

エコー拡散透かし手法における音質と埋め込み容量に関する検討

高 秉燮[†] 西村 竜一[†] 鈴木 陽一[†]

[†] 東北大学, 電気通信研究所/大学院情報科学研究科 〒 980-8577 宮城県仙台市青葉区片平 2-1-1

E-mail: †{kobs, ryou, yoh}@ais.riec.tohoku.ac.jp

あらまし 電子透かし方法として, エコーハイディング, スペクトラム拡散, マスキング, 位相変調などの手法が提案されてきた。そのうち, エコーハイディングは多くの利点がある反面, 攻撃者にとって検出が容易であるという欠点がある。我々は, この欠点を改善するために PN 系列を用いてエコーを時間軸で拡散させ, 情報を埋め込む方法を提案した。本報告では, エコー拡散手法の検出過程で用いる DFT の窓長と検出性能の関係から, 本手法における埋め込み容量とリサンプリングや圧縮などの信号処理に対する耐性について検討を行う。また, 一般に透かし情報を埋め込むことによって音質が劣化してしまうことから, エコー拡散手法による音質の劣化についても聴取実験を通して検討を行う。

キーワード 電子透かし, エコー拡散, 耐攻撃性, 音質

A study on sound quality and embedding capacity of Time-spread Echo Method for Digital Audio Watermarking

Byeong-Seob KO[†], Ryouichi NISHIMURA[†], and Yōiti SUZUKI[†]

[†] Research Institute of Electrical Communication/Graduate School of Information Sciences, Tohoku University 2-1-1 Katahira, Aoba-ku, Sendai, Miyagi 980-8577

E-mail: †{kobs, ryou, yoh}@ais.riec.tohoku.ac.jp

Abstract Several methods for digital audio watermarking have been proposed, such as echo hiding, spread spectrum, masking, and phase modulation. Among these methods, echo hiding provides many advantages. However, there is a weak with this method from view point from protecting copyrights, because its decoding process is very lenient. To cope with this weak point of the echo hiding, we proposed a time-spread echo method using PN sequences that play a role of a secret key in decoding the embedded information. In this paper, we investigate the embedding capacity, the robustness against typical signal transformations, and the sound quality of the proposed method.

Key words Digital watermarking, Time-spread echo, robustness, inaudibility

1. はじめに

近年, 情報通信技術の発展により, デジタルメディア (動画像, 静止画像, 音楽など) の流通が拡大している。これに伴って, デジタルメディアの不法な利用及び流通が増加している。そのため, デジタルメディアを対象とした著作権の保護技術が必要とされている。このような保護方法としてコピー防止 (Copy Protection) とコピー検出 (Copy Detection) がある。しかし, コピー防止は一旦解読されるとその効力がなくなってしまう。また, CD や DVD のように, たとえデジタルコピー防止機能がついていても, DA/AD 変換 (例えば, スピーカによる再生とマイクロホンによる録音) によりコピーすることができる。

そこで, 最近では, コピー防止ばかりでなく, コピー検出も行うことができる電子透かし技術に注目が集まっている。電子透かしとは, 使用者に知覚できないようにデジタルメディアの中へコピー防止とコピー検出のための識別記号情報を埋め込む技術である。したがって, 疑わしいメディアを発見した時, その情報をメディアから取り出し, そのメディアに対する所有者の使用権を確認, あるいは, コピー防止ができる技術である [1], [2]。

音響信号のための電子透かし手法として, エコーハイディング [3]~[5], スペクトラム拡散法 [5], [6], マスキングに基づいた手法 [7]~[10], 位相変調による手法 [5], [11] などが提案されてきた。そのうち, エコーハイディングは多くの利点がある反面, 攻撃者にとって検出が容易であるという欠点がある。我々は,

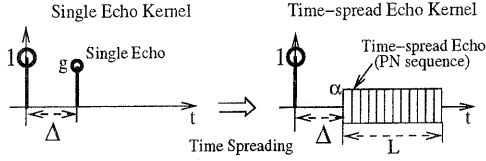


図1 エコーカーネルとPN系列を用いたエコー拡散カーネル

この欠点を改善するためにPN系列を用いてエコーを時間軸で拡散させ、情報を埋め込む方法であるエコー拡散透かし手法を提案した[12].

本研究では、まず、エコー拡散手法の各種パラメータ設定における検出性能について述べ、本手法による透かし情報の埋め込みデータ量について検討する。また、圧縮やリサンプリングといった攻撃に対する耐性と音質の劣化について聴取実験を通して検討を行う。

2. エコー拡散透かし手法

2.1 PN 系列を用いたエコー拡散カーネル

図1に、従来のエコーハイディングにおけるエコーカーネルと、PN系列を用いたエコー拡散カーネルを示す。エコー拡散カーネルは、エコーカーネルのエコーに対応する成分をPN系列により時間領域で、式(1)の関係を満たすようにして拡散することで得られる。ここで、使用されたPN系列は、著作権などの透かし情報を音楽から検出するための秘密鍵になる。

$$g^2 \approx \alpha^2 \cdot L \quad (1)$$

ここで、 g はエコーの振幅を調整する利得であり、 α ($0 < \alpha \ll 1$)と L はそれぞれPN系列の利得と長さである。

エコー拡散カーネルは、式(2)のように表すことができる。

$$k(n) = \delta(n) + \alpha \cdot P(n - \Delta) \quad (2)$$

ここで、 $P(n)$ は振幅が ± 1 のオリジナルPN系列であり、 $\delta(n)$ はディラックのデルタ関数である。また、 Δ は時間遅延であり、この時間遅延の値によって異なる情報がホスト信号に埋め込まれる。

エコー拡散カーネルは、 L が長くなると α が十分小さくなり、その結果、室内の残響に似た特性を有している[13]。そのため、ホスト信号が音楽の場合、このカーネルによるエコーは、従来のエコーカーネルより自然感が増し、聴取者に検知しにくくなると期待される。また、たとえ、検知されたとしても、音質の劣化は少なくなることが予想される。

2.2 検出過程

エコー拡散に基づく電子透かし手法の利点のひとつは、音楽から埋め込まれた情報を検出する際、エコーハイディングの検出方法であるオートケプストラム(ケプストラムの自己相関)[3]だけでは困難になることである。なぜなら、エコー成分がPN系列によって時間軸で拡散されたため、各エコー成分の振幅が極めて小さくなり、透かしが埋め込まれているのかの判断が困難になるからである。つまり、図1の Δ が決定できな

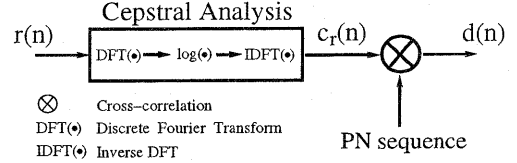


図2 エコー拡散手法の検出過程のブロック図

くなる。図2に、エコー拡散手法の検出過程のブロック図を示す。図2で、透かしが埋め込まれた受信信号は、チャンネルノイズを無視すると式(3)のように表すことができる。

$$r(n) = s(n) * k(n) \quad (3)$$

ここで、 $s(n)$ はホスト信号、 $k(n)$ はカーネル信号、 $*$ は線形畳み込みである。

まず、図2の中のケプストラム分析[3],[15]により、式(3)の $r(n)$ は、式(4)のようになる。

$$c_r(n) = \hat{s}(n) + \hat{k}(n) \quad (4)$$

ここで、 $\hat{s}(n) = \text{IDFT}(\log(\text{DFT}(s(n))))$ 、 $\hat{k}(n) = \text{IDFT}(\log(\text{DFT}(k(n))))$ である。また、 $\text{DFT}(\cdot)$ と $\text{IDFT}(\cdot)$ はそれぞれ離散フーリエ変換と逆離散フーリエ変換である。

検出信号は、式(4)と埋め込み過程で使用した元のPN系列との相互相関をとることによって得ることができる。すなわち、検出信号は、式(5)によって求まる。

$$d(n) = \text{xcorr}(c_r(n), P(n)) \quad (5)$$

ここで、 $\text{xcorr}(\cdot)$ は相互相関である。最後に、埋め込まれた情報は、従来のエコーハイディングの場合と同様に、 $d(n)$ のピークの位置を調べることによって検出される。

図3は、従来のエコーハイディングとエコー拡散手法による検出信号の例を示す。上から、エコーハイディングにより埋め込まれた信号に対する検出過程(オートケプストラム)による検出信号、エコー拡散により埋め込まれた信号に対してエコーハイディングの検出過程、すなわち、オートケプストラムにより得られた検出信号、最後に、エコー拡散手法により埋め込まれた信号を図2の検出過程によって検出した時の検出信号を表す。ここで、エコーの利得 g は、 $g = \sqrt{\alpha^2 \cdot L}$ になるように設定し、図の矢印は、透かし情報に相当するピークが出る時間遅延の位置を示す。これら検出信号の例から、エコーハイディングの検出過程で用いられるオートケプストラムでは、エコー拡散手法の埋め込みに対して検出が困難であることが分かる。したがって、エコー拡散手法は、検出過程に秘密鍵であるPN系列が必要になり、従来のエコーハイディングより秘匿性が高くなるという利点がある。

3. 検出性能と埋め込み容量

エコー拡散手法では、PN系列の利得 α や長さ L などのパラメータが検出性能に影響を及ぼす。表1は、検出性能を検討するために用いたパラメータとその設定条件を示す。表1で、ホ

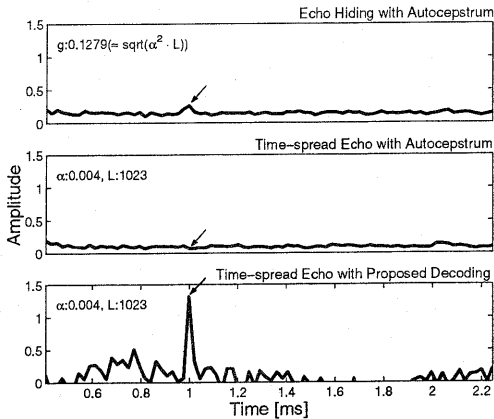


図3 従来のエコーハイドイングとエコー拡散手法による検出信号の例

表1 検出性能を検討するために用いたパラメータとその設定条件

パラメータ	条件
$P(n)$ の利得 (α)	0.002, 0.004, ..., 0.02. 0.002 刻み
$P(n)$ の長さ (L)	255, 511, 1023, 2047, 4095
DFT の長さ (F)	512, 1024, 2048, 4096, 8192
ビットレート (bps)	1, 4
ホスト信号の種類	5 (S01, S02, ..., S05)

スト信号は、標準化周波数 44.1 kHz, 16 ビットの量子化ビットで録音されたステレオ音楽信号である。また, S01 と S02 は、それぞれアメリカと日本のポップス, S03 はサプラノとピアノ演奏, S04 はピアノ演奏, S05 は、バリオリンやチェロなどの弦楽器によるフィランドのフォーク音楽である。

検出性能の指標として、信号検出理論の d' を使用する [16], [17]。 d' が大きくなればなるほど高い検出性能を表す。

図4は、表1に記した条件下における検出を行い、これらのパラメータのうち α と L 以外については平均を取った検出結果を示したものである。図4で、横軸はPN系列の利得 α 、縦軸は d' であり、黒いバーから白いバーまで色が薄くなる順番に従ってPN系列の長さ L が255, 511, 1023, 2047, 4096の結果である。図4から、今回実験を行った範囲では、検出性能はPN系列の利得 α に関して単調増加の関係にあることが分かる。しかし、2047, 4095 の場合には、PN系列の長さ L が必ずしもそのような傾向は確認されず、頭打ちの現象が見られた。

一般に透かしのパワーが大きくなるほど検出結果が良くなる。そのため、 α が大きくなるに従って、検出性能が向上することが期待される。一方、PN系列の長さが長くなることによって透かしのパワーが大きくなる。したがって、この場合も L の増加に伴って d' が上昇することが予想されるが、図4の結果は必ずしもそうではない。その原因として考えられることは、検出過程のDFT長 F の影響である。一般にスペクトラム拡散手法の場合、検出過程で分析長、すなわちPN系列の長さが長ければ長いほど検出性能が高くなる。これは、分析長が長く

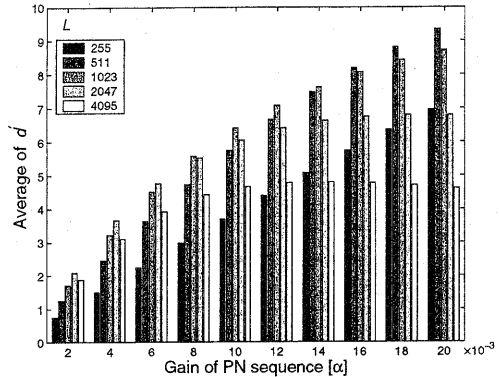


図4 パラメータ L と α に対する検出性能

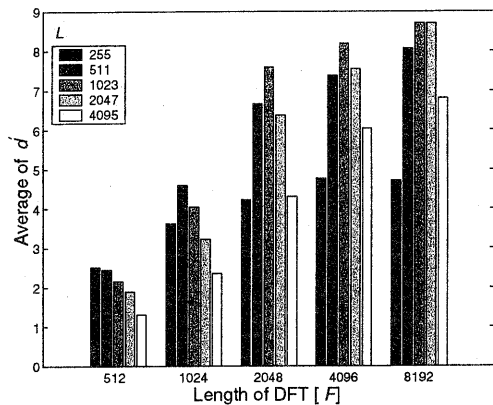


図5 DFT 長 F と PN 系列の長さ L をパラメータとした時の d'

なりと秘密鍵であるPN系列との相関のピーク値が高くなるからである [14]。したがって、エコー拡散手法に対しても拡散されたエコーに対して逆拡散する時、高い検出精度を得るためには、少なくとも拡散過程で使用したPN系列の長さに相当するDFT長が必要になると考えられる。

検出過程のDFT長 F は、埋め込む透かし情報の埋め込み容量に密接な関係がある。例えば、DFT長 F が8192で標準化周波数が44.1 kHzの場合、埋め込むデータが“0”と“1”のバイナリデータとすると埋め込み容量(ビットレート)は、約5 bpsとなる。そこで、 F と L のパラメータを変化させた時の検出性能を調べ、埋め込み容量について検討を行う。

図5は、DFT長 F とPN系列の長さ L をパラメータとした時の d' を示したものである。図5から、 L が同一であってもDFT長が長ければ長いほど検出性能が良くなることが分かる。また、DFT長が512の時は L が255, 1024の時は511, 2048の時は1023のように、DFT長 F がPN系列の長さ L の2倍程度になるときに検出性能が良くなることが分かる。しかし、長いDFT長を適用するのは、処理時間と透かしの埋め込み容量に対して望ましくない。そこで、検出過程に必要なDFT長は、PN系列の長さによって異なるので検出性能とPN系列の

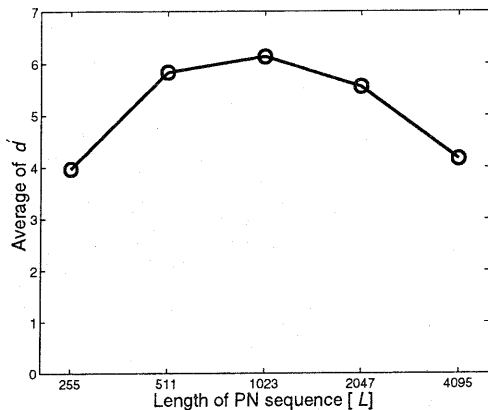


図 6 PN 系列の長さ L に対する検出性能

長さとの関係調べることによってエコー拡散手法における適切な埋め込み容量を検討する。

図 6 は、表 1 のパラメータと設定条件から得られた検出性能を PN 系列の長さ L に対して平均を取った結果を示す。この結果から、PN 系列の長さが 1023 の場合、検出結果が最高になることが分かる。ここで、満たすべき要件はアプリケーションに強く依存するが、PN 系列の長さが一般に短い場合には、生成される PN 系列、すなわち、秘密鍵の数が少ない。そのため、電子透かしの秘匿性の観点から見るとあまり望ましくない。また、 $L = 4095$ の場合は、先述の検討から 8192 の DFT 長が必要になり、多くの計算時間が必要になる。そこで、図 5 と 6 から、適切な DFT 長 F は、2048 と 4096 であると判断される。

したがって、エコー拡散手法による透かし情報の埋め込み容量は、標準化周波数が F_s の時、 F_s/F になり、 $F_s = 44.1$ kHz で $F = 2048$ の時、約 20 bps となる。また、エコー拡散手法は、音質の劣化が少ないので、PN 系列の「cyclic shifting」[10] のような方法によって多重ビット埋め込みが可能になり、埋め込み容量を高めることができる。

4. 信号処理に対する耐性

デジタル音楽信号は、計算機による編集、圧縮、DA (Digital to Analog) と AD (Analog to Digital) による再録音など様々な信号処理を受けることが多い。そのため、電子透かしは、そのような信号変化に対して耐性がなければならない。本節では、一般的な信号処理に対するエコー拡散手法の耐性に関する検討を行う。信号処理に対する耐性の指標として前節と同様に d' を用いる。

表 2 に実験で用いた信号処理の種類と条件を示す。ここで、各信号処理の条件は、STEP2000 [18] で採用された条件に基づいて設定した。この表のうち、「Multi-Attack」は、表の中の全ての信号処理を加えた場合である。また、検討に用いたパラメータの条件は、前章の図 6 で検出性能が一番高かった、 $L = 1023$ 、 $F = 2048$ とした。また、 α は 0.006 を選択した。図 7 は、処理に対するエコー拡散手法の耐性を表す。ここで、

表 2 実験で用いた信号処理と条件

信号処理		条件
Resampling	(RES)	16 kHz
Requantization	(AM8)	8 bits
MPEG 1 Layer III	(MP3)	128 kbps
MPEG 2 AAC	(AAC)	128 kbps
DAC and ADC	(DAD)	44.1 kHz, 16 bits
Multi-Attack	(ALL)	

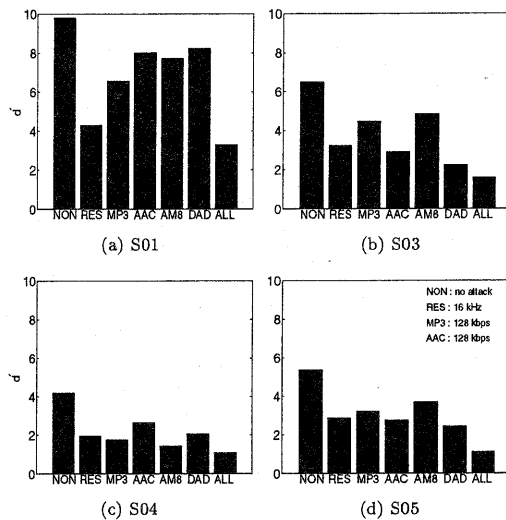
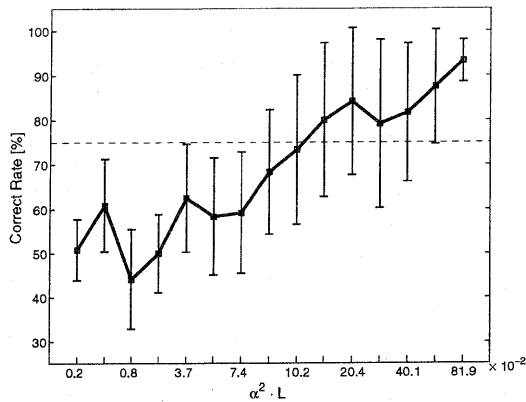


図 7 信号処理に対するエコー拡散手法の耐性。ここで、ホスト信号は、S01, S03, S04, S05 であり、「NON」は信号変化を加えなかった場合である。

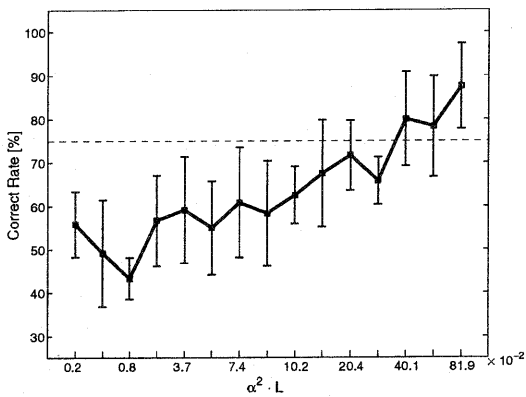
横軸は信号処理の種類、縦軸は d' である。また、「NON」は、信号変化を加えなかった場合である。すなわち、信号処理がなかった場合と加えた場合の検出性能を比較することによって耐性を検討する。「NON」の場合とできるだけ差が小さくなるほど、耐性が高いと考えられる。図 7 から、「ALL」の場合を除いて他の信号処理に対する d' は「NON」の場合の半分以下であり、「ALL」の場合でも d' が 1 以上になることが分かる。

図 7 から分かるように、信号処理を加えた場合、 d' が低下する。その原因としては、用いた PN 系列の周波数特性が白色雑音の場合と似た特性を持っているため、リサンプリングや圧縮のような処理過程で多くの周波数成分が無くなるからだと考えられる。また、再量子化の場合は、再量子化により元信号の波形のダイナミックレンジが狭くなり、エコー拡散カーネルによる振幅変化が少ないものなくなってしまう。そのため、検出過程で逆拡散する際に透かし情報として埋め込まれた PN 系列の情報が低下し、検出性能が低下したと考えられる。

図 7 の結果から、エコー拡散手法は、信号処理に対してもある程度耐性があると考えられる。また、元の PN 系列をそのまま使用するより、ホスト信号に似た周波数特性を持つように工夫をした PN 系列を適用することによって耐性が高くなるとい



(a) S01



(b) S04

図 8 聴取実験の結果. (a) S01, (b) S04

う報告 [19] がなされており、耐性をより高めることができるものと考えられる。

5. 聴取実験

本節では、エコー拡散手法によるホスト信号の音質に関する聴取実験について述べる。実験に用いたホスト信号は、ボーカルとパルス性の背景音で構成されたアメリカのポップス (S01) とピアノのソロ演奏 (S04) である。また、この二つのホスト信号は、図 7 に示した信号処理に対する耐性の結果で、「NON」に対する検出性能が四つのホスト信号のうち、一番高いホスト信号 (S01) と低いもの (S04) である。実験に使用した、パラメータは、 α が 0.002, 0.006, 0.01, 0.014, 0.02, L が 511, 1023, 2047 である。聴取実験には、20 代の男性 4 名と女性 1 名が参加した。

実験の方法は AXB 法を使用した。ここで A と B は、どちらかが透かしを埋め込んだ信号で、もう一方がオリジナル信号である。また、X は、A あるいは B のどちらかと同じ刺激である。AXB 法では、A-X-B の順番に被験者に聞かせて、X が A なのか B なのかを答えさせる。X には、A と B が同確率で出

現するように設定した [20]。

図 8 に、聴取実験の結果を示す。この図の横軸は、式 (1) の $\alpha^2 \cdot L$ であり、縦軸は、正答率である。また、図の点線は正答率 75 % を示し、刺激に対する検知限を表す [20]。すなわち、正答率がこの検知限以下になると、被験者はオリジナル信号と透かしが埋め込まれた信号の音質の差の区別が困難になることを意味する。

図 8 から、各ホスト信号に対する検知限は、S01 の場合で $\alpha^2 \cdot L = 0.102$, S04 の場合で、 $\alpha^2 \cdot L = 0.205$ であることが分かる。これは、各々 $\alpha = 0.01$ と $L = 1023$, $\alpha = 0.01$ と $L = 2047$ の組み合わせに相当する。

6. まとめ

本論文では、エコー拡散手法の埋め込む容量、信号処理に対する耐性、また、音質の劣化に関する検討を行った。

その結果、検出過程の DFT 長 F と L の関しては、高い検出性能のために一定の関係があることが判明した。また、この関係からエコー拡散手法における適切な埋め込む容量の限界について検討を行った。また、エコー拡散手法は、高い検出性能を持ちながら音質の劣化も少なく、様々な信号変化に対してもある程度の耐性があることが分かった。

謝辞

本研究はミレニアム技術開発提案公募プロジェクト (課題番号 12107) の一環として行われたものである。

文 献

- [1] Christine I. Podilchuk and Edward J. Delp, "Digital Watermarking: Algorithms and Applications," *IEEE Signal Processing Magazine*, pp.33-46, JULY 2001.
- [2] Steve Czerwinski, Richard Fromm, and Todd Hodes, "Digital Music Distribution and Audio Watermarking," *UCB 15 219(Tygar) Project Report*, Spring 1999.
- [3] Daniel Gruhl and Walter Bender, "Echo Hiding," in *Proc. Information Hiding Workshop*, pp. 295-315, Cambridge University, U.K., 1996.
- [4] Hyen O Oh, et al., "New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking," *ICASSP 2001*, Salt Lake, USA, MAY 2001.
- [5] W. Bender, D. Gruhl, N. Morimoto, and A.Lu, "Techniques for data hiding," *IBM Systems Journal*, Vol. 35, No. 3&4, 1996.
- [6] Ingemar J. Cox, Joe Kilian, Tom Leighton, and Talal Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *NEC Research Institute, Technical Report*, 95-10.
- [7] Laurence Boney, Ahmed H. Tewfik and Khanled N. Hamdy, "Digital Watermarks for Audio Signals," *Europ. Signal Processing Conf.*, Trieste, Italy, SEP 1996.
- [8] Mitchell D. Swanson, Bin Zhu, Ahmed H. Tewfik, and Laurence Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, Vol. 66, No. 3, pp. 337-355, 1998.
- [9] Nedeljko Cvejic, Anja Keskinarkaus, and Tapio Seppanen, "Audio Watermarking using M-Sequences and Temporal Masking," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp.227-230, New paltz, New York, USA, 21-24 October 2001.
- [10] Nedeljko Cvejic and Tapio Seppanen, "Improving Audio

- Watermarking Scheme using Psychoacoustic Watermarking Filtering," *Proc. 1st IEEE International Symposium on Signal Processing and Information Technology*, pp.169-172, Cairo, Egypt.
- [11] R. Nishimura, M. Suzuki and Y. Suzuki, "Detection threshold of a periodic phase shift in music sound," *17th ICA Proc.*, Rome, SEP 2001.
 - [12] Byeong-Seob Ko, Ryouichi Nishimura, and Yôiti Suzuki, "Time-spread Echo Method for Digital Audio Watermarking using PN sequences," *ICASSP 2002*, Orlando, Florida, USA, MAY 2002.
 - [13] W. G. Gardner, "The Virtual Acoustic Room," *S.B., Computer Science and Engineering MIT*. SEP 1992.
 - [14] Darko Kirovski and Henrique Malvar, "Robust Spread-Spectrum Audio Watermarking," *ICASSP 2001*, Salt Lake, Utah, USA, 2001.
 - [15] A.V.Oppenheim and R.W.Schaffer, *Discrete-Time Signal Processing*, Prentice Hall, Inc., NJ, 1989.
 - [16] Kenneth R. Boff, Lloy Kaufman, and James D. Thomas, *Handbooks of Perception and Human Performance*, Volume I, John Willey and Sons, 1986.
 - [17] David Heeger, "Signal Detection Theory," *Teaching Handout*, Department of Psychology, Stanford University, NOV 12, 1997.
 - [18] [Online], <http://www.jasrac.or.jp/release/00/10.html>
 - [19] Litao Gang, Ali N. Akansu and Mahalingam Ramkumar, "Security and Synchronization in Watermark Sequence," *ICASSP 2002*, Orlando Florida, USA, MAY 2002.
 - [20] 高橋晃, 西村竜一, 鈴木陽一, "繰り返し聴取が音楽信号の位相回転に対する検知限に及ぼす影響," *日本音響学会講演論文集*, pp.519-520, 2002 年 3 月.