry in the longer term. These strategies include reality television talent shows at one extreme and deliberate manipulation at the other. However, there are many strategies utilised which are not so direct. Examining the rise of the aforementioned Lily Allen and The Arctic Monkeys will serve to illustrate this idea.

Lily Allen[i] is now a well-known name in music circles, with several successful singles and two successful albums under her belt. She was originally signed to a record deal with the Warner Music Group. The deal expired without her ever having released any material. She then released into the 'MySpace' virtual world the original 'demo' songs that had prompted Warner and then Regal (part of EMI) to sign her.

As a result of interest generated in online communities and then amongst students (mostly in the UK), the music press started to review her material, creating enough exposure to warrant the mainstream media picking up the story. This prompted the record company, who maintained the rights to the recorded material, to release her album to huge success[i]. It is questionable whether any of this would have happened but for the online release of her original demo material. Having said that, the power of 'community' sites such as 'MySpace' has proven that the Internet equivalent of 'word of mouth' can be just as useful a marketing tool as real-world 'word of mouth' recommendations from trusted sources, despite the fact that most 'sources' on the Internet are virtually unknown, as are their motives.

The second example is the rise of The Arctic Monkeys, a UK 'indie' band who became the 'coolest' of the modern batch of bands during their development and then broke music Industry records after they released *'the fastest-selling debut album in UK chart history'* [**29**]. This followed on the heels of an ad-hoc, fan-powered campaign which saw their demos being ripped from the (few) CDs that the band had made themselves at home. The ripped files were uploaded onto file-sharing networks and made available to new fans after 'recommendations' online and by word-of-mouth. After a short time, the band's gigs were becoming well attended by fans that seemed to know the lyrics and sang along with

[i] While the story of Lily Allen's rise to fame should be treated with caution, most of the salient points are confirmed by the singer herself. An interview carried out in November 2006 for 'Pitchfork' can be found online at http://pitchfork.com/features/interviews/6476-lily-allen. Accessed 24th October 2009

the songs, songs which officially they could never have heard as there were no authorised copies available publicly. The band obviously did not mind their music being made available in this way – few 'unsigned' bands would object to the extra exposure. As a result of this, the band created enough interest at grass roots level to warrant interest from record label scouts, who the band allegedly shunned before eventually signing a deal with one of the lesser labels, the independently-owned 'Domino Records'. This is just one case of how illegal copying actually worked to the benefit of, rather than to the detriment of, the copyright owner. In this case, the artist in question (or an assignee) could have attempted to stifle such illegal copying but this could have proven detrimental to their development.

3.1.1 Popularity networks and social marketing

In both cases, the concept of popularity networks and concepts related to multiplicative stochastic[i] processes seem to be at work [**30**]. Word of mouth recommendations have always played a part in peer-groups. In recent years these peer-groups have become more than the handful of loosely affiliated music fans that previous generations might associate with the term. Certainly, they have a world-view of technological developments that far supersedes the previous generation. Their promotional power is unquestionable and, given the size of the group in question and by virtue of the fact that often they spend a lot of time online, they are very aware of what is going on in this arena.

It has even been suggested [ii] that even successful acts (or their associates) employ people to operate online and 'unofficially' promote them the same way as they would employ PR personnel to promote them to traditional audiences through traditional media. Online, there is little or no way of knowing if some 'recommendation' has been paid for as identities are obviously hidden, including the identity of the referring individual. The

[i] Briefly, stochastic processes are those that are non-deterministic and therefore unpredictable. The output of a stochastic process relies on the predictability of the system itself and some random input. In a multiplicative stochastic process, the 'randomness' leads to random outputs, returned as inputs to the process, thereby leading to more and more random outputs, essentially exponentially.
[ii] Anecdotal evidence collected privately through '*IrishUnsigned*'.

commercialisation of these concepts in the industry has led to the growth of online or virtual 'street teams' (defined below) which replicate the functions of real street teams on a contract basis. There are also virtual 'street teams' that perform the same function in the mobile phone environment [**31**].

'Street Teams' are groups of people who create an awareness of an act or campaign and perhaps execute purchases of current material in order to initiate sales for the purpose of promoting awareness of the act to industry or the public. This is perfectly legal and many campaigns have been started or extended using these techniques. In July 2003, as an experiment to analyse the accuracy of the Sales Chart reporting system, online artist collective '*IrishUnsigned'* released a compilation of material from 18 independent Irish acts. Through the use of street teams and online promotional techniques, the release eventually made the chart [**32**] albeit with lower reported sales than had actually transpired. The exercise showed that with only a small number of people, existing reporting systems can be manipulated to suggest an apparent market for a product, thereby creating the market. The same organisation has also been contracted to provide similar services to the major corporate record labels to ensure that releases by established artists are immediately successful, promoting awareness and leading to further sales.

Identified by marketing 'guru' Philip Kotler, the growing practice of marketing by use of 'popularity' and 'community' networking perfectly fits the description of 'Social Marketing' [**33**]. Social Marketing, as defined by Kotler, is '*the use of marketing principles and techniques to influence a target audience to voluntarily accept, reject, modify, or abandon behaviour for the benefit of individuals, groups or society as a whole*' [**34**]. There is no mention here of selling a product, but selling a philosophy, concept or idea (or the rejection of an idea) can be central to a 'Social Marketing' campaign. In terms of the music industry, this phrase may be paraphrased as 'the use of marketing principles and techniques to influence a target audience to voluntarily accept an act for the benefit of the act and the network or society in which the social marketing is performed'.

3.2 Industry indicators

3.2.1 Radio as a referrer

So, how does the concept of 'Social Marketing' raised in Section 3.1.1 concern radio broadcasts and the monitoring of output from radio stations? In the context of the development of any act, it can be seen that perhaps one of the most important things that can be done for the act is 'spreading the word'. The related functions of 'marketing' and 'promotion' are used to create an awareness of a product (in this case the act or the song) with a view to persuading the customer to 'buy' the product. Given that customers cannot exactly 'buy' an act, this might seem misleading. However, it is not.

Customers, in this case, might mean the record labels themselves. In Ireland, most developing acts would admit, privately at least, that they would like to be signed or licensed to a major record label, or their subsidiaries, in order either that they might reach a wider audience or they might be rewarded for their artistic endeavours (or both). 'Marketing' and 'promotion', therefore, could mean the process(es) of making potential customers (i.e. record labels and their scouts as well as the listening public) aware of the availability and qualities of a product (i.e. the act or the song). It could also mean the 'word of mouth' recommendations made by people to their friends and acquaintances, whether online or offline.

In its earliest days, Ireland's National radio and television broadcaster *'Radio Telefis Eireann'* (RTE) was a taste-maker. The station could, and still can, make the career of an act simply by allowing their material to be broadcast while limiting the broadcast opportunities for 'competitive' acts. Richard Pine [**35**] describes how RTE (or *'Radio Eireann'* as it was then) and its offshoot the *'Radio Eireann Studio Orchestra'*, afforded the new listenership of the post-war years the opportunity *'for making decisions – or at least establishing viewpoints – on the preferability of one cultural genre to another'*. While Pine was writing about a bygone era when radio was in its infancy as a communications medium in Ireland, RTE (indeed, all radio) still affords this opportunity to its listenership.

This is an accepted fact of life in the music industry in Ireland (perhaps worldwide) and the relationship between the musician and the broadcaster is one that still offers huge potential for the development of an act or, as Pine said, the chance for the public to *'judge the preferability of one act to another'*. It is in affording that opportunity that radio can best be of service to the act in its development. Insofar as the wider industry is concerned, radio also offers the opportunity for record labels and scouts to become aware of new talent and to *'judge the preferability of one act to another'* [**35**]. This is where airplay statistics are important. Even more important are airplay statistics that are accurate and representative

If it is possible to control what the airplay reports show, then it is possible to control which acts are given the exposure within the music industry that they need to promote their career. While it is fair to say that airplay statistics are post-event, in that they are produced after an act has had some broadcast exposure, it is also fair to say that these statistics provide the basis for at least some further opportunities. Industry-watchers use these reports as both a monitoring device to identify where their acts may or may not be gaining exposure and as a sort of barometer of taste, particularly at local stations, which tend to be more accessible to independent artists.

The advent and popularity of music video television certainly has made it more difficult for independent artists to become successful, mainly because a large proportion of the music-buying public now consumes more of its music on television, which is harder for unknown artists to break into. Radio, on the other hand, is still relatively easy to gain exposure on, particularly independent and/or local radio broadcasters. In Ireland, approximately 86% of adults listen to radio at some point every day according to official figures [**36**]. Radio is, therefore, still a major factor in the development of an artistic career, albeit a factor that might be expected to diminish over time as younger audiences who consume music on the Internet as much as anywhere else become adults and such listening habits become more prevalent in the survey demographic.

3.2.2 Sales charts

There are some common metrics that are used in the location and evaluation of potential new talent by a record label or their scouting networks. One such metric is the sales chart. When we hear that a record has reached a certain number in the charts, this is usually referring to the ‘official’ sales charts. Of course, the term ‘official’ may often be a misnomer. In Ireland, for example, the ‘official’ sales charts are compiled on behalf of the ‘Irish Recorded Music Association’ (IRMA), which is a representative body for the actual record labels who release the records. Given that a lot of radio play has been, and still is, based on sales charts, this gives rise to the opportunity for manipulating the sales chart in such a way as to give the impression that a song is performing better than it really is.

This manipulation of chart position may be designed to persuade record labels that the act in question is ‘popular’ on the underground scene. However, it may also be done by a record label. While a record label may not be manipulating sales in order to persuade anyone that the act has some ‘selling power’, they may wish to see their particular product reach a higher number on the chart so that the song is likely to generate more radio plays. This generates more awareness among the public who, in turn, buy the song because they (a)heard it more often and like it or (b)believe the act must be popular because it featured in the charts or the radio playlists. The act, then, becomes more popular and generates more sales which generate more awareness which in turn generates more popularity and so the cycle continues. More interestingly, because many commercial radio stations base their playlist at least in part on sales charts, and because higher positions in these charts lead to higher rotation on major broadcasters, higher royalty payments can be generated in the longer term.

An act may arrange the purchase of all of its own material released for sale. It may make a loss on this process (although the potential loss is much lower since the advent of charts that allow digital-only releases) but this loss can be offset by the fact that the higher sales figures will result in a higher chart placement and by extension to higher rotation on radio and higher royalties. The royalties may even pay for the losses incurred in the purchasing of the material while the whole process simultaneously increases public (and industry) awareness of the act. Of course, there is no reason why even established acts could not use such a process to ensure higher placement, high rotation and higher sales along with higher royalties. As mentioned in Section 3.1.1, an experiment was carried out by *'IrishUnsigned'* whereby a release was manipulated onto the official sales chart. This process was repeated by *'IrishUnsigned'* on behalf of other artists, both commercial and independent[i].

Other than manipulation of the data supplied to the public as the 'official' charts, there is another simple logistical and commercial problem faced by any developing act that has no connection to a record label. In Ireland, the sales data from approximately 285 retailers[ii] (6 of whom are online retailers) is compiled into the official charts, with various 'modifications' performed to the data (admittedly) by the organisation compiling the data for various reasons. This does not include all of the sales in all of the retailers selling into Ireland. Indeed, it is not even all of the sales in any given retailer. Instead, it is the number of qualifying sales in these retailers.

All sales must meet certain criteria to qualify for inclusion in the sales charts. The simplest one is that the physical item (CD, DVD) sold must be barcoded. Not only do many acts not realise this but, even if they did, it automatically disqualifies any sales of an item that was 'home made'. Moreover, not many acts are aware of the route they must take in order to register a barcode for their product. This raises a related problem: how do developing acts know which retailers fall into the category of those from whom sales data is collated and how can they then ensure that these retailers will stock the product, given that it is from an 'unknown' and not backed by a marketing campaign from a record label or distributor? It is logical to assume that record labels and their affiliates would know where

[i] While the information about the artists concerned in manipulation of chart position is provided confidentially, they include a number of highly successful acts.
[ii] From private communications with 'Chart Track UK', compiler of the official Irish music and games/software sales charts.

to stock their product (and how to achieve this), since they contract an outside third-party (UK based Chart Track Limited) to collate sales data and publish the chart.

3.2.3 Airplay charts

A second common metric used for the purposes of assessing the relative success of an act is the 'Airplay Charts' published privately to clients by Nielsen [**71**] who also publish the 'Radio and Records Airplay Charts' for Billboard magazine in the US [**37**]. Airplay charts are only used to evaluate the relative exposure of a song in the broadcast environment. However, this in turn reflects the overall popularity or performance of the act. Airplay charts are generally a report on the songs played on radio stations in a given period ranked according to how many times each song was played. The charts can be weighted according to the audience reach or territorial importance of the broadcaster on which it was played. Therefore, a play on a mainstream, national station would generate more chart points than a local, specialist or community station. While this may seem like a useful metric to use when monitoring the development and potential of an act, it is inherently limited and may actually be counter-productive. An understanding of the techniques used to monitor output will help to illustrate the problem.

One provider of broadcast monitoring services in Ireland is 'Nielsen Music Control'[i]. 'Nielsen Music Control', like IMRO, is not an official body. Instead, it is simply a private organisation set up as a result of a perceived commercial opportunity. The organisation monitors radio output and produces reports on material broadcast. However, it is in the way this is achieved that the limitations become obvious. In order for a piece to be tracked, it must first be sent to 'Nielsen Music Control' by the act (or their representative). It is then put through a proprietary fingerprinting process[i] to create a pattern and this pattern is added to their database. The output of most, but not all, Irish radio broadcasters is then monitored and the system attempts to pattern-match the output with the stored patterns, recording the time, date and source if successful.

---

[i] http://www.nielsenmusiccontrol.com/

‘Nielsen Music Control’ is not a well-known organisation and, other than the major record labels and their subsidiaries, plus a few well-informed independents in Ireland, few acts even know that this can be done, let alone how they would go about achieving it. In a survey of almost 500 acts, of which there were just over 100 respondents, less than 10 knew how they would get their music added to this monitoring system[i]. Furthermore, while there is no cost involved in getting a work added to the monitoring system, there is a cost involved for the provision of reports on the plays of those works so acts would not necessarily be able to pay for monitoring reports on a year-round basis. Those who have used the system have used it for short term purposes (i.e. monitoring the airwaves during the period immediately surrounding a release). Record labels and others in the music industry do, however, regularly pay for airplay statistics [**71**].

3.3 Impact of the current systems on developing acts and industry

As mentioned earlier, systems such as the ‘sales chart’ and ‘fingerprint recognition’ airplay chart that rely on some foreknowledge of both the existence and importance of the system, while at the same time being unable to take into account every possible release as it becomes available are inherently flawed. It is possible that these flawed systems of identifying new talent and of analysing the relative success of such new talent may be detrimental to the overall long-term health of the Irish music industry as both an employer and an export tool. Consider the following hypothetical scenario:

An artist at the early stage of their career releases their first record and achieves some exposure on local radio stations. The record is available in local ‘bricks and mortar’ retailers as well as online retailers. The awareness generated by local broadcast exposure along with word-of-mouth (one of the most important marketing techniques available to developing artists) means that the artist has soon reached the point where they are relatively well-known in a small geographic area. The artist then releases a follow-up record and

[i] Primary research conducted personally through *‘IrishUnsigned’*.

begins work on a full length album. The follow-up release also manages to achieve local broadcast exposure and the artist is even featured in-studio, perhaps in an interview or live performance of their work. Awareness of the artist is now much greater than it was, albeit still local.

The artist in question then manages to accumulate the finances needed to record, release and promote a full length album which will be likely to cost above €10,000. At this point, the artist really needs to be promoted to the rest of the broadcast and record industries. However, nobody outside his or her small geographic area has even heard of the artist because industry organisations do not routinely monitor the output of local radio broadcasters. To make matters worse, the artists does not appear on any airplay or sales report. This happens because their local broadcaster is not monitored by collective rights or airplay monitors and their sales, while comparatively good, were on non-affiliated retailers systems. As far as the wider industry is concerned, they simply do not exist. When the artist finally releases their album it is likely to sell relatively badly and cost them a heavy loss. This means the artist is less likely to try again so their career comes to a halt.

Had the above artist had the good fortune to have their two early releases and interview featured by a broadcast which was monitored for airplay or royalty purposes and had their sales been processed via only affiliated retailers, the artist may well have become relatively widely known and maybe even sparked some interest (or at least awareness) within the industry. Their album release might then have been more successful and this in turn might also have sparked interest from the industry, regardless of whether or not the artist desired to sign a contract. Selling as few as 2500 copies of an album would more than likely be enough to recover their costs[i] and encourage further artistic endeavour, as was envisaged by the early Copyright proponents. Moreover, even if their success had not materialised, they would still at least have generated some royalties from the exposure they

[i] Assuming an album costs €10,000 to produce and promote, and was sold for €10, the costs would be recouped as the artist's share of the sale of this album in stored would be c. €4 - €5. Therefore, 2500 sales would generate between €10,000 and €12,500.

did manage to achieve. In the above case, the hypothetical artist would receive no royalties whatsoever.

The scenario outlined above is depressingly common in reality, not only in Ireland but in most countries. In Ireland alone, for every relatively successful act such as ‘Republic of Loose’, ‘Director’ or ‘Delorentos’ there are hundreds of less successful acts like ‘Hoarsebox’, ‘Reemo’ or ‘Jaded Sun’ who have not had quite the same level of exposure or return on their investment. In hundreds of cases, acts have either had to persevere in the face of little encouragement or give up. The organisations which collect and distribute the tens of millions of euro paid in licensing fees by radio broadcasters every year, have the ability and the legal duty [**23**] to ensure that those artists who provide the material from which broadcasters generate their income streams are correctly rewarded.

3.4 Overview of royalty collection and distribution in Ireland

Irish artists are the third-highest selling category (by nationality) in world music sales. Irish-originated artists sell more music than all other nationalities, regardless of population or potential catchment, except artists from the United States and the United Kingdom [**35**] [**39**]. All this in an economic sector worth in excess of €100bn [**38**]. This is a point worth noting as it suggests that the current system is inherently unfairly weighted against Irish artists, particularly abroad, when the wider music industry is considered.

The three existing agents for the licensing of copyright music in Ireland are IMRO (the Irish Music Rights Organisation), RAAP (Recorded artists and Performers) and PPI (Phonographic performers of Ireland). Any outlet that plans to ‘make available’ any of the works that are protected under Copyright and are even partially owned by IMRO’s members must apply for an IMRO license to make those works available publicly. Similarly, outlets are required to obtain licenses from the PPI (and, through the PPI, from RAAP) for public performances. This is because IMRO members are not necessarily RAAP/PPI members and *vice versa.* IMRO members are songwriters/publishers and RAAP members are performers/musicians. Sometimes, of course, a writer may also be a performer. In this regard, music users likely to require a license to ‘make available’ copyrighted works are:

- Radio and television broadcasters, as well as internet-based webcasters;
- Owners of public houses, hotels, and other venues where music is performed live or pre-recorded whether by the owner of the music or other persons;
- Owners of stores, workplaces, restaurants, shopping centres and other locations where music is used in the background;
- Organisers of events in public areas, such as parks and streets (for example, local authorities) if the music is being made available to the public.

In fact, almost any type of outlet where music can be heard by any person(s), other than in the ‘domestic circle’ as defined in law, must all have permission from IMRO (and RAAP, and PPI as appropriate) to use the works of said members.

When an outlet pays for a license to use music owned by members, this money is then pooled and distributed to the owners of all music subsequently made available by all license holders [**25**]. The license fee received for a particular outlet over a given period should be distributed pro-rata amongst all the content owners who have had their works made available by that particular outlet. Essentially, if the work of one artist is played twice as often by license holders than another artist, then he or she should receive twice the amount of the royalty when it is distributed.

While this appropriation of license fees to owners may be a relatively simple task when a one-off license is sought, perhaps for a fireworks display that is set to music, or for a concert where the music is known and consistent, it is not feasible for IMRO or any other Organisation to keep a complete and perfect record of all the pieces used in all of the outlets for whom it has issued licenses. Therefore, they cannot distribute the license fees received as royalties to all of the correct owners. IMRO readily admits this limitation[i]. There are, however, systems and processes in place to generate playlist data from many outlets such as radio and television and then extrapolate the overall distribution ratios using the limited sample data collected.

3.4.1 Limitations in the current system

Unfortunately, the systems and processes used by organisations like IMRO are by definition likely to overlook a certain section of their members when distributing royalties. This leads to an ever-increasing disadvantage to this section in a permanent downward spiral. It is, moreover, this particular section that the entire concept of Copyright was evolved to protect. As explained in Section 2.2.1, a work may be broadcast numerous times and receive no royalty fees whatsoever if its use is not reported in sampled data while another work, broadcast once, may receive far more royalties that it is entitled to, including some royalties due to the unreported work.

This is of course an extreme example but it is certainly indicative of the way that distributions are calculated that this does, to some extent at least, happen in the Irish territory. This can be seen from the information provided in Table 2.1 in Section 2.3. There is no reason to believe that the same does not happen to some degree elsewhere.

In terms of the more accessible ‘community radio’ sector, the fees generated are only a very small amount in comparative terms. The difficulty in collating play data is exacerbated in the case of community broadcasters by the time, effort and cost that can be involved in collating such data. IMRO find it is not economically viable to collect play data from smaller community broadcasters. Instead they simply add the fees generated from community broadcasters to the comparatively large and thinly-spread RTE pool of license fees. RTE is the biggest broadcaster with the widest range of broadcasted material in the territory. This, of course, means that the plays that developing artists and writers *do* receive,

[i] From private communications with IMRO.

which very often happen on easily-accessible community and local radio stations, will be overlooked. Instead, the royalties due to these developing artists will be added to the RTE pool and paid to the very well-known and usually very industry-aware artists played on RTE's various channels, thereby compounding the problems already faced by developing artists in this area.

It is reasonable to presume that the more organised, informed and commercially aware artists, writers and publishers will all be able to navigate these obstacles and will take steps to ensure that their works are properly recorded and reported to the licensing agent. Moreover, these commercially-aware sections of the membership are likely to be involved on a regular basis in the music industry as this is the only real way that such concepts come to light as far as artists are concerned. Newcomers to the industry are often underinformed about Copyright[i] and how it is administered. Even when they are informed enough to have their works copyrighted and register as members of organisations like IMRO their rights cannot be administered perfectly. They simply cannot be and IMRO acknowledge as much in its members handbook[ii].

It is up to the artist to ensure that they are paid the correct royalties from the 'pool' of any given license fee received. Of course, like IMRO and its peer organisations, artists are unable to monitor all radio and television output, as well as all publicly-heard music all over the (licensed) world and are therefore at the mercy of the agents. Given that more commercially-aware members will be actively submitting data to the agent, and that the agent's collection and distribution system is already (accidentally, at least) inherently weighted against the uninformed developing Artist, the disadvantage suffered through lack of awareness is magnified, often to the long-term detriment of artists who see no return even when relatively successful in terms of the Irish marketplace.

---

[i] From primary personal research conducted through *'IrishUnsigned'*.
[ii] From private communications with IMRO. Also, see the IMRO members' handbook, Section C, paragraph 2(c).

In order to overcome the problems caused by the systems currently used to collate information, and to make the entire process of collection and distribution of royalties fairer (albeit never perfect), in particular for works played less often and for specialist or developing musical shows, a complete and accurate list of *all* works broadcast at *all* times on *all* stations needs to be made available. This is unlikely to ever be possible, but modern technology should be able to produce a system fairer than it currently is. Currently, as illustrated by the information shown in Table 2.1 in Section 2.3, the distribution of the entire license fee paid by a broadcaster may be distributed in its entirety to the owners of less than 8% of the pieces played.

Of course, non-broadcast outlets such as shops and factories simply cannot be monitored at all so estimates and extrapolations are used to distribute the license fees collected from these groups of outlets. This is sometimes done using the same metric derived from the extrapolated data generated by incomplete radio and television playlist data. This assumes that a large number of outlets will be using radio and television for public performance, without regard for those that might use CD or those that never play music other than live music. The pool of fees generated from these sectors is therefore distributed as unfairly as that of broadcasters. There are always going to be limitations like this, but it is obvious that the more complete the available data is, the more accurate the distribution will be.

IMRO generally collects a standard 'blanket' license fee from most broadcasters[i] as does the PPI along with myriad fees of all sizes from other users of music for public performance (see *Appendix 1*). This may include record stores, live music venues, cinemas, hotels, and restaurants where music is used as part of the 'ambience' and so serves a commercial purpose. It also includes anywhere that music is listened to, outside the 'domestic circle', including workplaces, factories, and retailers and extends in theory even to buses and cabs. Each type of music use is treated in a different manner, depending on the

[i] Some broadcasters, such as Community Radio Stations are either exempt from having to pay a license or may not be exempt but are assessed to pay a license fee of zero

use of such music, what purpose it serves and what revenue it generates. Some users, such as restaurants and hotels, are charged a flat fee dependent on size and potential audience. Others (for example, radio stations) are charged a percentage of their income if they rely on music to generate that income. The entire pool of license fees collected, after deduction of overheads, administration and other costs, is then distributed to Copyright owners according to whatever data IMRO has been able to collate from broadcasters. Given that IMRO alone generates c. €36 million in license fees (increasing year on year) [**26**], the revenue is not insubstantial.

3.5 The importance of fair and equitable rights administration

Some relevant statistics might serve to illustrate the reasons for, and the potential scale of, the long-term benefit to organisations like IMRO, its members and the Irish economy in general:

1. The royalty collection industry in Europe alone is worth an estimated €5bn in 2004 [**8**] and is increasing in size. IMRO's own revenue increased 10% in 2007 [**26**]. This figure of €5bn excludes the rest of the world and of course excludes significant public performance royalties that might accrue from the US.
2. IMRO collected only approximately €3.5 million in royalties from affiliated organisations in the rest of Europe (including the UK) [**26**], suggesting that Irish-originated IMRO members accounted for only a miniscule amount of radio/television output and live performances. Given the success of Irish-originated artists [**39**] [**40**], this is obviously open to debate. IMRO itself admits that it faces major obstacles trying to recover royalties for its members from public performances abroad because it cannot quantify them and because the US continues to delay introducing legislation to protect public performance royalties within its borders [**26**].
3. In the UK music sales market, 'Irish-originated' artists were ranked a cumulative third place in the overall UK sales statistics [**39**] in all of the years from 1986 – 1994, except 1989 (when Irish artists were ranked 4^{th}) in terms of sales. If radio

output reflects, to an extent, sales statistics then it would be logical to assume that Irish-originated artists made up a large proportion of public-performances. From IMRO's annual results, it is apparent that, even many years on from these statistics, income from UK public-performance royalties paid to IMRO for use of its members' works in the UK should be a lot higher than the €1.5 million shown in their 2007 results [**26**].

4. According to the Music industry governing body, *'Revenues from public performance and broadcasting income grow incrementally every year'* [**40**] so there is ever more reason for both more accurate public-performance playlist data.
5. *'The fact that Irish people use English is often cited as increasing our vulnerability to Anglo-American mass culture. This is so, but it also increases our opportunities in the vast English-speaking market, the most affluent in the World'* (former Irish Cultural Minister Michael D. Higgins) [**41**]

If the Irish royalty-collection industry were to take a lead in the development of an open-source, transparent and accountable measuring and reporting system, then Irish artists (meaning, by default, their more established and well-known members) could easily benefit exponentially from being at the forefront of this development by generating an increased inflow of revenues from European and US royalty revenues, notwithstanding that this takes no account of other markets, such as China, which is apparently moving towards providing more protection for intellectual property of all types [**42**] [**43**].

The problem in implementing a system to identity all musical works made publicly available by all licensed broadcasters is not one of technology or of identification of the beneficial owner of the piece. Instead it is one of logistics and motivation. The system under consideration in this document is one that deals with automated identification of the work 'on the fly' in a realistic radio or TV scenario. It also addresses the very difficult logistics of monitoring use of Copyrighted material in non-broadcast outlets such as retail, industrial and tourism. It would be a simple step then to assign the correct royalty payment to content owners according to a more fully-monitored, accurate, open and transparent reporting mechanism.

3.6 Identification of a unique piece of audio in the Music Industry

In order to perform the function of this system and identify any particular piece of music, the identifier used must be capable of being applied to only one piece of music. While the initial choice might be to use a barcode-based system, this soon proves inadequate and outdated. In reality, a barcode identifies a physical object. In terms of music - released publicly or otherwise - this barcode would generally be assigned to a physical CD or DVD containing the music, rather than to the actual music itself. This is an important distinction for two reasons. First, a CD can hold more than one song and all songs would be represented by the same barcode. Second, physical CDs are likely to eventually become as rare as vinyl recordings and it is now common practise for artists and songwriters to release songs in a digital-only format which owners can then add to a CD if they choose. Songs released only to the music industry or broadcasters, for the purposes of promotion or radio-play are also, in the current climate, delivered on CD. There is generally no barcode for a digital release.

On hearing a track, most people are very capable of identifying an artist. They might know the artist and track. They might know the album it was included with. They might even know the year it was released. However, this is only useful as a means of identification when communicating such details to other people as it is possible to correct any assumptions or misunderstandings and incorrect decision-making is easy to rectify.

In the music industry, something more specific and less error-prone is required in order to ensure that royalties are not paid simply to the first or most well-known artist who might have performed a song. Similarly, even a single work can have multiple Copyright owners as described in Section 2.3

Every recording of every version of every work ever released to the public, whether released by a record label, an artist, a singer, a manager, a performer or any other person or organisation whatsoever, for any reason, should be uniquely identifiable in a manner that makes it clearly different than every other work. Moreover, it should be possible to differentiate it from every version of the same work, or every version that includes portions of the work in question (such as is the case with remakes and sampling, for example). In fact, even if the recording contains no audio, or no deliberate audio, it would need to have the capacity to be catalogued, as evidenced by John Cage's *'Four minutes thirty-three seconds of silence'*, discussed in Section 2.2.

3.6.1 The ISRC Code

Fortunately, there is no need to invent such a cataloguing system as such an identifier is already in existence and internationally recognised. It is called the *'International Standard Recording Code'* or ISRC. This code has been in use for almost 20 years by the membership of the *'International Federation of the Phonographic Industry'* (IFPI) which is the representative body for almost 1500 record labels in around 70 Countries. As its name suggests, the ISRC code is an International Standard, published by the *'International Organisation for Standardisation'* (ISO). The current revision of the ISRC Standard is ISO 3901:2001 [**44**]. An ISRC code identifies a song by using four segments as follows:

- A 2-character Country Code (e.g. 'IE' for Ireland)
- A 3-character Registrant code for the record label([i]) that assigned the ISRC
- A 2-digit year of the assignation
- A 5-digit sequential number of the recording within the year

[i] Any individual or organisation that issues a recording has the right to assign an ISRC code to that song, including the artist themselves. Their only requirement is that they are assigned a registrant code from the PPI. The PPI should then be informed of all releases from the Registrant.

Consider as an example the Irish record label *'Vinyl Destination Records'*. It has been assigned the three letter registrant code 'AVY' for audio releases and 'VVY' for video releases. All of the audio releases from this content producer would begin with 'IE-AVY' followed by the year and the sequential release number within that year. Therefore, since *'Vinyl Destination'* released a song entitled *'Angel from Heaven'* by a band called *'DeXtra'*, which was the first public release from this label in the year 2004, its ISRC code is 'IE-AVY-04-00001'.

A song's ISRC code is a unique identifier and can be used to identify the song by referencing a database of ISRC codes stored for such a purpose. In Ireland, this task falls to 'Phonographic Performance (Ireland) Ltd' (PPI). The example song can therefore be identified by referring to PPI with the above-mentioned ISRC code for cross-referencing against their national database. One example use of the ISRC code is in ensuring that only material with an assigned ISRC code actually appears in the sales charts. Multiple remixes of the same song should have multiple ISRC codes but may be counted together for sales-reporting purposes.

3.7 Summary

This chapter described some metrics by which the early career of an artist is measured by the public and the music industry. Explanations of the impact their incorrect or inefficient production can have on the artist's career were explored. The difficulties that can sometimes be encountered when attempting to uniquely identify artistic works, particularly in the broadcast environment, were outlined. The standard identifier currently in use was then explained. In the next chapter, the human auditory system is introduced and sound processing explained. This will provide the level of understanding necessary to fully understand the effectiveness of the developments described later in this work.

Chapter 4: Sound, hearing and data-hiding

In this chapter, the human auditory system, and how it responds to sound waves, is described. Psychoacoustic concepts are introduced and explained briefly, concentrating on those concepts that might be of relevance in the current work. Data-hiding concepts are introduced later in the chapter in order to provide grounding and to differentiate data hiding from other forms of data security. Finally, basic digital signal processing concepts are described and an overview of how they apply to broadcast monitoring is provided.

4.1: The Human Auditory System

Sounds heard in the real world are continuous waves of varying amplitude and frequency. Waves with higher frequency are described as being 'higher pitched' than those with lower frequency, while waves with greater amplitude are louder than those with lower amplitude. All sounds can be described using combinations of frequency and amplitude values. Even very complex sounds such as an orchestral piece involving thousands of instruments can be described in this way. The overall sound is created by the sound waves of individual instruments 'superposing'. Superposing is essentially the same as overlaying or, in the case of sound waves, adding the waves together. In reverse, given a complex sound, it is possible to calculate the individual waveforms of the component parts [**74**].

In general terms, sound is considered to be the perceived impact of a disturbance on a medium. The physiological manner in which the human hearing system converts waves into nerve impulses to be perceived as sound is beyond the scope of this thesis. However, by way of illustration, if we clap our hands (this causes the disturbance) we create a movement in the air between our hands (the air is the medium) which propagates away from the source of the disturbance, diminishing as it goes. We hear these sounds because they cause the ear-drum to vibrate (see Figure 4.1). The ear drum converts vibrations into nerve impulses, with the responding nerves reacting to predefined frequencies. The frequency components are transmitted to the brain for processing which is done by analysing which nerves have fired and in what order.

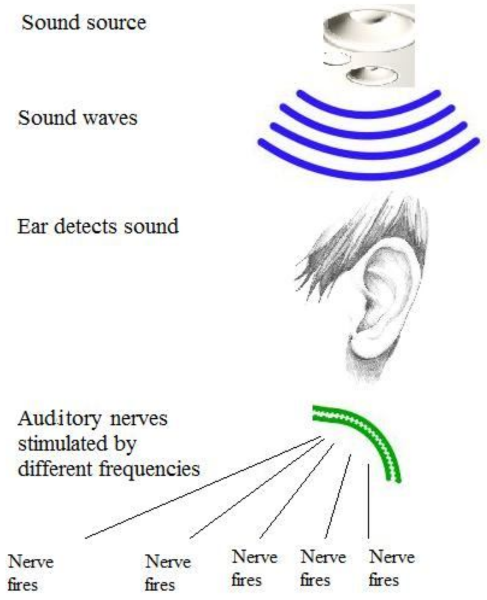


Figure 4.1: Representation of the way in which humans process sound.

4.1.1 Visualising sound waves

Normally, when a sound wave is considered, the intuitive (graphical) representation of the wave is in terms of a wave-form along the timeline as produced, for example, by the sound of human speech (Figure 4.2a) or a heartbeat represented in a heart monitor (Figure 4.2b).

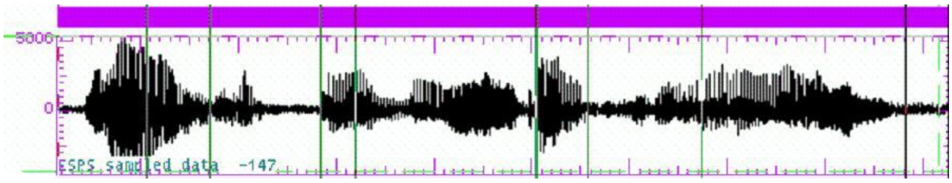


Figure 4.2a: A waveform representation of a voice saying 'Will you have marmalade or jam?' [45]

Figure 4.2b: A sample heartbeat output on a heart monitor [46].

However, there is no reason why the information contained in the representation cannot be represented in other formats. In fact, it is often more useful and easier to visualise sound in other domains, particularly the frequency domain. In this way, the frequency spectrum of the sound can be analysed. The frequency-domain representation obtained via the Fourier transform presents the signal in terms of the sum of a number of simple sinusoidal waves of different frequencies, amplitudes and phases which make up the overall sound [**54**].

The principles that allow the conversion of continuous analogue sounds from the time domain into the frequency domain facilitated the evolution of digitised sound into modern formats. One such commonplace facility is the compression of sound from CD-format ‘Pulse Code Modulation’ (PCM) into a more useful compressed format (e.g. MP3) which has become the *de facto* standard for modern digital distribution of sound files. The theory behind the transformation from time to frequency representation (and, perhaps more importantly, back again) was first demonstrated in a paper submitted by Jean Baptiste Joseph Fourier in 1822 [**47**]. The wider area of Fourier theory is so great that it is beyond the bounds of this paper. Having said that, it is referred to continuously as there is really no way of talking about the relationship between time and frequency domains without using Fourier theory and the ideas that were developed using it.

Sound, as we perceive it, is made up of complex collections of less complex sounds which can theoretically be decomposed to the lowest possible level: individual sinusoids of different frequencies, amplitudes and phases. The complex sound can, in fact, be described as a series of individual waves of different frequencies and phases, interacting with each other. Figure 4.3a shows the fundamental frequency of ‘middle C’ at 440 Hz, plus its first 4 harmonics at 880 Hz, 1320 Hz, 1660 Hz and 2200 Hz. Figure 4.3b is the combined sound generated by these five components.

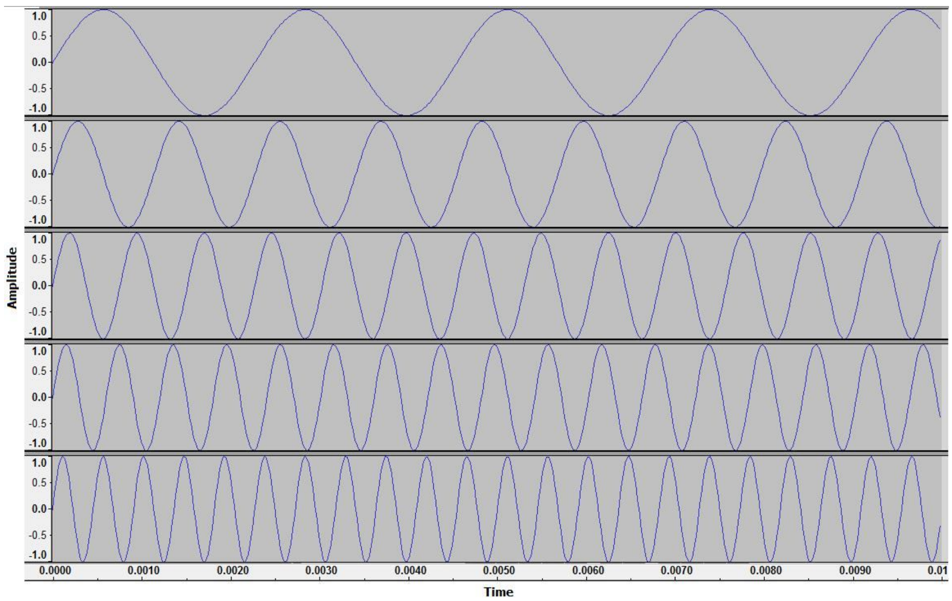


Figure 4.3a: Synthesised fundamental frequency and first four harmonics of 'Middle C'.

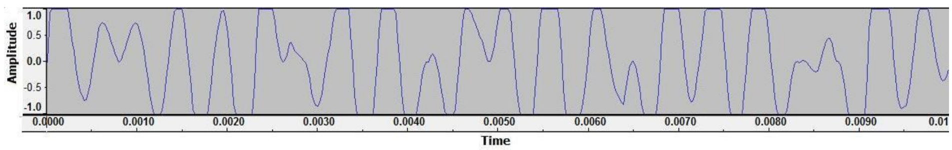


Figure 4.3b: The composite wave generated by adding the fundamental and the next four partial frequency waves together.

Any periodic sound can be described in terms of a set of underlying components. For example, the complex periodic sound in Figure 4.3b can be described in terms of the individual amplitudes and frequencies of the components in Figure 4.3a.

4.2 Psychoacoustics

It is generally accepted that humans hear frequencies in the range of 20 Hz to 20 kHz, deteriorating with age. Similarly, it is accepted that we hear amplitudes (or 'loudness', 'sound volume' etc) in the range of -6dB to 130dB (the latter of which is generally considered *painfully* loud). Ordinary daytime noises generally vary between 0dB and 100dB. We can also detect sounds more accurately the longer they are played.

There is, it must be pointed out, a very distinct difference between what can be *heard* and what can be *detected.* The priority for most professional users of music, from artists to record labels and TV/radio broadcasters is **-** correctly **-** on audio reproduction that is as close to being ‘imperceptibly different’ from the original as possible. Any difference that is perceptible will obviously alter the reproduced sound in a way that makes it unacceptable for professional use and so should therefore be avoided at all costs if the resultant sound file is to be used in a professional manner.

4.2.1 Auditory masking

When a sound that can normally be heard by a listener in isolation is no longer audible in the presence of another sound, it is said to be masked. There are various forms of auditory masking and they are employed for a variety of tasks, often completely unrelated to the field of audio recording. Sound masking principles are employed to create either a passive or active barrier to unwanted or distracting sounds. These may be, for example, traffic noise, building noise or machinery noise [**48**]. Imagine, for example, even a relatively quiet intermittent beeping sound in a quiet room. It will obviously command the attention of most listeners and become a distraction. However, if there is an underlying sound between the listener and the beeping sound, it will often shield the ear from the distraction. These techniques are used in areas where sound travels beyond an acceptable distance, either for reasons of confidentiality or interference, and where intermittent sounds may cause unwanted distraction.

One simple example of the masking concept masking is amplitude masking. While a listener may comfortably hear an Opera singer from the stage in an Opera house, the same singer performing at the same volume (or amplitude) might be impossible to hear standing in front of a Jet engine. The reason is simply that the amplitude (loudness) of the engine is much higher than that of the singer and therefore drowns it out. Similarly, normal conversation may be impossible to hear even in the presence of a rumbling diesel engine, a much quieter masking sound than a jet engine but nonetheless louder than normal speech. This form of masking is referred to as ‘simultaneous masking’ [**49**] as it happens when two sounds are simultaneous to each other.

However, there are also non-simultaneous masking phenomena. These occur when a ‘strong’ sound masks a weaker sound that precedes or follows closely in time. If the weaker sound is masked by the stronger sound that follows, this is said to be backward masking because the strong sound masks backwards. Conversely, if the stronger sound masks a sound that comes after it, this is said to be forward masking. Both are examples of ‘temporal masking’ or ‘time-masking’ effects.

4.2.2 Threshold of hearing

As mentioned in Section 4.2, the human auditory system is capable of detecting sounds in the range from 20 Hz to 20 kHz, with most of us capable of hearing a much narrower range of frequencies. However, it should be noted that ability to hear a given sound is not consistent across the range of frequencies. At the lowest frequencies, most sounds need to be very loud to be heard. In the mid-range, in which our hearing system is designed to be most effective, a much quieter sound can be detected. Finally, at the upper reaches of the range, high frequency sounds must be played at a louder volume than mid-range sounds in order to be detectable. There are accepted levels at which various frequencies will be considered to be equally loud, as illustrated in Figure 4.4 which ‘*specifies combinations of sound pressure levels and frequencies of pure continuous tones which are perceived as equally loud by human listeners*’ [**50**].

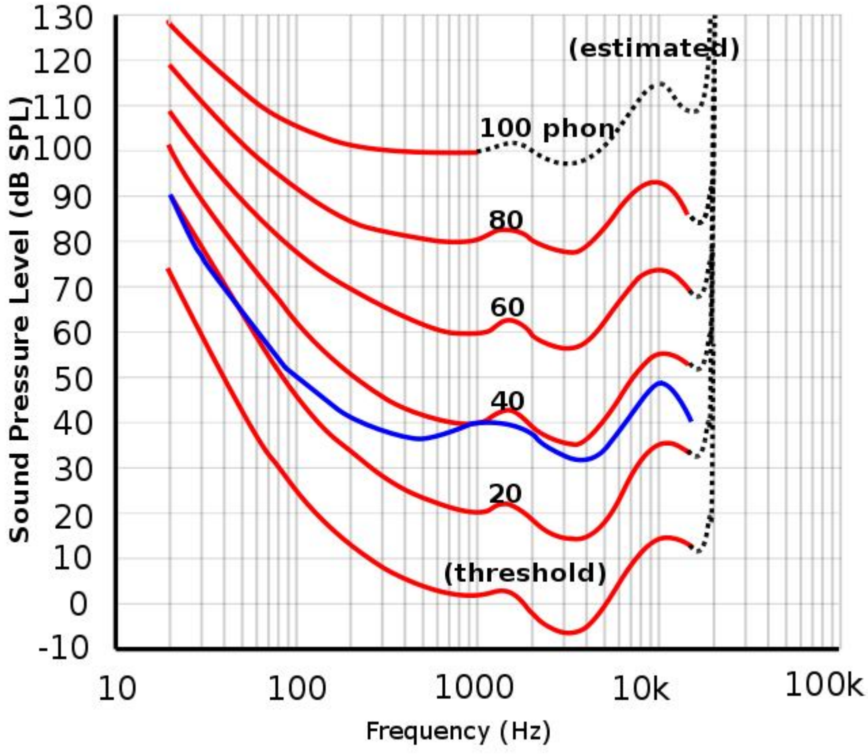


Figure 4.4: Equal loudness curves, indicating the sound pressure levels (SPL) at which tones of a given frequency (Hz) are described as 'equally loud'.

Figure 4.4 represents the sound pressure level or 'loudness', displayed in decibels (dB SPL[i]) on the *y* axis, at which tones at frequencies on the *x* axis are perceived to be equally loud in normal hearing (identified in 'phons' in this graph [**51**]). The 'threshold' curve indicates to lowest SPL at which a given frequency can be detected by the human ear. For example, at very low frequencies up to 100 Hz the curves are comparatively high and these tones must be played comparatively loud in order to be perceived to be just as loud as tones in the mid-range (e.g. between 1 kHz and 6 kHz) played relatively quietly.

Another way to analyse the information in the equal-loudness curves is to examine a horizontal line across from a chosen sound-pressure level on the *y* axis. Take for example, the value of 20dB on the SPL scale. As the horizontal line extends from left to right, the frequency of the tones become higher pitched. Note that the equal-loudness curves vary

---

[i] dB SPL is defined as twenty times the log of the ratio between the measured sound pressure level and a reference point. This reference point normally is defined as 0.000002 Newtons per square meter, the threshold of hearing [51].

from very far above the 20dB line, to somewhere just below it in the mid-range, before increasing above it again as it approaches the 10 kHz frequency range. This indicates that lower frequencies will be difficult to detect at 20dB while mid-range frequencies of the same SPL will be relatively easy for most people to detect. At frequencies above approximately 15 kHz they become almost impossible to detect. The curves represent equal loudness levels for a person with normal hearing [**50**].

4.3 Data-hiding concepts

Understanding of the concepts outlined in Section 4.2 helps to enable the development of a strategy to hide sounds in other sounds. Whether deliberate hiding or masking of one sound by another, or accidental obfuscation of a sound, the effect is the same. The question should be asked: why would we want to cover up the presence of one sound using another? Unlike subliminal visual messaging, as employed for example for a short while in television commercials, it is unlikely that many listeners would be able to detect a sound masked by another sound in the same way that a subliminal video message would be sometimes detected although apparently unnoticed. Also, when data is hidden in audio by the use of additional sound waves, the data is in a format indecipherable to humans so even if it was detected it would have no discernable meaning.

For millennia, people have sought ways to hide information from prying eyes. Using encoding of one sort or another ('cryptography') to change a message is a technique that has been common-place, particularly in military circles, for thousands of years [**52**]. In business and commerce, encoding and hiding information from unauthorised sources has become hugely important as a result of industrial espionage. Even in the entertainment industry, preventing unauthorised access or distribution is becoming increasingly important as a result of the recent rise in albums and movies made available on the Internet even before they are officially released. Protecting such content from access is a very important consideration for respective sectors.

The purpose of cryptography is to make information unreadable, or 'indecipherable', to anyone not intended as a legitimate recipient of the information. The message may even be visible, but in a way not easy to understand. For example, one of the simplest forms of cryptography is to assign numeric values to the letters of the alphabet according to their position, add or subtract a value across all the characters in the message and re-write it with the characters that are at the new value. Any unintended recipient of the new message will have no way of reading it as it no longer makes sense. The intended recipient - assuming they know the numeric value that has been added or subtracted - can simply reverse the process and decrypt the message. The process can be illustrated simply as follows[i]:

1. Start with the original message

'THIS IS A MESSAGE'

2. Write the message in numbers according to letter-values in the alphabet becomes:

T H I S I S A M E S S A G E

20, 8, 9, 19, 9, 19, 1, 13, 5, 19, 19, 1, 7, 5

3. Adding a value (for example, 1) to each digit produces:

21, 9, 10, 20, 10, 20, 2, 14, 6, 20, 20, 2, 8, 6

4. This, in turn, according to the alphabet placement of the new numbers above, becomes:

UIJTJTBNFTTBHF

To decrypt, the recipient needs to know what process was performed on the letters in order to be able to decrypt the message by reversing the process. The process of encryption is performed using a 'key' which, in this case, is 'add 1'. In the above example, the string '**UIJTJTBNFTTBHF**' would be decoded by subtracting 1 from its numerical alphabetical values and the new characters deduced from the new numbers.

[i] This simple encryption employs a form of 'shift cipher' or Caesar cipher'

4.3.1 Steganography

One of the major disadvantages of cryptography is the obvious one: if an unintended recipient intercepts the message after it has been encoded, they can then set about decoding it. Another early historical development to hide information from unintended recipients was the lesser-known technique of '*Steganography*'. While the purpose of cryptography is to change the content of a message so that interceptors could not read it, the purpose of steganography is subtly different: to hide the fact that there *is* any message. In its simplest terms, steganography requires that the sender hides the fact that he/she is sending any message and the intended recipient is the only one who knows that (a)the message has been sent and (b)where and how to find it.

Steganography means 'hidden writing' in Greek, as does cryptography. However, the difference is in transmission of the information. As far back as 440BC, according to Herodotus, details of a future attack were sent to Greece by Demeratus by writing them on a wooden panel and then covering the writing in wax [**53**]. Since wax tablets were a common writing implement, on which writing was clearly visible and this particular tablet had no visible writing (the message was under the wax), no suspicion was aroused. Another example from ancient Greece tells of Histiaeus shaving the head of a trusted slave and tattooing his scalp with a message, completely hidden when the hair grew back. In this scenario, perhaps even the intended recipient did not need to know of the existence or location of the message, only the carrier did.

In order to be able to transmit a message using steganographic techniques, the sender will usually need to have some form of 'carrier'. In the case of Demeratus the carrier was the tablet disguised as a blank wax tablet while in the case of Histiaeus it was a person. In both cases, and ideally with all steganography, nobody intercepting the carrier would be aware that there *was* a hidden message being transmitted so would not attempt to decode it. Of course, combining the two techniques would seem to be logically more effective – hiding an encrypted message rather than a plain message – but if the steganographic process is successful, the message does not need to be encrypted since the *presence* of the message is not discovered in order for the *content* to be intercepted. Nevertheless, given that most encryption can be broken, steganography seems a sensible measure to take where possible, and *vice versa*.

In recent times, cryptographic and steganographic techniques have been applied to the transmission of digital data to protect it from unintended recipients and unintended uses. *'Digital Signal Processing'* (DSP) techniques are used to alter the host digital file so that an alternative, but invisible (or inaudible), set of data can be embedded within it and later recovered and deciphered if necessary. In steganographic terms, the original digital file (image, music, voice, video etc) becomes the carrier analogous to Histiaeus' servant while the information embedded is hidden 'in plain sight' as the file is used publicly. The key difference between cryptography and steganography is simply one of visibility: in cryptography, the encoded or encrypted data is often visible or audible while in steganography it is not.

4.4 Digital Signal Processing concepts

While modern recording techniques may use digital formats for the production, storage and duplication of sounds, this has not always been the case. Previously, analogue systems (such as vinyl records and magnetic cassettes) were used for these tasks and there are pre-existing catalogues and libraries of musical pieces that are still in analogue formats. Like continuous time analogue signals, these can be converted (*digitised*) to be processed within a digital domain.

Digitising, in the audio context, is the process of converting continuous (analogue) signals into information that can be used by a computer. In its simplest terms, this process of converting, or digitising, a signal is achieved by taking a series of time-separated 'samples' of the waveform of the audio at a pre-determined time increment. This process is known as 'sampling'. At each time-segment the audio waveform at that point is given a value according to a scale. Once the whole signal has been 'sampled', the series of values at each time-step is used to reproduce the new, digital, signal.

Of course, the actual process is more complex than suggested above. For example, caution must be exercised when deciding how many samples are taken in any given time period as well as the 'scale' used to record the amplitude (or 'height') of the current waveform position at any given time. The following example (Figure 4.5a) shows a very simple segment of continuous waveform and is used to illustrate the main problems with digital sampling.

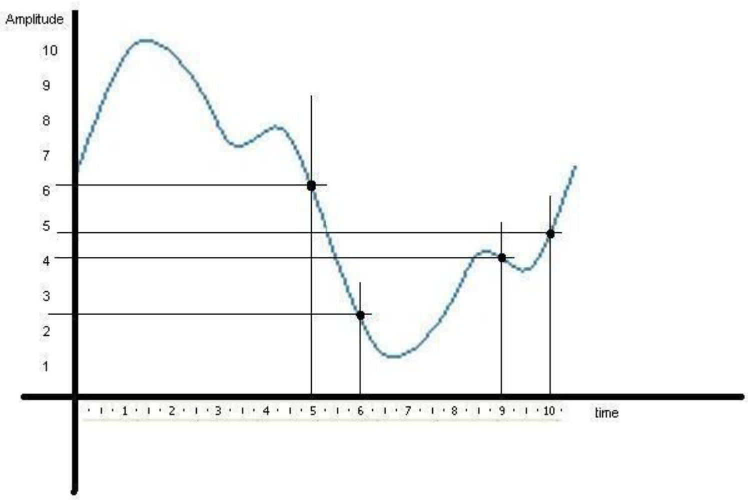


Figure 4.5a: A simple waveform showing selected points sampled at time-intervals 5, 6, 9 and 10.

If the waveform is 'sampled' at time-intervals represented by 1-10 on the horizontal axis, and a value taken from the corresponding intersection on the vertical axis, then the following values result for selected points:

Time:	Approx value:
5	6.1
6	2.4
…	…
9	4.1
10	4.8

Problems arise when the end result of the ‘sampling’ process is examined. Looking at a ‘digital’ version of the above waveform, it becomes apparent that it is likely that some of the ‘finer detail’ between samples can be lost. For example, the sample values for time-points 9 and 10 are 4.1 and 4.8 respectively. When plotted onto a graph they appear linear, with an upward bias between points 9 and 10 (4.1 to 4.8). However, looking at the original waveform, there is a downward wave segment towards an amplitude value of approximately 3.9 at time-point 9.5 (Figure 4.5b) before the increase towards an amplitude of 4.8 at time-point 10. If no record is taken of the amplitude values in between points 9 and 10, then there is no record available of the values in this region and the lower amplitudes in the waveform will be impossible to recreate.

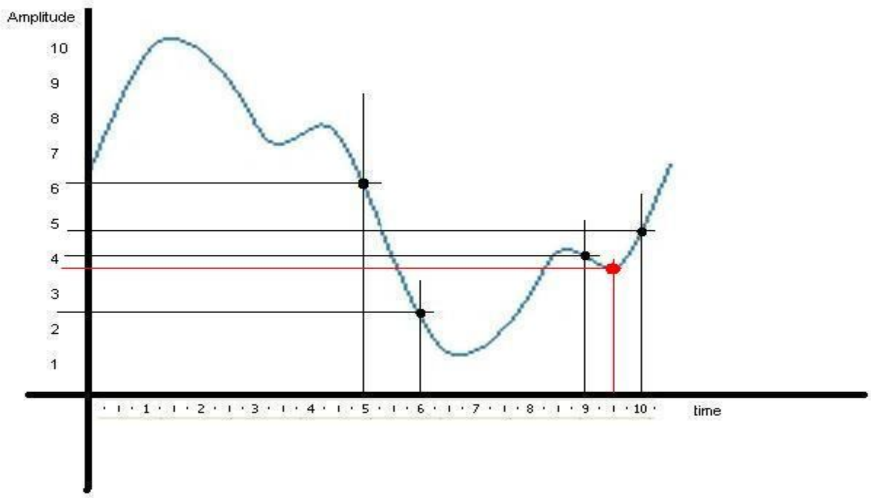


Figure 4.5b: The coloured point shows extra information gained by taking samples more often.

It is relatively simple to correct the problem of potentially missing out on important samples and that is to increase the sampling rate (i.e. the number of samples taken in a given time segment). If this was done in the above example, and twice as many samples were taken between time points 1 – 10, then the sample taken at 9.5 would record the lower value of the sound wave at that point, shown in figure 4.5b, and it could be more accurately reconstructed. Put simply, the more samples that are taken in any given time period, then the more accurately the record of the wave can be made and the more accurate the reconstructed wave would be. Of course, there are trade-offs. More samples means more information, more processing, more delay and more data storage required.

Problems also start to arise when examining exactly *how* to record the values for each 'intersection' of time and amplitude (or each 'sample'). Consider the sample at time-point 6 in Figure 4.5b. A decision must be made as to whether to record this with amplitude of 2 or 3. Assuming the value would be rounded to the nearest whole number, the recorded value would become '2'. Similarly, rounding up or down the rest of these selected values would result in the following table of values:

Time:	Amplitude value:
5	6
6	2
…	…
9	4
10	5

This 'rounding' of values is called 'quantising' and it can introduce errors to the digitised signal [**54**]. It cannot be avoided completely so must be minimised. One way of minimising the 'quantising error' is to use a smaller increment between the values on the vertical amplitude (y) axis. If, on the scale shown on the y axis in the Figure 4.5b, each whole number increment was sub-divided 5 times, the quantising error would be reduced as there would be values closer to the original value that can be rounded to. As mentioned, this would not completely solve the problem and quantising errors will always be present in the final output but at least the difference between the original and the digital signals would be smaller and therefore less noticeable.

As with increasing sampling rate, increasing the number of increments between quantising values will lead to more accurate data recording and more accurate reconstruction. However, there will again be a corresponding increase in the amount of data recorded and therefore the processing load and storage requirements. This can become particularly important at the time the signal is reconstructed from its digital record. There is no ‘ideal’ sampling rate or quantising step. There are, however, values that result in ‘acceptable’ audio reconstructed from the sampling process.

4.5 Digital audio files

When modern music is digitised for CD recording purposes it is sampled at a frequency of 44,100 Hz, meaning samples of the sound are taken 44,100 times each second This rate is chosen because it is considered the lowest sampling rate that meets the requirements of exceeding the Nyquist [**54**] rate, with added ‘headroom’ for safety. The Nyquist rate is defined as being at least twice the frequency of the sound that needs to be sampled. Since human hearing has a theoretical maximum frequency of 20 kHz, then the Nyquist rate for the highest perceptible frequency is 40 kHz. Since the amplitude value is stored for every sample, this equates to more than 9 million amplitude values per three-and-a-half minute pop song or more than 12 million for John Cage’s ‘silent’ piece referred to in Section 2.2.

The level of accuracy of the stored value can also affect the sound quality on reconstruction, as mentioned in Section 4.4, due to quantising. For modern digital systems, there are 16 bits allocated to store the amplitude value per sample. A 16-bit sampling system can store up to 65,536 (or 2 to the power of 16) amplitude values. An 8-bit system would be limited to 256 possible amplitude values. It is therefore evident that a 16-bit sample can store amplitude values to a far higher degree of precision than an 8-bit sample. This is, however, at the cost of higher computational cost.

Given that a 16-bit sampling process will store 2 bytes of information per sample (a byte is 8 bits, so 2 bytes are required for 16-bits), and the above sampling rate is 44,100 samples per second, it follows that each second of sampled audio results in 88,200 bytes of information. Across a three-and-a-half minute pop song this equates to over 18 million bytes. This is then doubled as there are left and right channels in stereo music, to give over

36 million bytes or more than 30Mb. A rule-of-thumb equivalent of 10 megabytes of storage per minute of audio is indicated for standard CD quality.

In recent years, the digital music world, in terms of both the PC and the Internet, has converged towards accepting compressed file formats as the *de facto* standard in audio transmission. While the MP3 is by no means the best format for storing digital audio, it might be said to be one of the best known. The MP3, or *'MPEG-1, Layer 3'* [**55**] format is also quite old in digital terms. It is based on *'Audio Spectral Perceptual Entropy Coding'* (ASPEC) and allows for high-quality sounds compressed even to bitrates of as low as 64kbps (kilobits per second) in its official standard definition [**56**].

Recent advances in processing power and capability have led to the same MP3 standard being used to generate extremely high fidelity audio files of 320kbps while maintaining relatively small file sizes and, more importantly to broadcasters, relatively fast processing for 'almost real-time' broadcast (or 'streaming'). In terms of psychoacoustics, it has been demonstrated that human perception does not permit noticeable differences between the reproduced sound of files coded to 192kbps and those coded to higher bitrates [**56**]. It has also been also suggested in the same publication that humans cannot differentiate between audio coded at 192kbps and that coded at the CD-quality 44,100 kHz, at least in everyday use. This means that an audio file greater than 30Mb in size can be reduced to less than 5Mb without any perceptible 'loss' of quality to listeners. While there might be some argument about the perceived difference between 192kbps MP3 and CD-quality 44,100 kHz encodings, using a 320 kbps encoder would limit any possible difference through higher bit rate. These high quality 320kbps compressed files offer a compression ratio of 4.4 : 1[i], meaning a 30Mb track would require less than 7Mb for the highest possible quality MP3.

[i] CD audio has a bit rate of 1411.2 kbps (16 bit/sample × 44100 samples/second × 2 channels / 1000 bits/kilobit). Therefore, a 320 kbps bit rate indicates a compression ratio of 1411.2 : 320 or 4.41 : 1.

4.6 Automatically identifying audio in the digital arena

As explained in Section 3.6.1, every song in recent years, and every variation of it, can be assigned its own standardised unique identifier which can then be used to identify the song if encountered in the real world. Pre-existing works can even have ISRC codes assigned retrospectively by the Registrant that beneficially owns them. Many uses of ISRC codes already exist in industry. For example, in order to ensure that the 'official' charts published in Ireland (and other countries) are restricted in some way to 'legitimate' and identifiable releases, the chart compilation system operated by 'Chart Track UK' on behalf of IRMA insists that only those digital sales of works with an issued ISRC code are included. While unfair to developing, and perhaps uninformed, artists this does attempt to ensure that digital sales are of some reasonable quality to qualify for chart status. ISRC codes are used in other areas and at the time they were created it was envisaged that they would also be used in broadcast monitoring [**57**].

4.3.3 Digital Rights Management

There are two distinctive techniques utilised for the purposes of 'Digital Rights Management' (DRM), namely 'audio fingerprinting' and 'audio watermarking'. The area of DRM includes, but is not limited to:

- Validation and authentication.
- Verification of ownership.
- Copy protection, copy-prevention.
- Broadcast monitoring.
- Identification of illicit distribution and unauthorised access.

Broadcast monitoring is perhaps the least prioritised of these areas despite being of potential value in an economy worth more than €5bn in Europe alone [**8**]. However, broadcast monitoring systems are not just of potential value for identification and calculation of royalty distribution. They could also be useful as a barometric tool, being implemented in the TV domain to monitor the effectiveness of advertising campaigns, the re-use and re-broadcast of news items and for many more tasks [**58**].

Instead of providing accurate payments to those whose works were used, thereby adding '*the inducement of private emolument*' [**6**] to an author's other potential rewards, today's royalty distribution systems often penalise developing and unrepresented artists while over-compensating well-established artists, corporate publishers and Copyright owners. It is clear from an examination of the information provided in Table 2.1 in Section 2.3 that at least some of the royalties that should be paid to developing artists is instead likely to be paid to well-established corporate content owners and successful artists. Any redistribution of this revenue, no matter how small, might be expected to reduce the urgency within the corporate sector of the Music industry that which might otherwise be attached to such technological development.

4.7 Fingerprinting and Watermarking

Research efforts into the areas of Audio Fingerprinting and Audio Watermarking for music classification, identification, search and retrieval etc, date back to the early days of digital signal processing [**59**]. Before that, in such areas as speech recognition, researchers were aware that audio could be modelled in some way and later a candidate audio sample could be recognised as being of the same source. In *'Signal Modeling Techniques in Speech Recognition'* [**60**], Picone outlines various techniques that had been published in the preceding half-decade, dating back to the mid- to late-1980's, which dealt with speech recognition problems and solutions. It is clear that much research had already gone into the area since the 1970's.

Most of the early speech-recognition research was based on recognition of the specific characteristics of the voice of individual speakers rather than being independent of the individual who was recorded. Attempts were then made to model the characteristics of speech, including its spectral composition, in order that future techniques and technologies may be more extensible and may prove more useful in the wider world. Even at this relatively early stage in the research, however, Picone admits that *'There are far too many algorithms in use today to make an exhaustive survey feasible'* [**60**]. These early efforts did lead to the adoption of a number of techniques to identify what Picone called *'perceptually meaningful'* parameters that could be used for describing and therefore identifying human speech. He points out that the existing *de facto* techniques were attempting to find some manner in which normal human speech could be parametrically represented in order that a system could *'emulate some of the behavior observed in the human auditory and perceptual systems'* [**60**].

It follows that if a person's speech patterns can be shown to be *individual*, while at the same time there could be shown to be a sort of collective standard of parametric characterizations of speech, then it would be possible to record and analyse some candidate speech, reproduce it in its parametric representation and then compare it to stored/known speech characterisations of individual speakers There would be a means, therefore, by which it could be determined if the candidate speech was from the same source as the reference speech. In his paper, Picone illustrated the 6 major spectral analysis algorithms, reproduced in Figure 4.6. It is worth noting that many of the techniques used for producing representations of human speech can be used, as one would expect, to represent any other form of audio, including music. Many of the same techniques made their way into later research in the general area of audio characterisation and recognition. Research in the area of characterisation of human speech is still topical [**61**] and it is likely this will continue for some time because, despite advances in the area, it is still a difficult task.

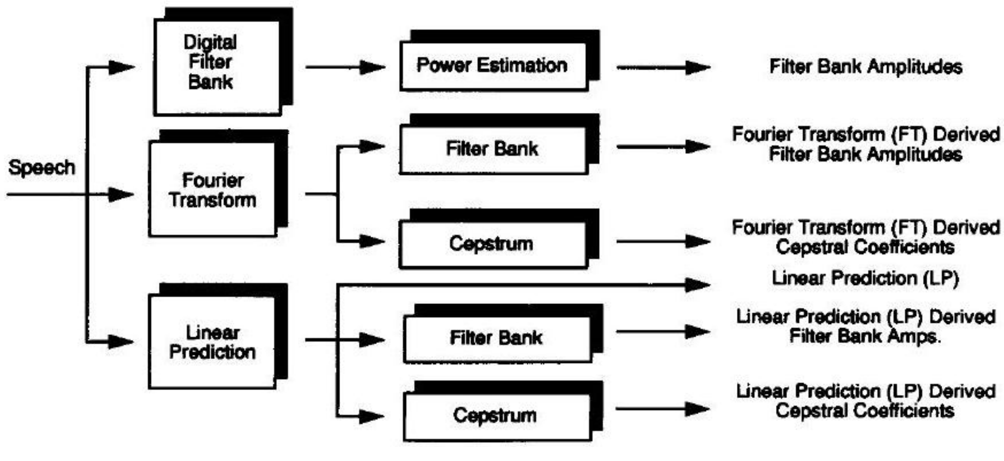


Figure 4.6: The six major spectral analysis algorithms [60].

From the turn of the 21st century, there has been a marked increase in research in the area of digital fingerprinting and watermarking. As the Internet became more prevalent and users began to realise they could digitally transmit files of any type, it soon became clear that illegal file sharing, particularly of music, was going to be the 'battlefield' on which industry and technology were going to become legally entangled [**62**].

The success of the file-sharing facilitator paradigm embodied by 'Napster' (1999 – 2001) soon caused issues to come to a head. The music industry began to take seriously the dangers posed by unlimited and uncontrolled access and to take legal action to try to prevent technological facilitation of illegal distribution [**62**]. While music has been the main area of threat to and from industry, recent developments in provision of high-speed bandwidth to Internet users and of high-quality compression techniques for movies are likely to see the movie industry, previously protected by the comparatively huge size of movie files, drawn into the 'battle' on a more regular basis [**63**]. The illegal release of the unfinished version of the movie '*Wolverine'* in early 2009 seems to have been a particularly eye-opening incident for the movie industry. Nevertheless, at the turn of the century it was predominantly the music industry that was suffering from illegal distribution.

4.7.1 Audio fingerprinting and broadcast monitoring

As explained in Section 1.2.1, '*an audio fingerprint is a compact content-based signature that summarizes an audio recording'* [**64**]. Fingerprinting involves performing a digital analysis of the audio to serve as the 'fingerprint'. In order to attempt to identify candidate audio, it is again fingerprinted in the same manner as the reference fingerprint was derived and some form of pattern matching against the reference fingerprint is performed. If the two fingerprints are identical, a match is reported and the audio is identified. The process is outlined in Figure 4.7.

When a musical work is finalised (i.e. a song is mastered) a digest, known as a fingerprint, is generated based on the particular algorithm used. The fingerprint itself is then added to a database before the song is 'released'. It is important to note that the source audio is unaltered in any way by the process. This is of course a concern for content owners who would want to ensure that their creative works are not altered after they have been released. However, while it might be cited as an important consideration, the fact is that most content is altered by many of the uses to which it is put, including modern digital radio and TV broadcasts. What is most important, in fact, is that any process that is applied to content will have no *perceptual* impact on it. In other words, the user can identify no difference between the unprocessed and processed content.

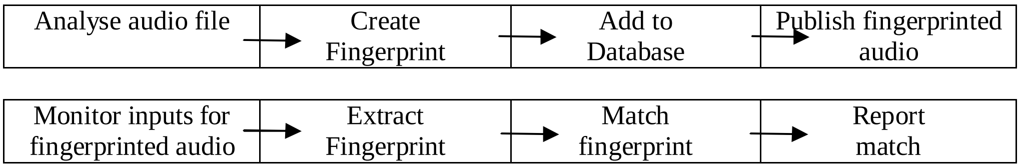


Figure 4.7: Basic fingerprinting process.

Audio Fingerprinting is not a particularly new technology and has seen relatively widespread Industry uptake over the years as it is useful in various scenarios. While commonly used to compare the fingerprint of a file ‘in the wild’ (i.e. in the real world) with a fingerprint of the original file, to ascertain whether they are from the same source (or copies thereof), fingerprinting has also many uses in other situations. It is based on the premise that all files (in this case, audio signals) are inherently different in terms of their audio content and as such would have differing audio fingerprints. A further assumption is that, given two copies of the same file, and identical fingerprinting algorithms, the resultant fingerprints are identical. Furthermore, it should be impossible to generate a fingerprint that is identical to another fingerprint, except by way of using the same original audio and the same fingerprinting algorithm.

Fingerprinting can be used to automatically identify digital files across Peer-to-Peer (P2P) sharing networks and help to identify possible areas for further investigation, such as automatically preventing copying across networks, or collecting appropriate fees and royalties. It is not likely to be a useful technique in identifying whether a song has been illegally copied or whether it has been legally copied by a rightful owner for ‘fair use’ in creating a personal backup. As a result of the legal issues surrounding personal copying (‘fair use’) [**65**], along with the obvious fact that *‘audio will eventually be played in an unscrambled or decrypted format’* [**66**] and could therefore be re-recorded without protection, the systems currently in use are not of much protection against Copyright infringement by determined attackers.

A form of compression and subsequent error-checking known as ‘hashing’ is similar to fingerprinting [**67**] as is the file-checking ‘Cyclic Redundancy Check’ technique [**68**]. These two techniques are similar in that they first create a ‘reference’ of the original file and, later, to compare a suspect copy to the original, they again perform the routine to generate a reference. Simple comparison of the two references will confirm if they are the same. However, audio fingerprints tend to be more complex as they must survive all forms of tampering – both malicious and accidental – such as compression, cropping, format-changing, broadcasting and editing.

It is not intended to examine the technical difficulties experienced by fingerprinting technologies. The most obvious limitation is one of access. Even assuming a content owner knows that there is a requirement or facility for fingerprinting their works, they must then decide which fingerprinting technique to employ and which database(s) to add their works to.

Consider *‘Shazam’*, a freely available iPhone application from Apple’s iTunes App-store. It records sounds through the microphone of a device such as the *iPhone*, processes the recordings and compares them to its own stored database of fingerprints. The candidate audio must already be in the *‘Shazam’* database, which consists of 8 million tracks [**69**], in order to be recognised. An online/mobile application which provides similar services is *‘Midomi’*. It recognises tracks by recording through a microphone and comparing the recording to data from its own database [**70**]. The candidate audio must already be in the *‘Midomi’* database in order to be recognised. *‘Nielsen Music Control’* provides TV and Radio broadcast monitoring services using proprietary fingerprinting technology to various sectors of Industry, including over 1000 record labels in the music industry. It does this by comparing processed output from broadcasts to its database [**71**].The candidate audio must already be in the *'Nielsen'* database in order to be recognised. Other broadcast monitoring services include *'MusicTrace'* [**137**]. The candidate audio must already be in the *'MusicTrace'* database in order to be recognised

These are just a few of the service providers using fingerprinting technologies. Since these providers, as well as most other providers, use proprietary fingerprinting techniques and maintain proprietary databases, each work released publically must be fingerprinted by all providers and added to their databases in order to be recognised.

Fingerprinting techniques used in broadcast monitoring, as mentioned in Section 1.2.1, have some important limitations that are specific to broadcast monitoring for royalty reporting. If a piece of audio is made publicly available before being fingerprinted and added to the database, then the track will not be identified in the broadcast environment. As mentioned in Section 1.2.1, the official partner of the PRS for provision of royalty distribution information, 'Nielsen Music Control', has a database of only 500,000 pieces. This means, obviously, that if there are more than 100 million tracks available[i], Nielsen's official UK procedures could only possibly identify less than 0.5% of them, assuming it had perfect accuracy of identification.

It is still the less established, less well-informed and less well resourced developing artists who are likely to be left out of the distributions in a system that is based on the recognition of audio fingerprints compared against a database. This is simply because developing artists are less likely to have the knowledge or the finances required to have their works included and monitored as part of the myriad databases against which the comparisons might be performed.

---

[i] Online music information provider *'Gracenote'* claims to have a database containing over 100 million tracks. http://www.cddb.com/business_solutions/music_id.
Accessed 25th October 2009.

4.7.2 Audio watermarking and broadcast monitoring

The process of ‘audio watermarking’ attempts to add a message or some other data to the original audio. For example, in a bank note, the watermark hidden in the paper itself adds the message that the note is not likely to be a forgery. When watermarking of paper was first introduced in the thirteenth century, its prime purpose was to differentiate the products of different paper manufacturers [**72**] by adding a mark identifying the producer, in a manner similar to Silver and Gold hallmarks. The purpose of a watermark, therefore, is simply to add information to an original for a reason set out by any person who adds the watermark. In recent developments, digital watermarks are often used by the owners of copyrighted material such as images/videos on the internet, to make them less attractive to casual and opportunist would-be Copyright infringers. Watermarks, however, are not solely designed to prevent or hamper copying and they are not insurmountable. For example:

- Bank notes with high quality watermarks, although forgeries, are available.
- A digital image that bears the legend *‘specimen’* might irritate an opportunist but is not going to be much of a barrier to a determined user with even a standard graphic manipulation software package.
- The explosion in illegally shared video content would seem to indicate that video files with embedded visible Copyright warnings are accepted by the opportunist/average user as an ‘acceptable’ quality of copied material.
- Low quality audio, even with audible inconsistencies, was not a deterrent to early Internet-based copying or file-sharing. This is evidenced by the fact that, even though low bit rate MP3 files are likely to be perceived as lesser quality than higher bit rate files [**55**], they were being illegally copied on file-sharing services such as Napster. One of the earliest *legal* music download providers, ‘MP3.com’, provides access to .mp3 files compressed at a ratio of approximately 11.5 : 1 [**73**], which is equivalent to a bit rate of approx 128 kbps, far short of the perceptually transparent 320 kbps or acceptable quality 192 kbps [**56**].

The most important aspects of, and the form taken by, a watermark are generally defined by their intended use. For example, banknote watermarks should be as secure as possible against copying while watermarks embedded in images should generally be semi-transparent to allow viewing of the image but should be difficult and/or time-consuming to circumvent.

When a watermark is used to promote the owner or producer of a piece of video, rather than to prevent or inconvenience its illegal copying, the watermark could be deliberately made visible as a form of subliminal advertising. In most cases, however, it is preferable if the watermark cannot be detected as users would then not be aware of its existence and less likely to attempt to remove it. With audio watermarking, almost all forms of watermarking should be transparent to the user. There are limited occasions where an audible watermark might be of interest to an owner or producer of a piece audio. For example, a piece that is going to be sent to a very limited set of people for their review/analysis (i.e. a CD pre-release sampler) might be audibly watermarked and the process for removing/circumventing the watermark sent by separate cover, so as to make copying worthless to the end user.

Not only should public users of the audio be unable to hear the watermark but there should also be no obvious sign, even on closer examination, that a watermark has been applied. This is to prevent attempts to remove the watermark for malicious reasons. This requirement makes watermarking, in audio implementations, more difficult as there is a trade-off between hiding the watermark effectively and affecting the integrity (and overall perceived quality) of the audio it is embedded in. Where the watermark is simply an identifier, however, the question of deliberate ‘attacks’ to remove the watermark are less of a concern. After all, malicious copiers of a CD are unlikely to be interested in taking the time to remove the identifier from the audio as there is no reason to.

Even such watermarks, however, are likely to come under attack, albeit incidental attacks, such as those related to compression and broadcasting. The effects of these attacks on the watermark and its subsequent identification and recovery are of more interest in the case of a watermark technique that is used simply to identify of a piece of audio. A watermarking scheme that adds the watermark message as some form of post-processing ‘addition’ to the audio is likely to be defeated by perceptually-based compression. Similarly, while such a scheme might be of value in the digital environment, where ‘headers’ can be transmitted alongside the file, they will not survive in the analogue domain (e.g. traditional radio transmission). For the purposes of broadcast monitoring, therefore, a watermarking scheme should embed the watermark message as *part* of the audio, rather than alongside it, and should do so previous to public release if possible.

4.8 Summary

In Chapter 4, the human auditory system was introduced and a brief overview provided of how humans process sound. Psychoacoustic concepts that will prove of interest in formulating an effective audio watermarking system were discussed. Concepts relating to the hiding of data were introduced in order to frame the problem and potential solution. *‘Digital Signal Processing’* techniques and considerations were also discussed, along with considerations relating to digital audio file formats. A discussion of the relative advantages and disadvantages of audio fingerprinting and audio watermarking followed, specifically as they relate to broadcast monitoring. In the next chapter, the focus will turn to reviewing the systems and schemes already proposed in literature in the area of digital watermarking.

Chapter 5: Existing watermarking techniques

In this chapter, various general and specialised digital audio watermarking schemes will be described. These schemes propose alternative means of embedding watermarks in audio, for various applications and in various domains. The proposals will be described and, where appropriate, their relative advantages and disadvantages will be outlined. As a broadcast monitoring tool a standardised watermarking scheme could be implemented to facilitate worldwide recognition, tracking and reporting of public performances of watermarked audio. This could include any audio recording, such as political speeches, advertising and news reports. However, for the purpose of this thesis the following example is intended to be applied to music.

5.1 Application of watermarking to broadcast monitoring

The ISRC code of a piece of audio recorded or produced for intended release could be embedded as a watermark continuously throughout a piece of audio at the time of production. Indeed, to protect somewhat against it being stripped out either accidentally or by 'pirates', it is suggested that the code be made part of the audio waveform rather than being an addition to it. It is suggested that this would make the watermark harder to identify and harder to remove without degrading the audio quality. Once embedded throughout the audio rather than in some discrete section such as a 'header', the ISRC is permanently available to any system capable of distinguishing it from the 'host' or 'cover' signal (i.e. the actual audio file in which the watermark is embedded).

A system could then be designed to monitor the output signal from broadcasters to 'read' and decode the ISRC from the audio along with the output source and the time/date. A report of ISRC codes found could then be created and presented to the rights agency in order that they might be able to correlate these codes against the national repositories and/or the central ISRC repository and distribute royalties accordingly.

Since some royalty payments are distributed according to number of plays, a simple tally of identified ISRC codes would suffice with regard to producing a royalty payment report. Therefore, the output could be monitored on a continuous basis in order to detect when a new ISRC code was detected. However, where royalties are paid according to duration of play(s) rather than a simple count, some means is needed of identifying how long a piece was played for. This might also make the system more efficient as it means that a signal need not necessarily be monitored continuously. Once an ISRC code is identified, the signal could then be ignored for a specified amount of time, and the monitoring system could then monitor another signal for a short time.

For example, if it is decided that royalties are to be paid for each 30 seconds of play (or part thereof), a broadcast output signal could be monitored until a single instance of an ISRC code is identified. This particular output does not need to be monitored again for 30 seconds. If the same ISRC code appears in the subsequent analysis after 30 seconds or more, then the play has exceeded 30 seconds in duration and the signal can then be re-analysed a further 30 seconds later. Further monitoring of the output from this source could then be undertaken only ever 30 seconds until the ISRC extracted from the source has changed, at which time the cycle repeats.

In attempting to understand how a digital watermarking or fingerprinting scheme might work, it is useful to have at least a conceptual understanding of the domain of digital signal processing in general, and an understanding of Fourier theory as it is implemented in this area would be useful in this regard. The essence of Fourier theory as it applies to sound waves is that any sound can be 'decomposed' down to a set of components of individual frequencies [**74**].

It has been explained in Chapter 4 that the human auditory system is not perfect in that it does not allow us to hear every sound exactly as it was generated. We know that our hearing is often overwhelmed by the presence of conflicting sounds and the auditory system has to 'choose' which sound is delivered to the brain. We know also that we can be deceived into hearing things that are either not present or are very different than we perceive them to be. Examples of these auditory phenomena include those presented by Professor Diana Deutsch, Professor of Psychology at the University of California, San Diego, who has written extensively on subjects related to the psychology and physiology of sound, music and hearing [**75**]. Research into the physiological workings of the 'Human Auditory System' (HAS) in the past few decades has led to clear understanding of how we process sound and how we can manipulate sonic components to take advantage of the limitations of our hearing system [**75**].

With these concepts in mind, the next step is to devise a means in which digital signal processing techniques can be utilised to perform a deliberate manipulation in order to manage human perception, by adding sounds to other sounds in such a way as that they cannot be detected. There are many different techniques and technologies researched and investigated in the area of digital audio watermarking, which of course has different criteria than the areas of image or video watermarking. The techniques used are often dependent on the expected implementation domain of the technique or application. For the purposes of the current research and in the intended application domain of broadcast monitoring, the two major considerations for a watermarking scheme are as follows:

Robustness

- A watermark is called **fragile** if it is altered and cannot be detected after the host audio in which it is embedded has been subject to any form of modification, deliberate or otherwise. Fragile watermarks are commonly used for proving the integrity of a candidate sample. If the watermark is intact, the audio has not been tampered with in any way [**76**].
- A watermark is called **semi-fragile** if it resists some permitted modifications, such as transmission interference, addition of channel noise and so on, but is noticeably corrupted after unauthorised modification [**14**].
- Conversely, a watermark is said to be **robust** if it is unaltered after modifications to the watermark itself or to the host in which it is embedded [**77**].

Perceptibility

- A watermark is called **imperceptible** or **perceptually transparent [77] [78]** if the original host audio and the watermarked audio are perceptually indistinguishable from each other in the domain in which the scheme is to be implemented. Generally, this means that human listeners cannot tell the two apart.
- A watermark is called **perceptible** if its presence is noticeable to either deliberate or casual observers [**78**]. This might mean that the watermark itself, which is simply an added signal of some sort, is noticeably audible in the watermarked host audio or if the original audio and the watermarked audio are noticeably different.

There are, of course, many other considerations when evaluating a watermark embedding scheme, including capacity (data payload) and computational complexity, not to mention decoding techniques and associated issues. Proponents of digital audio watermarking schemes must also exercise caution in addressing the various permitted or accidental attacks that will be allowed against the watermarked audio in the intended application domain. In this regard, the most likely attacks to be faced by a watermarking scheme which serves only to allow the automatic identification of audio in a broadcast environment will be those associated with transmission and compression. If a watermark is robust against these attacks, in this domain, it will be successfully implemented. Notwithstanding this consideration, the actual process of creating and separately embedding the watermark can be achieved using many techniques, each with their own advantages and disadvantages. Some techniques are discussed in the next section.

5.2 Literature review of watermarking techniques

5.2.1 Least significant bit modification embedding techniques involve manipulation or replacement of the least significant value of each byte that represents information about the ‘cover’ or ‘host’ signal (the signal in which the watermark is to be embedded) in order to create a predefined bit sequence that, upon decoding, will recreate the hidden message. This is a simple enough technique, described as the *'simplest way to embed data into other data structures'* [**79**] and, while it has some uses, it has some major problems in that the least-significant value of a signal is the value most likely to be altered by typical signal manipulations expected by the authors of [**79**]. Even those 'attacks' which are simply incidental to transport are more likely to impact on the least significant bit [**80**] as well as more direct manipulation, such as compression, DA/AD re-encoding and so on. Some effort can be made to ensure that the sequence of least significant bits, if altered, can still be recovered. One simple trick is to repeat the watermark numerous times and recreate the message from the most commonly recovered bits in a form of *Mode* operation. Even if most bits are altered, it is unlikely that they will always be altered in the same way, in the same sequence, so recovery accuracy is increased by comparing the repeatedly recovered bit-streams against each other and selecting the most common bit at each index.

Least-significant bit (LSB) modification is a common technique in image-processing but, as explained by Kaliappan Gopalan in [**81**], which deals with utilising bit-modification techniques for covert communications, the same technique as used in image processing cannot easily be transferred to the audio domain because the human visual system is not as sensitive to minor variations as the human auditory system. The added data would be more likely to be perceptible to the end user or to an unintended interceptor. Gopalan suggests that a human listener can detect changes in an audio file (in comparison to an original) of one part in 10 million [**81**]. However, there is cause for concern in using a technique that might work well in one domain, and transferring it to a domain where it would appear to be significantly less effective and more likely to aid unauthorised detection.

Finally, as Gopalan also points out [**81**], individual bits cannot be correlated to particular frequency components in the cover audio, so any chance of using amplitude masking techniques to aid perceptual hiding would be accidental at best and counter-productive at worst. Gopalan is also concerned with embedding actual human speech into a cover signal [**127**], and as such, even if the speech is not completely recovered, it would be possible (to acceptable levels of effectiveness) to recover the main thrust of the 'message'. While this might be acceptable for embedded human speech it would not be adequate for a stream of embedded bits, representing the unique identifier, because even a single missing bit would render the message illegible. Similarly, a bit that is recovered incorrectly might make the message meaningless or, perhaps worse, incorrectly decoded.

Taking a different approach, manipulating bits that are more perceptually significant would increase the accuracy of recovery of the data, since these are less likely to be corrupted accidentally. However, such manipulation would be also more likely to lead to perceptibility problems. Having said all his, the technique is still being implemented at present, particularly in image watermarking [**82**] as it does show promise for some applications, particularly where the problem of added noise is not a major consideration. It is also said to be resistant, even from its earliest investigation by Bassia et al [**83**], to expected signal manipulations and compression techniques. Bassia claimed in 2001 that the technique was resistant to time-shifting, which is one of the attacks in the audio domain that could adversely affect echo-hiding techniques, another relatively uncomplicated watermarking technique.

5.2.2 Echo-hiding consists of a technique for embedding the watermark into the host or cover audio by taking advantage of the human auditory system's inability to detect certain very short echoes in a sound. As a simple explanation, a signal might be broken into non-overlapping frames of user-defined length before the encoder adds a delayed version of a candidate frame (or even just some components from the frame), delayed by, say 0.005s to represent a '0' and 0.008s to represent a '1' bit. In theory, even neglecting to add a delay for a '0' bit would be potentially useful but more likely to increase incorrect detection as there will obviously be times when an echo is either present or absent in the original audio and this might be confused with deliberately-controlled echo.

The added echo also has its amplitude weighted against the original host audio so that it is less perceptible. According to [**83**], echoes less than 1ms will 'fuse' with the original audio, from which the echo is derived and the listener will not be able to hear the echo, it being perceived as the part of the original. Based on the belief that the human auditory system cannot distinguish a very short echo, various researchers have proposed different echo durations that would allow the watermarked audio to pass as indistinguishable from the original. Application-specific requirements may determine the delays to be used.

In order to embed the watermark as a sequence of echoes, the segmented host audio is first duplicated with delay *d1*. It may be duplicated again with delay *d2* if dual echoes are required for '0' and '1' bits. Then, according to the sequence of bits to be embedded, the system applies the relevant delayed audio, and adds it to the original audio. The end result is described by equation (1):

$$y(n) = \begin{cases} x(n) + w\,x(n - d_0) & \text{if bit} = 0 \\ x(n) + w\,x(n - d_1) & \text{if bit} = 1 \end{cases} \qquad (1)$$

where d_0 and d_1 are the delays introduced by the scheme to represent a '0' and '1' bit respectively and *w* is a weighting factor. Figure 5.1 illustrates the two variants of echo-hiding, with the first variant leaving the audio unaltered to represent a '0' bit while the second variant adds a different delay for '1' and '0' bits respectively. Note that the use of two delays increases the technique's robustness and accuracy but also add to its complexity.

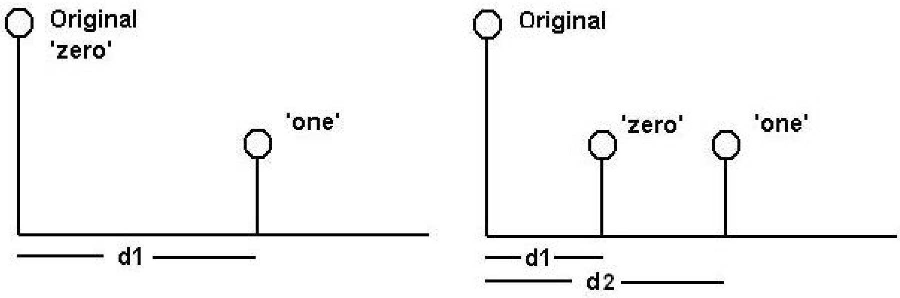


Figure 5.1: Two variants of echo-hiding watermarking techniques.

One major issue in decoding echo-coded watermarks is that any sort of time-shift (such as stretching or compressing the signal in the time domain, even accidentally) will result in the delay (*d1* or *d2)* between the original and echoed signal being incorrect, and the watermarked bit could therefore be incorrectly decoded. As far back as 1996, early efforts at research into watermarking described the successful of hiding bits in a media stream in a perceptually acceptable manner [**79**]. It was not described as perceptually invisible, as the addition of echo often results in a noticeably different resonance between the original and the watermarked audio, with the echo making the audio more 'full' sounding [**79**]. In order to alleviate the resonance addition caused by the echo, the work suggests that the encoding mechanism should set the '*initial amplitude and the decay rate below the audible threshold of the human ear'*, thereby making any echo perceptually invisible.

Conversely, making these settings so low would mean the echo is less likely to be recoverable, being susceptible to added noise from the transmission channel or the environment. Various supplemental techniques are suggested [**79**] to ensure effectiveness of the echo-hiding technique, but, as with all such supplemental techniques, each adds its own distinct disadvantage. Redundancy and error-prevention techniques might increase recovery rate but might also increase likelihood of detection, fragility and susceptibility to accidental and deliberate attacks while simultaneously decreasing capacity. For example, if a code is repeatedly embedded in a cover signal, this aids robustness because it can be decoded multiple times and correlated to increase recovery precision. However, this also means that a signal of a given length has a lower capacity as it must repeat the watermark multiple times within the same signal.

Later research, from both Ko *et al* [**84**] and Kim & Choi [**85**], who introduce forward-echo or *pre-echo* (which they also describe as 'virtual echo') in addition to 'normal' echo, developed the echo-hiding concept in such a way that the echoes that are added are of much lower amplitude and are simultaneously less perceptible and more effectively recovered. This makes the technique much more viable as a practical watermarking technique.

Another issue of concern with echo hiding, particularly in schemes where the process of 'blind decoding' is important, is that the original unwatermarked audio is required in order to identify where echoes occur that are not simply part of the original audio or added as a result of the transmission environment (e.g. added 'reverberation') or some form of attack on the signal. Also, production of a corrupted version of the audio would be possible simply by applying an echo delay across the whole signal, thereby making it impossible to detect the echoes that are part of the watermark and those that are added as part of the attack.

5.2.3 Amplitude masking consists of a process of embedding the watermark into the host audio in the form of an additional audio signal at very weak power. This technique utilises the known masking effect of sounds on other sounds, as described in Section 4.2. Masking of one sound by another is dependent on various parameters, including but not limited to, the frequency distance between the components, the amplitude or magnitude difference between the two components and the individual magnitude of the components themselves. Low-powered components may simply be too quiet to mask another component. Conversely components may be below the threshold of hearing (shown in figure 4.4) and therefore the presence of other components is irrelevant. Masking can be described as:

$$W(c_2) - W(c_1) \approx 1 \tag{2}$$

where $0 < W < 1$, W is the independent level of audibility of a component C_1 or C_2 and W approaches 1 as C_2 gradually becomes more audible in the presence of C_1.

The audibility of individual components is a function of the frequencies and amplitudes of these components. Three conditions must be satisfied for masking to occur. Firstly, the frequency separation between the components *f1* and *f2* must be below a frequency-dependent threshold T_1

$$| f_2 - f_1 | < T_1(f) \tag{3}$$

and magnitudes A_1 and A_2 must be separated by a threshold T_2

$$T_2(A) \tag{4}$$

$$(A_2 - A_1) > T_2(A)$$

and the magnitudes of both components must be above their audibility threshold

$$T_3(f_1, f_2) \tag{5}$$

$$A_{1,} A_2 > T_3(f_1, f_2)$$

This latter requirement in equation (5) simply means that if two component frequencies are present but are so quiet as to be independently inaudible to humans, then they cannot mask each other since the concept of masking is that one component makes the other inaudible in its presence. If either is inherently inaudible, masking does not occur. Similarly, if one component is audible but the other is not, then neither can act as a masker.

The watermark signal, representing a bit-sequence, might be converted into a sinusoid in such a way that the components of the sinusoid and their relationship to each other (or the cover audio) are controlled to represent the appropriate bit [**86**]. Adding the watermark to the host audio will of course result in potentially audible artefacts, even if added at low power, especially where the host audio has low-powered components. This can be mitigated by ‘perceptually shaping’ the watermark signal either using a well-defined psychoacoustic shaping model or by weighting the watermark components against the relative power of some or all of the components of the host audio at the point of embedding.

In [**86**], Gopalan and Wenndt attempt to take advantage of knowledge of the limitation of human hearing and the masking effect of components against each other. As with Gopalan’s work using bit-modification techniques outlined in section 5.2.1 [**82**], this work relates to embedding a watermark message into a cover signal consisting of human speech. The technique uses two components and manipulates their power against each other in order to create a relationship that represents ‘1’ and ‘0’ bits. The components chosen are in the lower end of the human hearing range as human speech is, itself, in the lower half of the hearing range, generally not exceeding the mid-hearing range.

Decoding the watermarked audio is simply a matter of analysing these components in the candidate audio and recreating the bit depending on the relationship of the components to each other. It is a computationally simple and elegant solution and it appears to work quite well in cover audio consisting of human speech. There are, however, problems with this technique when implemented using cover audio that consists of music rather than speech, partly as music uses a much wider range of frequencies than speech.

Gopalan and Wenndt also noted the problem of cover audio that already *has* some components at the frequency that is about to be manipulated: if a frequency *f1* is added that should be lower than another frequency *f2* but if there is already a higher-power component at *f1* the decoding would be inaccurate. Finally, while low-powered components would be imperceptible to listeners, the pattern of these manipulations would likely be visible in a spectrum analysis such as a spectrogram, making the 'covert' nature of the watermarking technique open to question [**86**]. As a watermarking technique, it works well, although embedded messages would probably also need to be encrypted to increase security.

5.2.4 Phase coding techniques work by substituting the phase of one piece of audio by the phase of another, or simply by altering the phase of the cover audio to represent some binary value. In [**79**] the phase coding method is outlined in theoretical terms and it points out that its success as a watermarking technique in certain situations is a result of the inability of the human auditory system to detect phase, within certain limits. A binary message is represented by a series of $+\pi/2$ or $-\pi/2$. The cover audio is modified so that its phase represents the message. The absolute phase of the cover audio's initial frame is substituted by a 'reference' phase, artificially constructed, and subsequent frames have their phase set relative to this reference phase, depending on the binary bit to be embedded.

The phase coding process is illustrated in the Figure 5.2 [A - H] and the algorithm is also presented, reproduced from [**79**].

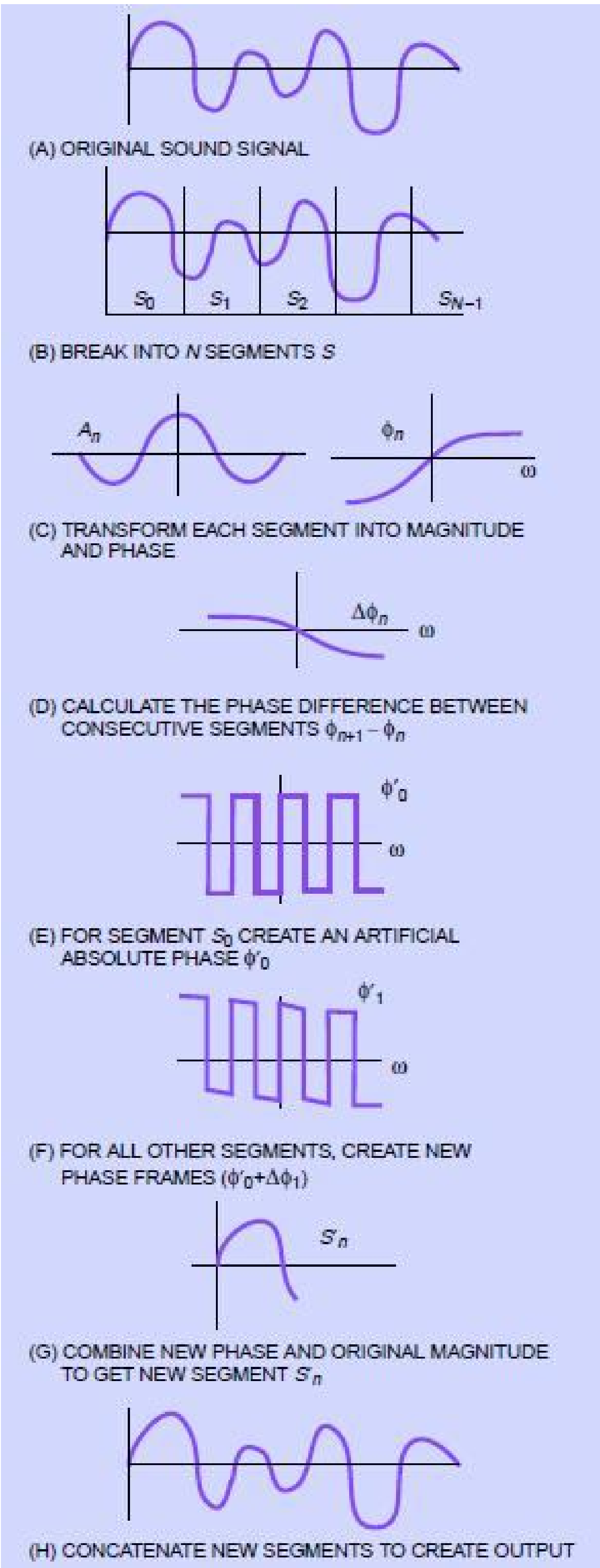


Figure 5.2: The phase coding process.

1. Break the sound sequence $s[i]$, $(0 \le i \le I - 1)$, into a series of N short segments, $sn[i]$ where $(0 \le n \le N - 1)$ [Figure 5.2.A and 5.2.B]

2. Apply a K-points discrete Fourier transform (DFT) to n-th segment, $sn[i]$, where ($K=I/N$), and create a matrix of the phase, $\phi_n(\omega_k)$ and magnitude $A_n(\omega_k)$ for $(0 \le k \le K-1)$ [Figure 5.2.C]

3. Store the phase difference between adjacent segments for $(0 \le n \le N - 1)$

$$\Delta\phi_{n+1}(\omega_k) = \phi_{n+1}(\omega_k) - \phi_n(\omega_k) \qquad \text{[Figure 5.2.D]}$$

4. A binary set of data is represented as

$\phi_{data} = \pi/2$ or $-\pi/2$ for bit '0' or '1' [Figure 5.2.E]

5. Re-create phase matrixes for $n > 0$ by using the phase difference [Figure 5.2.F]

$$\begin{bmatrix} \phi'_1(\omega_k) = \phi'_0(\omega_k) + \Delta\phi_1(\omega_k) \\ \cdots \\ \phi'_n(\omega_k) = \phi'_{n-1}(\omega_k) + \Delta\phi_n(\omega_k) \\ \cdots \\ \phi'_N(\omega_k) = \phi'_{N-1}(\omega_k) + \Delta\phi_N(\omega_k) \end{bmatrix}$$

6. Use the modified phase matrix $\phi'_n(\omega_k)$ and the original magnitude matrix $A_n(\omega_k)$ to reconstruct the sound signal by applying the inverse DFT [Figure 5.2.G and 5.2.H]

Embedding a watermark by altering the phase of components within the cover signal can be troublesome as, while the human auditory system is generally not able to detect absolute phase, any sharp or radical alteration of phase from one frame to the next may result in audible phase inconsistencies. In [**87**], Lipshitz *et al* state that *'Even quite small midrange phase nonlinearities can be audible on suitably chosen signals'*. While this contention might be open to debate, it seems to have been supported by research in 2000 by Koya [**88**]. Therefore, phase changes from one frame to the next must be more gradual and this causes reduced capacity of the scheme. Having said that, no watermarking scheme is of much use if it introduces unwanted audible artefacts, so capacity reduction is an acceptable trade-off to reduce such artefacts.

Decoding of the watermark in a phase coding scheme requires information about the length of each segment and the number of DFT points ('K' in step 2 above) and perhaps some additional information. It is therefore not a blind decoding technique but rather a semi-blind or informed technique.

This technique was further developed over intervening years. Work by Yardimci *et al* [**89**] in 1997 took a slightly different approach to phase modification, utilising all-pass filters with different phase characteristics to represent the bits to be embedded. However, these filters had the effect of adding disturbances to the cover audio and altering its characteristics. In comparing watermarked to original audio, this would make the presence of the watermark easy to discover and therefore limit the usefulness of the technique.

Takahashi *et al* [**90**] published a work that used dynamically varying phase characteristics in order to embed watermark data by altering the phase of the cover audio. Takahashi describes a time-varying *'Finite Impulse Filter'* (FIR) setup which is controlled by a sinusoidal function to have smoothly-changing coefficients. The filter described by Takahashi is referred to as a 'Phase-Modulation Filter' [**91**]. The technique described allows for watermark data to be embedded using amplitude shift keying, frequency shift keying, phase shift keying or any combination of the three. The phase-shifting process allows for the inter-channel phase difference to be used for binary bit representation. However, as Takahashi explains, a simple down-mix from stereo to mono will result in the phases being altered in such a way as to be impossible to detect the inter-channel phase difference.

5.2.5 Multiplicative watermarking is also referred to as **transform-domain watermarking** and refers to the embedding of the watermark data after the cover audio has been *transformed* in some way. In recent years there has been an increasing number of publications of watermarking schemes based on manipulation in a transform domain of some sort, including but certainly not limited to the *Discrete Fourier Transform* (DFT) domain, *Discrete Cosine Transform* (DCT) domain, *Integer Wavelet Transform* (IWT) domain, the *Modified (or Modulated) Complex Lapped Transform* (MCLT) domain and *Discrete Wavelet Transform* (DWT) domain.

There are also various modifications and adaptations of these techniques along with a multitude of other transforms of various types, each with its own uses, advantages and disadvantages. None of them are, in themselves, a watermarking technique. Instead, they are a primary step to perform before the digitised audio is in a form that can then be watermarked. The actual watermark message itself, in any of these transform-domain techniques, could be in any of a number of forms (e.g. pseudorandom sequence or even a simple unencrypted bit sequence). A brief description of some of the common transforms follows, with reference to watermarking or related works that implement them.

- ***Discrete Fourier Transform.*** As explained by [**74**] and [**91**], the '*Fourier Transform'* is a means whereby the relative strengths of the various components (frequencies) inherent in a given signal can be calculated. The *'Discrete Fourier Transform'* (DFT) can be implemented using a *'Fast Fourier Transform'* (FFT), a fast implementation of the DFT, on a sampled signal. The transform performs this calculation on discrete segments (also known as frames or windows) of the signal, one at a time, treating them as if they were infinitely periodic in order to satisfy the requirements of Fourier concepts. The result of the DFT expresses the signal in terms of complex exponentials, capturing information about the magnitude and phase of its inherent components over the whole frame. The output of the transform on a given audio signal is the frequency representation of the signal but only for the frame under consideration as if it was completely independent of the rest of the signal. Where the frequency changes continuously, or occasionally but for a short time, the Fourier transform is inadequate [**92**]. Finally, since it provides information only about enough of the components to be able to replicate the signal there may well be components in the signal for which no information is captured. These considerations limit the use of the FFT for signals such as music and suggest that increased usefulness is achieved by using smaller and smaller time segments (windows) thereby increasing resolution of the analysis output but also increasing computational complexity. Nevertheless, the transformation of a signal by DFT is a first step in many watermarking schemes including [**93**] [**94**] and [**95**].

- ***Discrete Cosine Transform***: The Discrete Cosine Transform is similar to the DFT. Like the DFT, it represents or decomposes a signal so that its components might be described. Unlike the DFT, the DCT uses only *cosine* waves. This is an efficient algorithm because the DCT is equal to a DFT or approximately twice the length and so it is an important development in terms of computational complexity in signal processing tasks. As there are multiple types of DCT with different input and parameter values, the DCT is flexible in its applications.

The DCT has many of the same advantages as the DFT, but also many of the same disadvantages related to non-stationary signals. However, its complexity is much less than that of the DFT so much smaller windows can be analysed at the same computational cost. Note also that the DCT transformation is more useful and more prevalent in the transformation of image and video signals than audio signals, perhaps because the granularity of the higher-frequency components in images and video is less important than in audio signals. Examples of audio processing after DCT transformation include [**96**], which interestingly uses an embedded watermark as an audio quality test. In this work, the authors embed the watermark after a DCT transform, and then allow the signal to undergo various forms of attack. An analysis of the watermark after attacks is then performed on the assumption that the watermark will undergo the same corruption as the rest of the signal. DCT-based watermarking schemes are still very topical and recent publications include [**97**] in image watermarking and [**98**] in the audio watermarking domain.

- ***Wavelet Transforms*** have become more commonplace in recent years because of their versatility. Essentially, a Wavelet is a 'small wave' that compromises multiple frequency components. The 'C*ontinuous wavelet transform'* was discovered by George Zweig [**99**], a physicist and neurobiologist, during research into the human auditory system in the mid 1970's and was initially called the 'cochlear transform'. However, it is in the signal processing arena as well as the area of human physiology that wavelet transforms have recently generated much research interest.

In Fourier analysis, *sine* and *cosine* waves are used as the basic components with which to decompose the input signal of complex waveforms into something more fundamental. The same purpose is served by groups of related *wavelets* in this area of transformations. Essentially, in all forms of wavelet transform a tightly defined 'reference' wavelet is created and the input signal is manipulated with regard to both this wavelet and others closely related to it in order to identify certain inherent characteristics of the input signal.

The design and specification of the wavelet is paramount to its usefulness and it is therefore application-dependent and purpose-dependent. For example, in simple terms, a wavelet comprising frequencies oscillating at 1 kHz, 2 kHz and 3 kHz may be created and applied to the input signal in order to ascertain if and where there are any components at these frequencies. Of course, any number of wavelets comprising different frequencies may be utilised and this would be sufficient to provide an analytical breakdown of the input signal.

This is the fundamental difference between *sine*-based or *cosine*-based analysis and wavelet analysis. A *sine* or *cosine* wave is a single-component continuous and periodic signal that stretches from minus infinity to infinity whereas a wavelet is generally short-lived, limited time and of multiple components. This distinction is illustrated in Figure 5.3.

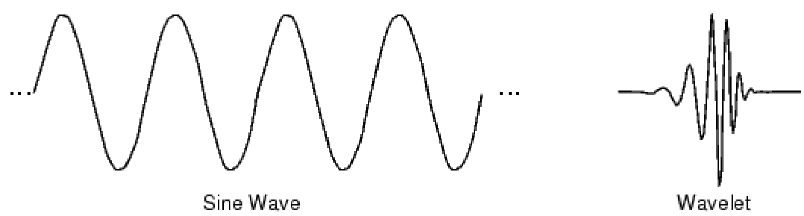


Figure 5.3: A periodic and stationary sine wave comprising a single component alongside a non-stationary wavelet comprising multiple components[i].

Wavelets are also more useful than *sine* and *cosine* waves as the basis for decomposing signals that have sharp changes, very short duration components and localised features [**100**] [**101**].

Other transforms that might be used prior to the subsequent embedding of watermarks include the *'Integer Wavelet Transform'* (IWT) [**102**], the

[i] Image from the *MathWorks* tutorial *'What is Wavelet Analysis?'* http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/ch01_in9.html

‘Modified (or Modulated) Complex Lapped Transform’ (MCLT) [**103**] and transform to the *‘Cepstrum’* domain [**104**].

It is worth repeating that none of these techniques is, in itself, a watermarking technique. These few transform examples serve to illustrate the various means by which a cover signal can be transformed into another domain before the watermark data is added.

5.2.6 Spread spectrum watermark embedding techniques were developed out of the use of spread spectrum techniques in military communications beginning after the Second World War [**105**]. This, in turn, derived from Nikola Tesla’s research into frequency hopping, apparently prompted when he encountered problems trying to demonstrate the remote control of a radio-controlled boat [**106**]. The central concept behind spread spectrum encoding is that the message to be embedded can be ‘spread’ across a wide spectrum of the frequency range, thereby adding some advantages.

Firstly, and most relevant for the current research area, if a watermark is spread across the entire spectrum of a cover signal, the watermark appears to observers to be no more than noise. Secondly, the watermark can be detected (as long as the detector has a pre-defined decoding ‘key’) even in the case that the carrier signal is very weak at the receiving end. This was of course very useful in long-range radio communications. Thirdly, the watermark can be embedded in the significant components of the cover signal, as well as the less-significant, and this leads to increased robustness against attacks which often concentrate on perceptually less significant components. As a result of this, spread spectrum embedding is very robust against ‘jamming’ [**107**], whereby an opponent may not know whether a message is present or not, but simply ‘jams’ or tries to corrupt the entire signal.

Like other watermarking techniques, spread spectrum watermarking is a technique for actually embedding the watermark data, not necessarily encoding or cryptographically enciphering it. Therefore, as a steganographic tool, its usefulness is a function of whether or not an intercepted cover signal is believed to include a watermark. Having the watermark data spread across the entire spectrum can help or hinder the potential attacker.

In a 2002 paper [**108**], Kirovski and Melvar introduced an *'asymmetric direct sequence spread-spectrum'* watermarking technique (although there are many alternatives to *'direct sequence'*). The embedding in spread-spectrum watermarking techniques is achieved by first applying a pseudo-random signal, generated using a 'key' that becomes the decode key, to the watermark data before actually adding the newly-modified signal to the cover audio. Assuming the pseudo-random 'key' represents a filter of sorts that has a frequency response that is flat across the frequency range (for example, a chaotic or white-noise type signal) the newly-modified watermark signal generated by multiplication of the watermark and the pseudo-random key is then added to the cover audio and would be identified by attackers as noise, but could be reproduced using the same 'key' in decoding.

In their system, compromising by an attacker of a single client (e.g. a computer, a music player or a video player) would not lead to a breach of security of all watermarked cover video. They state that a single two-hour high-definition video file had the capacity to protect as many as 900,000 individual devices. This might seem a relatively small number, considering that Apple have sold over 7 million devices in the *iPhone* range alone (not including video-enabled iPod devices) in one quarter-year of 2009 [**109**].

In [**66**] Kirovski & Malvar make a point that is often overlooked but should be noted in relation to *all* watermarking schemes: *'audio will eventually be played in an unscrambled or decrypted format'*. What this means is that no watermarking scheme is of any practical use unless the watermark message is, and remains, part of the actual audio rather than just an addition. Otherwise, in order to produce a version of the signal with no detectable watermark, all the attacker has to do is play the signal through normal speakers and allow it to propagate through the air before re-recording. On that point, spread spectrum watermarking has some scope for use as it covers the whole spectrum of the audio, so it would seem it could be used to embed a watermark throughout the frequency range in such a way that it becomes part of the signal and not just something to be removed from it, thereby reverting to the original cover audio.

5.2.7 Patchwork methods were initially proposed for use in the area of image watermarking [**79**] but could potentially be of use in other areas such as video and audio watermarking. The key concept of patchwork watermarking is that a single sample of the cover signal has some statistical modification performed on its mathematical representation. This could be achieved by modifying the value of a parameter in an individual sample by a constant value while modifying another sample so that the relationship between the two modified samples is known.

For example, the brightness of a pixel in an image could be increased while being simultaneously decreased in another pixel that is mathematically related to the first in some pre-defined manner. Alternatively, the value of a Fourier coefficient for some selected pair of frequencies could be simultaneously modified by some constant, either by addition/subtraction of a constant value or by multiplication by +1 or -1. In [**110**], Kim describes the publication of a development on the idea by Arnold in 2000 [**111**] as a '*landmark'*. The process is described as a 'blind detection' system but it is worth noting that this assumes certain limitations in the modifications to the watermarked signal. It is also a relatively low-capacity system. Arnold's 'landmark' system managed a capacity of only 1 bit in more than 1 second of audio (1bps).

5.2.8 Interpolation methods are another alternative watermarking method, this time said to have good capacity and resistance to DA/AD conversion and some common attacks [**112**]. They include ‘spline’ interpolation and ‘polynomial’ interpolation [**113**]. The latter is perhaps the best-known interpolation method for one-dimensional signals such as audio. The term ‘interpolation’ means to define a data point using other data points to calculate it. For example, if there are two points in space, separated along an axis, then theoretically, any point along the line between these points can be calculated by interpolation, as illustrated in Figure 5.4.

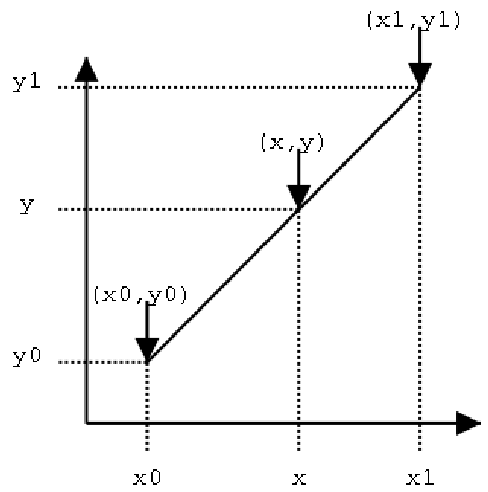


Figure 5.4: Illustration of linear interpolation concept.

In the example in Figure 5.4, assuming the co-ordinates of the point (x0, y0) and (x1, y1) are known, any point along the line between these two co-ordinates, such as at the point (x, y), can be calculated. This is, of course, a very simplistic example but the concept can then be expended so that any ‘new’ data point can be constructed by using the locations or values of known data points in a given set, such as sample points in audio.

It is worth noting that the data point that is to be constructed may be a ‘virtual’ point because it does not inherently exist in any way in the sample set it is being interpolated into. This can have benefits and can also have a very obvious downside: additions to a sound can be problematic in terms of perceptibility of any artefacts created. A very recent publication by Fallahpour and Megas [**114**] claims a capacity of almost 3kbps, which is an extremely high capacity for a watermarking system, by using ‘spline interpolation’ techniques. The technique is also claimed to be stable, robust to attack and perceptually transparent.

5.3 Additional watermarking schemes

The foregoing examples of watermarking techniques represent some common approaches to the task. However, since the earliest forays into digital audio watermarking, there have been a number of unusual and innovative approaches to the problem – some more successful than others. Included amongst this group would be the following:

5.3.1 Muteness based audio watermarking is an interesting, if somewhat limited variation on the watermark embedding scheme. In this scheme, Kaabneh and Youssef [**115**] analyse audio to identify very short periods of muteness. Their definition of muteness is essentially experimental and in reality would rely heavily on the audio that is being analysed. Opera pieces, for example, along with spoken word and many other types of audio, would have a comparatively large number of mute periods, while heavy Rock music or even orchestral classical music would be expected to have less. Nevertheless, this is not a criticism of the scheme. The success of most digital audio watermarking schemes could be said to be reliant to a greater or lesser extent on the type of audio that is being used to hide the watermark.

In this scheme, the watermark is added to the audio by first creating a record of the mute periods and their lengths before deciding to increase by a small amount the length of any given mute period or to leave it untouched to represent an embedded bit. Leaving the mute period untouched to represent a ‘0’ and extending it slightly to represent a ‘1’ is just one option. Alternatively, the mute period could be increased by a value for a ‘0’ and a different value for a ‘1’. With the mute periods altered, the watermark is detected by analysis of the candidate audio, and comparison of the lengths of the newly-analysed mute period sequence against a record of the original mute periods in the unwatermarked audio. It is not necessary to have access to the original audio at the time of decoding – all that is required is a record of its mute periods. The scheme therefore provides for a semi-blind decoding process.

This is an unusual approach to watermarking and at first glance it would seem to be quite limited in scope. The main advantage of the technique is that periods of relative quiet or silence in an audio host would be considered inherently important to the audio and so would generally be assumed to survive compression and perceptually-focussed attacks on the watermarked audio. The authors did concede in [**115**] that at high MP3 compression rates (down to 16 Kbps) there was deterioration in the recovery rate but this would be expected in most watermarking schemes **-** MP3 compression to 16kbps is quite destructive and is rarely used. Less drastic compression was accompanied by much higher recovery of the embedded data.

Another advantage that this technique has is that it is a time-domain embedding technique. Band-pass filtering, another common attack on audio signals, was carried out by the authors with a cut-off frequency of 2205 Hz and had no adverse impact on the recovery rate. Resampling from the original 44.1 kHz down to 22.05 kHz and 11.025 kHz, before being reverted back to 44.1 kHz led to a noticeable distortion in comparison to the original audio but still allowed for a recovery rate of more than 98%.

One apparent disadvantage of the scheme is in the definition of muteness. As mentioned, it is dependent on the host audio and this in itself is not a major problem except insofar as capacity of individual audio would be concerned. However, it would appear likely that transmission across noisy channels, the addition of random or white noise and even atmospheric interference might be issues that would negatively affect the recovery rate by introducing noise where previously there had been a mute period. Having said that, there may be error-correction techniques that could overcome some or all of these obstacles and it remains a simple and elegant technique.

5.3.2 Sinusoidal pattern watermarking. In attempting to address some of the problems experienced in the use of pseudo-random sequences embedded as watermarks in spread spectrum watermarking techniques, Liu and Inoue [**116**] proposed the use of sinusoidal patterns rather than pseudorandom sequences as the material to be embedded. According to the authors, there are a number of generic problems associated with spread spectrum watermarking, including

- The requirement in many cases for access to the original audio, to be subtracted from the candidate audio to accurately determine the watermark.

- The need for near-perfect synchronisation between the pseudorandom sequence and the watermarked audio in decoding means that cropping of the audio or any form of time-shifting will make synchronisation, and therefore detection, much less effective.

- The detection of the watermark in these schemes can be adversely affected by short pseudorandom sequences. Longer sequences are used to ensure effective correlation. This is not a major problem in image or video watermarking but in audio watermarking it is more restrictive because the Human Auditory System is more sensitive than the Human Visual System, therefore limiting the components in which watermarks can be embedded transparently.

There have, of course, been efforts made to address these issues. However, in their scheme, Liu and Inoue took an unusual approach [**116**]. First, they created a set of sinusoidal 'patterns' consisting of a small number of frequencies, each representing a given value. These were then modified according to a psycho-acoustic model in order to ensure they fell below an audibility threshold before being embedded in the host audio. Performing this step in advance meant that it was a one-time, computationally inexpensive process. The amplitudes of the concatenated sinusoidal patterns, constructed to represent a message sequence, were then modified against the power of the host audio in which they were to be embedded **-** thereby ensuring they were not loud enough to be audible above the host. The sinusoidal pattern included a signal that represented its own start point, thereby making it self-synchronising and overcoming the problem of synchronisation before decoding.

Detection was a matter of computing the correlation between step-wise blocks of candidate audio against the sinusoidal patterns. This made the process a semi-blind detection scheme as it does not require the host audio, merely some information about the embedding process or data such as the initial group of sinusoidal patterns. As explained by the authors, some of the problems with this scheme included the negative effect of cropping and time-shifting [**116**]. Down-mixing from stereo to mono might also be a potential problem as the scheme relied on the correlation of timing marks in the left and right channels to enable it to self-synchronise. However, the authors did achieve some excellent results and the scheme would certainly be computationally much less expensive than some other schemes.

5.3.3 Linear chirp watermarking. In **'***A Robust Audio watermark Representation Based on Linear Chirps'* [**117**], the authors proposed a scheme which embedded data in the host using 'linear chirps', with the composition of the chirps designed to represent various values. A linear chirp is a sound that has a constantly increasing (or decreasing) frequency and it therefore has a 'slope' that is constant. Essentially, the variation of the frequency from one time-segment to the next is the 'slope'.

This work is actually concerned more with the composition of the watermark itself, rather than how or where it is embedded. In their experiments, the authors state that they used spread spectrum techniques for embedding the watermark in the host audio but also suggest that their 'linear chirp' approach to construction of the watermark message could be extended to most other types of embedding algorithm.

What was significantly noticeable from this work was the recovery rate after hostile attacks on the watermarked audio. The authors state that they recovered the watermark with an extraction rate of 100% as long as the Bit Error Rate (BER) did not exceed 20%. They also state that in their robustness testing after signal manipulations, the BER did not exceed 11.36%, even after attacks on the signal [**117**]. The authors also state that they were able to *'detect the message when even half of the message bits are consecutively in error'*, assisting greatly in the robustness of the watermark, at least when using the particular spread spectrum embedding mechanism that they did. As mentioned, this work was focused on the actual construction and subsequent extraction of the watermark. The authors contend that their watermark construction scheme is extensible to not only many different embedding techniques but also to other domains such as image and video watermarking. This makes the technique very useful but it also suggests that it requires much experimentation to categorise the relative effectiveness of the watermark detection process using different embedding techniques and in different domains.

5.3.4 Watermarking with 'MPEG 1 Layer 3' compression. In attempting to address some of the potential problems facing any watermarking scheme in the use of digital audio, Megias *et al* [**118**] focussed on trying to identify where in the frequency spectrum the watermark should be placed by using watermarking based on MPEG Layer-3 compression**.** They achieved this by analysing which components of the host audio would be susceptible to removal or alteration by modern 'MPEG 1 Layer 3' (better known as 'MP3') compression. Essentially, their scheme processed the unmarked host audio with a compression algorithm before reconstituting the original from the compressed audio. This might seem like a redundant or pointless process but what it did was identify which components would be fully recovered after compression / decompression. It is worth noting that the quantity of fully-reconstructed components would be a factor of the compression bit rate but even if an extremely high bit rate was used, some components would still be altered due to the nature of the MP3 compression algorithm.

At this point, the authors proceeded to embed their watermarked message in the frequency components which were left unaltered (or altered only to a predefined threshold). This served to ensure that, in the event that the watermarked audio would be compressed /

decompressed using an MP3-compliant algorithm, the watermarked components would be assumed to be recreated intact. The scheme then proceeded to increase or decrease the magnitude of the chosen components, obtained by computing the FFT of the host audio. The increase or decrease factor that was applied to the component by the authors was variable and could be 'tuned' as required. On performing an *'Inverse Fast Fourier Transform'* (IFFT), the altered signal was then transformed back to the time domain.

Watermark decoding was achieved by computing the FFT of the candidate audio, after MP3 conversion / compression if it was in another format. The process of identifying the potentially watermarked frequency components relied on a comparison against the original, meaning that the decode process was not a blind-decode process but rather an 'informed' process. The major advantage of the scheme was in its inherent natural ability to survive many compression type attacks.

The authors reported that, as would be expected, the watermark was fully recovered when the decode process used the same compression bit rate as the embedding phase. However, they also reported that even with re-compression of the candidate audio to MP3 with bit rates much lower than the embedding bit rate, recovery rates were promising. They used bit rates of 128kbps for embedding and were able to recover the watermark after any level of up-compression (to higher bit rates) of the candidate audio. This might be expected, of course, but they also had promising results in the order of 97% recovery when the candidate audio was subjected to down-compression to bit rates as low as 64kbps.

It would seem likely that there is scope for developing this scheme for indentifying watermark embedding regions. One obvious option is to use lower bit rates for MP3 compression before watermarking. This would mean less unaltered components to be used for embedding (i.e. lower capacity) because more components from the original audio would be affected by lower bit rate compression. However, it would conversely mean more robustness of the message after compression even into the lowest bit rates.

While the requirement to have access to the original audio for decoding is one limitation of the scheme, a potentially more serious disadvantage is that of conversion across codecs. Given that the published MPEG standards do *not* include instructions or requirements for the design of an MPEG-compliant file encoder but proscribe only the content of the MPEG-compliant file *after* encoding [**55**], providers are free to create their own encode/decode mechanism. How this encoding is achieved is entirely up to the designer of the encoder. If a particular MP3 codec was used for identifying the relevant areas used to represent the watermark, and the marked file was then cross converted either by another MP3 codec or amongst any of the myriad compression codecs (not just MP3) that are in modern circulation the possibility is that some of the components that would otherwise be preserved unaltered might be altered beyond recovery of the watermark. Nevertheless, the scheme did prove to be highly robust against compression, at least insofar as it was tested, and seems like a promising approach to ensuring robustness of a scheme against modern compression attacks.

A similar compression-based watermarking scheme was presented by [**119**]. In this scheme, instead of adding the watermark to a signal and *then* compressing it, the watermark is spread using spread-spectrum modulation before being added to the cover signal while it is undergoing compression. In essence, the cover signal is passed through a perceptual model to ascertain the appropriate compression parameters. The same analysis parameters are then used to transform the watermark signal. Since the spectral values of the watermark and cover signal are the same, they are then simply added together on a line-by-line basis. This work was again robust against compression attacks but could be susceptible to the same issues as other spread spectrum schemes, as outlined in Section 5.2.6.

5.3.5 High capacity watermarking schemes.

Fujimoto *et al* proposed a scheme which reports extremely high capacity [**120**]. Normally, one might expect capacity to be in the order of a few bits per second (i.e. a small number of bits of a watermark sequence per second of host audio) and sometimes lower, depending on the intended application domain. Fujimoto *et al* reported capacity more than 1kbps, which is extraordinary, especially in the context of a claimed transparent scheme.

Imperceptibility and transparency are often mutually conflicting, so achieving high capacity in a transparent manner is noteworthy. This work has been developed since, and the authors and others have produced some promising results. In a recent publication, Fallahpour and Megias [**121**] reported data capacity of about 3kbps '*without significant perceptual distortion*' along with good robustness to common attacks. Their scheme differed from Fujimoto's by using a variation on the concept and interpolating FFT samples to represent the watermark. They claim that capacity and imperceptibility could be improved depending on the intended application and domain.

5.3.6 Audio watermarking for live performances.

Watermarking of live performances is an area of little research interest, perhaps because of the limited financial losses incurred by the majority of the music industry or because it is actually the artists and performers who lose out on illegally copied live performance content, rather than the corporate sector of the industry. Whatever the reason, the output from research into this area is quite small. In one of the comparatively few publications in the area, Tachibana [**122**] reports on experiments to mix a watermark into a live performance that is either fed through a mixing desk and mixed therein with the output of the performer or is instead mixed 'in the air' with the sounds of the performance and, presumably, the venue.

It is an interesting approach to the field of audio watermarking and it is also an area that may have increasing usefulness in the future as more and more of an artist's or performer's revenue, particularly for the more established names, is derived from live performances and sales of recordings of these. The purposes of watermarking in this context is twofold: first, to identify the origin of illegally-recorded material in order to try to identify means of preventing them and second, to perhaps add a watermark that might make the recorded material lose audio quality when played back elsewhere, thereby making it less attractive to organised illegal copiers.

In his conclusion, Tachibana included a very perceptive and otherwise apparently unconsidered issue, namely that of ownership of the watermark itself. The statement '*if a* *person plays music and another person generates a watermark sound, an ownership conflict could occur'* raises a very important question in terms of the legal position relating to the Copyright and ownership of watermarks, especially if they have been added to content owned by another person and more importantly, especially if this is done without consent such as when added by a radio or TV broadcaster. The signal that contains the watermark is, by definition of the legislation, the property of the watermark creator. This point might only be of theoretical legal interest but it is a perceptive insight nonetheless.

5.3.7 Chaos-based watermarking scheme.

In an attempt to develop a watermarking scheme that addresses some of the issues relating to the *creation* of the watermark by the use of pseudo-random sequences, Mooney and Keating proposed the use of a chaos-based watermarking process [**123**]. While directed at the area of image watermarking, the concept would appear to be as valid in watermarking video and audio in that it is one of watermark *creation* rather than *embedding* and the watermark generation phase in most schemes, particularly those in the transform domains, is actually independent of the embedding phase.

The scheme described in [**123**] provides for a semi-blind decoding phase since the chaotic function used to generate the watermark is based on an initial value and a seed value. Each subsequent product of the function using the same two values will result in the same output. The detection algorithm needs only these two values in order to recreate a chaotic sequence and then examine candidate signals for the presence of the sequence before decoding it back the watermark.

Having two separate input parameters means having two separate, and independent, 'keys'. Without both keys, rather than one pseudo-random value as a key, there is a reduced possibility of the watermark being decoded by an unauthorised recipient. Even having discovered *one* of the keys, an attacker will not be able to decode the watermark. This makes the process more secure against unauthorised decoding.

The advantage of chaos-based watermark creation over pseudo-random techniques is primarily one of robustness against bandpass filtering and, secondarily, the information required to enable watermark detection. As an extension of the robustness of chaos-based watermark creation, there is also scope for controlling the watermark so its resistance against filtering is application and/or domain dependent. Initial value and seed value for the chaos algorithm can be selected and adapted depending on the intended application domain, so as to be most robust against likely attacks (common or accidental) in *that* domain.

While there seems to be little in the literature relating to the use of chaotic techniques in *audio* watermarking, adaptation of the technique for use in this related area might warrant further examination since the generation of a watermark and its subsequent embedding in a host signal are independent steps that are not necessarily co-dependent.

5.4 Summary

As evidenced by the various disparate techniques outlined in this chapter, digital audio watermarking is an area that has seen much research in recent years. Researchers have approached the problem of transparently hiding data in a cover signal in many different ways. Some schemes are more successful in achieving their goal than others. Some are perhaps more useful in very specific problem domains or for very specific applications. Some are more complex or computationally costly than others. A comparison of some of the more commonly researched techniques is shown in Table 5.1.

Technique	**Advantages**	**Disadvantages**
Least significant bit.	Computationally simple. Mature technique in image processing. Resistant to time-shifting of the audio. Can be decoded without original or watermark.	Least-significant bit is most likely to be altered by typical signal manipulations. Not suitable to audio processing due to sensitivity of auditory system. Masking techniques would be inconsistent.
Echo-hiding.	Human auditory system often cannot perceive very short echo. Audible added echo can cause sound to be perceived as	Weak against time-domain compression and time-shifting. Low-amplitude echo susceptible to interference from

	'warmer'. Addition of 'pre-echo' can improve effectiveness.	channel noise. Original or watermark required for decoding.
Amplitude masking.	Computationally simple. Ineffective in signals with low-powered components.	Not suitable for real music. Problems encountered with interference from inherent components.
Phase coding.	Human auditory system generally unable to detect phase.	Phase inconsistency between components must be controlled. Easy to identify the presence of a watermark.
Multiplicative techniques: **Discrete Fourier Transform** **Discrete Cosine Transform** **Wavelet Transform**	Transformations of cover audio that take place before a watermark is then added, allowing for different types of watermarking to take place.	These are not watermarking techniques as such. Instead, they are a primary step to perform before the digitised audio is in a form that can then be watermarked.
Spread spectrum.	Watermark spread across spectrum so appears as noise. Robust technique, even against signal corruption. Can be embedded in significant components, which are more likely to survive attacks.	Susceptible to corruption by compression. Requires the signal to remain synchronized.
Patchwork method.	Can be implemented as a blind-decode system. Can be designed to be robust in given domains.	Low capacity. Computationally complex.
Interpolation method.	Good capacity. Resistant to Digital/Analogue conversion and transmission.	Added 'virtual' components can cause perceptibility issues.

Table 5.1: Comparison of some common watermarking techniques

Insofar as broadcast monitoring is concerned, no particular technique is right or wrong. If a watermarking scheme is successful in transparently embedding a hidden message in cover audio in such a manner that it can be accurately recovered after the types of attacks that would be common in a broadcast domain (e.g. format conversion, analogue transmission, compression), then it would be suitable for the application under consideration. In the next chapter, such a watermarking scheme is proposed and developed.

Chapter 6: Low Complexity Watermarking Scheme

In this chapter, an audio watermarking scheme is proposed and developed. The intention from the beginning was to address the specific requirements of a broadcast monitoring application rather than a generic audio watermarking application. The watermarking scheme should fulfil the requirements of robustness against common attacks likely to be faced in the broadcast domain, particularly analogue transmission. In addition, the watermark should also be perceptually transparent and difficult to remove without adversely affecting the quality of the audio.

The procedure investigated in this research is the embedding of a unique identifier in a host audio file in such a way as to be inaudible, easily recovered and robust. Any type of information can be hidden in sound if it can be converted into sound itself. The principle behind this is simple: some sounds cannot be heard by humans if they are 'masked' by other sounds. Similarly, some sounds are simply outside the hearing range of humans, such as high-frequency dog whistles. Further, some sounds are within the hearing range of only the most effective ears, such as the 'mosquito-buzz' deterrents used by Police forces and Local Authorities to target loitering groups of teenagers [**124**], which cannot be heard by adults as their hearing has deteriorated slightly.

Psychoacoustics, as mentioned in Section 4.2, is the area of research devoted to these concepts and the MP3 audio format is just one well-known technological advancement that has been developed using the results of psychoacoustic research. If a watermarking scheme is designed to add some signal to the cover audio, then care should be taken to ensure that audibility is avoided. Knowledge of psychoacoustic principles underpins decisions taken when designing such schemes. On the other hand, if a scheme does not actually *add* anything to the cover audio but instead modifies it slightly so the watermark can be represented by some characteristics of the cover audio itself, then the considerations relating to audibility are likely to be less of a problem. While psychoacoustic modelling is an additional step that can be applied to watermarks before embedding them into the host or cover audio in a watermarking scheme, it may not always be necessary.

As discussed in Section 4.4, one major advantage of watermarks over fingerprints in terms of their application is that the decode process often does not need to have access to the original audio or any digest of it. There are many different forms of watermarking decoding technique and all rely on their corresponding embedding technique. However, all decoding techniques fall into one of three types.

When the decoding phase of a watermarking scheme does not need access to the original host audio, or to the form of the actual watermark before it was embedded, it is said to be a ‘blind’ decode process [**125**]. Conversely, if the decoding phase requires knowledge only of the watermark message or some criterion that was used in the watermark embedding process, such as the frequency range in which the watermark was embedded, it is said to be ‘semi-blind’. Finally, if the decode phase requires knowledge either of the host audio prior to watermarking or to the watermark itself, in order to successfully decode, it is said to be ‘informed’ decoding. This latter type of decoding is less useful in the case of broadcast monitoring as it suffers many of the same limitations as fingerprinting.

Generally, a self-contained blind decode watermarking scheme could probably be adapted to most purposes, including those that could otherwise be achieved using a semi-blind or informed scheme. The reverse is not the case. In short, a blind watermarking scheme that provides acceptable levels of transparency, robustness and security as well as reasonable capacity is likely to be most useful in the vast majority of applications, all other things being equal, including in a broadcast monitoring scheme.

A blind or semi-blind decoding technique removes the need for a central publicly-accessible data store of watermarks or audio files for comparison purposes. A watermarking scheme that can accomplish decoding without any external input or additional resources, other than the candidate watermarked audio, is the ultimate goal of most audio watermark research, albeit with acceptable levels of transparency, robustness, security and capacity.

6.1 First phase of the development of the watermarking scheme

The purpose of this research is to attempt to define a blind or semi-blind audio watermarking scheme with applications in the domain of monitoring of radio and television broadcasts with a view to transparent and accurate reporting of broadcast output. The primary motivation is one of equitable administration and distribution of Copyright royalties [**126**]. The initial hypothesis for such a watermarking scheme is outlined in Figure 6.1 and was inspired by previous research by Gopalan *et al* [**86**] [**127**]. It initially revolved around the well-understood *'Dual Tone Multi-Frequency'* (DTMF) standard [**128**] as used in touch-tone and mobile telephony where two tones are combined to represent a single piece of information. The initial decision to embed the watermark by manipulation of the DTMF frequencies was prompted by the possibility of the scheme being implemented using modern telecommunications systems that are already DTMF-compliant. It was believed that this would enable candidate audio to be identified over a standard telephone system without any additional hardware.

The initial idea was to reduce the data to be watermarked (the unique identifier, in this case the ISRC code, as described in section 3.6.1) to a series of bit-representations of its ASCII codes. Every alpha-numeric character has a unique ASCII code and since all of the characters of the ISRC are simply letters and digits, it made sense to use the decimal ASCII values, converted into binary representation, as the basis for the watermark. Once the Binary sequence was created, it was then used as the pattern for the creation of a pair of pure sinusoidal waves using the combined DTMF standard frequencies for '1' and '0'. The process is illustrated in Figure 6.1a and the spectrogram of the resultant signal is illustrated in Figure 6.1.b. The DTMF tones consist of [**128**]:

DTMF '1' tone: 697Hz and 1209Hz combined
DTMF '0' tone: 941Hz and 1336Hz combined

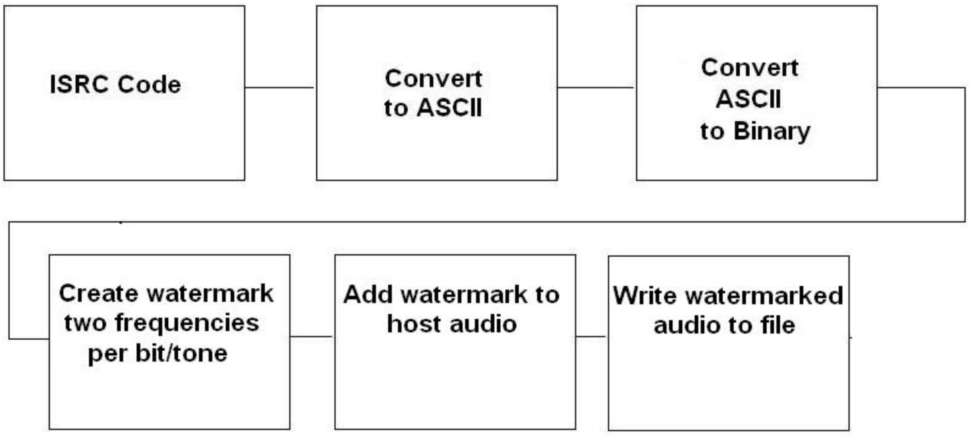


Figure 6.1a: A simple block diagram of the watermarking scheme.

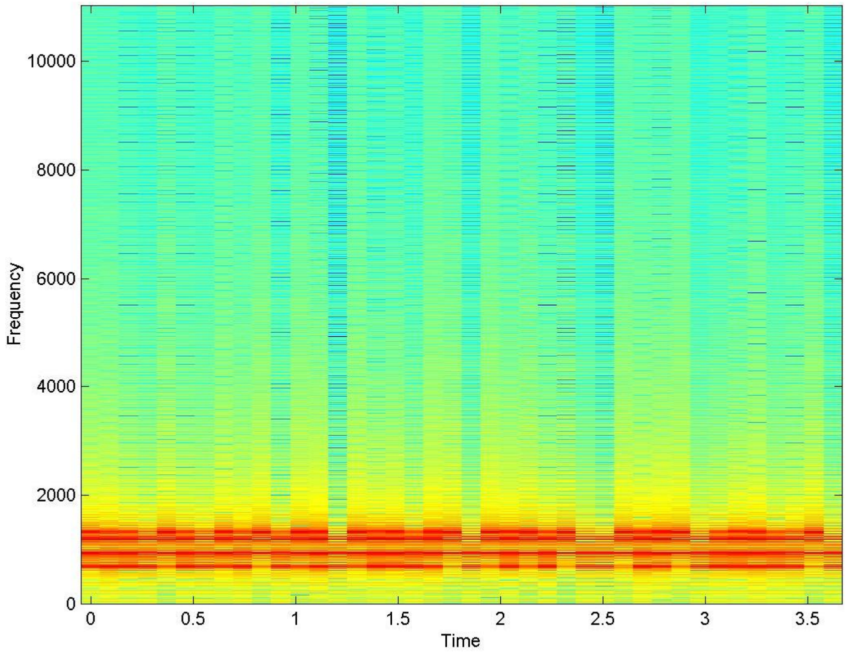


Figure 6.1b: Spectrogram of a section of audio illustrating the DTMF-based watermark pattern. In this case, the frequencies used are 697 Hz, 9412 Hz, 1209 Hz and 1336 Hz.

To ensure that no waveform discontinuities occurred when concatenating the sine waves the instantaneous frequencies of the pattern were first created, integrated to generate the phase, and then the *sine* of this phase was taken resulting in a smooth waveform. To illustrate, denoting the two possible tone combinations for a '0' and '1' respectively as f_0 and f_1, and given a bit sequence of '101' (sequence length = 3), with the tone for each bit lasting for *N* samples, the watermarked signal is described as:

$$y[n] = \sin(\phi[n]) \tag{6}$$

where $0 \leq n \leq 3N - 1$ and the generated phase is the integral of a normalised sequence of frequency values, with sampling frequency F_s

$$\phi[n] = \int 2\pi[\, f_1(n_1),\ f_0(n_2),\ f_1(n_3)]n/F_s \tag{7}$$

and
$$0 \leq n_1 \leq N - 1 \tag{8}$$
$$N \leq n_2 \leq 2N - 1 \tag{9}$$
$$2N \leq n_3 \leq 3N - 1 \tag{10}$$

The sequence was generated for 12 ASCII characters of 8 bits, each lasting 20ms (total watermark length = 1.92 seconds). Finally, after storing the watermarked audio as a file (WAV format) and reading it in again before decoding the watermark, this initial proof-of-concept experiment resulted in 100% recovery of the watermark as would be expected in the absence of any attack on the signal.

The length of each tone in this case was set at 20 milliseconds although this length is not necessarily fixed. The duration of each tone can be reduced to any length as long as it can still be detected by the decoding process after the candidate audio is sampled appropriately. It is worth noting that shorter tone duration would logically increase the capacity of the watermarking scheme. However, reducing to the absolute minimum may increase the likelihood of missed tones and false decoding of the embedded message. The process of embedding and decoding a single '0' bit is illustrated in Figure 6.2a.

In an effort to aid robustness of the watermarking scheme, it was decided to implement a *Mode* operation in the decoding phase. This operation would increase the likelihood of the watermarked message being recoverable. In order for the mode operation to be successful, the watermark message, which has been converted into a bit sequence, would have to be embedded repeatedly throughout the length of the cover signal. The number of loops of the watermark was therefore a function of the watermark length and the signal length.

Upon decoding the bit sequence embedded in candidate audio, it contained multiple loops of the watermark and so it was first broken into subsets of the length of the watermark. Each subset was then broken into sequential 8-bit lengths and converted back to decimal. At this point, the mode of the set of values encountered at each index of the watermark was chosen as the watermarked value. It was felt that, in the event that one instance of the watermark in some part of the signal was corrupted, the remaining loops of the watermark would be unlikely to be corrupted in the same manner. Therefore, taking the mode, or the most common value in a given index of the watermark, would increase the likelihood of identifying the correct watermarked value. The process is shown in Figure 6.2b. Initial experiments to embed/decode the watermark independent of the host audio file were very successful, as would be expected, returning 100% precision.

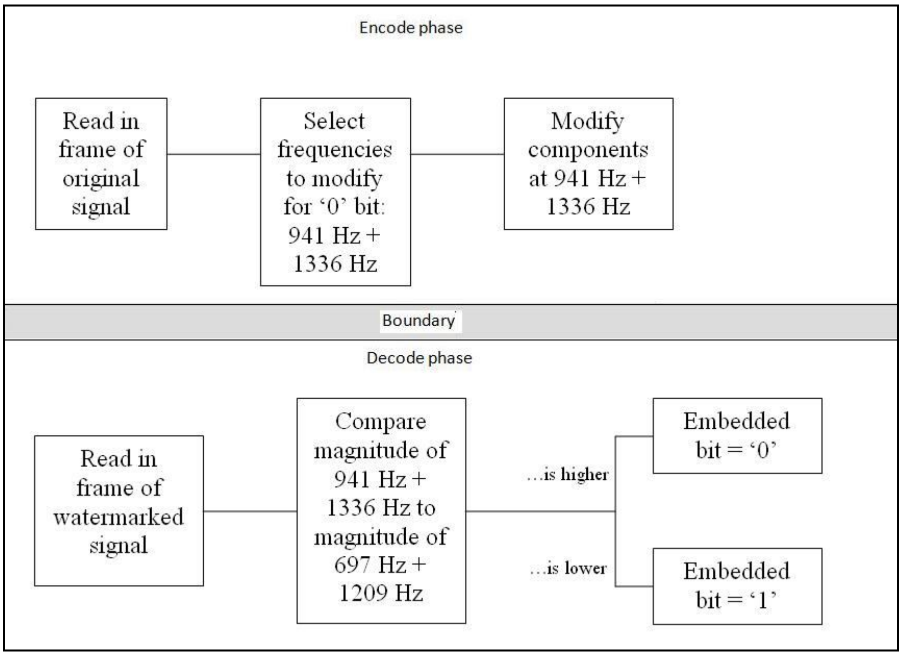


Figure 6.2a: Block diagram illustrating the encoding and decoding steps.

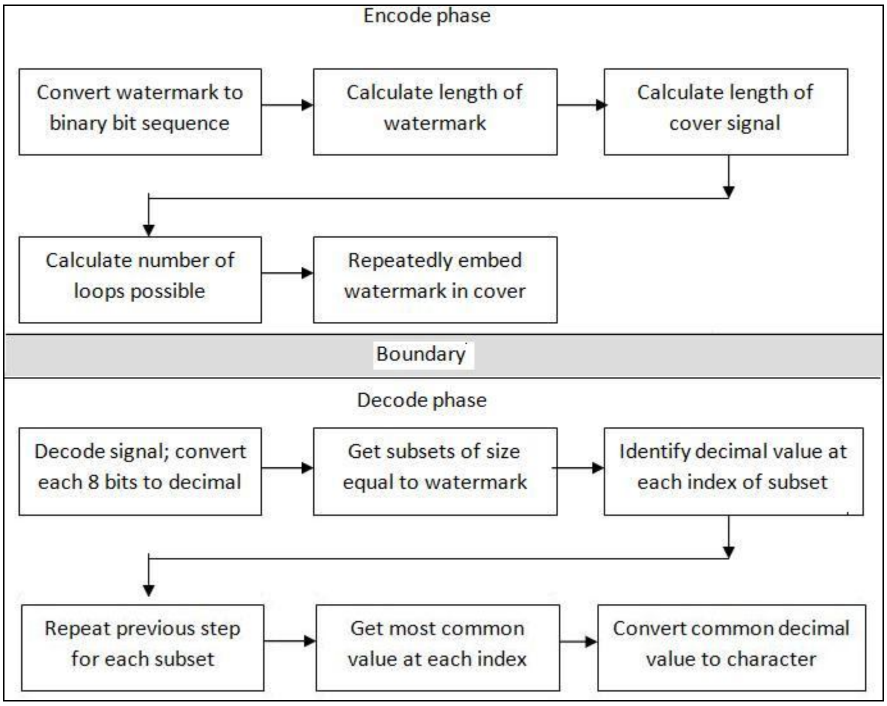


Figure 6.2b: Block diagram illustrating repeat embedding in the encode phase and the mode operation in decode phase to aid robustness of the scheme.

These experimental watermarked audio files were written to CD and played to listeners in various environments. As the purpose of the listening test at this point was not intended to be definitive, none of the volunteer listeners were required to be ‘expert’ and the equipment used to reproduce the audio differed in each case. The audio files used for the casual listening test were chosen at random from audio files in a variety of genres.

Once informal listening tests were carried out, the limitations of the system started to become obvious. Initially, embedding the watermark was achieved by simple addition of one sound (the watermark signal) into another (the cover audio). This resulted in the watermark being audible. Since inaudibility of a watermark in the presence of a host signal is a constraint of any useful watermarking system, it was decided to reduce the amplitude of the signal representing the watermark in order that the host audio might mask its presence in the manner of auditory masking described in Section 4.2.1. Using this relatively basic method of amplitude reduction until the watermark signal became inaudible, it was embedded satisfactorily into cover signals again and a ‘straw-poll’ of listeners to ten watermarked tracks suggested that the watermark could not be readily detected. Again, listening tests at this point were not intended to be definitive so no particular measures were taken to control the test.

6.1.1 Identifying components present in the signal

The problem of isolating and subsequently decoding the watermark then became obvious. The initial intention was to analyse the candidate audio file in which a watermark was believed to be present in order to identify its actual frequency content. A simple iterative check of each 20ms 'block' of audio would determine the presence or absence of the frequencies sought. Identification of *both* frequencies meant that the corresponding bit ('1' or '0') was present. Eventually, a sequence of bits would be found using this method and decoded back from binary to ASCII and ultimately to alphanumeric ISRC characters.

The first major problem encountered was to identify which actual frequencies were present in the candidate audio. This was not as simple as might be thought. For example, the *'Fast Fourier Transform'* (FFT) method of decomposing complex audio signals into fundamental components is limited in that is does not analyse a signal for the presence of *particular* frequencies. Rather, because of its uniform sampling of the complete frequency spectrum, it will only plot the relative strengths of components that may be close to but not exactly the frequency locations of interest. This problem was partially overcome by using the Goertzel algorithm instead. This is because the Goertzel algorithm can be used to identify components energies that exist at *specific* frequencies in a signal, whereas the FFT identifies components across a bandwidth. The Goertzel algorithm is also more efficient than the DFT [**129**]. It essentially defines a high Q, narrowband, second order IIR filter. Given an input $x(n)$, the first stage of the Goertzel algorithm produces an output according to:

$$s(n) = x(n) + 2\cos(2\pi k/N)s(n-1) - s(n-2) \qquad (11)$$

where N is the length of the input,
k is a frequency index value with $0 \le k \le N-1$
and $s(-1) = s(-2) = 0$.

The second stage then gives

$$y(n) = s(n) - e^{-j2\pi k/N} s(n-1) \tag{12}$$

The only output of interest from the algorithm is $|y(N)|^2$ which is the energy of the component at the frequency of interest and is derived according to the following formula:

$$|y(N)|^2 = s^2(N) + s^2(N-1) - 2\cos(2\pi k/N)s(N)s(N-1) \tag{13}$$

6.1.2 Synchronising the watermarked bits

Another issue that arose in the decode phase and needed to be addressed was identifying where the bit sequence started and ended. Some form of synchronisation had to be used. It was originally decided to use the DTMF tone for the '*' key (1209 Hz and 941 Hz combined [**128**]) as a reference point to signal the start of the bit sequence since it would never need to be used for the bit sequence that represents the ASCII code that makes up the watermark message. Additionally, it was necessary to identify where one bit ended and the next began because if the monitoring of the host audio did not commence from the very beginning of the track (which would be unlikely in a broadcast scenario) the simple iteration over the candidate audio and analysis of each 20ms block would be useless as a measuring scale. Furthermore, a repeated '1' or '0' bit in the sequence might not be easily identifiable as the decoding process had no way of discerning if the decoded tone represented only one bit or two instances of the same bit in sequence. It was decided to add the DTMF tone for the '#' key (1477 Hz and 941 Hz combined [**128**]) in between every bit in the bit-sequence. While this made it easy to identify where one bit-tone ended and the next one began, it also doubled the length of the watermark, thereby halving the capacity of the scheme. However, this was not an issue at this point since capacity was not being considered as a priority.

6.1.3 Problems encountered with pre-existing components

Once these changes were made, it became easier to identify the segments of audio that represented the bits in the watermark and this led to a much improved decoding phase. However, there was another issue that needed to be addressed, namely what would happen in the case where the host audio, before being watermarked, already had high-power components at the same frequencies as used for the watermark tones. The components of five separate frequencies (697 Hz, 941 Hz, 1209 Hz, 1336 Hz and 1477 Hz) would be required to be added to the cover audio in various combinations to represent the watermark bit, the separator tones and the start point. However, these components may already be present in the cover signal.

This would not be a problem if the current frame had large-magnitude components at the frequencies needed to represent the bit that was to be added. However, if it contained components of the frequencies for the *other* bit, the separator tones or the synchronising tone, this could lead to inaccurate decoding. By way of illustration, consider if the bit to be embedded was a '0', comprising the frequencies 941Hz and 1336Hz and the frame of the host audio already had strong components at frequencies 697 Hz and 1209 Hz, there could be a false positive found in decoding. The decode process would evaluate the power of the various components in the frame and report the stronger of the two (in this case the 697 Hz and 1209 Hz components) as the bit embedded. In this example the stronger components would be those of the wrong bit but the decoding process would naturally return them as an incorrect value at that point in the binary sequence, thereby producing an inaccurate result.

6.1.4 Analysis of representative audio for inherent components

At this point, since the matter had become an issue worthy of consideration, it was decided to analyse audio in various musical genres. A statistical analysis of the *actual* frequency content of 100 audio files at specific frequencies of interest was performed. This was achieved by applying a 1 Hz width bandpass filter at the DTMF frequencies followed by the Goertzel algorithm to try to ascertain whether there was any correlation between the powers of the frequencies already inherent in general audio and the DTMF frequencies.

As can be seen from the illustrations in Figure 6.3a and 6.3b, the energy of the components inherent in general audio that are at the DTMF frequencies that separately identify the '1' and '0' from each other is very low. In most cases it was zero or almost zero. Nevertheless, there is sometimes energy at each component so no guarantee could be given before watermarking that the component being added would not be replaced in decoding by a component that was inherent in the host. It was clear that this method would be unsuccessful in a practical application.

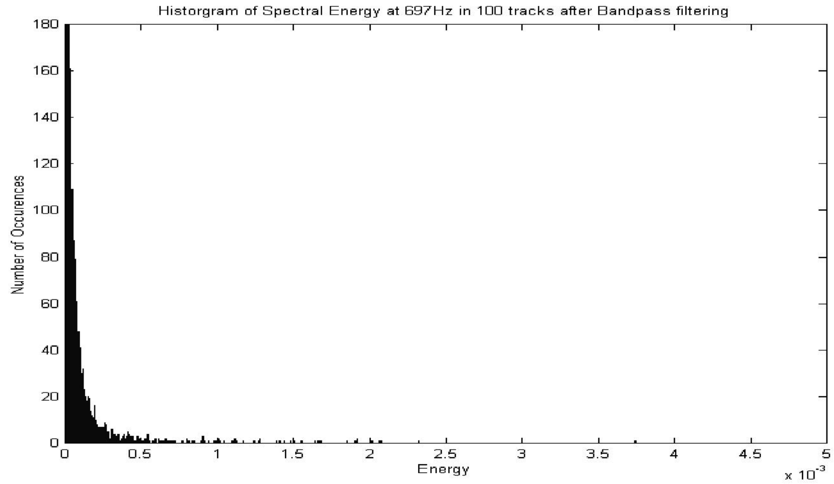


Figure 6.3a: Histogram Plots of Spectral Energy at 697 Hz determined by bandpass filtering and Goertzel analysis of 100 audio files.

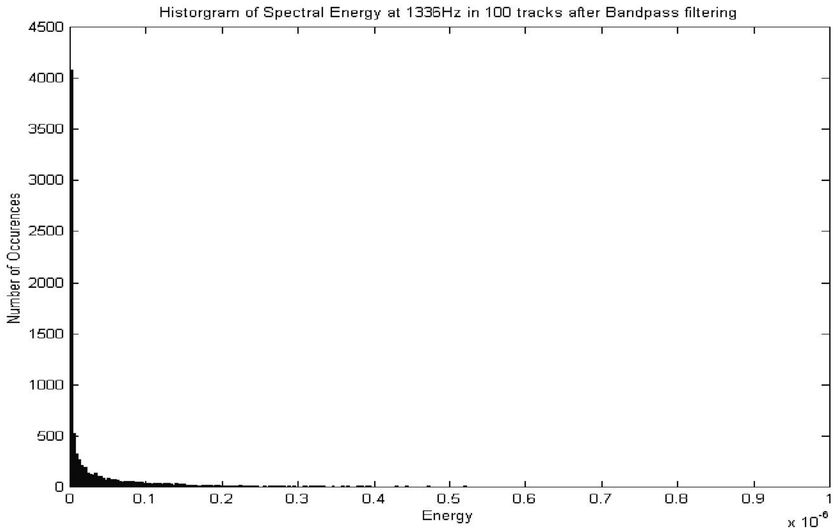


Figure 6.3b: Histogram Plots of Spectral Energy at 1336 Hz determined by bandpass filtering and Goertzel analysis of 100 audio files.

6.1.5 Weighting of frequency pairs

The next development attempted was to weight the two tones representing the '1' and '0' bit against each other so that both the frequencies for the '1' and '0' were embedded in the host audio simultaneously, rather than separately, but at powers that made it easy to decipher which one was the required bit. This experiment was based on work by Gopalan and Wenndt [**86**] [**127**] which illustrated the use of a single frequency whose individual powers were weighted by the total power of the frame into which they were being embedded. Thus, if it was desired to embed a '1' then, denoting the power at any frequency f in the k^{th} frame by $P_k(f)$, the algorithm adjusts the two component magnitudes at frequencies f_0 and f_1 so that the following relationship is satisfied:

$$\left(P_k(f_1) \Big/ \sum_f P_k(f) \right) > \left(P_k(f_0) \Big/ \sum_f P_k(f) \right) \tag{14}$$

Otherwise, if it is desired to embed a zero, the inverse relationship is specified as follows:

$$\left(P_k(f_1) \Big/ \sum_f P_k(f) \right) < \left(P_k(f_0) \Big/ \sum_f P_k(f) \right) \tag{15}$$

In using two components instead of one, the components were also controlled in such a way that their magnitudes met predefined minimum relational criteria to each other as well as the overall power of the frame. In the above example, the two modified components simply had to fulfil the criteria of one being of bigger magnitude than the other. However, this was then adapted so that the magnitudes of the components must differ by a threshold or margin. Various thresholds were considered and it was found that there was little difference between them, as long as the ratio of the powers of the *desired* bit to the *undesired* bit was significant enough to be detectable while the undesired bit was sufficiently powerful to be independently detectable. The process was therefore adapted to:

$$\left(P_k(f_1) \Big/ \sum_f P_k(f) \right) > c + \left(P_k(f_0) \Big/ \sum_f P_k(f) \right) \tag{16}$$

where a ‘1’ bit was to be embedded and

$$c + \left(P_k(f_1) \Big/ \sum_f P_k(f) \right) > \left(P_k(f_0) \Big/ \sum_f P_k(f) \right) \tag{17}$$

where a ‘0’ bit was to be embedded. The value c represents a threshold between the magnitudes of the two components and it is constant from frame to frame. Different values for c were experimented with and there was little difference in the decode result as long as c was above a minimum value. A sample of the weighting ratios used in modifying the component pairs, along with results from encode/decode trials, is shown in figure 6.4.

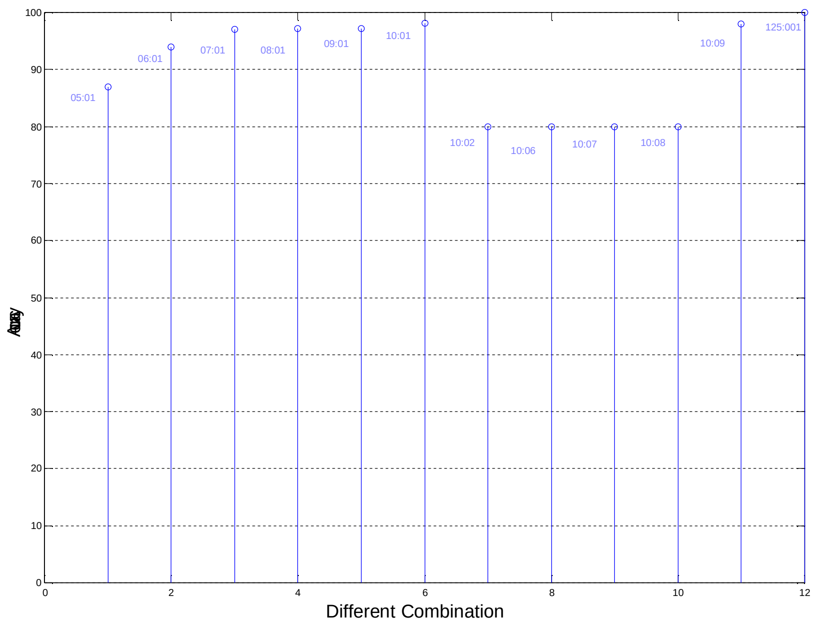


Figure 6.4: Results of watermark embed/decode cycle when component pairs are weighted.

Figure 6.4 illustrates the precision of recovery of the watermarked message when using different combinations of weighting ratios between the two components and the average power of the frame in which they are embedded. The ratios noted at the top of each stem are the ratios between components representing desired and undesired bits. The most accurate recovery was achieved by using a weighting ratio of 125:1, shown at number 12 on the horizontal axis, meaning that both components were added to the cover signal with the component representing the desired bit being set 125 times more powerful than that of the undesired bit. While this might then seem to be an ideal ratio to use, it raises the issue of weighting such components against the frame average.

The 125:1 ratio could be maintained by making the desired component comparatively large in relation to the frame average, meaning that the other component would be large enough to be independently detectable. However, it could mean that the component representing the desired bit might be of too large a magnitude in comparison to the rest of the frame's components, leading to audibility problems. Alternatively, if the component representing the desired bit was set to a low enough magnitude to minimise audibility, the other component may be of such a small magnitude as to be undetectable.

Various weighting ratios were compared and it was found that those where the component representing the desired bit was set between 0.07 and 0.1 of the average power of the frame and the other component was set to a detectable difference from this value, the precision of recovery was comparatively good. Note from Figure 6.4, for example, that numbers 3, 4 and 5 resulted in similar precision with weighting of 7:1, 8:1 and 9:1 respectively. Weighting of 10:1 produced results slightly higher but it is notable that a similar precision is achieved with ratio of 10:9. This serves to confirm the contention that the magnitude difference between the two components does not need to be very large, as long as it is enough to permit the decode process to calculate a difference between them.

6.1.6 Embedding at higher frequencies

When the weighted frequency-pairs were added to the host audio the resultant decoding was significantly more successful than previously, as shown by Figure 6.4. However, the issue still remained that the energy of some components inherently present in the original audio could impact on ability to correctly identify the embedded bit. In the event of the components representing the ‘0’ bit being present in a frame that was to have the ‘1’ bit embedded, and vice versa, they could negatively affect the result. Slight alterations were unsuccessfully attempted to circumvent this issue arising. The frequencies chosen to represent the ‘1’ and ‘0’ bits were changed from the DTMF tones previously used to higher frequencies, on the premise that higher frequencies would be easier to identify.

Using much higher frequencies for the embedded tones did in fact increase the detection reliability of the system. Conversely, this increased the likelihood of very high frequencies – and therefore the watermark embedded at those frequencies – being lost in incidental ‘attacks’ such as format conversion, FM transmission and perceptual encoding (e.g. MP3). This is because high frequency components outside the range of normal human hearing would be removed or altered by such attacks. Frequency pairs across a wide range of values were evaluated and successful decoding rates were compared.

Eventually, given the intended application domain for the proposed scheme, frequencies ranging from 8 kHz to 15 kHz were chosen for practical comparison. It was decided not to exceed a ceiling of 15 kHz to minimise the potential for the watermark to be lost, accidentally or deliberately, in perception-based attacks, as mentioned above. The equation for the algorithm as it was implemented at this point was:

$$\left(P_k(f_1) \Big/ \sum_f P_k(f) \right) > c + \left(P_k(f_0) \Big/ \sum_f P_k(f) \right) \forall \, \{8000 < f_1, f_0 < 15000\} \quad (18)$$

where the bit to be embedded is a ‘1’ and

$$c + \left(P_k(f_1) \Big/ \sum_f P_k(f) \right) > \left(P_k(f_0) \Big/ \sum_f P_k(f) \right) \forall \, \{8000 < f_1, f_0 < 15000\} \quad (19)$$

where the bit to be embedded is a ‘0’.

Despite achieving relatively good results from the watermarking scheme at this point, it was reasoned that the implementation outlined here would be inadequate in a practical watermarking scheme intended for broadcast monitoring. The most notable consideration was the frequency range chosen for the embedding components. If set too low, the watermark could be corrupted upon decoding by the presence of low-power components inherent in the cover signal as explained in Sections 6.1.3 and 6.1.4. Conversely, if the components selected for modification were at much higher frequencies, these components would be more likely to be lost in lossy compression based on perceptual models.

6.2 Second phase of the development of the watermarking scheme

6.2.1 Application of a notch filter

By way of evolution of the technique, it was decided to attempt the removal of any existing frequency content inherent in the host at the same frequencies as used by the watermark, before actually adding the watermark to the host [**130**]. Essentially, the host would have a 'hole' or 'notch' created at the desired frequencies, setting the power of these components to almost zero. The watermark would then be included at those frequencies, weighted against the frame they were embedded and also weighted against each other to create the desired relationship.

A *'notch filter'* was designed that was as narrow as possible in the range of frequencies it attenuated but those that it *did* attenuate were reduced to almost nil. In terms of filters, a notch filter is a type of band-stop filter. A band-stop filter will pass most frequencies but lower the magnitude of specified frequencies to very low levels. Notch filters may be visualised as the opposite of a band-pass filter. Where a band-pass filter allows frequencies in a band (or range) to pass, a band-stop filter prevents these components from passing. A notch filter is a derivation of this technique with *extremely* narrow range and such filters are often designed to attenuate only those frequencies in a one or two hertz range. An example of a notch filter is shown in figure 6.5.

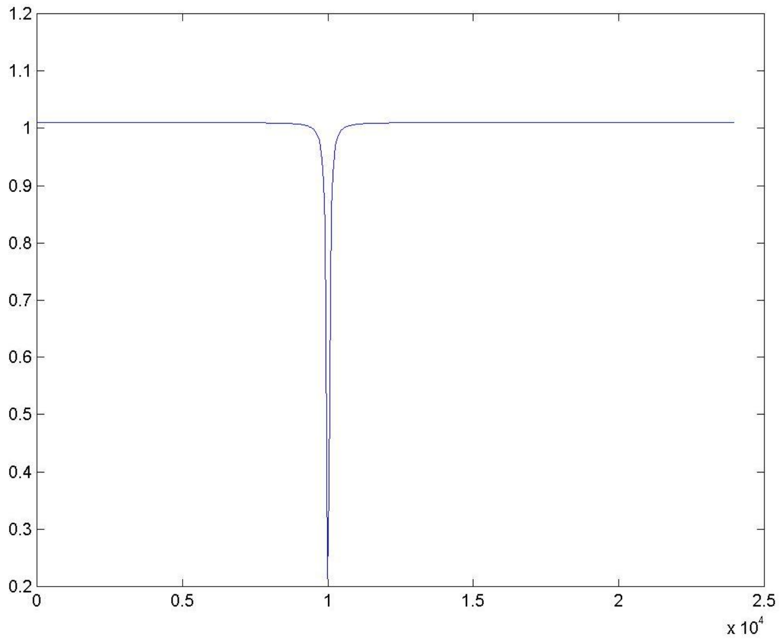


Figure 6.5: An example of a notch filter.

The magnitude response of the filter shown in Figure 6.5 attenuates the frequency centred at the middle of the notch (i.e. at 1.0 on the horizontal/frequency axis). It will have no impact on the frequencies in the lower or upper section of the frequency range but those nearer to the middle of the notch (0.95 to 1.05) will be minimally affected. Those in the middle of the filter will have their magnitude attenuated to almost nil. The transfer function of a notch filter is of the form [**131**]

$$H(z) = \frac{1 - 2\cos(\theta)z^{-1} + z^{-2}}{1 - 2\beta\cos(\theta)z^{-1} + \beta^2 z^{-2}} \qquad (20)$$

where θ determines the frequency of the notch and β determines the width of the notch.

In the application of notch filters to the watermarking scheme to attenuate the components at the embedding frequencies, it was hoped to remove any possibility that components inherent in the cover audio could negatively impact on the decode results. If such components were removed by the filter, and then manually set according to the desired relationship, this problem would be overcome without significantly affecting audio quality.

Figure 6.6a shows an example of a frame of audio under two conditions: (a) the original frame and (b) following a notch filter being applied to the frame and subsequent inclusion of the watermark signal at 10500 Hz and 11500 Hz. Note that the components at these two frequencies appear marginally different after the watermark has been added. Figure 6.6b shows a section of audio displaying the visible artefact of ‘notching’ the original signal in two places as a pair of horizontal lines across the signal at specific frequencies. Figure 6.6c shows the signal after subsequent addition of the watermark. Note that the pattern in these visible lines is the same as the pattern in the watermark bit sequence and so these lines provide a visible confirmation that a watermark may be present as well as initial clues for decoding attacks.

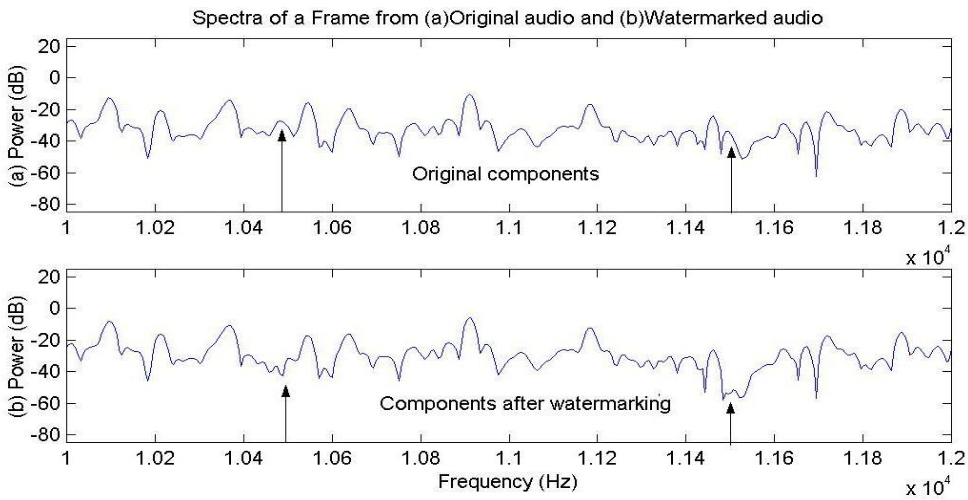


Figure 6.6a: Spectral profile of a segment of audio before and after it is watermarked.

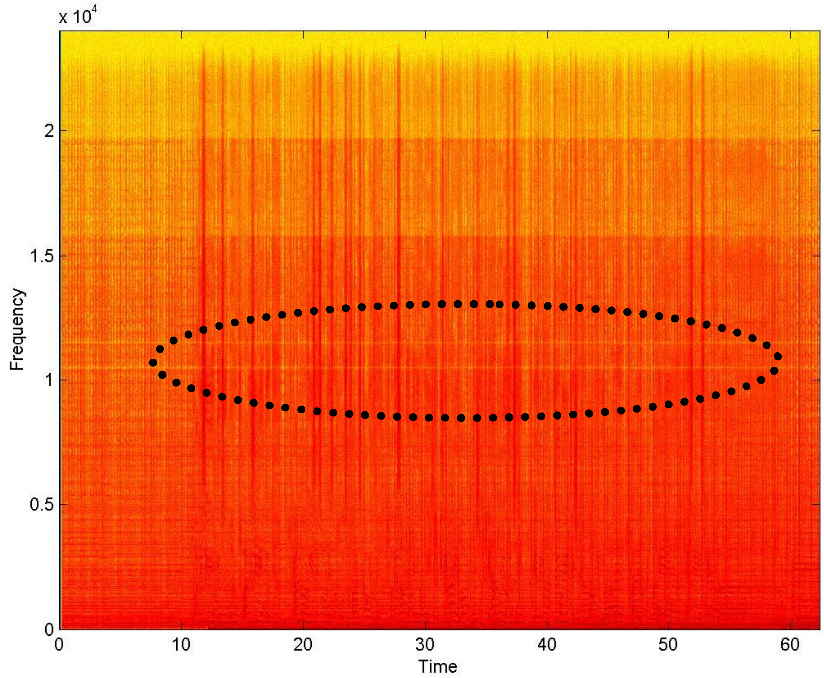


Figure 6.6b: A visible ‘artefact’ of the notching process can be distinguished as a pair of horizontal lines at frequencies of 10500 Hz and 11500 Hz, highlighted by the dotted ellipse.

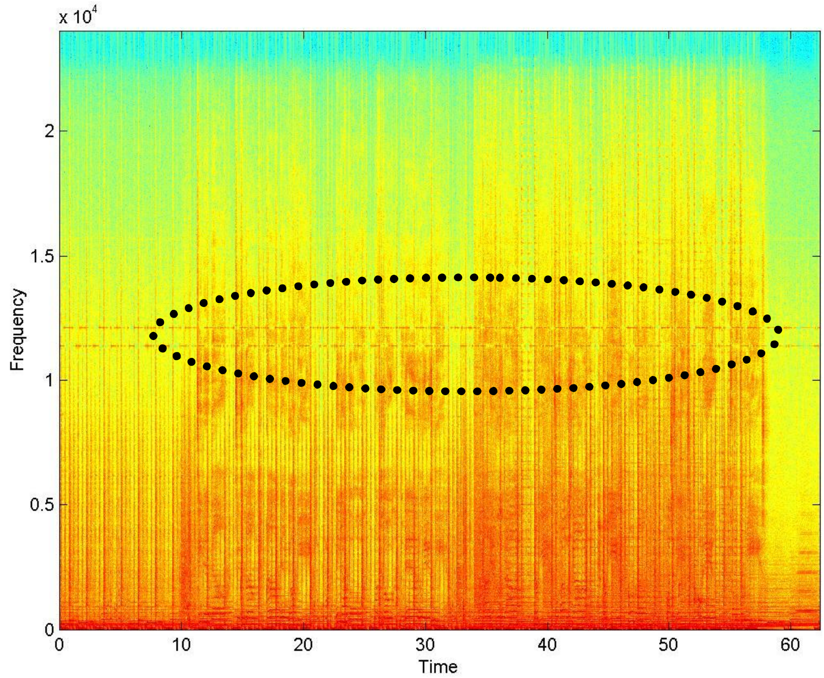


Figure 6.6c: A visible ‘artefact’ of the watermarking process can be distinguished as a pair of horizontal lines, highlighted by the dotted ellipse, which may offer some indication as to the watermark’s presence.

6.2.2 Optimising the watermarking scheme

Once the above process was completed, the files were analysed for watermarks, and the identified watermark decoded. Initial results were promising and various modifications were made to increase the success rate. The most effective modification, as mentioned earlier, was the use of higher frequencies to create the watermark tones. Frequencies in the region of 12 kHz to 15 kHz produced significantly better success, all other parameters being equal. Unfortunately, audio watermarked at the higher frequencies proved to have an unacceptably high level of audible artefacts so they were discounted.

The most acceptable compromise was in the frequency range of 9 kHz to 12 kHz which, under optimum conditions (in terms of the additional parameters) led to successful identification of the watermark for more than 99% of the tracks analysed (343 tracks correctly identified from 347 attempted). Moreover, the watermarked files did not display any watermark artefacts. In other words, it was not obvious which was the original and which the watermarked. Trial and error experiments were then performed on small groups of 25 - 50 audio files to ascertain which embedding parameters were promising enough to warrant full-scale analysis in order to identify optimum values.

A subset of the results of a series of experiments, showing the parameters that resulted in the highest precision, is provided in Table 6.1. The full set of results from approximately 11,500 iterations of the embed/decode cycle is presented in Appendix 3, Table A3.1 for completeness. Note that the columns represent:

Samples (S) = million samples decoded from the beginning of the audio file.
Frequency (F) = base frequency from which two embedding frequencies were derived.
Tone Length (L) = length of each tone per bit and separator in the watermark.
Files Decoded = total number of files watermarked and decoded with these parameters.
Percentage Accuracy (P) = accuracy of correctly recovered watermarks (rounded down).

The frequency value ‘F’ is the base frequency, intended to be a user-defined parameter. The two embedding frequencies were then calculated as ‘F + 1000 Hz’ to represent a ‘0’ bit and ‘F + 2000 Hz’ to represent a ‘1’ bit. The tone length ‘L’ defines the length of the overall watermark (one tone per bit plus one tone per separator), which is then looped repeatedly, dependant on the length of the cover signal.

No.	Samples (S)	Frequency (F)	Tone (L)	Files Decoded	Accuracy % (P)
1	4	9500	25	347	99
2	3	9500	30	346	99
3	3	9000	25	300	99
4	4	9000	25	694	99
5	4	9500	30	346	99
6	2	9000	25	135	98
7	3	9500	25	281	98
8	3	9500	20	347	98
9	4	7000	25	50	98
10	3	7000	25	50	98
11	2	9000	25	100	97
12	4	7000	20	50	96
13	3	6000	25	50	95
14	2	9500	25	347	95
15	3	7000	20	50	95
16	2	9000	25	100	95
17	1	12000	20	50	94
18	2	9500	20	694	93
19	1	12000	25	50	92
20	4	7000	16	50	92
21	3	7000	16	50	90
22	1	12000	10	50	90
23	2	13000	15	397	90
24	2	9000	25	347	90
25	2	12000	15	350	90
26	1	11000	13	25	88
27	4	7000	10	50	88
28	1	11000	13	25	88
29	2	11000	15	347	87
30	2	11000	15	347	87

Table 6.1: Results of the embed/decode cycle using different values for samples, tone and frequency parameters, with resultant decode precision.

Table 6.1 shows that there are a number of permutations of input parameters that each independently contribute to the accuracy of the decoder. Also, the decode phase itself can sometimes provide identical results using different numbers of samples. These observations will be discussed shortly. The results in Table 6.1 are a subset of the results from more than 11,500 encode and decode cycles of a group of approximately 350 audio files of varying genres from ‘Classical’, ‘Acapella’ and ‘Instrumental’ to ‘Rock’, ‘Pop’ and ‘Punk’. The complete results of 150 encode/decode experiments is included in Appendix 3.

The three parameters that contribute most strongly to the precision of the watermark embed/decode cycle are base frequency (F), tone length per bit (L) and the number of samples available at the decoder (S). Optimal values for these parameters are obtained according to:

$$\underset{\mathrm{s}}{Min}\ \mathrm{d}(\mathrm{S},\mathrm{L},\mathrm{F})=\{\mathrm{P}:\mathrm{P}\geq 99\%\}\ \forall\{6000\leq f\leq 12000\} \qquad (21)$$

where S = number of samples decoded; L = length of watermark tone per bit; F = base frequency manipulated to generate tones; P=percentage of watermarks decoded accurately.

In equation (21), *d* is a function that returns a value of P greater than or equal to 99% for the lowest value of S with input parameters S, L and F and with the constraint that F (frequency) is constrained to be in the range 6-12 kHz.

Note that watermarks were only considered to be accurately recovered if the system returned the *complete* identifier identical to that originally embedded in the host and that partially-correct identifiers were completely discounted as incorrect. Rather than calculating the number of *bits* that were accurately identified, the result was only considered successful if a complete identifier comprising almost 200 bits was successfully recovered. It was felt that this was a more useful metric than simply the number of bits that were accurately recovered as recovery of a string of bits would only be useful if actual watermarks could be decoded correctly from them.

6.2.3 Minimising the computational cost of decoding

The value S represents the number of samples read from the beginning of the audio file that are available for decoding to provide the result for P. In this regard, S has an increasing computational cost associated with it as it increases. However, the value of **S** in producing a more accurate P does not increase continuously or consistently so it is useful to find out at which value of S it ceases to have a beneficial effect on P so as to minimise unnecessary computational cost in decoding. Since the decoding process is intended to be performed in real-time or better, the number of samples required for decoding, while maintaining an acceptable level of accuracy, should be as low as possible. While the maximum value of S is constrained by the length of the candidate audio signal, the watermark length is far smaller so it warrants minimising of S to only decode the number of watermark loops that are required to provide acceptable precision.

This process of minimising the decode parameter S is preferred over minimising input parameters L or F since the encoding process is only performed once and so has comparatively little effect on overall computational complexity. However, as L increases, so the length of the watermark increases, meaning there will be a lower number of watermarks for the same number of samples decoded. The number of watermarks decoded can impact on the accuracy of the decoding process and so while it is true that **S** has more cost associated with it, L also has *some* cost-benefit trade-off in the decoding phase.

The choice of frequencies F to manipulate has no apparent effect on the computational cost of decoding. The choice of base frequency should be left to the user, within the constraints mentioned in equation (21), so as to allow for private watermarking by using the value of F as a form of ‘private key’ required for recovering the watermark.

The value for L is the length in milliseconds of the tone for each binary bit. The length of the watermark is a product of the value of L, where watermark length in ms = L x 193 tones. This affects the number of loops of the full watermark that can be embedded in a given host audio length. The decoding process has a number of samples (S) available from which to decode the watermark. The number of loops of the watermark that could theoretically be embedded in S is a product of S divided by the sampling rate (48,000 Hz), then divided by length of the watermark.

According to equation (21), the optimal set of parameters is defined as being the lowest value of S that returns P equal to or above 99%, with constraints setting 6000 < F < 12000. It was decided also to insist on the condition that L *(tone length per bit)* should be greater than 3ms in order to ensure it was not too short to identify in decoding without lowering the precision P. Other values of F and L are included in Table 6.1 and Table A3.1 in Appendix 3 for illustration purposes. In implementing a lower limit for L, equation (21) is now re-written as:

$$\underset{S}{Min}\ \mathrm{d}(\mathrm{S}, \mathrm{L}, \mathrm{F}) = \{\mathrm{P} : \mathrm{P} \geq 99\%\}\ \forall \begin{Bmatrix} 6000 \leq f \leq 12000 \\ \mathrm{L} > 3 \end{Bmatrix} \qquad (22)$$

The number of samples available for decoding the watermark is an important factor in producing accurate decoding results but, again, only up to a point. Not only is it not correct to say more samples always equals higher accuracy but a larger value for S equates to a much more computationally costly decode process so S must be minimised for an acceptable result of P.

For comparison purposes, an excerpt from Table 6.1 is shown in Table 6.2. Note that the results indicate that decoding 2 million samples of signals watermarked with F = 9000 and L = 25 provides 98% accuracy (line 6). Decoding 3 million samples, (line 3) produces the same accuracy as 4 million samples (line 4). However, note that cycles using the same parameters in lines 3, 16 and 24 provide markedly different results, suggesting that the cover audio or some unknown variable has some impact on the result. It is not correct, therefore, to say that a combination of three input parameter values will always result in the same decode precision.

No.	Samples (S)	Frequency (F)	Tone (L)	Files Decoded	Accuracy % (P)
3	**3**	**9000**	**25**	**300**	**99**
4	**4**	**9000**	**25**	**694**	**99**
	…	**…**	**…**	**…**	**…**
6	**2**	**9000**	**25**	**135**	**98**
	…	**…**	**…**	**…**	**…**
16	**2**	**9000**	**25**	**100**	**95**
	…	**…**	**…**	**…**	**…**
24	**2**	**9000**	**25**	**347**	**90**

Table 6.2: Excerpt from table 6.1.

It would appear from the results of over 11,500 iterations of the cycle that a watermark with L=25ms tone length is optimal for accuracy at an acceptable level of 99%. However, using the same tone length in watermarks where the frequency manipulated is in the 6000 Hz range produces unacceptable results (Table 6.1, line 13). This suggests that the tone length value is only optimal when combined with certain other embed/decode parameters. Longer tones (e.g. with L = 30ms) can also result in precision of 99% (table 6.1, line 5) but setting L = 30 would mean 20% less watermark loops in a given cover signal, compared to L = 25, for no gain in precision.

6.2.4 Subjective listening tests

Listening tests were carried out to investigate whether or not the method described here was perceptually transparent. The listening tests were performed on a small group of listeners representing various levels of expertise. The test subjects included:

- One studio/live music sound engineer
- One production sound engineer
- One record company A&R representative
- Three experienced musicians
- One music critic/reviewer
- Five music consumers (e.g. no professional connection to the music industry)

Test subjects were presented with only a small subset of randomly selected audio files in order to avoid listening fatigue or boredom. The tests were carried out in a variation of the double-blind AB-X test [**132**]. This test was intended to identify whether listeners could state with any certainty whether a candidate audio file was the same as either one of two other files. A single blind test is one where the test subject has no way of knowing if the candidate file (‘X’) was identical to the first or second comparison files (‘A’ or ‘B’). A double-blind test is similar except that, to aid confidence in the results, the files chosen were randomly selected by computer from almost 350 watermarked files in order to prevent any bias from the tester. In each case, the file chosen for ‘X’ could equally have been the original or watermarked version of the file and the tester would not know which was selected. Either ‘A’ or ‘B’ could be the original with the other being the watermarked file. In all cases, one of ‘A’ or ‘B’ was to be the original and the other to be the watermarked file. None of the tests was carried out with identical files selected for ‘A’ and ‘B’. Testing was carried out as follows:

1. Random file selected.
2. Subject listens to sample ‘A’: original OR watermarked audio.
3. Subject listens to sample ‘X’: either original or watermarked audio.
4. Subject listens to sample ‘B’: watermarked OR original audio, depending on step 2.
5. Subject listens to sample ‘X’ again.
6. Subject was then permitted to listen to any sequence of files to aid comparisons.
7. Subject selects which of ‘A’ and ‘B’ is identical to ‘X’.

Since these listening tests were not intended to be definitive, it was not necessary to use specialist equipment or perform the tests in a specialist environment. Tests were carried out in an informal setting, using the following equipment:

Technics SL-HD350 CD player.

Technics SE-HD350 amplifier.

Technics SB-HD350 stereo speakers with 60 watt output and 6 Ω impedance.

Philips SBCHC8440 series FM headphones with frequency range 20-20,000 Hz [**133**].

* Note that subjects were permitted choose to listen with or without headphones and to alternate between headphones / speakers at will.

In these early listening tests, listeners were presented with ten audio comparisons each. Thus, a total of 120 AB-X comparisons were made. They were not informed whether sample ‘A’ or sample ‘B’ was the original. They were asked to ascertain if the ‘X’ audio sample was the same as either of the other samples. In all but a few cases, the test subjects could not generally identify the watermarked audio but the fact remained that at least some of the watermarked files were identifiable to some listeners as a result of perceptual differences between the two samples. This, of course, was not a positive development as it indicated that the modifications to the signal to embed the watermark were sometimes audible, even if only when compared directly to the original audio.

However, what was slightly more encouraging was that there was no consensus in the comments from the test subjects about whether or not any differences they may have noticed made the watermarked audio sound worse or better. As with echo-coding, where the addition of echo to a cover audio can make the sound ‘warmer’, there were some watermarked files that sounded better to listeners than their original counterparts. Nevertheless, while the results were generally encouraging, if the scheme made the original and watermarked audio perceptually different it was cause for concern.

6.2.5 Summary of scheme based on notched and weighted frequency components

The watermarking technique discussed in the section 6.1 and 6.2 does provide a platform for a computationally simple watermark encoding/decoding system that could use a set of standard parameter values to enable semi-blind decoding of the watermarked message much like a public-key decryption system will allow holders of the public key to decipher the encrypted message. Similarly, the use of private values that are not made publicly available would facilitate covert communications. Conversely, a completely blind decoding system could be derived from this semi-blind process by using signal analysis techniques to identify patterns in candidate audio and therefore to derive potential values for L and F in the decode phase. The results of the watermarking scheme, at this stage in the research, was the full and complete recovery of the watermark embedded in an encouraging 99.4% of the 350 watermarked audio files after approximately 11,500 embedding and decoding cycles with varying input parameters.

6.3 The Complex Spectral Phase Evolution method

6.3.1 Description of CSPE and its advantages

Notwithstanding the satisfactory outcome of the experiments at this stage, the publication of a recent technique for analysis of audio at super-fine resolution led to further evolution of the watermarking scheme [**134**]. The Complex Spectral Phase Evolution (CSPE) method of signal analysis is described as*'a tool to analyze and detect the presence of short-term stable sinusoidal components in an audio signal. The method provides for super-resolution of frequencies by evaluating the evolution of the phase of the complex signal spectrum over time-shifted windows. It is shown that this analysis, when applied to a sinusoidal signal component, allows for the resolution of the true signal frequency with orders of magnitude greater accuracy than the DFT'* [**135**].

The CSPE input is a signal frame of user-defined length, the amount of zero padding and the window type. As the CSPE technique, when applied with carefully chosen parameters, allows for identification at much greater levels of accuracy of the frequency components of the signal under investigation it could be used to implement an almost-ideal notch filter. It was decided to develop the watermarking scheme described in Section 6.1 and 6.2 in order to take advantage of the benefits of this new technique.

As with the earlier method, a component value is first chosen which is used as the basis for calculating two other components to modify to hide the message. The initial component choice may be dependent on various factors, such as the type of audio used as host/cover. For example, human speech generally consists of lower frequency components than a modern Rock or Pop song so hiding data in a recording of speech would naturally limit the component of choice. However, even in such a limited range, there are still thousands of values to choose from due to the super-resolution capabilities of the CSPE method. It is worth pointing out that the choice of initial component value does not need to be from those components inherently present in the audio. In fact, the value chosen is merely a reference or key value.

The value of the chosen component becomes, in effect, a *private key* and this value is needed in order to decode the watermark – assuming that the *presence* of the watermark has previously been detected, which is not as likely as it would be in the earlier method. This private key adds to the security of the technique when used in an environment where security of the *content* of the hidden message is an issue.

The signal intended as the cover or host audio is segmented into frames of uniform length and the frame is then analysed using CSPE techniques to identify the presence and magnitude of its inherent components. The principal of CSPE algorithm can be described as follows:

An FFT analysis is performed twice, firstly on the signal of interest and the second time upon the same signal but shifted in time by one sample. Then, by multiplying the sample-shifted FFT spectrum with the complex conjugate of the initial FFT spectrum, a frequency dependent function is formed from which the exact values of the frequency components it contains can be detected. The procedure of the CSPE algorithm is depicted in block diagram form in Figure 6.8:

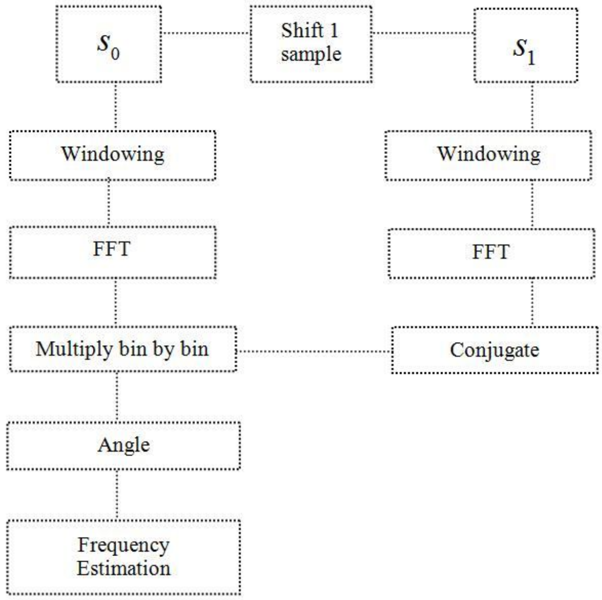


Figure 6.8: Block diagram illustrating the CSPE process [134].

Mathematically, the algorithm can be described as follows. Assume a real signal S_0, and a one-sample shifted version of this signal S_1. Say that its frequency is $\beta = q + \delta$ where q is an integer and δ is a fractional number. If b is an initial phase, w_n is the window function used in the FFT, F_{ws_0} is the windowed Fourier transform of S_0, and F_{ws1} is the windowed Fourier transform of S_1, then, from [**135**], we find

$$D = e^{\frac{j2\pi\beta}{N}} \tag{23}$$

The frequency dependent CSPE function can be written as

$$CSPE_w = F_{ws_0} F^*{}_{ws_1} = (\frac{a}{2})^2 \begin{bmatrix} D^* \left\| F_w(D^n) \right\|^2 \\ + 2\,\mathrm{Re}\left\{ e^{j2b} D F_w(D^n) \otimes F_w{}^*(D^{-n}) \right\} \\ + D \left\| F_w(D^{-n}) \right\|^2 \end{bmatrix} \tag{24}$$

The windowed transform requires multiplication of the time domain data by the analysis widow, and thus the resulting transform is the convolution of the transform of the window function w_f with the transform of a complex sinusoid. Since the transform of a complex sinusoid is a pair of delta functions in the positive and negative frequency positions, the result of the convolution is merely a frequency-translated copy of w_f centred at $+\beta$ and $-\beta$. Consequently, with a standard windowing function, the $\| F_w (D^n) \|$ term is only considerable when $k \approx \beta$ and it decays rapidly when k is far from β. Therefore, the analysis window must be chosen carefully. It has been shown that apodization (the use of an appropriate windowing function) can be used to increase the resolution of the CSPE algorithm under certain conditions [**136**]. The use of a Nuttall window is suggested as a good choice in this regard.

If the analysis window is chosen so that it decays rapidly to minimize any spectral leakage into adjacent bins it will render the interference terms, i.e. the second and third terms, to be negligible. Thus, from equation (24) the CSPE for the positive frequencies is as follows:

$$CSPE_w \approx \frac{a^2}{4} \left\| F_w(D^n) \right\|^2 D^{-1} \tag{25}$$

From equation (25) we find the CSPE frequency estimate

$$\begin{aligned} f_{CSPE_w} &= \frac{-N\angle(CSPE_w)}{2\pi} \\ &= \frac{-N\angle\left(\frac{a^2}{4}\left\|F_w(D^n)\right\|^2 D^{-1}\right)}{2\pi} \\ &= \frac{-N\angle(\frac{a^2}{4}\left\|F_w(D^n)\right\|^2 e^{-j\frac{2\pi}{N}\beta})}{2\pi} = \frac{-N(-\frac{2\pi}{N}\beta)}{2\pi} = \beta \end{aligned} \tag{26}$$

The frequency dependent function produces a graph with a staircase-like appearance where the flat parts of the graph indicate the exact frequencies of the components. The width of the flat parts is dependent on the main-lobe width of window function used to select the signal before FFT processing. An example of the output of the CSPE algorithm is shown in Figure 6.9.

Consider the signal S_t which contains components with frequency values (in Hz) of 17, 293.5, 313.9, 204.6, 153.7, 378 and 423. The sampling frequency is 1024 Hz. A frame of 1024 samples in length is windowed using a Blackman window and is padded using 1024 zeros. As shown in Figure 6.9a, the exact frequency of each component in the signal can be calculated using the CSPE algorithm and these are identified with an arrow in the graph. The largest error among all the estimates of the components frequencies is approximately 0.15 Hz.

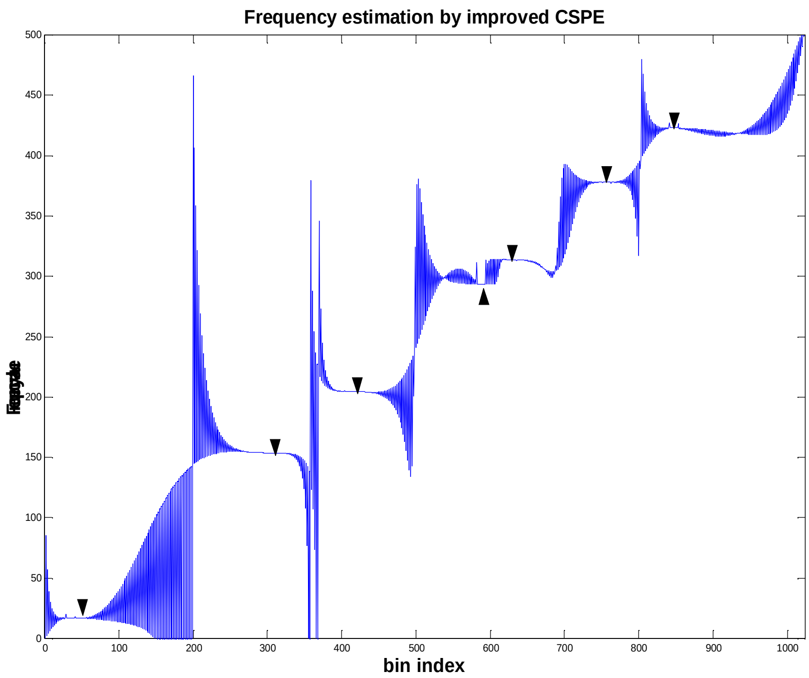


Figure 6.9a: Frequency estimation of signal S_1 by CSPE.

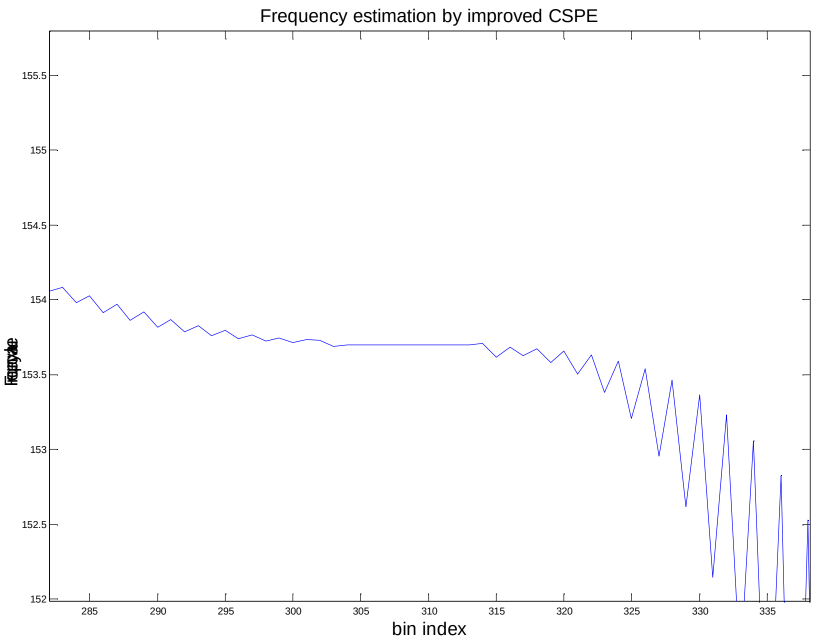


Figure 6.9b: A close-up view of a section of the graph in Figure 6.9a shows the flat area denoting the presence of a frequency component at 153.7 Hz.

Notice too in Figure 6.9a that the widths of flat sections where the arrows point are related to the width of the window's main-lobe in the frequency domain. A closer view of one of these sections is shown in Figure 6.9b where it can be seen that there is a flat section, and therefore a frequency component, at 153.7 Hz.

In addition, with the CSPE technique, we can get the amplitude and phase of the *k*th frequency component using the following equations, where *W(ω-fcspe(k))* is the Fourier Transform of window function which has been shifted to *fcspe(k)* in frequency domain.

$$Amp_k = \left\| \frac{2 * F_{ws_0}}{W(\omega - fcspe(k))} \right\| \tag{27}$$

$$Phase_k = \angle\left(\frac{2 * F_{ws_0}}{W(\omega - fcspe(k))} \right) \tag{28}$$

6.3.2 Experimental evaluation of the CSPE algorithm

Experiments were designed to evaluate the performance of the CSPE algorithm in correctly identifying frequency components within a multiple-component signal [**134**]. In each set of experiments, a total of 500 signals with sampling frequency 44100 Hz and containing components across the human hearing range of 100 Hz to 20,000 Hz were generated. Each signal contained many equally spaced frequency components. The number of components in each generated signal was not consistent. For each signal, there was a randomly-generated interval ranging from 169 Hz to 668 Hz, which defined the space between two neighbouring frequency components of the signal, making every input signal to the analysis unique.

Equation (29) and (30) were designed to assess CSPE accuracy in frequency estimation. Denoting $Freq_{estk}$ as the value of kth components of the signal; $Freq_{orgk}$ as the original value of the kth component of the signal; M_k as the number of frequency components contained in the signal; *FreqError* as the frequency estimation error between $Freq_{est}$ and $Freq_{org}$ of the signal and $MeanError_{cspe}$ as the mean error of the CSPE frequency estimation over N signals. For this experiment, N = 500, M changes with the randomly-generated signal interval as outlined above.

$$FreqError_k = \frac{\sum_{i=1}^{M_k} \left| Freq_{estk}(i) - Freq_{orgk}(i) \right|}{M_k} \qquad (29)$$

The estimation error of 500 signals is computed using equation (30) and shown as Figure 6.10 while the distribution of frequency estimation error (*FreqError)* is shown in Figure 6.11:

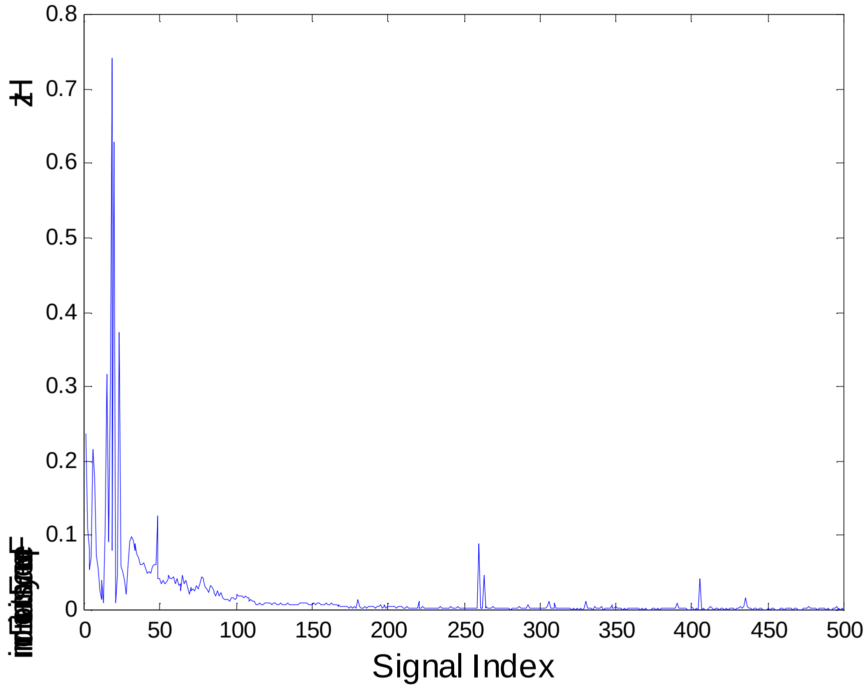


Figure 6.10: CSPE estimation error for each signal.

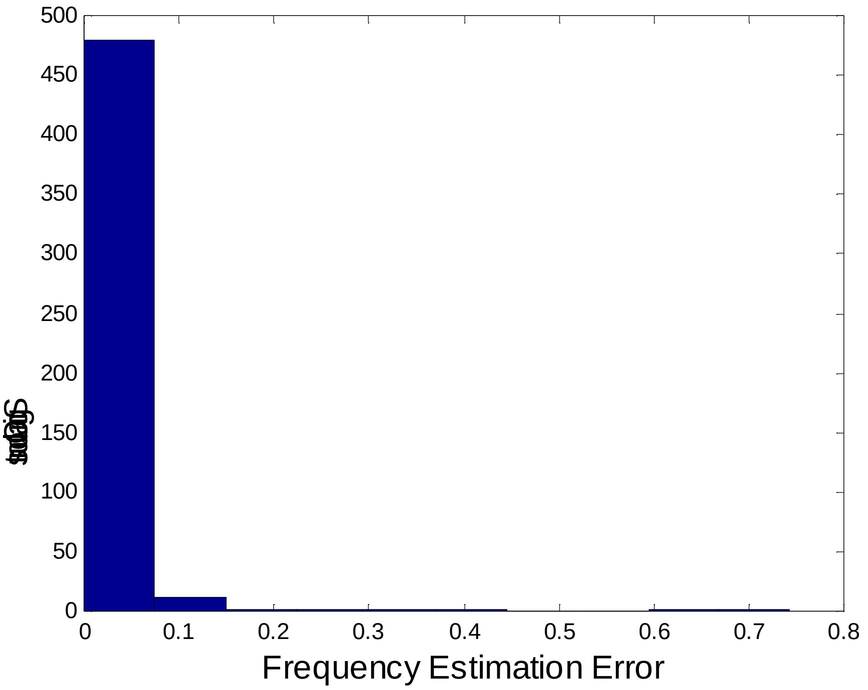


Figure 6.11: The distribution of frequency estimation error.

The mean error is calculated according to equation (30)

$$MeanError_{cspe} = \frac{\sum_{k=1}^{N} FreqError_k}{N} \tag{30}$$

By data analysis, we note that 97.8% of signals analysed using the CSPE algorithm resulted in a $FreqError$ value of less than 0.1 Hz, and the $MeanError_{cspe}$ is 0.0174 Hz, meaning that the algorithm identified the component to within 0.1 Hz in almost all cases. It can be concluded from these results that the CSPE is extremely accurate in frequency estimation for signals containing constant frequency signal components. With accurate estimation of the frequency, the amplitude and phase can then be estimated using equations (27) and (28).

6.3.3 Modifying components

Once the user-defined base component has been identified in the signal by the CSPE algorithm, its magnitude is calculated. It is then a matter of modifying the magnitude of this component, weighting it against a second value from within the signal, in order to represent a single bit '1' or '0'. We may choose to weight the user-defined components against the average power of the frame in which the bit is to be embedded. Recall that this was the procedure followed in both the earlier method described in Sections 6.1 and 6.2 as well as the work by Gopalan referenced earlier [**86**].

6.3.4 Dynamically selecting components

It was decided to make the process of choosing the candidate component(s) for modification as flexible as possible by making this a dynamically chosen *pair* of values. The initial reference or key value can be selected by the user and it need not necessarily be an inherent component of the signal.

Two components are then selected for modification dependent on the user-defined value but also dependent on the signal under consideration. Components chosen for modification were those identified by the CSPE method as being the nearest components *above* and *below* the user-defined value by more than a calculated threshold as illustrated in equation (31) where *compA* is the highest CSPE-detected frequency component that is lower than the user-defined component *u*, by more than the threshold *c* while *compB* is a CSPE-detected frequency component above the user-defined component by the same threshold amount

$$(compA < (u - c)) < u < (compB > (u + c)) \qquad (31)$$

What is interesting to note, using the formula in Equation (31) for defining which component we need to modify, and in which frames of the cover signal to perform such modification, is that only approximately half of the frames will require *any* modification. This is because the relationship between the values of the two chosen components in any given frame may *already* fit the criteria used for representing a '1' or a '0'. In this case they would not have to be modified in any way. This consideration makes this method far more favourable than earlier efforts. Another very interesting observation, from the perspective of security of the watermark, is that during the formulation of such a selection algorithm it became obvious that it is unlikely that any two subsequent frames would have the same component modified. This means that any attempt to by an attacker to use visualisation or other techniques to discover the components in a frame that have been modified, in order to allow complete decoding, would be unsuccessful. This would overcome the problem referred to by Gopalan [**86**].

The system would then compare the magnitude of both components (*compA* and *compB*) in any given frame before deciding if any modification would be required in order to satisfy the embedding criteria, depending on the bit to be embedded and the magnitudes of the two components in that particular frame. If they are already in the correct relationship, no modification is required. If, however, they are not in the correct relationship, one of the components must be modified. The decision that it is necessary to modify a component leads to another issue being raised. As mentioned earlier, the CSPE algorithm can be used to accurately identify a component within a signal, and from there its phase and amplitude can be calculated. Assuming that the magnitude of *compA* is lower than that of *compB*, in a frame in which it needs to be of a higher magnitude to represent a '1' bit, there are two possible methods available to achieve the desired relationship. First, the magnitude of *compA* can be increased until it is higher than that of *compB*. Secondly, the magnitude of *compB* can be lowered until it is below that of *compA*. Each has its advantages and disadvantages:

If a component's magnitude is increased, it increases the likelihood that the component will play more of a role in the perceptual evaluation of the audio by listeners. In other words, if *compB* occupies an important position in the audio, in terms of perceptibility, and *compA* is increased to be of a bigger magnitude, then the increased component may become too loud and therefore be audible as an artefact. Therefore, if this method is used, care must be taken to ensure that the modified component does not have too large of a magnitude.

The alternative, lowering a component's magnitude, might therefore seem the obvious choice as it means that, regardless of the magnitude of one component, if the other component is set to a lower magnitude it would not produce an audible artefact. This method is also, however, also problematic. If a component is chosen that is quite low-powered in the frame, then lowering another component might make it so small as to be undetectable by the CSPE algorithm. As a result of these considerations, limiting the scheme to selecting candidate components only within certain thresholds would seem like an appropriate compromise. The algorithm can be further strengthened in terms of its transparency if the components so chosen are not amongst the higher-powered components in a frame.

A set of rules is defined that would lead to the modification of only one of the components (*compA* or *compB*) in approximately half the frames. The rules are as follows (where 'Amp' refers to amplitude of the component)

Let Amp(*compA)* > Amp (*compB*) + *margin where bit = 1*
Let Amp (*compB*) > Amp(*compA)* + *margin where bit = 0*

In order to increase the magnitude of a particular component in the cover signal $s(t)$, another component is added at a defined magnitude and matched to the phase of the component it is being combined with, as illustrated in Equation (32):

$$s(t) = s(t) + (rAmp - lAmp + threshold)\cos(2\pi(compA)\,t + lp) \qquad (32)$$

where $rAmp$= amplitude of *compB*

$lAmp$= amplitude of *compA*

$compA$= Frequency of *compA*

lp = phase of *compA*.

This is equivalent to an ideal notch filter where an individual component can be altered without impacting on any other component. Conversely if it is decided to reduce the magnitude of a component $s(t)$ so that it satisfies the requirements for embedding a '1' bit, this is achieved by reducing the component to the *right* of the user-defined component value, by adding in a component that is 180° out of phase with the original component in the signal as follows:

$$s(t) = s(t) + (rAmp - lAmp + threshold)\cos(2\pi(compB)\,t + \pi - rp) \quad (33)$$

where *compB* and *rp* define amplitude and phase of *compB*.

6.3.5 Decoding

In order to process candidate audio for detection and decoding of a potential embedded watermarked message, the system must first be provided with the user-defined value used as a basis for calculating the embedding values, along with the above rules that define a '1' bit and a '0' bit. The candidate audio signal is then segmented into frames using the same frame size as was used for embedding. The system calculates the magnitude of the components identified using the rules and after CSPE analysis, and performs a simple comparison. From this comparison the watermarked bit sequence can be recreated. It is a comparatively simple matter of repeating the CSPE algorithm to identify the two components *above* and *below* the user-defined value by more than a pre-defined threshold. These two components would then have their magnitude compared and a ‘1’ or a ‘0’ bit would be determined according to the rules used in their embedding.

6.4 Evaluation of the CSPE-based watermarking scheme

A series of experiments was carried out to evaluate the performance of this codec, based on the same 500 synthesised signals as introduced in section 6.3.2. For each signal, a randomly generated binary bit-sequence of length 150 was embedded by means of increasing the magnitude of components as described. The system then decoded the modified signal in order to detect the watermarked code.

The difference between these two code sequences can be calculated in terms of equation (34) below, where *DCode* denotes the code sequence obtained in the decode process, *ECode* denotes the code sequence initially embedded in the signal, *CodecPrecision* denotes the precision of the decode process with code length *L* for signal *k* and *MeanPrecision* denotes average error of the decode process over *N* signals. In this experiment, *L* and *N* are set to 150 and 500 respectively.

$$CodecPrecision_k = \frac{L - \sum_{i=1}^{L} \left| DCode(i) - ECode(i) \right|}{L} \tag{34}$$

$$MeanPrecision = \frac{\sum_{k=1}^{N} (CodecPrecision_k)}{N} \tag{35}$$

The results of this experiment are depicted in Figure 6.12.

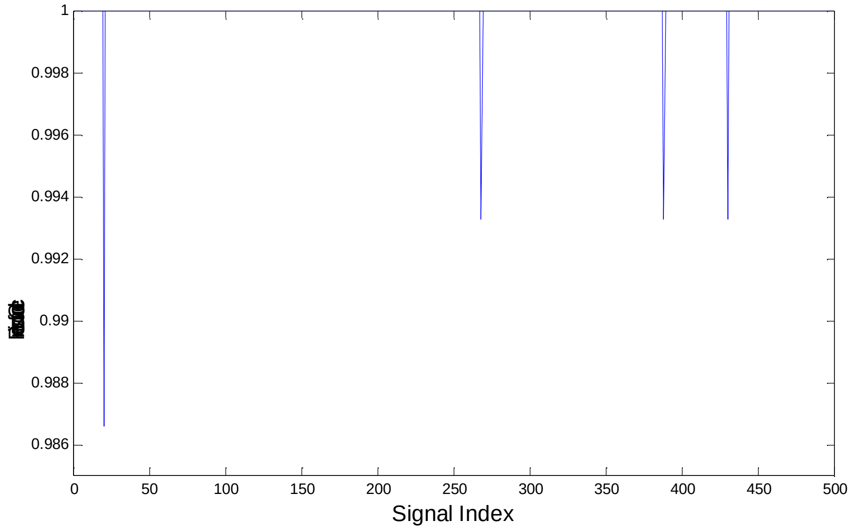


Figure 6.12: Precision of codec for each signal.

The distribution of *CodecPrecision* is shown in Figure 6.13.

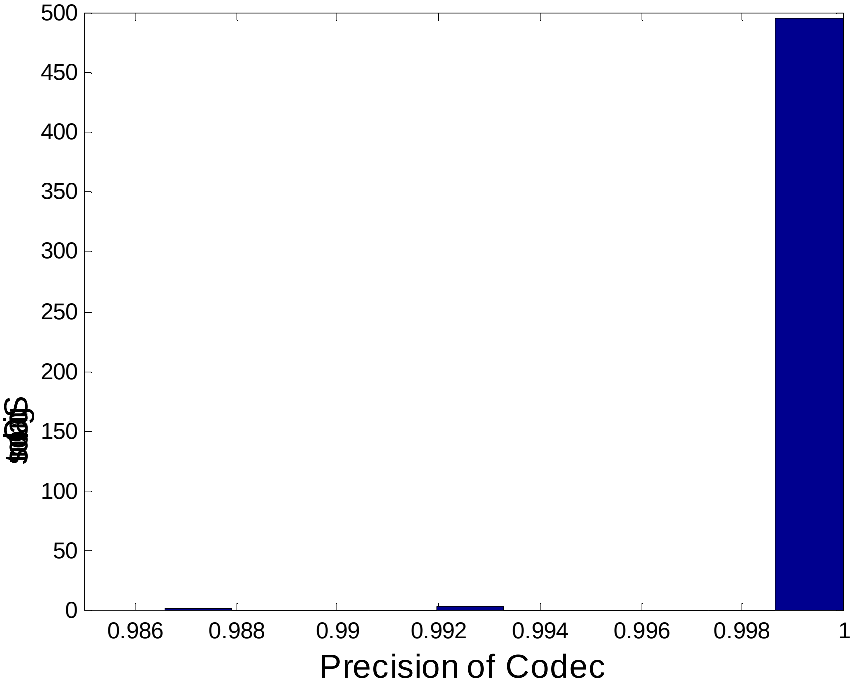


Figure 6.13: The distribution of precision of codec.

From the experiment results, it can be seen that 99.2% of signals produce a *CodecPrecision* value of 1 (100%). This means that, from 500 randomly generated signals with multiple components of different frequency spacing, watermarked with a binary bit-sequence of 150 bits, 99.2% of these signals were decoded to the *exact* 150 bit sequence. Only 0.8% (a total of 4 from 500 signals) was not decoded perfectly. Of those *not* perfectly decoded, the bit sequence recovery rate was above 98.66%. The *MeanPrecision* computed using Equation (35) is 0.9999 (99.99%). Therefore, the performance of this codec is almost perfect for this experiment, albeit with synthesised signals.

What should be remembered is that the decode experiment in this case represented a single iteration of a bit sequence over the length of a signal. Given that any realistic use of such a scheme would embed the message repeatedly in a cover signal, sometimes hundreds of times, it would be possible to increase the effectiveness of the decode process by, for example, repeated decoding and using the mode of the results.

6.5 CSPE-based watermarking of real signals

Simulated signals, albeit with many different components at many different frequencies and with many different magnitudes are not real signals. Real signals usually contain hundreds, if not thousands, of continuously varying individual components. This is a much more difficult proposition for signal analysis. As mentioned in the discussion of DFT techniques in Section 5.2.5, the ubiquitous FFT-based analysis techniques used in digital signal processing cannot decompose a complex signal into its component parts [**93**]. The Goertzel algorithm used in the early phase of this watermarking scheme outlined in Section 6.1 is not capable of exactly identifying the components inherently present in a given complex signal at any given point.

Initially, as discussed above, attempts to develop the CSPE algorithm into a watermarking scheme using synthesised signals proved to be encouraging. The technique was then applied to real signals, namely 20 music tracks of varying genres. After changing the frame size, it was possible to embed a series of watermarks in music tracks of varying genres and then decode them to a high degree of accuracy. As described in Section 6.3.1, the CSPE watermarking of synthesised signal was achieved with a frame size of 1024. However, given that the CSPE algorithm is less effective at identifying components in small frames, the frame size was increased to 8192 and padded by the same amount. This meant that more components would be identifiable providing more candidate components for modification.

6.5.1 CSPE watermarking by magnitude increase

In order to test the overall usefulness of the algorithm in isolation a series of 20 audio tracks was watermarked with a randomly generated bit sequence 700 bits long. This was done without any form of error-correcting being employed. In decoding the watermark from these 20 tracks perfect recovery of more than 97% of the 700-bit sequence without any additional processing was achieved, as shown in table 6.3:

Track	Precision	Track	Precision	Track	Precision	Track	Precision
1	0.9846	6	0.9881	11	0.9849	16	0.9748
2	0.9757	7	0.9749	12	0.9902	17	0.9826
3	0.9909	8	0.9973	13	0.9914	18	0.991
4	0.9836	9	0.9782	14	0.9761	19	0.9783
5	0.9889	10	0.9823	15	0.976	20	0.9792

Table 6.3 shows the precision values for decoding of 20 music signals when encoded and decoded using the CSPE-based magnitude-increase algorithm without error correction.

The process was then further modified in an attempt to develop it into a practical application-dependent watermark embedding and decoding scheme. The first step was to reduce the embedded bit-sequence that represented the watermark from 700 bits long to the 100 bits needed to represent the same 12 character identifier that was previously embedded using earlier methods described in Section 6.1 and 6.2. This shorter sequence was then embedded repeatedly in the cover signal. A single frame is shown in Figure 6.14 to illustrate the minute change made to the component in the algorithm when it increases the magnitude of one of the two components either side of a user defined value.

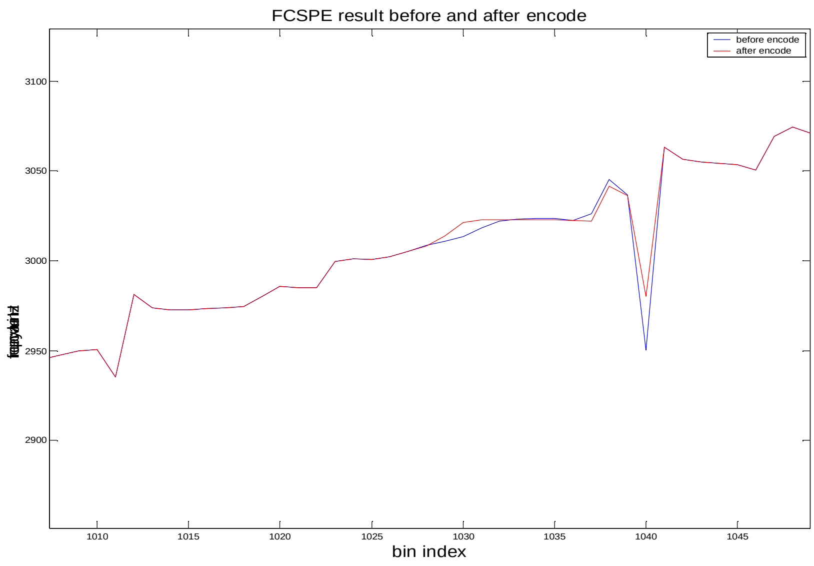


Figure 6.14: A frame showing the change from original to the watermarked signal when processed with the magnitude increase algorithm.

Note from Figure 6.14 that the left component modified is in bin 1017 on the *x*-axis and the right component modified is in bin 1036. A larger image is included in Appendix 2 for closer examination.

Upon decoding the repeated watermark, the CSPE-based scheme proved to be apparently perfect in its operation. In performing the embed/decode cycle on the same 20 files that had on average produced 98.35% precision as shown in table 6.3, the full 100-bit sequence was successfully decoded in *all* of the test cases. Of course, this was for only 20 real world sound signals but it was certainly encouraging.

6.5.2 CSPE watermarking by magnitude reduction

One discouraging result of the work to this point was that the watermarking process introduced some audible artefacts in some frames. Of course, this would make the scheme unacceptable in a broadcast monitoring scheme. The cause of these artefacts was that the magnitude of some components was quite high after they were altered to reflect the desired relationship. The solution was to modify the embedding phase of the algorithm so that instead of *increasing* the magnitude of any component, the magnitude of the other component was *reduced*. If a relationship of $A>B$ was required for a particular bit and the components were not in such a relationship, the magnitude of A had previously been increased until $A>B$. Now, instead, the magnitude of B would be reduced until it was less than the magnitude of A, thereby satisfying $A>B$. Therefore, under none of the component modifications would a magnitude increase be introduced. A single frame is shown in Figure 6.15 to illustrate the changes made to the components in the algorithm when it reduces the magnitudes of one of the two components either side of a user defined value. A larger image is included in Appendix 2 for closer examination.

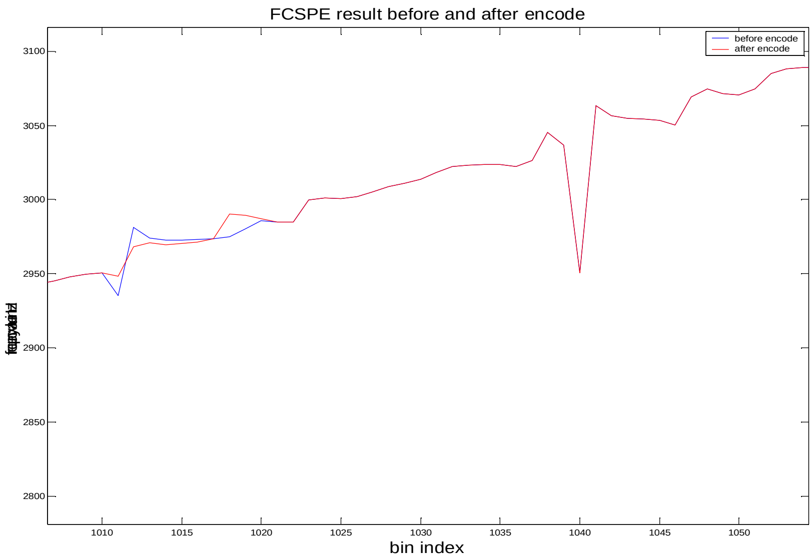


Figure 6.15: A frame showing the change from original to the watermarked when processed with the magnitude reduction algorithm.

Note from Figure 6.15 that the left component is noticeably altered. The process of embedding and decoding 20 real signals was performed again, this time by component magnitude reduction rather than increase. The result was that there were no audible inconsistencies or artefacts introduced by the process. However, the accuracy of the system dropped dramatically, with some files reporting watermark extraction barely above 50% as shown in Table 6.4.

Track	Precision	Track	Precision	Track	Precision	Track	Precision
1	0.8438	6	0.8438	11	0.7708	16	0.7500
2	0.8854	7	0.7292	12	0.7604	17	0.8333
3	0.8750	8	0.5521	13	0.8021	18	0.8750
4	0.8542	9	0.7708	14	0.7396	19	0.8542
5	0.7917	10	0.8542	15	0.8750	20	0.8438

Table 6.4 shows the precision values for decoding of 20 music signals when encoded and decoded using the CSPE-based 'magnitude-reduction' algorithm.

The reason for this was that the components chosen for manipulation were the first components that the CSPE could detect on either side of the key component. When two components were identified, often they were of such small magnitude that when one was decreased in magnitude it became almost zero. Ironically, it was the precision of the CSPE technique that allowed the identification of components of such small magnitude but it was not practical then in the decoding phase to identify components with magnitudes reduced to near-zero.

6.5.3 CSPE watermarking by magnitude reduction and swapping

The final experimental adaptation of the CSPE-based watermarking system was as follows. In the case where one component was greater than the other, and the '1' or '0' bit being embedded required that the relationship was the reverse, the magnitudes of both components was recorded, reduced to zero and then swapped. The result of this process on a single frame is shown in Figure 6.16. A larger image is included in Appendix 2

.

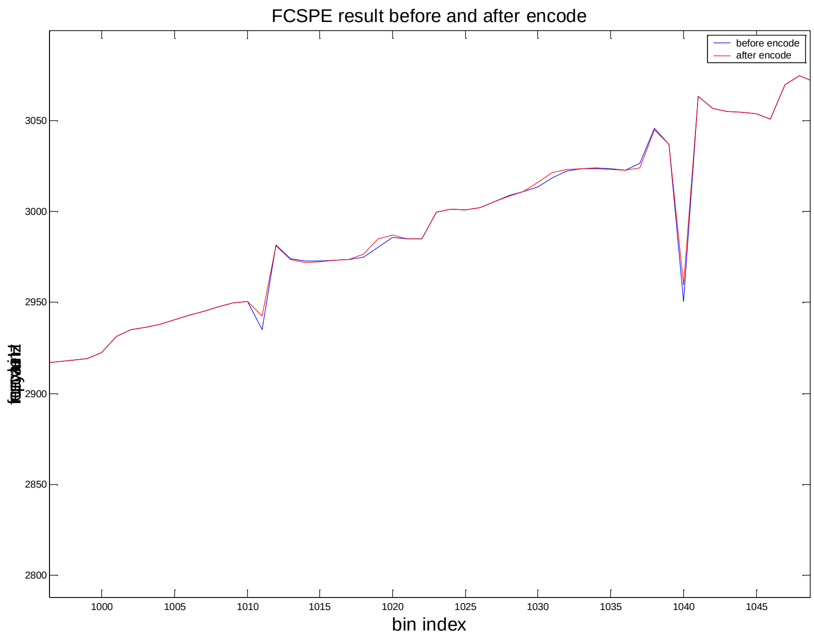


Figure 6.16: A frame showing the change from original to the watermarked when processed with the magnitude 'reduce and swap' algorithm.

What this meant in practise was that *compA* was re-introduced into the signal but with the magnitude that *compB* had previously had, and *vice versa*. This process did, of course, produce the relationship that was required since the components would then have a relationship opposite to that which had previously existed. As explained in section 6.3.4, component magnitudes were only altered in the case that the components were not already in the correct relationship. The results of this 'reduce and swap' algorithm are shown in table 6.5.

Track	Precision	Track	Precision	Track	Precision	Track	Precision
1	0.9896	6	0.9896	11	0.9896	16	0.9792
2	0.9792	7	1	12	0.9896	17	0.9896
3	0.9896	8	0.7708	13	1	18	1
4	0.9896	9	0.9479	14	0.9896	19	0.9792
5	0.9792	10	1	15	1	20	0.9896

Table 6.4 shows the precision values for decoding of 20 music signals when encoded and decoded using the CSPE-based 'magnitude reduction and swap' algorithm.

As can be seen when comparing Tables 6.1, 6.2 and 6.3, the precision for each file is uniformly worst in the 'magnitude reduction' algorithm and consistently best in the 'magnitude reduce and swap' algorithm, which gives slightly better results than simply increasing magnitudes.

There were still a number of alternatives and modifications that could be investigated in order to achieve the desired result of an inaudible but accurate watermark embedding. One of these modifications was simply to alter the magnitude of not only the first component to the left and right of the key component that were identified by CSPE, but two. This would serve the purpose of doubling the likelihood of correct comparison. If the first of these components was reduced to such a small magnitude as to be undetectable, then the second could possibly still be detected. Of course, the second component might *also* be undetectable but this modification would at least increase likely overall detection rates without increasing audibility of the watermark.

Another modification to be investigated was to select as embedding candidates only the components that were above a certain minimum magnitude but within a certain range of each other. If two components above a certain magnitude were slightly modified, even if one was reduced to below that magnitude, the likelihood is that CSPE-based analysis would then be still able to find the component in the decoding phase.

Still another modification that could be examined was to alter the magnitude of *both* candidate components to reflect the desired relationship. For example, if we need *compA* > *compB* and if *compA* is 0.02 and *compB* is 0.08 then, rather than increasing *compA* to 0.09, or reducing *compB* to 0.01, we instead do both at the same time, but only apply half as much change to each. In this case we would increase *compA* by 0.035 extra (half the difference, plus a small increase) to 0.065 and at the same time reduce *compB* by 0.03 (half the difference) to 0.055 thereby creating the desired relationship in such a way as to minimise the possibility of making either component audible or undetectable. More comprehensive testing of these possible developments would need to be undertaken before any conclusions could be drawn on their precision and transparency.

6.5.4 Capacity of the scheme

Watermarking schemes range in capacity from less than 1 bit embedded per second of cover signal [**110**] to more than 3 kbps [**121**]. Capacity is usually a secondary consideration of the scheme, being less important than robustness and transparency yet affected by both of these considerations. In the scheme proposed here, the capacity is a direct function of the frame size used in the CSPE algorithm. In Section 6.3.1, the frame size that produced perfect precision on synthesised signals was 1024. However, in order to overcome some of the shortcomings of CSPE analysis on small frames, the frame size was increased to 8192 when the scheme was applied to real signals, as described in Section 6.5. Capacity of the system is calculated by dividing the sampling rate by the frame size:

$$cap = Fs / frame \tag{36}$$

In the watermarking of music files, capacity is therefore a comparatively low 5.85 bps (48000/8192). However, the capacity can be increased in two ways. Firstly, multiple components can be modified on either side of the user-defined base component. For each additional component modified, the capacity doubles. Caution should be exercised, however, as each additional modification to the cover signal increases the chances of perceptible artefacts. If the scheme is used as a covert communications method this might not be a primary consideration. However, when implemented in a broadcast monitoring scheme, perceptual transparency is paramount.

The other way of increasing capacity is simply to reduce the frame size. A smaller frame size will result in a higher value for *cap* in equation (36). In using synthesised signals, experimental results were generated and analysed with the CSPE algorithm. Frame sizes as small as 256 could be adequate for some types of signals (although not music). This means that capacity would increase 32-fold in an application that utilised deliberately created synthesised signals for steganographic purposes. The capacity for a system that sampled the cover signal at 48 KHz and used a frame size of 256 would be 187.5 bps.

6.6 CSPE and watermark security

In the earlier watermarking scheme outlined in Sections 6.1 and 6.2, one noticeable result of the algorithm to notch the cover audio and then insert two components in the required ratio is that it would lead to some visible pattern when the signal was analysed using the spectrogram. Experienced practitioners who know what to look for would, in theory, be able to identify the presence of some repeated pattern and this *could* lead to breach of the security of the content of the watermarked message, assuming it had not been enciphered prior to insertion. Similarly, if a watermarking scheme was being employed for covert communications or any critical data-hiding purpose, the visible pattern itself could theoretically be decoded even without any knowledge of the watermarking technique.

The CSPE-based watermarking algorithm described above does not introduce any such visible patterns. This is because the CSPE algorithm does not alter the magnitude of the same component in each frame. Instead, the proposed scheme modifies the magnitude of one component on each side of a reference component (the so-called private key). If the key component is, say, 3000 Hz, the CSPE-based scheme might alter the components perhaps at 2997 Hz and 3002 Hz in one frame and it might alter 2999 Hz and 3010 Hz in the next frame. In fact, it would appear that the choice of components to modify in a real signal would almost be random due to the non-stationarity of the content of real signals.

It is a positive characteristic of the proposed CSPE-based scheme that it is actually more useful when implemented using real signals than synthesised signals because this variation in components modified from frame to frame provides additional security. Without having access to an unwatermarked version of the original cover audio, no attacker trying to decipher the content of the watermark, even if its presence *could* be perceived, would know which components in each frame had been modified, if any. This quality of CSPE-based watermarking adds to the steganographic nature of this particular technique.

In fact, if a non-watermarked signal and a watermarked signal were to be decoded without knowledge of the initial key component chosen at the embedding stage, their output would probably appear equally random. This contention will have to be investigated in some detail, particularly from the perspective of probabilities and statistics, in order to state with any degree of certainty that the content of the watermark, again even if its presence could be perceived, would be safe against unauthorised decoding as long as the 'private key' was kept private. A CSPE-based scheme that embeds a bit-stream in the cover audio would seem to be extremely secure in this context because the output of *every* signal, watermarked or otherwise, *will* be a sequence of bits.

6.7 Summary

While there is no significant different in precision between the two variations of watermarking proposed in Chapter 6, the development of the CSPE-based watermarking algorithm outlined in Section 6.3 led to a significant improvement on the techniques in Section 6.1 and 6.2. One of the limitations of the early technique was that it was not technically a steganographic scheme because spectral representations of the watermarked signal could be employed to identify a pattern created by the process to notch the cover signal and then add components. This limitation was removed in the CSPE-based scheme as there is no notch performed on the cover audio and, more importantly, the components modified in each frame are different, regardless of which magnitude modification technique is employed. There is therefore no pattern to discern. In steganographic terms this is a major improvement.

The scheme is almost perfect in its precision and, with further development it may be made perfect. While a complete listening test is yet to be carried out, informal initial listening suggests that the watermark is perceptually transparent. Only one single file presented any form of audible distortion and it is not clear if this is a function of the track or the watermarking process. A complete encode/decode of the collection of almost 350 audio files in various genres will be carried out to validate the transparency of the scheme.

No claim for robustness is made at this stage as full tests have not been carried out. However, if the algorithm is modified so components modified in the CSPE-based scheme are chosen carefully to avoid using components that might be perceptually insignificant, then robustness against perceptually based compression should be ensured.

Chapter 7: Conclusions

It is worth recalling that the primary motivation for this research was to identify problems with the current administration of Copyright, particularly as it pertains to public performance copyright. The intention was to attempt to address the problems inherent in these systems which lead to a negative or detrimental effect on developing artists – exactly the demographic that Copyright was envisaged to protect from exploitation, even if that exploitation was accidental and/or legal. As the 'Statute of Anne' was intended to address what was, at the time, the *legal* copying of works by book publishers [**5**], it might be said that the purpose of this research was to address the legal situation that can be shown to also be exploiting the less well informed artists to the unwarranted benefit of the better informed, mostly commercially aware, artist.

The early stage of the research was spent analysing the impact of collective rights licensing at the grass roots level, including but obviously not limited to the direct financial impact on artists who should have seen reasonable royalty payments and who did not. The restrictions on career development caused by any alleged underpayment(s) are obviously impossible to judge but there are a number of case studies which illustrate the comparatively huge impact that royalty payments can have on an artist's development. Most notable among artists who have seen such positive development as a result of receiving royalty payments they *were* entitled to, are Irish rock band '*Jaded Sun'*. For a single live performance in May 2006, the band received royalty payments that covered the cost of recording their first full album, '*Gypsy Trip*'. On the strength of this album, Jaded Sun toured Europe and were offered the opportunity to develop their career to the point where they were hailed as one of the best rock bands on the continent by many publications. While most artistic endeavour is undertaken, at least at first, whether there is ever any *'promise of emolument'* [**6**], this single example is used to illustrate the effect on an artistic career that can be achieved with correct and accurate distribution of royalties that are actually due to an artist.

There are dozens of techniques that can be applied to the area of audio manipulation for broadcast monitoring and they can be described in two main categories: fingerprint based and watermark based. Fingerprint-based techniques are those where a copy or digest of audio is made and stored before being compared to a candidate signal in a broadcast monitoring environment. Watermark-based techniques are those where some identifier or other data is embedded within the audio and broadcasts are monitored for the existence of these marks within the signal. Given that this research was approached with no preconceived ideas about the 'right' way to solve the problems in the administration of collective rights licensing and with no *a priori* knowledge of the techniques and technologies already applied to this and related areas, there was no preference for any one over technique over any other. Each has its advantages and disadvantages but the disadvantages of fingerprint-based monitoring for developing or less successful artists would actually be advantages for the more commercially aware, successful artist. These disadvantages, and others, are the primary reasons why fingerprinting techniques were discounted as unsuitable for broadcast monitoring and audio identification in this research.

There are a comparatively small number of broadcast monitoring schemes based on watermarking techniques. For example '*Audio Auditing*' utilises echo watermarking (see Section 5.2.2). On the other hand, various digital watermarking techniques exist in image or video manipulation. The increased sensitivity of the human auditory system compared to the visual system means that watermarking in audio was a tougher prospect than in images.

Given that watermarking was the preferred choice of technique to implement a fair, accessible and transparent broadcast monitoring system that would not unfairly weigh against any sector, research was undertaken into the myriad forms of watermarking that were already in existence and were continuing to be proposed. In order to understand the various techniques available and their relevant advantages and disadvantages, watermarking schemes based on echo-insertion, phase modification, amplitude modification, multiplicative or transform-based embedding, spread spectrum embedding and chaos-based watermarking techniques, amongst others, were investigated,.

A form of amplitude modification was first attempted and while this was very successful when implemented with synthesised signals, it soon became apparent that more realistic signals would be much less likely to be easily modified using a simplistic amplitude modification approach, most notably because of the pre-existing content of the host signal. Some analysis of the inherent content of the host audio sample set was undertaken to try to identify any relationship between various components in the signals. However, while there was no real identifiable relational pattern between components, and the relative strengths of the components that had previously been chosen for modification were not extreme at either end of the scale, it was felt that the approach needed to be changed in order to maximise decoding accuracy.

Various alterations to the basic amplitude modification scheme were proposed and attempted, some with more success than others. One successful alteration was to modify the amplitude of higher-frequency components as these components tended to have relatively higher strengths within a given frame of audio. However, it was also noted that selecting components above certain upper bounds would mean that these components, and therefore any watermark embed by their modification, would be more likely to be altered or even removed by attacks such as compression or perceptual coding in any form. Components in the mid-range of the human auditory system were chosen for more in-depth experimentation. Eventually, recovery rates of more than 99% were achieved with perceptual transparency of the watermark achieved in almost all cases.

Another successful alteration to the embedding phase of the scheme was the use of a notch filter to reduce to almost zero the power of the component in the host audio that was at the point where the watermark was to be embedded. In this way, inherent components that were higher-powered in the host signal than they should be after watermarking, were first reduced so as not to be too strong on decoding, spoiling the result.

A concerted effort was then made to identify the relative contribution of each of the watermark input parameters to the overall output result. Using the results of approximately 11,500 individual watermark embedding/decoding processes on hundreds of real audio signals taken from commercial releases in various genres, the relationship between the input parameters and the result was evaluated. A set of optimal embedding parameters was defined for most cases.

The discovery of the Complex Spectral Phase Evolution (CSPE) technique for identifying components in an audio signal to a very high degree of accuracy led to a further alteration to the process. Since CSPE-based analysis could be used to identify components at frequencies down to a fraction of a Hertz, the embedding phase of the scheme was evolved to make use of this technique. Rather than applying a notch filter to the host signal in order to create a 'hole' into which the components used to represent the watermark would be added, the CSPE technique was used to identify components and they could then be manipulated directly.

It was also found that the CSPE-based scheme had two major advantages over the previously proposed scheme. First, the ability to identify components very accurately made the selection of components for modification more fine-grained than previously and this led to the CSPE-based scheme being potentially more secure because the number of fractions of frequencies that could be used for watermarking was higher. The likelihood of identifying these components was subsequently much less. Secondly, modification of different components in each subsequent frame (rather than the same components in frame after frame) meant that the possibility of identifying a pattern and therefore identifying the modified components was reduced drastically. This alone makes the process a steganographic one, as there is no way of identifying whether or not a candidate signal has been watermarked.

With these observations, the value of the CSPE-based watermarking scheme as something other than a simple broadcast-monitoring tool became evident. A robust watermarking scheme that is perceptually transparent has many uses. One that also produces a watermark that is undetectable, indecipherable and resistant to common attacks such as perceptual coding or compression and achieves all of this with inherent pseudo-security has many more uses. Furthermore, the characteristic of the proposed scheme that allows the modification of components at a high degree of granularity means that there is every opportunity to embed more than one watermark with no discernible loss of audio quality. This enables its use *simultaneously* for more than one purpose. Some of the potential applications for which such a watermarking scheme can be utilised include:

- Covert Communications.
- On-the-fly music identification.
- Additional revenue generation for broadcasters & content producers.
- Watermarking individual performers.
- Tracking unauthorised distributions.

An explanation of each of these potential applications is provided in Appendix 4.

7.1 Step by step description of a proposed broadcast monitoring process

Notwithstanding the multiple uses for which an effective watermarking scheme can be deployed, the purpose of the research undertaken and described in this document was to provide a system for broadcast monitoring that was efficient, accessible, accurate, transparent and robust. By considering the process that would be undertaken by a content producer from the beginning, it can be shown that these requirements have been met. Consider the following scenario:

An artist, producer or other content creator completes their work. Before making any copies available to any third party the piece is watermarked. The watermark is embedded with a ‘public key’ value. This value is standard across the industry and issued to content creators along with the three-letter ISRC ‘Registrant’ code described in Section 3.6.1 by any of the various organisations that provide these codes under the auspices of the IFPI. The ISRC system already exists in the industry so it is a simple matter to ensure all new Registrant are given the ‘public’ watermarking value. The content creator can choose to add additional watermark(s) if required. The content is then released to interested parties or the public.

A means of monitoring radio and television broadcasts is assumed to be in place. Such monitoring is already performed for audio-fingerprinting tasks by organisations such as 'Nielsen' [**71**], 'MusicTrace' [**137**] and others. These systems could be adapted to analyse the audio signal for watermarks instead of, or in addition to, fingerprints. Alternatively, collective rights societies could implement an industry-wide broadcast monitoring system in a collaborative development.

Once identified, watermarks are decoded back to an alphanumeric ISRC code and a table created of instances where each ISRC code is identified. At the end of each reporting period, which could be daily, this data is used to query the database of ISRC codes already in place in the various ISRC national registers and the IFPI. The national registers then publish, extrapolate, condense or otherwise manipulate this data to increase its usefulness.

Finally, and most importantly for this research, the collective rights organisations use the reports of plays over a period of time to calculate the distribution of royalties based *exactly* on the number (and length) of plays identified for each watermarked work. Monitoring can also be carried out by content owners as the values used to watermark the audio will be known to them. The issues of awareness and accessibility remain for the industry to address. It is, of course, impossible to ensure every artist is made aware of the availability of such a system. However, if implemented as outlined here, then every artist who is ever made aware of, and requests, an ISRC code would be made aware of the advantages and process of watermarking their works.

…ergo:

The development of an accessible, perceptually transparent, robust and adaptable watermarking scheme has been achieved in this research. Moreover, it more than satisfies the primary goal of the research: namely an open and accessible broadcast monitoring scheme that could be used to provide accurate and correct records to enable the efficient and equitable calculation of royalties due for public performance of copyrighted works.

Chapter 8: Future Work

The discovery of the relatively newly developed CSPE algorithm for super-high resolution analysis of signal components is likely to lead to significant development in any area that relies on accurate signal representation. The watermarking scheme outlined in this work will be developed further and will be extended to different domains. In the first instance, analysis of the output of the CSPE algorithm will be performed in order to identify where its use in the area of digital audio watermarking might be improved.

The precision of component analysis, along with the capacity of the system, are areas that would benefit from further investigation. Precision of analysis is, as mentioned in Section 6.3, a function of frame size when performing CSPE analysis. However, capacity is also a function of frame size. These two characteristics of the algorithm as applied to an audio signal are counter to each other and have an inverse relationship. This means that a smaller frame size will lead to increased capacity of the system but simultaneously lead to decreased precision in identifying suitable candidate components for modification to represent the watermark. The converse is also true, that a larger frame size will lead to greater accuracy but also to lower capacity. The use of larger windows allied to smaller frame sizes is one area that warrants further attention as it seems to promise increased capacity without loss of precision.

A full scale test will be carried out to ascertain if the results achieved in section 6.3 can be replicated in a large number of audio files in a wide variety of genres. Wide ranging listening tests will also have to be carried out to investigate the perceptual transparency of the scheme more comprehensively. Similarly, objective testing will also be carried out in order to evaluate how the scheme compares to other watermarking schemes.

The ability to embed information *within*, rather than alongside, the signal means that the watermarked message is more likely to survive attacks, particularly accidental attacks such as transmission channel noise and interference. A future development of this technique will be to apply it to the sound channels that are carried alongside video in every television / cinematic environment. There are a number of applications for which this might prove useful, including but not limited to the following.

A watermark scheme can be developed to embed a television programme's script within the sound channel of the programme. This would not be too difficult to achieve and would be useful for several purposes. First, it could easily replace the traditional teletext-based subtitling provided for deaf or hard-of-hearing viewers. Approximately 10% of the population suffers some form of hearing loss and subtitling for this demographic sector has been provided for some years. However, subtitling of a programme takes time to arrange and many programmes are left without subtitles. The '*holy grail*' of subtitling is to be able to subtitle live television News broadcasts. With a watermarking scheme that uses the information produced for display on the newsreaders 'teleprompter' as a message source, and a decoder in a set-top box or a microphone in a modern mobile phone, this could be achieved relatively easily and inexpensively.

An added bonus of such a scheme is that automated processing would mean that subtitling would become relatively inexpensive. Currently, subtitling is done 'by hand' which is costly in terms of man-hours, or by a form of speech recognition which is inherently inaccurate. Finally, in this regard, such a scheme would also widen the provision of subtitle display to the handheld arena. More and more, users are consuming their television on mobile and/or internet connected devices. These devices are generally not teletext-enabled and as such are incapable of displaying the subtitles. If the subtitle was actually part of the television signal, embedded in its sound channel, mobile devices could easily be programmed to decode them via their microphone and display them on screen.

While subtitles are also a feature of cinema and DVD movie releases, they present a different challenge and are produced differently. However, a watermarking scheme for film might prove very useful for various purposes. The adaptation of the proposed scheme to embed script, director's notes, associated information and dialogue within the sound channel of a film would provide great flexibility in terms of searching film. In order to search through a film for a small section of dialogue, users must currently either know the general location of the dialogue in the film's timeline, or must scan through the entire footage. Even having access to a typed script of the film will only help narrow down the search as scripts are not time-stamped. If the dialogue was watermarked into the sound channel, a basic search for a line of dialogue could direct the user to the film footage exactly at the right time segment. It would be a one-time task to produce the necessary watermark, embed it in the footage and make it available for more efficient and effective research in future. The same technique could be applied, for example, to animations, recordings of political speeches, parliamentary proceedings and even Court proceedings. Since the watermark is embedded in sound, not video, these techniques can be applied wherever audio can be recorded.

Appendices:

Appendix 1: Example license costs for the Phonographic Performers of Ireland (PPI) and Performance Rights Society UK (PRS) public performance license.

Appendix 2: Large scale images of frames before and after modification using three variations of the CSPE algorithm outlined in Section 6.3

Appendix 3: Complete results of trial and error experiments performed to identify optimal encoding parameters in phase 1 of the watermarking scheme.

Appendix 4: A series of suggested applications that could be implemented using the watermarking scheme proposed in Chapter 6.

Appendix 1: Example costs for the Phonographic Performers of Ireland (PPI) and Performance Rights Society UK (PRS) public performance license.

Annual Income Band (EURO)			Band Amount (EURO)	Individual Band Tariff (%)	Individual Band Tariff (EURO)	Cumulative Top Of Band Tariff (EURO)	Tariff Amount As % Of Income
0	-	105,999	106,000	1.30	1,378	1,378	1.30
106,000	-	211,999	106,000	1.95	2,067	3,445	1.63
212,000	-	317,999	106,000	2.60	2,756	6,201	1.95
318,000	-	527,999	210,000	3.25	6,825	13,026	2.47
528,000	-	843,999	316,000	3.90	12,324	25,350	3.00
844,000	-	1,266,999	423,000	4.55	19,247	44,597	3.52
1,267,000	-	1,689,999	423,000	5.20	21,996	66,593	3.94
1,690,000	-	2,112,999	423,000	5.85	24,746	91,338	4.32
2,113,000	-	2,639,999	527,000	6.50	34,255	125,593	4.76
2,640,000	-	3,166,999	527,000	7.15	37,681	163,274	5.16
3,167,000	-	3,693,999	527,000	7.80	41,106	204,380	5.53
3,694,000	-	4,220,999	527,000	8.45	44,532	248,911	5.90
4,221,000	-	4,747,999	527,000	9.10	47,957	296,868	6.25
4,748,000	-	5,274,999	527,000	9.75	51,383	348,251	6.60
5,275,000	-	6,332,999	1,058,000	10.40	110,032	458,283	7.24
6,333,000	-	8,441,999	2,109,000	11.05	233,045	691,327	8.19
8,442,000	-	10,550,999	2,109,000	11.70	246,753	938,080	8.89
10,551,000		plus		12.35			

Table A1.1: PPI (Ireland) license costs for independent commercial radio stations [138].

'Net Broadcasting Revenue' (NBR)	Rate
Below £578,877	3%
£578,878 - £1,157,754	4%
£1,157,755 or more	5.25%
Where the total music use is less than 15% of the broadcast then regardless of the level of the NBR the percentage rate to be applied will be	1%

Table A1.2: PRS (UK) license costs for independent commercial radio stations [139].

Note: The above rates are provided for illustration purposes and reflect the royalty rates payable by commercial radio broadcasters. The PPI and PRS have a range of license tariffs for different users

Appendix 2: Large scale images of frames before and after modification using three variations of the CSPE algorithm outlined in Chapter 6, Section 6.3.

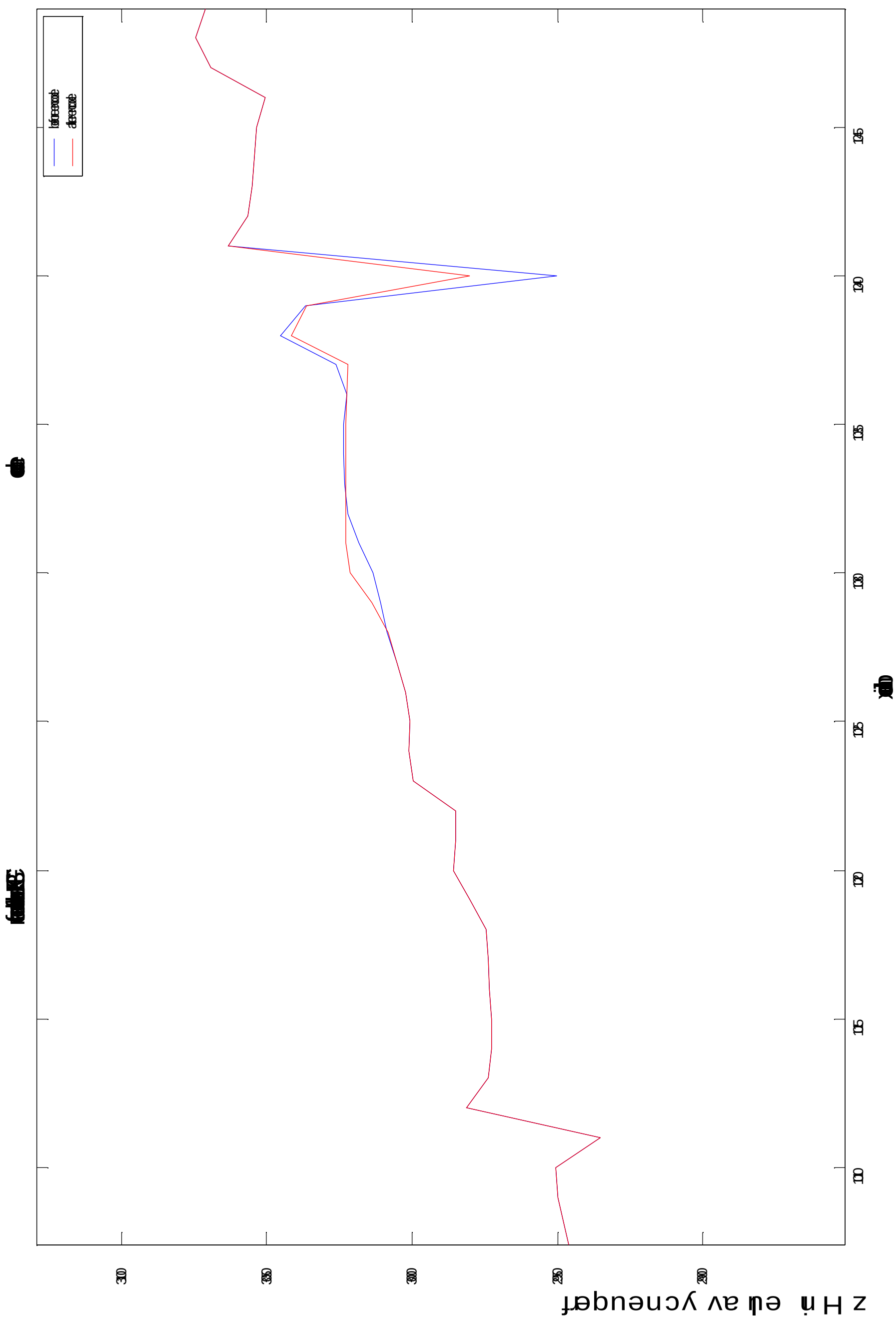


Figure A2.1: A large scale image showing the change from original to the watermarked when processed with the magnitude increase algorithm.

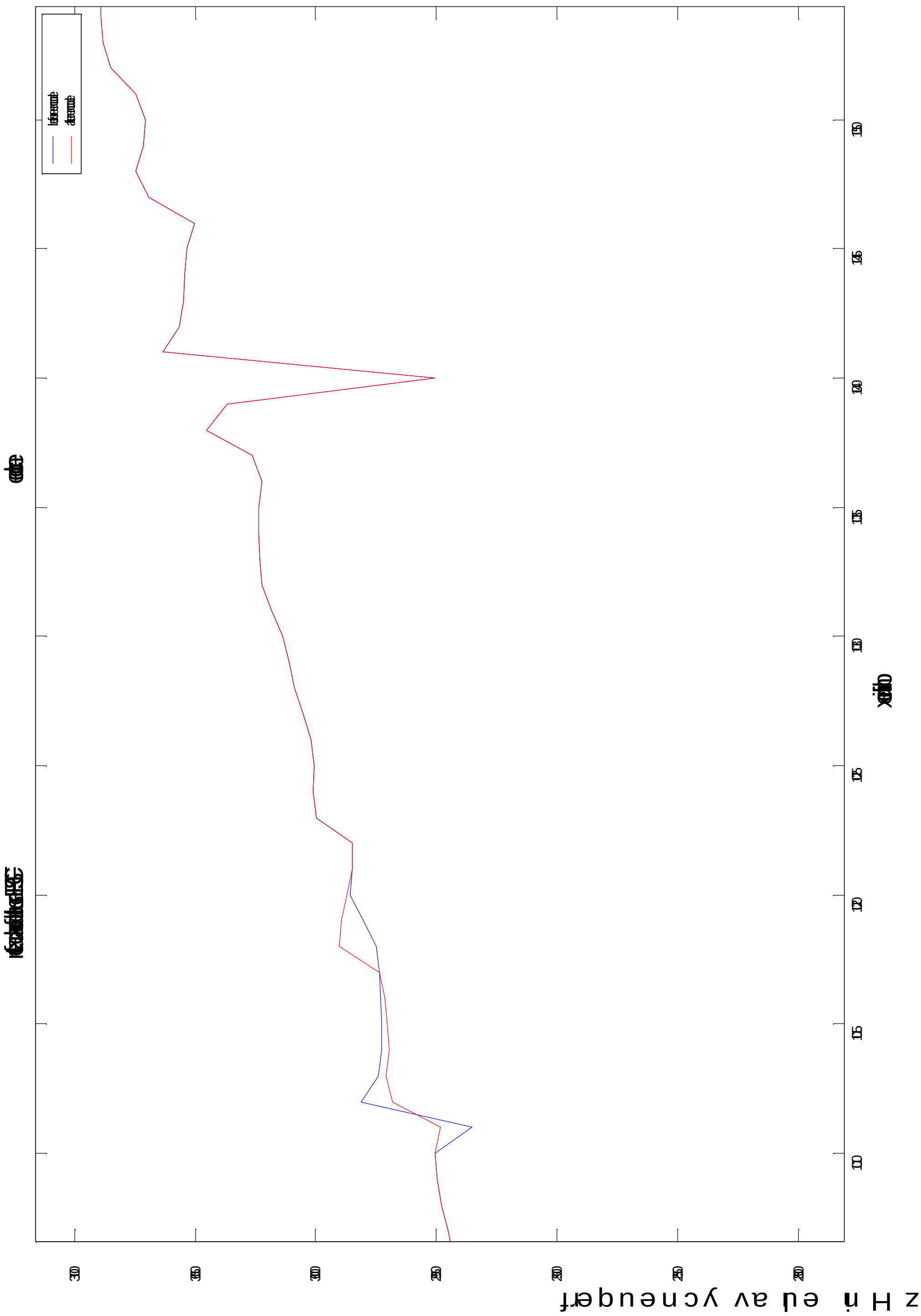


Figure A2.2: A large scale image showing the change from original to the watermarked when processed with the magnitude reduction algorithm.

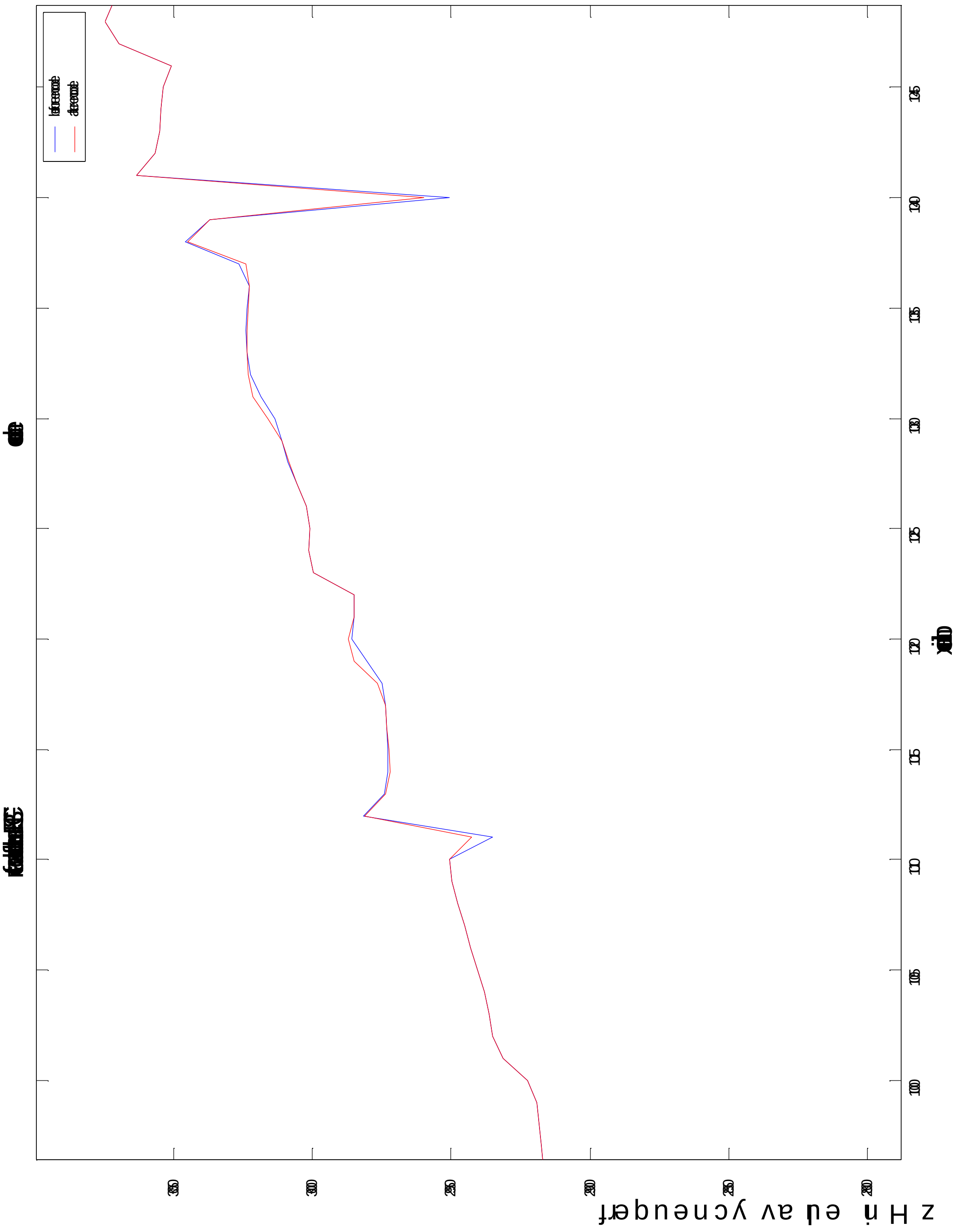


Figure A2.3: A large scale image showing the change from original to the watermarked when processed with the magnitude reduce and swap algorithm.

Appendix 3: Complete results of trial and error experiments performed to identify optimal encoding parameters described in Chapter 6, Sections 6.1 and 6.2.

No.	Samples (S)	Frequency (F)	Tone (L)	Files decoded	Accuracy % (P)
1	4	9500	25	347	99
2	3	9500	30	346	99
3	3	9000	25	300	99
4	4	9000	25	694	99
5	4	9500	30	346	99
6	2	9000	25	135	98
7	3	9500	25	281	98
8	3	9500	20	347	98
9	4	7000	25	50	98
10	3	7000	25	50	98
11	2	9000	25	100	97
12	4	7000	20	50	96
13	3	6000	25	50	95
14	2	9500	25	347	95
15	3	7000	20	50	95
16	2	9000	25	100	95
17	1	12000	20	50	94
18	2	9500	20	694	93
19	1	12000	25	50	92
20	4	7000	16	50	92
21	3	7000	16	50	90
22	1	12000	10	50	90
23	2	13000	15	397	90
24	2	9000	25	347	90
25	2	12000	15	350	90
26	1	11000	13	25	88
27	4	7000	10	50	88
28	1	11000	13	25	88
29	2	11000	15	347	87
30	2	11000	15	347	87
31	2	11000	15	347	87
32	1	10000	21	25	86
33	1	11000	9	25	84
34	1	11000	9	25	84
35	4	10000	25	50	84
36	2	10000	15	348	84
37	2	10000	15	347	84
38	1	10000	33	25	82
39	1	10000	13	25	82
40	1	10000	25	25	80
41	1	10000	29	25	80
42	2	9000	15	347	80
43	1	10000	17	25	78
44	1	8000	17	25	78
45	1	9000	29	25	78
46	1	10000	9	25	76
47	1	10000	27	25	76
48	4	5000	16	50	76
49	6	5000	16	50	76
50	1	9000	13	25	76
51	2	9000	15	281	74
52	1	11000	15	25	74
53	1	9000	17	25	74
54	1	10000	31	25	74
55	1	11000	15	25	74
56	4	7000	5	50	72
57	1	9000	27	25	72

58	1	9000	31	25	72
59	1	8000	29	25	72
60	1	10000	35	25	72
61	1	9000	9	25	72
62	1	8000	13	25	72
63	1	9000	21	25	70
64	1	9000	33	25	70
65	1	9000	25	25	70
66	1	11000	11	25	68
67	1	10000	19	25	68
68	1	9000	23	25	68
69	1	8000	9	25	68
70	1	11000	11	25	68
71	1	7000	17	25	66
72	1	8000	25	25	66
73	1	8000	21	25	66
74	1	7000	13	25	66
75	1	9000	35	25	66
76	1	8000	35	25	66
77	1	7000	21	25	64
78	1	10000	23	25	64
79	1	7000	29	25	64
80	1	9000	19	25	64
81	1	7000	10	50	64
82	1	8000	23	25	62
83	1	8000	31	25	62
84	1	8000	33	25	62
85	1	7000	33	25	62
86	1	10000	15	25	62
87	1	8000	27	25	62
88	1	11000	5	25	62
89	1	11000	5	25	62
90	2	9000	5	117	60
91	1	9000	11	25	60
92	1	7000	9	25	60
93	1	8000	19	25	60
94	1	8000	15	25	60
95	1	11000	7	25	60
96	1	10000	11	25	60
97	1	11000	7	25	60
98	1	7000	27	25	58
99	1	9000	15	25	58
100	1	9000	5	25	58
101	1	7000	25	25	58
102	1	10000	5	25	54
103	1	7000	31	25	54
104	1	8000	11	25	54
105	1	7000	15	25	52
106	1	6000	9	25	52
107	1	10000	7	25	50
108	1	7000	23	25	50
109	1	7000	19	25	50
110	1	5000	16	50	50
111	1	7000	35	25	50
112	1	6000	13	25	46
113	1	6000	21	25	46
114	1	6000	31	25	46
115	1	6000	33	25	46
116	1	6000	17	25	44
117	1	7000	5	25	44
118	1	9000	7	25	44
119	1	6000	35	25	44
120	1	5000	21	25	42
121	1	6000	29	25	42

122	**1**	**7000**	**11**	**25**	**42**
123	**1**	**8000**	**5**	**25**	**42**
124	**1**	**6000**	**25**	**25**	**40**
125	**1**	**6000**	**27**	**25**	**40**
126	**1**	**8000**	**7**	**25**	**38**
127	**1**	**6000**	**19**	**25**	**38**
128	**1**	**6000**	**15**	**25**	**38**
129	**1**	**5000**	**29**	**25**	**36**
130	**1**	**6000**	**23**	**25**	**36**
131	**1**	**5000**	**33**	**25**	**36**
132	**4**	**12000**	**3**	**50**	**34**
133	**1**	**5000**	**35**	**25**	**34**
134	**1**	**5000**	**15**	**25**	**34**
135	**1**	**5000**	**13**	**25**	**34**
136	**1**	**5000**	**23**	**25**	**32**
137	**1**	**5000**	**19**	**25**	**32**
138	**1**	**5000**	**17**	**25**	**32**
139	**1**	**6000**	**11**	**25**	**32**
140	**1**	**5000**	**27**	**25**	**30**
141	**1**	**6000**	**5**	**25**	**30**
142	**1**	**5000**	**31**	**25**	**30**
143	**1**	**5000**	**25**	**25**	**28**
144	**1**	**5000**	**9**	**25**	**28**
145	**1**	**5000**	**11**	**25**	**24**
146	**1**	**7000**	**7**	**25**	**24**
147	**1**	**5000**	**5**	**25**	**18**
148	**1**	**12000**	**3**	**50**	**12**
149	**1**	**6000**	**7**	**25**	**8**
150	**4**	**7000**	**3**	**50**	**6**
151	**1**	**5000**	**7**	**25**	**6**

Table A3.1: Complete results for the experiments outlined in section 6.1 & 6.2.

Appendix 4: Suggested applications for the watermarking scheme proposed.

While the work described in this document was intended from the outset to be a digital audio watermarking scheme with an express focus in the domain of broadcast monitoring for the purposes of accurate airplay monitoring and equitable royalty distribution, it is by no means restricted to this realm. Almost any form of digital audio watermarking scheme will be applicable to a number of different domains and for a number of uses. A scheme that offers robustness against accidental or deliberate attack, combined with variable levels of security and a blind or semi-blind decoding phase will have a wide variety of potential uses, some of which are briefly mentioned here.

Covert communications

Almost since the dawn of indirect communications (i.e. communications other than face-to-face) there has been a place for secret transmission of messages. The domains of military, industrial and private communications have always been the most likely to warrant attempts to protect the communications from unauthorised interception. Whether it is a wax seal or some variant on a letter delivered by an intermediary – still used in some modern communications channels - which indicates unauthorised interception of the message content with a broken seal, or a fragile watermarking system which 'breaks' itself if any attempt is made to remove it, content protection is still an imperative in some communications. The types of indicators mentioned (fragile watermarks or wax seals) do not actually *prevent* the interception of messages. Instead they merely indicate when the messages *have* been intercepted or 'broken'. They are, in effect, after the fact and are of little use in protecting the content of the message. The only means by which a sender can guarantee no unauthorised interception of their message is by limiting any knowledge of the fact that such a message exists.

Covert communications will traditionally be viewed as the realm of cryptography rather than steganography. As mentioned earlier, cryptography is the process of encoding a message so it (hopefully) cannot be deciphered if discovered, while Steganography is the process of hiding content to prevent it being discovered in the first place. If there is no knowledge of the actual *presence* of a message then there is no chance of it being intercepted. Combining the two related fields of steganography and cryptography is perhaps the most effective means of covert communications. If both methods are used simultaneously it would suggest that there is an expectation that the steganographic phase of the security is going to be inadequate. Nevertheless, caution is to be commended and there is nothing stopping an enciphered message being embedded in a steganographic process. If a message is hidden and its presence is unknown, it will be safe from unauthorised interception. However, if it *should* be discovered or, for example, leaked in a malicious incident, having it encrypted before embedding will add an additional layer of security to the message.

If an attacker obtains an original version of the host audio signal as well as a watermarked version, one can be compared to the other to identify the watermark (plain, transformed or enciphered). One way of avoiding this, in terms of military or industrial covert communications, is to embed the watermark in audio signals that are *not* publicly available without the watermark. In this way, no attacker would have an original to compare the candidate watermarked audio signal against. This is not a difficult thing to achieve. A simple recording of some cover audio, even speech, made by the sender that is then embedded with the watermark, before deletion of the original, would suffice. In a situation where a covert message was embedded into a voice (or other) recording, no attacker or interceptor would have any indication that there was an embedded watermark within the audio. Nor would they be able to compare the candidate audio to an original if they did.

Some watermarking schemes, however, process the host audio in such a way as to leave a visible pattern or other noticeable artefacts when embedding the watermark. As Gopalan mentioned in the work **[86][127]** that originally inspired the early approach of the proposed watermarking scheme, a spectrogram or histogram could provide visual clues of the presence of a watermark thereby allowing attackers to focus their attempts to intercept the message on relevant signals and areas. In a watermarking process such as the one proposed in the third phase of the proposed scheme outlined in Section 3.4 above, there would be no pattern identifiable in a visual representation of the candidate signal. This is because the algorithm designed does not manipulate or modify the same frequency components in each subsequent frame. In fact, the choice of components to be modified in one frame is completely independent of the choice in any other frame and relies solely on the ‘key’ or ‘reference’ value chosen by the embedder. Without this value it would be impossible to know which components were modified.

Of course, this key value chosen as the first input parameter to the scheme could be guessed or leaked. Similarly, it might be suggested that a ‘brute force’ attack could simply iterate through all possible values to find the correct parameters. This is correct but it omits the fact that an attacker would not have any idea what to look for in each of these brute force iterations and so would not know if they had found the correct parameters or not. Essentially, the choice of components to modify in a given frame (if any) is almost random, within a range of values. This includes fractions of values and the precision of the selection depends only on the input parameters used in the CSPE frequency analysis. The process is outlined in Figure A4.1 below.

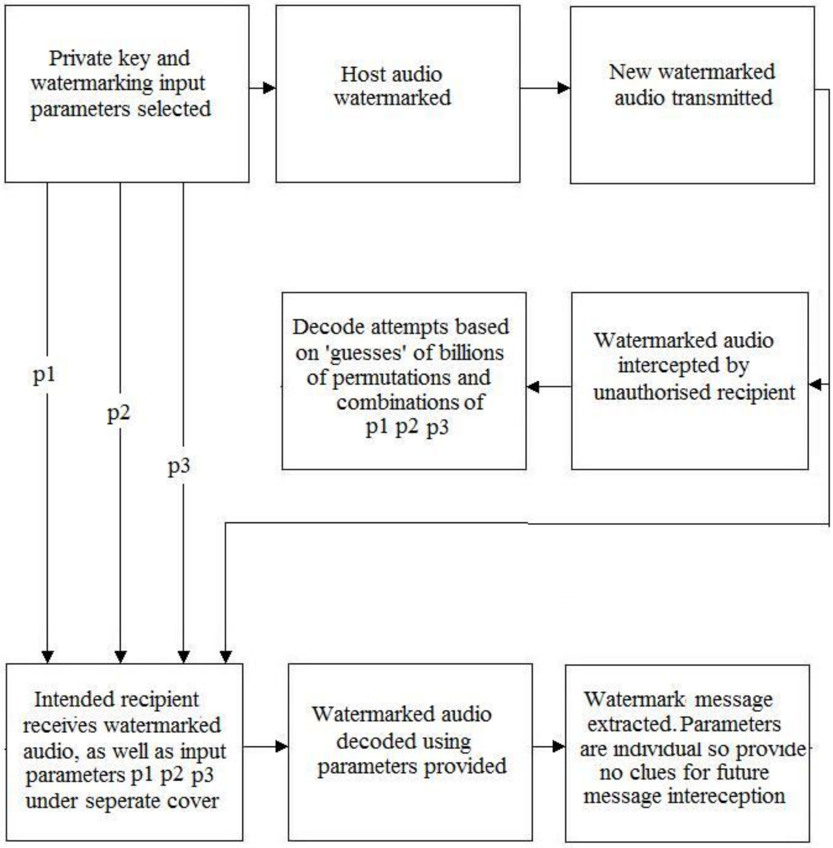


Figure A4.1: Block diagram illustrating the covert communications watermark scheme.

Given that the attacker is unlikely to know the length of the frame used to embed the bit, the length of the watermark (which is repeated any number of times within the bit sequence) or whether every individual bit is part of the actual message and not a decoy injected into the bit sequence before embedding, the chances of discovery of the *presence* of a message, let alone its content, is very slim indeed. In the event that an attacker was otherwise *informed* of the presence of a watermark there would still be the issue that the brute force attack would have no way of identifying when it has discovered a potential pattern in order to make an attempt to decipher it.

On-the-fly music identification

When a track is played on radio that a listener is unfamiliar with, he or she is at the mercy of the presenter or producer of the radio show for identification of the track. However, in many cases there is no identification given. This occurs when a radio presenter plays a number of tracks in sequence and does not identify any/all of them, or when a radio station is using a computerised delivery platform – particularly in the so-called ‘dead hours’ where there is no human involvement in the broadcast. It can be a frustrating situation when a listener hears a track they are interested in and wants to know more about it but is given no information with which to identify the performer or track title. It is also a potential sales opportunity lost. Assuming the listener wants to find out more about the artist, and may even want to buy the material they are listening to, there is no current way of facilitating this discovery short of, perhaps, writing some lyrics and searching for them on the Internet at a later time.

Digital audio watermarking at source could easily, quickly and – most importantly – at little or no added cost in the production chain - overcome this limitation and not only offer listeners the information they need to research the artist or buy the material, but also offer an ‘added value’ technology for a manufacturer of both domestic and in-car entertainment devices as a user-centred selling point [**140**]. It is not a complex development of the watermarking scheme proposed to design a system to ensure an audio watermark is embedded into songs, preferably at the time of production but perhaps even later, which will allow the real-time identification of the performer, track title and/or other content such as publisher details, ownership details etc. This information can be extracted and displayed on existing screens in both domestic and in-car audio entertainment systems, in a manner similar to the way in which time or station identification (known as Radio Data System or RDS) is currently extracted and displayed. The information will be part of the actual audio content, rather than meta-data such as MP3 headers, so would be transmitted with the audio even over an analogue transmission channel.

Existing technologies allow the on-screen display of information about artist and title but only from a digital source or from an external source manually entered into the broadcast environment and transmitted alongside the audio. The scheme proposed here could easily be adapted so this may be achieved in a traditional analogue radio environment or in situations where an audio source (such as an ‘iPod’ etc) is being transmitted via FM to a local audio device. This is a common setup in car audio setups and becoming more common in domestic scenarios where (for example) internet radio is received and wirelessly transmitted to a home entertainment system. Moreover, since the embedded information is actually *part of* the audio, rather than being transmitted alongside it, even the common action of recording the audio for later use on another device will preserve the watermarked message. Assuming the device used to play the recorded audio is designed to identify and display watermarks, it will still be present and displayed to the end-user.

Another area in which the proposed watermarking scheme could prove of benefit to content creators and content owners is in on-the-fly content identification. This domain is very topical and there are many applications that purport to offer the end-user the chance to identify the music they hear, for example, in a local store or on the street. The microphone in modern mobile phones can be used as a recording device, as is the case with the Apple's 'iPhone', Nokia's mobile range and RIM's 'Blackberry' amongst others. Once sampled, the audio is then compared to existing databases of audio and identification attempted. One of the better-known of these providers is 'Shazam', which illustrates the popularity of the concept by its 20 million users in 60 countries on A45 network carriers. The major problem with these applications, as is the case with fingerprint-based broadcast monitoring applications, is that they rely on a database of files to compare the candidate to and also that they require some sort of data connection – which can incur a charge.

Using a watermark-based system, where the identifier is embedded in the audio by the content owner or producer, these limitations can be completely negated. First, an application could easily be developed for modern smart-phone clients such as the iPhone, Blackberry and Nokia/Symbian operating system that would simply analyse the audio being recorded via the in-built microphone and perform a CSPE-type frequency analysis before outputting the identifier. No data connection would be needed and no fingerprints would need to be compared to a database. An illustration of the process is given in Figure A4.2:

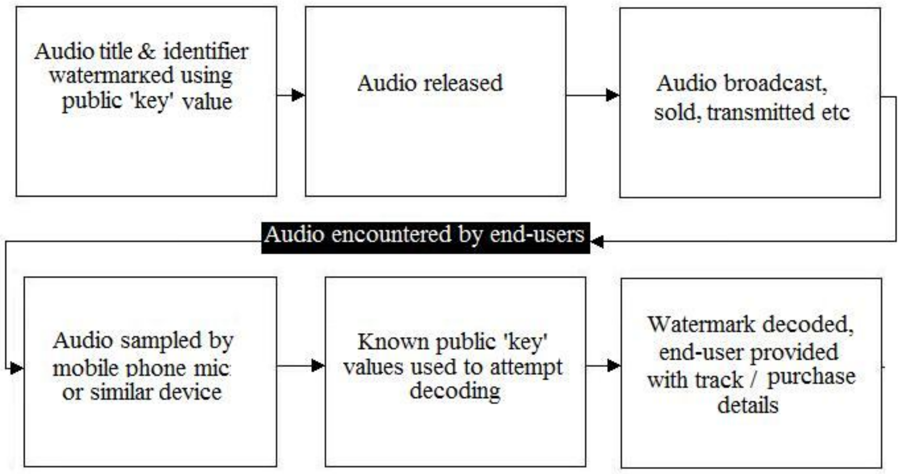


Figure A4.2: Block diagram of on-the-fly decoding of embedded watermarks.

The widespread use of existing (limited) applications shows that there is a viable and exploitable market for them [**141**]. Content creators could maximise their sales potential by utilising watermarking techniques in their production processes and this would encourage widespread use of the end-user application. More identifiable audio would make the applications more attractive to the customer. As with traditional broadcasters, watermarks could be used to embed sponsor messages alongside identification information so even if the user does not decide to buy the track, it creates a completely new revenue stream for content producers. These additional watermarks could be added by using a different key value from which the modified components are derived. As mentioned in Section 6.5.4, capacity could also be doubled by using two components on either side of the base frequency. The final point to note is that in the case where a work is watermarked at the point of creation, it does *not* have to be done by a record label, publisher, or other corporate sector of the music industry. Like the recording itself, the process could easily be undertaken by the artist independent of the rest of the creation process. In fact, the ubiquitous 'bedroom artist' would be just as capable of watermarking their information into their work as any commercial artist. This, of course, would help in some small way to level the playing field for developing artists.

Additional revenue generation for broadcasters and content producers

In the commercial broadcasting environment the financial imperatives are at least as important as the quality and quantity of audio that is actually broadcast. Some would suggest that they are in fact more important. A digital audio watermarking scheme implemented by a broadcaster could offer three attractive opportunities for the broadcaster to optimise their existing revenue stream and develop a complementary source of additional revenue. A diagram outlining the various watermarking options that could be employed in a broadcast scenario is provided in Figure A4.3 below. It should be remembered that the embedded watermark in the proposed scheme is, and must be, completely inaudible to any human listener so it could not possibly impact on the quality of the material broadcast.

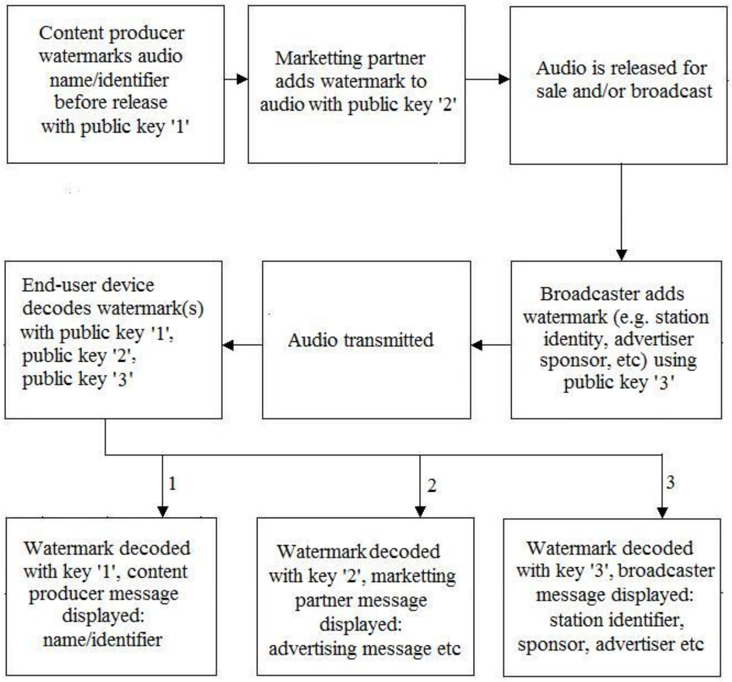


Figure A4.3: Block diagram illustrating proposed scheme for a broadcast environment to enhance value for advertisers, broadcasters and content providers.

Note that each of the three watermarks embedded at the production phase (indicated by the use of *public key 1, 2* and *3*) are independent and can be included or omitted. In the case that one of these watermarks was not applied, the result of the decode phase using that particular value would be null as no watermark would be found. Therefore, no message or information would be displayed. Alternatively, if all watermark phases were utilised, all three messages would be displayed wither simultaneously, one demand or in sequence depending on the end-user's device.

In the first case, a broadcaster could embed a watermark in their output that acts like the aforementioned Radio Data System. Rather than simply identifying the broadcaster or displaying the time, however, it could be used to identify existing output. The watermarking mentioned up to this point has dealt mostly with watermarking the music that these broadcasters transmit. However, there is no reason why they cannot themselves employ a similar method of watermarking audio adverts with additional information from the advertiser, such as special offers, web address and so on. Since adverts on radio are traditionally 'record once, play many times', there is limited scope for adding to them once they have been produced, short of appending some information at the end of the advert. If a broadcaster had an in-house watermarking system, which would not in any way be a difficult or expensive setup to achieve, they could offer a service where the advertiser could contact them at any time and have additional information embedded as a watermark in the broadcast over their advert, which would in turn be displayed on the device used by the end-user.

Second, a variation of the same technique could be used to embed *content-sensitive information* that could be charged for. There are two obvious types of content-sensitive information envisaged. In the first place, broadcasters playing work by a particular artist could embed details of their upcoming live appearances in their broadcast catchment area along with details of ticket availability. Artists and promoters already pay for this type of advertising and it would make commercial as well as common sense to have the details visually displayed while the artist in question is being broadcast. Listeners who hear the audio and like it are the prime targets for this additional information. Embedding the information into the audio to be displayed when the artist in question is being broadcast is a classic case of product placement which is a common marketing mechanism.

In the second variation of *content-sensitive information*, the broadcaster could provide an opportunity for advertisers who already use music in their advertising to have their advertising used *in the music*, thereby reinforcing the association and brand recognition. To illustrate this, consider for example the track *'Let's Dance'* by Paula Flynn. This track was originally released by David Bowie in 1983 but Ms. Flynn's version of it was featured in a television advertising campaign for Vodafone in the last few years. This exposure led to the track being heavily featured on radio and becoming a sort of 'cult' track for some time. It was regularly referred to as 'the Vodafone song', even on radio and led to Ms Flynn developing a comparatively successful music career that otherwise might not have been as productive.

There are myriad examples over the last few decades of such radio and television advertising campaigns leading to successful releases of golden oldies by famous artists as well as works by almost unknown artists. In many cases, careers have been made realistically commercially viable by such exposure. Bringing this into the present, note the current *Budweiser* campaign featuring '*All together now*', which was written originally by The Beatles for *'Yellow Submarine'* but recently recorded by 'The Hours', in a house and in one take, specifically for the campaign. This version is now so popular, particularly on the Internet where it has apparently been downloaded tens of thousands of times from various providers in a few short weeks since the campaign launch, that the act in question ('The Hours') are recording a full version for commercial release. Similarly, French newcomer 'The Do' has seen their profile skyrocket when their track '*Stay Just a Little Bit More'* was featured in O2's *'Turtles'* advertising campaign. Bell X1 might be just one act who are uneasy at the fact that '*The people from the mobile phone company say who gets to play and who gets to not*' ('The Great Defector') but there is no doubt that marketing departments have become tastemakers in the world of music over the last number of decades. There is also no doubt that advertising campaigns can, and do, make an artist's career. Given this, there should be every reason to make them complimentary so every party benefits.

Broadcasters could offer advertisers the opportunity to have their advert embedded into the broadcast *when the track in question is being played* and transmitted to the screen on whatever receiving device the end-user is using to receive the broadcast. Extending this concept to allow any combination of the content owner, record label, publisher, advertiser and the broadcaster co-ordinate to present a highly complementary synergetic package would also be very easy to achieve. This offers opportunities for all, not least the artist and the broadcaster but the advertiser would benefit from being associated completely and intimately with the work or the artist. It is for this reason that major corporations sponsor large-scale tours and cultural events.

The third opportunity that broadcasters could avail of as a revenue generation stream is that of station-sponsorship or individual show sponsorship. Notwithstanding the regulatory requirements for such arrangements, broadcasters could contract a sponsor for a particular show (if not for the entire station). These arrangements already exist in both television and radio. In the case where an in-house watermarking process is used, the broadcaster would easily embed the sponsor's name into their broadcast of the show. Even talk-shows and current affairs shows which do not tend to employ a lot of music content would be attractive vehicles for watermarking a sponsor's name or message wherever an advertiser's watermark was not being broadcast and displayed. Since it is obviously not possible to advertise in the traditional sense while a broadcast presenter is talking, advertising is currently limited to short blocks of time. In many cases, these are prompts for listeners to tune to a competing channel. Implementing advertising 'streams' as a text-on-screen addition to current methods will enable more advertising, longer messages and less separation between the advertiser and the broadcast. This would be easily achievable with an in-house watermarking scheme.

None of these applications of the basic watermarking scheme is inherently difficult. Environment-specific considerations might lead to modifications to the system. It is worth noting that if the in-house watermarking scheme used, for example, the same initial input parameter (the 'key' referred to earlier) to embed their watermark into the audio before broadcast, it would effectively overwrite the watermark that might have been present in the audio. Broadcasters could instead use a unique key for their own watermarks which would leave any inherent watermark untouched. Care should be taken in this regard to ensure that not too many watermarks were embedded as, while multiple watermarks *could* be embedded using the proposed scheme, each one might impact slightly on the perceptual quality of the audio and the combined effect might be noticeable.

Finally, in relation to broadcasters and other owners of archives or previously released audio, there is no difficulty in watermarking pre-existing audio. A simple once-off batch process could be employed to ensure any material already in their possession is also watermarked. In essence, there is no audio that could not be watermarked by the proposed system. It is entirely possible that what humans consider silence (which is not, of course, always silence [**24**]) could even be watermarked using this method.

Watermarking individual performers

Another area in which watermarking techniques can assist the artist, performer, collective rights societies and the Music industry is by watermarking individual performances *within* a work. Rarely is there a publicly released piece of music which contains only one individual track. In the recording process, individual instruments, vocalists and additional production 'effects' – often the creation of the Producer - can be assigned to individual tracks within the studio or live-recording environment. Once mixed down to the stereo 'master' for duplication and release, there are only two tracks – left and right channels. In the recording process, each individual performer (guitarist, drummer etc) is assigned a separate track. Similarly, vocalists and backing vocalists can be assigned separate tracks. This is important from the point of view of both performance royalties and 'sampling'. A watermarking scheme employed in the studio could embed a track with unique watermarks per each contributor, at the time of their recording, including the producer if appropriate. This process is described in Figure 4.4 below. The process of individually watermarking performances, separately from the watermark of the end-product, would ensure that there would never be a version of the audio without the performer's identifier and this in turn would ensure correct and proportionate distribution of royalties. It would furthermore prevent any conflict at a later date about the contributions any individual that is used in the final 'mix'. This might seem like a trivial matter but in the world of the creative artist this identification of ownership is paramount and is, in fact, one of the main reasons why Copyright exists.

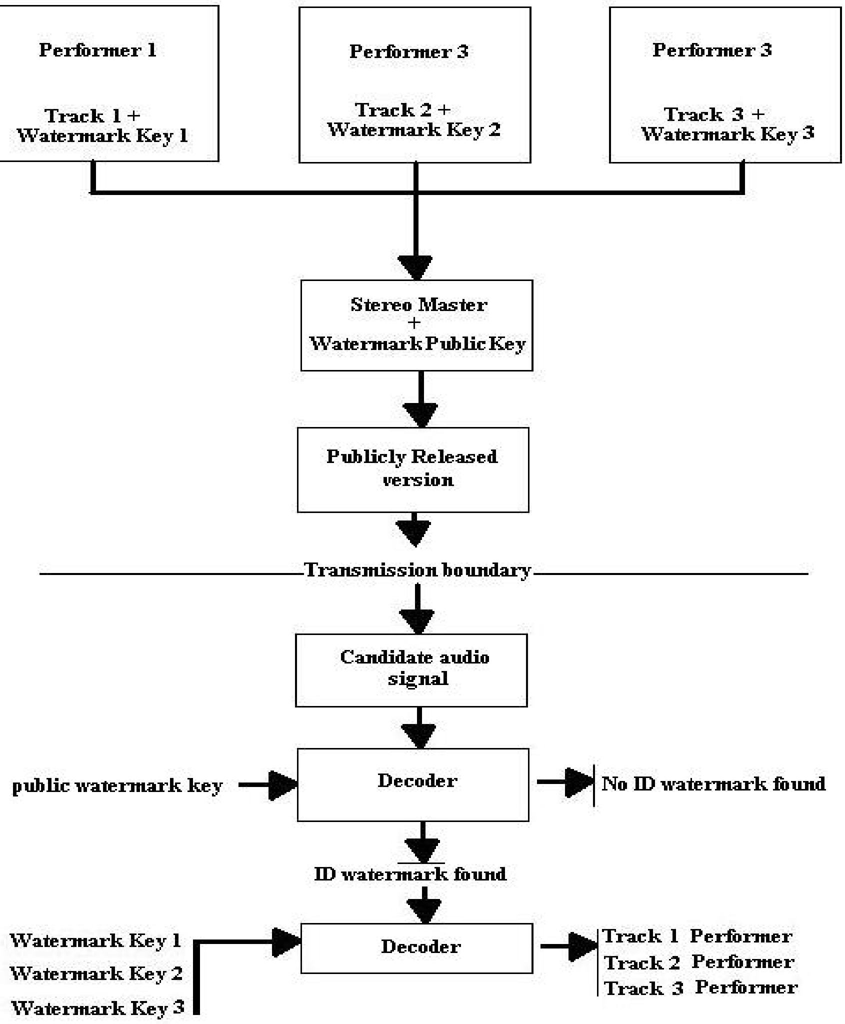


Figure A4.4: A block diagram illustrating the process of watermarking individual performers' recordings separately from the end-product recording.

As can be seen from Figure A4.4, the individual performances recorded for each performer on a work, which may be created on different dates, in different studios sometimes in different countries, are assigned an individual 'private' watermark. The 'mix' down to a stereo master is then performed and this master is separately watermarked with, perhaps, an industry-standard ‘public’ key before being released.

When a candidate signal is analysed for the presence of a watermark, an attempt is made to decode the watermark with the standard ‘public’ key. If there is no watermark founding this way, then it is likely the track is either not watermarked or its watermark is corrupt or damaged. If this is the case, there is no point in attempting to decode individual ‘private’ watermarks as these are either similarly not present or damaged. If the candidate signal *does* have a standard watermark identified using the ‘public’ key, the track is identified and its identification appropriately reported. The signal can then be decoded using any of the ‘private’ keys that are potentially embedded within it. While in most cases, all of the individual performances will be identified, this process would identify even those remixes and versions of a work that have had some of the individual performances omitted.

Tracking unauthorised distributions

There are, in fact, a number of modifications to this CSPE-based watermarking technique that could allow it to be adapted and specialised for different requirements. A watermark could, for example, be designed to be audible (and irritating) when played in any digital music or video device that was not authorised to play it. Authorised players could have the relevant parameters input as part of their authorisation process or the parameters could be communicated to the authorised device via a networked on-demand ‘handshake’ arrangement. This would perhaps be overkill for music and video files. However, it could be useful for pre-production evaluation copies of digital files where the owner might want to allow certain authorised devices unrestricted access while making it unattractive to copy the file for unauthorised distribution.

Having the watermark embedded as part of the actual material would mean that simply playing it through speakers and re-recording it would not, as mentioned above, remove such a watermark. This is a critical and valuable difference between pre-release watermarking and post-release watermarking. In the case of an audible watermark, playing the audio through speakers and then re-recording it would replicate the audible and irritating sounds added by the watermark. Such a system, if designed for tracking unauthorised ‘leaks’ of new music or films could easily achieve this by personalising the watermark-embedding parameters for the few authorised recipients.

An alternative to this concept could be to embed the watermark inaudibly, per recipient and with individual private keys, in order to identify suspected leaks. If a version of the material were to appear in the public domain with no audible artefacts, this would suggest it contains no watermark, the watermark is damaged or it is a pre-release version. If it is suspected that the candidate signal is a pre-release version, distributed without authorisation (i.e. ‘leaked’), steps could be taken to identify the leak. The candidate signal could be analysed for the presence of the watermark and the parameters used in any watermark that was discovered would prove that it originated at the source to which the key was allocated. This process is described in Figure A4.5.

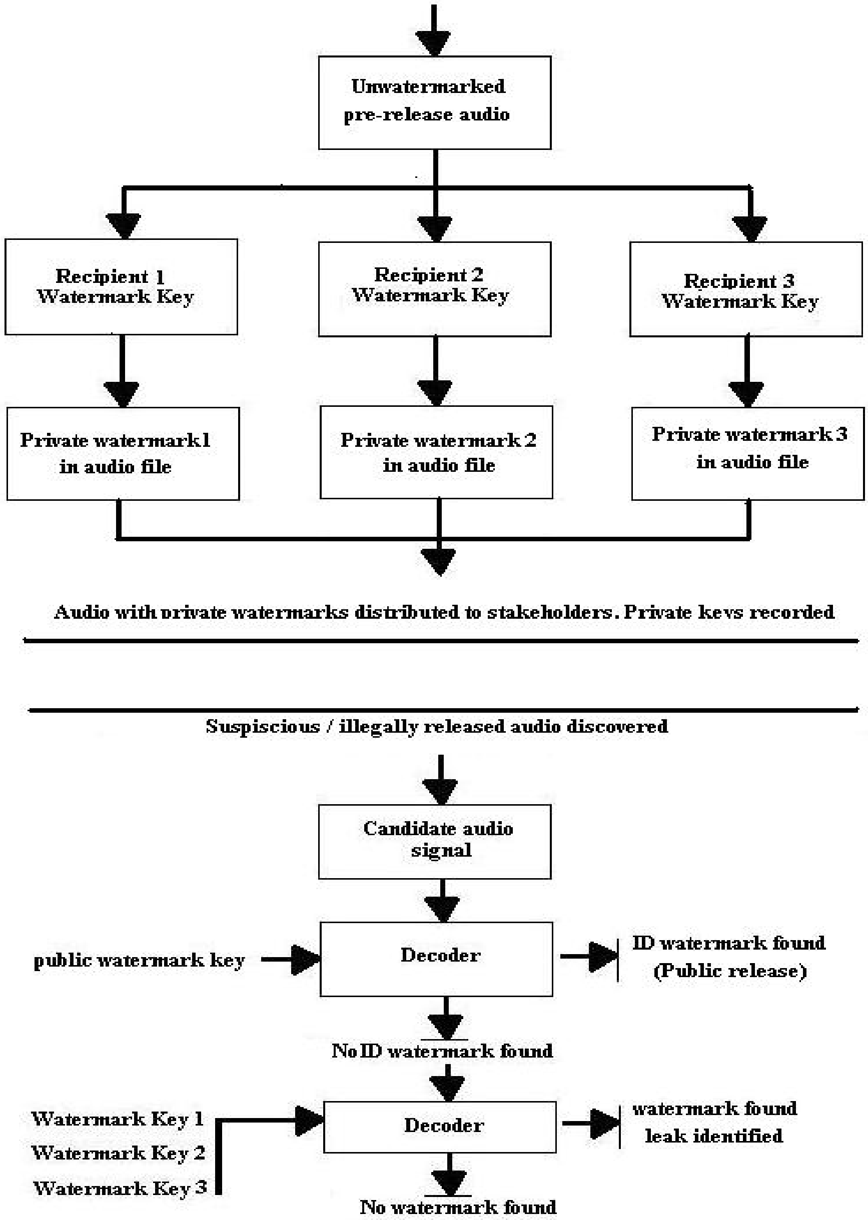


Figure A4.5: Block diagram of a proposed watermark system designed to identify and track illegal distribution of content.

The source audio that is presented in the first step of the process shown in Figure A4.5 is unwatermarked. This means it has not been prepared for public release. Content owners can easily produce a small number of private-distribution versions of the audio to stakeholders and interested parties, each watermarked with using a separate key value. Each version of the audio would sound identical to stakeholders as the watermark is inaudible and has no impact on audio quality. These pre-release versions would obviously be expected to remain private and never to be found in public circulation.

However, should the content owner or monitoring agent be suspicious of a candidate audio found in public circulation, there would be a simple two-step process to identify the source of the leak. The first step is to ensure that the candidate is not a public-release version. This is achieved by decoding using the standard key that all public-release versions would be watermarked with. If this key is not present, it is either damaged or was never present. An additional step could easily identify if the watermark is partially present and damaged. Otherwise, the candidate is decoded using the private values assigned to each recipient of the pre-release version. If the watermark is successfully decoded using any of these values, the recipient to whom that value was assigned is the likely source of the leak.

References

1 The MP3 History, Fraunhofer Institute, Germany.
http://www.iis.fraunhofer.de/EN/bf/amm/products/mp3/mp3history.
Accessed 22 September 2009

2 'Future of secure digital music initiative grim', USA Today, April 29th, 2002. Online source:
http://www.usatoday.com/tech/news/2002/04/29/sdmi.htm.
Accessed October 23rd 2009.

3 Craver, S. A., Wu, M. and Liu, B. 'Reading between the lines: Lessons from the SDMI challenge'. 10th USENIX Security Symposium. Washington, DC, 2001.

4 The 'Recording Industry Association of America'.
http://www.riaa.com/physicalpiracy.php.
Accessed 21 September 2009

5 Tallmo, K.E. 'The History of Copyright: A Critical Overview with Source Texts in Five Languages' Online pre-release version.
http://www.copyrighthistory.com.
Accessed 23rd October 2009.

6 Donaldson v. Beckett, 'Proceedings in the Lords on the Question of Literary Property, February 4th through February 22nd, 1774'.
http://www.copyrighthistory.com/donaldson.html.
Accessed 21 September 2009.

7 'Changes Coming to the iTunes Store', Apple press release (paragraph 3), January 6th 2009.
http://www.apple.com/pr/library/2009/01/06itunes.html.
Accessed 21 September 2009.

8 'The Economy of Culture in Europe, Study prepared for the European Commission (Directorate-General for Education and Culture)', October 2006. Page 246.

9 'Frontline: The way the music died', PBS Documentary, May 27th 2004.
http://www.pbs.org/wgbh/pages/frontline/shows/music.
Accessed 21 September 2009.

10 'Stairway to Heaven' as featured on 'Sold on Song', BBC Radio 2.
http://www.bbc.co.uk/radio2/soldonsong/songlibrary/indepth/stairway.shtml.
Accessed 25th October 2009.

11 Citizendium entry for 'Stairway to Heaven' online source at:
http://en.citizendium.org/wiki/Stairway_to_Heaven#_ref-22.
Accessed 25th October 2009.

12 'Monitoring deal announced between the MCPS-PRS Alliance and Nielsen Music Control'. PRS Press Release, 19 May 2008,
http://www.prsformusic.com/aboutus/press/latestpressreleases/mcpsprsalliance/Pages/Monitoringdealannounced.aspx.
Accessed 21 September 2009.

13 Image sourced from the 'GDI+' project at 'codeproject'
http://www.codeproject.com/KB/GDI-plus/Watermark/watermark_final.jpg.
Accessed 21 September 2009.

14 Ho, A. 'Semi-fragile Watermarking and Authentication for Law Enforcement Applications'. Second International Conference on Innovative Computing, Information and Control, pp. 286, Kumamoto, Japan, 2007.

15 Beauget, S., van der Veen, M., and Lemma, A. ‘Informed detection of audio watermark for resolving playback speed modifications’. Proceedings of the 2004 Workshop on Multimedia and Security, Germany, 2004.

16 Craver, S., ‘Zero Knowledge Watermark Detection’. Third international workshop on Information hiding, Germany, 1999.

17 ‘Paris Convention for the Protection of Industrial Property’, World International Property Organisation
http://www.wipo.int/treaties/en/ip/paris/trtdocs_wo020.html.
Accessed 23rd October 2009.

18 ‘Berne Convention for the Protection of Literary and Artistic Works’, World International Property Organisation
http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html
Accessed October 23rd 2009.

19 Tom Kitt T.D., Minister for State. Dáil debate, 11.11.1999.
http://www.oireachtas-debates.gov.ie/D/0510/D.0510.199911110005.html.
Accessed 11th October 2009.

20 ‘Title 17 of the United States Code’, Cornell University Law School
Online source at: http://www.law.cornell.edu/uscode/17/1003.html.
Accessed 23rd October 2009.

21 Lunney, Glynn S. ‘The Death of Copyright: Digital Technology, Private Copying, and the Digital Millennium Copyright Act’. Virginia Law Review 87 (September 2001) pp. 813-920.

22 ‘What Is Meant By A Public Performance?’
http://www.imro.ie/faq/music_users.shtml (Item 4).
Accessed 11th October 2009.

23 ‘Copyright and Related Rights Act, 2000’
http://www.irishstatutebook.ie/2000/en/act/pub/0028/index.html.
Accessed 14th October 2009.

24 ‘BBC orchestra silenced at the Barbican & on Radio 3’. BBC Press Office, 12th January 2004.
http://www.bbc.co.uk/pressoffice/pressreleases/stories/2004/01_january/12/john_cage.shtml.
Accessed 23rd October 2009.

25 IMRO Fact Sheet: ‘What is payment run 05P041’.
http://www.imro.ie/music_makers/distribution_05P041.shtml.
Accessed 14th October 2009

26 **‘**IMRO Annual Report and Accounts, 2007’.
http://www.imro.ie/docs/pdfs/IMRO_Annual_Report_07.pdf.
Accessed 23rd October 2009.

27 ‘R.A.A.P. pays out over €3 million in 2005 to Artists’. RAAP News.
http://www.raap.ie/musician-performing-rights-news.html.
Accessed 23rd October 2009.

28 ‘U2 move their rock empire out of Ireland’, Irish Independent, 6th Aug. 2006. Online source at:
http://www.independent.ie/national-news/u2-move-their-rock-empire-out-of-ireland-133364.html.
Accessed 23rd October 2009.

29 ‘Arctic Monkeys make chart history’. BBC News 26.01.2006.
http://news.bbc.co.uk/1/hi/entertainment/4660394.stm.
Accessed 22nd October 2009.

30 Papadopoulos , T., and Rolston, C.P., ‘Stochastic Demand and Sound Recording Pricing’. The MEIEA Journal, Volume 6, Number 1, 2006.
http://www.meiea.org/Journals.html.
Accessed 24th October 2009.

31 ‘Record labels street teams now online’. CNET News.
http://news.cnet.com/2100-1027_3-5486129.html.
Accessed 23rd October 2009.

32 ‘Scene Unheard gets into the charts... just’.
http://www.irishunsigned.com/fullstory.php?id=37.
Accessed 7th October 2009

33 Kotler, P., and Roberto, E., ‘Social Marketing: Strategies for Changing Public Behavior’, The Free Press, 1989.

34 Kotler, P. ‘Social Marketing: Strategies for Changing Public Behavior’. Sage 2002

35 Pine, R. ‘Music and broadcasting in Ireland’, Four Courts, 2005

36 Joint National Listenership Research, August 20, 2009.
http://www.bci.ie/documents/jnlr_july08_jun09.pdf.
Accessed October 7th, 2009

37 ‘Radio and Records’, a subsidiary of ‘Billboard’.
http://www.billboard.biz/bbbiz/radioandrecords.html.
Accessed 25th October 2009.

38 ‘Recorded music - driver of a US$100 billion economic sector’. IFPI news, June 6th 2006.
http://www.ifpi.org/content/section_news/20060622.html.
Accessed 26th October 2009.

39 Clancy, P. and Twomey, M., ‘Clusters in Ireland: The Irish Popular Music Industry’. National and Economic Social Council, Research Series Paper No. 2, Dublin 2007.

40 ‘The threat becomes the opportunity’, speech by John Kennedy, CEO and Chairman IFPI, MIDEMNET Keynote speech, 22 January 2005.
http://www.ifpi.org/content/section_views/view013.html.
Accessed 11th October 2009.

41 ‘Counting the beat : culture, democracy and broadcasting. A selection of papers and issues from Counting the cultural beat: a symposium of broadcasting policy in New Zealand held in Wellington on 31 August 1999’. Institute of Policy Studies, Victoria University of Wellington, NZ. 1999. Quote from Keynote Speech by Michael D. Higgins.

42 Zhang, F. and Xie, D. ‘Chinese Copyright Protection Has Storied History, Strong Future’. White paper published by Sourcetrix Corporation.

43 Zhang, N. ‘How to Reach the Cross-Over Point of Maskus Curve in the Chinese Intellectual Property Law Enforcement’. Socio-Legal studies Association Annual Conference, UK, 2009

44 The ISRC standard.
http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=23401&ICS1=1&ICS2=140&ICS3=20.
Accessed 11th October 2009.

45 ‘Guidelines for ToBI Labelling’, The Ohio State University Research Foundation.
http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/ToBI/ToBI.1.html.
Accessed 10th October 2009

46 Image source: ‘CyclingMind.com’.
http://www.cyclingmind.com/HeartRateMonitors.html
Accessed 28th October 2009.

47 Fourier, J. ‘The Analytical Theory of Heat’. Cambridge University Press (reissued in 2009)

48 You, J., Yong Jeon, J., ‘Sound-masking technique for combined noise exposure in open public spaces’. 9th International Congress on Noise as a Public Health Problem (ICBEN). USA, 2008.

49 Ibrahim, K.F., ‘Newnes Guide to Television and Video Technology’, pp. 77. Elsevier, 2007.

50 ‘Acoustics - normal equal loudness level contours’. ISO Standard 226:2003
http://www.iso.org/iso/catalogue_detail.htm?csnumber=34222
Accessed 28th October 2009.

51 Equal loudness contours image reproduced from http://en.wikipedia.org/wiki/File:Lindos1.svg
Accessed 28th October 2009.

52 ‘Steganography and Steganalysis: An Overview’, SANS Institute InfoSec Reading Room.
http://www.sans.org/reading_room/whitepapers/stenganography/steganography_and_steganalysis_an_overview_553.
Accessed 23rd October 2009.

53 Petitcolas, F.A.P., Anderson R.J., Kuhn, M.G. ‘Information Hiding: A survey’. Proceedings of the IEEE (special issue). 1999.
http://www.petitcolas.net/fabien/publications/ieee99-infohiding.pdf
Accessed 3rd October 2009.

54 McClellan, J., Schafer, R., Yoder. M. ‘DSP first : a multimedia approach’. Prentice Hall, 1999.

55 ‘MPEG standards: coded representation of audio and video’.
http://www.iso.org/iso/prods-services/popstds/mpeg.html.
Accessed 24th October 2009.

56 Rumsey, F., ‘The audio workstation handbook’, Oxford, Boston, Focal Press, 1996.

57 ‘The International Standard Recording Code’. IFPI handbook (Section 3.2. ISRC Principles).
http://www.ifpi.org/content/section_resources/isrc_handbook.html#Heading117
Accessed 11th October 2009.

58 TeleTrax global television tracking and reporting.
http://www.teletrax.tv
Accessed 11th October 2009.

59 Wold, E., Blum, T., Keislar, D., Wheaton, J. ‘Content-Based Classification, Search, and Retrieval of Audio’, IEEE MultiMedia, vol. 3, no. 3, pp. 27-36. 1996.

60 Picone, J. W., ‘Signal Modeling Techniques in Speech Recognition’, Proceedings of the IEEE, vol. 81, no. 9, pp. 1215-1247. 1993.

61 Kelly, A., Ní Chasaide, A., Berthelsen, H., Campbell, C., Gobl, C. ‘Corpus Design Techniques for Irish Speech Synthesis’. Proceedings of the China-Ireland International Conference on Information and Communications Technologies, NUI Maynooth, Ireland, 2009.

62 ‘A&M Records Inc’ versus ‘Napster, Inc’. United States Court of Appeals for the Ninth Circuit. Decision February 12th 2001. Harvard law School online.
http://cyber.law.harvard.edu/~wseltzer/napster.html.
Accessed 24th October 2009.
63 ‘International & Swedish Music industry welcomes Pirate Bay judgment’. International Federation of the Phonographic Industry news release**,** 17th April 2009.

http://www.ifpi.org/content/section_news/20090417.html.
Accessed 24th October 2009.

64 Cano, P., Batlle, E., Kalker, T., and Haitsma, J., ‘A Review of Audio Fingerprinting’, The Journal of VLSI Signal Processing, Volume 41, Number 3. 2005.

65 ‘Apple DRM is illegal in Norway, says Ombudsman’. Published online in ‘OUT-LAW News’, Pinsent Mason, 24/01/2007.
http://www.out-law.com/page-7691.
Accessed 24th October 2009.

66 Kirovski, D., and Malvar, R., ‘Robust Covert Communication Over a Public Audio Channel Using Spread Spectrum’, Proceedings of the 4th International Workshop on Information Hiding (IH 2001). USA, 2001.

67 ‘MIT's Introduction to Algorithms, Lecture 07’, Online M.I.T. video lecture.
http://video.google.com/videoplay?docid=-727485696209877198.
Accessed 24th October 2009.

68 Peterson, W.W. and Brown, D.T. ‘Cyclic codes for error detection’, Proceedings of the Institute of Radio Engineers, January 1961.
http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4066263.
Accessed 24th October 2009.

69 Shazam website, ‘About Us, Background’ section.
http://www.shazam.com/music/web/pages/background.html.
Accessed 24th October 2009.

70 Midomi website, ‘About Us’ section:
http://www.midomi.com/index.php?action=main.about_us.
Accessed 24th October 2009

71 Nielsen Music Control.
http://www.nielsenmusiccontrol.com.
Accessed 24th October 2009.

72 ‘Watermarking world’
http://www.watermarkingworld.org/faq.html
Accessed 24th October 2009.

73 Burke, A., Montgomery, C., ‘You Say You Want a Revolution? A Case Study of MP3.com’. International Journal of Entrepreneurship Education (page 111). Senate Hall Academic. 2002.

74 Bernsee, Stephen. ‘The DFT à Pied: Mastering The Fourier Transform in One Day’.
http://www.dspdimension.com/admin/dft-a-pied.
Accessed 9th October 2009.

75 Prof. Diana Deutsch, Professor of Psychology. University of California, San Diego.
http://deutsch.ucsd.edu/psychology/deutsch_research1.php.
Accessed 9th October 2009.

76 J. M. Blackledge, ‘Digital Watermarking and Self-Authentication using Chirp Coding’, ISAST Transactions on Electronics and Signal Processing, Vol. 1, No. 1, 2007

77 Cox, I.J., Miller, M. and Bloom, J. ‘Digital watermarking and steganography’, Morgan Kaufmann Publishers. 2008.

78 ‘Principles of Digital watermarking’. InfoSecurity Magazine, August 2009.
http://fanaticmedia.com/infosecurity/archive/Aug%2009/Digital%20Watermarking.htm.
Accessed 24th October 2009.

79 Bender, W., Gruhl, D., Morimoto, N., Lu, A. 'Techniques for data hiding'. IBM Systems Journal, Vol 35, No. 3 & 4, 1996

80 Acharya, R.U., Suri, J.S., Spaan, J.A.E. and Krishnan, S .M. 'Advances in Cardiac Signal Processing'. Springer Berlin Heidelberg, 2007.

81 Gopalan, K. 'Audio Steganography Using Bit Modification', Proceedings of the 2003 IEEE International Conference on Acoustics, Speech & Signal Processing. China. 2003 (cancelled).

82 Yang, H., Sun, X. and Sun, G., 'A semi-fragile watermarking algorithm using adaptive least significant bit substitution'. Information Technology Journal, 2009.

83 Bassia, P., Nikolaidis, N. 'Robust audio watermarking in the time domain', IEEE Transactions on Multimedia, Volume 3, Issue 2, Jun 2001 Page(s):232 - 241

84 Ko, B.-S., Nishimura, R., and Suzuki, Y., 'Time-spread echo method for digital audio watermarking using pn sequences'. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, 2002, pp.2001–2004.

85 Kim, H.J. and Choi , Y.H. 'A Novel Echo-Hiding Scheme With Backward and Forward Kernels', IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, NO. 8, August 2003

86 Gopalan, K. & Wenndt, S. 'Audio Steganography for Covert Data Transmission by Imperceptible Tone Insertion', Proceedings of the IASTED International Conference on Communication Systems And Applications, Canada, 2004.

87 Lipshitz, S.P., Pocock, M., and Vanderkooy, J. 'On the Audibility of Midrange Phase Distortion in Audio Systems'. Journal of the Audio Engineering Society, volume. 30. 1982.

88 Koya, D. 'Aural Phase distortion Detection'. MSc Thesis in Music Engineering Technology submitted to the University of Miami, USA. 2000.

89 Yardimci, Y., Cetin, A.E. and Ansari, R. 'Data Hiding in Speech Using Phase Coding'. 5th European Conference on Speech, Communication and Technology, Greece. 1997

90 Takahashi, A., Nishimura, R., Suzuki, Y., 'Multiple Watermarks for Stereo Audio Signals Using Phase-Modulation Techniques', IEEE Transactions on Signal Processing, Volume 53, No. 2, February 2005

91 Morris, D. 'An Intuitive but Not-All-That-Mathematically-Sound Explanation of the Fourier Transform' (online tutorial)
http://techhouse.brown.edu/~dmorris/projects/tutorials/fourier_tutorial.pdf
Accessed 27th October 2009.

92 'FFT window functions: Limits on FFT analysis', Bores Signal Processing.
http://www.bores.com/courses/advanced/windows/files/windows.pdf.
Accessed 25th October 2009.

93 Wu, C.P., Su, P.C., and Kuo, CCJ., 'Robust Audio Watermarking for Copyright Protection', 44th Annual Meeting, Advanced Signal Processing Algorithms, Architectures and Implementations, USA, 1999.

94 Tan, W., Yang, S., Chen, Y. and Zhou, J. 'Research on DFT Domain Digital Audio Watermarking Algorithm Based on Quantization', First International Workshop on Education Technology and Computer Science, China. 2009.

95 Singh, J., Garg, P. and Nath, A., 'Audio Watermarking Using Spectral Modifications', International Journal of Signal Processing, Volume 5, Number 4. 2009

96 Arya, A.K., lather, J.S., Dewan, L., 'A Tool for Audio Quality Evaluation Under Hostile Environment', International Journal of Computer and Information Science and Engineering, 2008

97 Yan, Y., Rong, H. and Mintao, X. 'A Novel Audio Watermarking Algorithm for Copyright Protection Based on DCT Domain', Second International Symposium on Electronic Commerce and Security, China, 2009.

98 Liu, J. and Lu, Z. 'A Multipurpose Audio Watermarking Algorithm Based on Vector Quantization in DCT Domain', World Academy of Science, Volume 55, July 2009.

99 Biography of George Zweig at Wolfram Research.
http://scienceworld.wolfram.com/biography/Zweig.html.
Accessed 25th October 2009.

100 Lang, W.C and Forinash, K. 'Time-frequency analysis with the continuous wavelet transform', American Journal of Physics, Volume 66, Issue 9, pp. 794-797. 1998.

101 Lei, W., Xue, X., Li, X. and Lu, P. 'A novel feature-based robust audio watermarking for Copyright protection', Proceedings of the International Conference on Information Technology: Coding and Computing [Computers and Communications], USA. 2003

102 Delforouzi, A., and Pooyan, M., 'Adaptive Digital Audio Steganography Based on Integer Wavelet Transform'. Proceedings of the 3rd International Conference on International Information Hiding and Multimedia Signal Processing, Vol. 02 (2007).

103 H. Malvar, 'A Modulated Complex Lapped Transform and Its Applications to Audio Processing'. Proceedings of the International Conference on Acoustics, Speech and Signal Processing. USA. 1999

104 Shenghong , L., Cui, L., Choi, J. and Cui, X. 'An Audio Copyright Protection Schemes Based on SMM in Cepstrum Domain', published in 'Structural, Syntactic, and Statistical Pattern Recognition', Springer Berlin / Heidelberg. 2006.

105 Sterling, C.H. 'Military communications: from ancient times to the 21st century', pp 427. ABC-CLIO, 2008
http://books.google.ie/books?id=RBC2nY1rp5MC.
Accessed 25th October 2009.

106 Peterson, G.L. 'Nikola Tesla - A Man of Comprehensive Solutions'. 21st Century Books.
http://www.tfcbooks.com/articles/ntbio.htm.
Accessed 25th October 2009.

107 Schilling, D.L., Pickholtz, R.L. and Milstein, L.B. 'Spread spectrum goes commercial', IEEE Spectrum, Volume 27, Issue 8. 1990.

108 Kirovski, D., and Malvar, H.S., 'A dual watermarking and fingerprinting system'. Proceedings of the IEEE International Symposium on Information Theory, Switzerland. 2002.

109 'Apple Reports Fourth Quarter Results'. Apple press release, October 19, 2009.
http://www.apple.com/pr/library/2009/10/19results.html.
Accessed 25th October 2009.

110 Kim, H., J., Xiang, S., Yeo, I.K., Maitra, S., 'Robustness Analysis of Patchwork Watermarking Schemes,'. Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarking. Information Science Reference. 2007

111 Arnold, M., 'Audio Watermarking: Features, Applications and Algorithms'. IEEE International Conference on Multimedia and Expo. USA, 2000

112 PhD Thesis of Shijun Xiang (awarded as the 2007 Guangdong Excellent) 'Robust Audio Watermarking Algorithms', Sun Yat-Sen University, Guangzhou, China, 2006

113 Powell, M.J.D. ‘Approximation theory and methods’, Cambridge University Press. 1996. http://www.worldcat.org/oclc/174425607.
Accessed 25th October 2009.

114 Fallahpour, M., and Megias, D., ‘High capacity audio watermarking using FFT amplitude interpolation’, IEICE Electron. Express, Volume 6, No. 14, pp.1057-1063, June 2009

115 Kaabneh, K.A. and Youssef, A. ‘Muteness Based Audio Watermarking Technique’, Proceedings of the International Conference on Distributed Computing Systems. USA. 2001.

116 Liu, Z. and Inoue, A., ‘Audio Watermarking Techniques Using Sinusoidal Patterns Based on Pseudorandom Sequences’. IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, No. 8, August 2003.

117 Erküçük , S., Krishnan, S., and Zeytino glu, M., ‘A robust Audio watermark Representation Based on Linear Chirps’, IEEE Transactions on Multimedia, Vol. 8, No. 5, October 2006.

118 Megias, D., Herrera-Joancomarti, J., Minguillon, J., ‘A Robust Audio Watermarking Scheme Based on MPEG 1 Layer 3 Compression’, Seventh IFIP Conference on Communications & Multimedia Security. France. 2003

119 Siebenhaar, F., Neubauer, C. and Herre, J. ‘Combined Compression/Watermarking for Audio Signals’. Audio Engineering Society 110th Convention. Holland. 2001.

120 Fujimoto, R., Iwaki, M., and Kiryu, T. 2006. ‘A Method of High Bit-Rate Data Hiding in Music Using Spline Interpolation’. Proceedings of the 2006 international Conference on intelligent information Hiding and Multimedia. USA. 2006.

121 Fallahpour, M., and Megias, D., ‘High capacity audio watermarking using FFT amplitude interpolation’, IEICE Electronics Express, Vol. 6, No. 14, pp. 1057 - 1063, June 2009.

122 Tachibana, R., ‘Audio Watermarking for Live Performance’, Proceedings of Security and Watermarking of Multimedia Contents. USA, 2003.

123 Mooney, A., and Keating, J., ‘Generation and Detection of Watermarks Derived From Chaotic Functions’, Proceedings of Opto-Ireland, Dublin, Ireland. 2005

124 ‘Teen trouble beaten by mosquito'. BBC News, 31 March 2006.
http://news.bbc.co.uk/2/hi/uk_news/england/derbyshire/4866730.stm.
Accessed 25th October 2009.

125 Dittmann, J et al ‘Theoretical framework for a practical evaluation and comparison of audio watermarking schemes in the triangle of robustness, transparency and capacity’, Transaction on Data Hiding and Multimedia Security I; pp. 1-40. Springer. 2006

126 Healy, R. and Timoney, J., ‘300 years of copyright: have we gone full circle? On the use of technology to address limitations in distributing public performance broadcast royalties’, Socio-Legal studies Association Annual Conference. UK. 2009
http://www.irishunsigned.com/papers/slsa09.pdf
Accessed 23rd October 2009.

127 Gopalan, K., et al, ‘Covert Speech Communication Via Cover Speech By Tone Insertion’, Proceedings of the 2003 IEEE Aerospace Conference. USA. 2003.

128 Schenker, L, ‘Pushbutton calling with a two-group voice-frequency code’, The Bell system technical journal 39 (1): pp. 235-255. 1960.

129 Ifeachor, E. C. and Jervis, B. W. ‘Digital Signal Processing: a Practical Approach’. 2nd. Pearson Education, 2002.

130 Healy, R., and Timoney, J., ‘An Application of Digital Audio Watermarking to Broadcast Monitoring’, Paper accepted for publication at the China-Ireland International Conference on Information and Communications Technologies, China, 2008. On file at NUI Maynooth, Ireland.
http://www.irishunsigned.com/papers/ciict08.pdf
Accessed 20th October 2009.

131 Orfanidis, S. ‘Introduction to signal processing’, Prentice Hall, 1995

132 Explanation of ‘blind’ and ‘double blind’ testing
http://www.hydrogenaudio.org/forums/index.php?showtopic=16295.
Accessed 25th October 2009.

133 Philips SBCHC8440 user manual with specifications
http://www.p4c.philips.com/files/s/sbchc8440_05/sbchc8440_05_dfu_eng.pdf.
Accessed 25th October 2009.

134 Wang, J. and Healy, R., ‘Digital Audio Watermarking by Magnitude Modification of Frequency Components Using the CSPE Algorithm’. Proceedings of the China-Ireland International Conference on Information and Communications Technologies, NUI Maynooth, Ireland, 2009.
http://www.irishunsigned.com/papers/ciict09.pdf
Accessed 20th October 2009.

135 Garcia, R.A., and Short, K.M., ‘Signal Analysis Using the Complex Spectral Phase Evolution (CSPE) Method’, Audio Engineering Society 120th Convention, France, 2006.

136 Wang, J., Timoney, J. and Hodgkinson, M. ‘Using Apodization to improve the performance of the Complex Spectral Phase Estimation (CSPE) Algorithm’. Proceedings of the China-Ireland International Conference on Information and Communications Technologies, NUI Maynooth, Ireland, 2009.

137 MusicTrace (a spin-off of the Fraunhofer Institute for Integrate Circuits)
http://www.musictrace.de.
Accessed 26th October 2009.

138 PPI Ltd. license tariffs.
http://www.ppiltd.com/tariffs.htm.
Accessed 26th October 2009.

139 PRS UK license tariffs.
http://www.prsformusic.com/users/broadcastandonline/Radio/Pages/default.aspx.
Accessed 26th October 2009.

140 Healy, R. and Timoney, J. ‘Digital Audio Watermarking with Semi-Blind Detection for In-Car Music content Identification’. Proceedings of the AES 36th International Conference: Automotive Audio—Sound in Motion. USA, 2009.
http://www.irishunsigned.com/papers/aes36
Accessed 24th October 2009.

141 ‘Shazam surpasses 50 million global users and attracts investment from the backers of Google and Amazon’, Shazam news release.
http://www.shazam.com/music/web/newsdetail.html?nid=NEWS089.
Accessed 25th October 2009