

RK-005

# 蝸牛遅延に基づいた電子音響透かし法とそのブラインド検出の実現

## Study on non-blind detection method for digital-audio watermarking technique based on the cochlear delay characteristics

鵜木祐史\*  
Masashi Unoki

宮内良太\*  
Ryota Miyauchi

### 1. はじめに

近年、情報通信技術の発達や迅速なインフラの整備により、インターネットを利用したマルチメディア情報(テキスト、音声・音楽、静止・動画像など)の利用が盛んになっている。これらの利用の安全性を高めるために、コンテンツ保護や機密情報保護の中核的な技術が必要である。マルチメディア情報ハイディング(MIH)技術は、暗号技術とは異なる情報保護技術として、最近注目を集めており[1]、デジタルコンテンツの新しい著作権保護技術である電子透かし法[2]として実用化されはじめてている。

この技術は、利用者に知覚されないように不可分な情報として著作権情報をコンテンツ自体に埋め込み、それらを検出することで違法コピーや違法配信を防ぐ、あるいは追跡することを可能とする。静止・動画像の研究分野では、数多くの成功例があり、広く普及しはじめている。一方、音(音楽)コンテンツに関する電子透かし法に関しては、静止・動画像に関するものと同じように成功しているわけではない。STEP2001のコンペティション[3]以降は、その研究事例も激減しており、この分野における新しいブレイクスルーが起こらない限り、音情報に関するMIH技術の進展は非常に難しいものと思われる。

一般に、完全な電子音響透かし法(図1)を確立するためには、次の三つの条件:(a) 知覚不可能性(埋め込み情報が利用者に知覚されず、埋め込みによる原信号の知覚可能な歪みが一切生じないこと)、(b) 頑健性(通常の信号変換処理や悪意のある攻撃に対して影響を受けないこと)、(c) 秘匿性(情報が埋め込まれていることに気づかせないこと、気づかれてもその情報を容易に検出されないこと)を満たす必要がある[3, 4]。

現在までに、様々な形態の電子音響透かしの手法が提案されている(サーベイ等の詳細は、文献[4, 5]を参照されたい)。代表的な手法として、符号化/量子化レベルで透かしを埋め込む方法(例えば、LSB(Least Significant Bit replacement)法[4])や原信号の広範なスペクトルに情報を埋め込む方法(例えば、DSS(Direct Spread Spectrum)法[6])がある。また、位相に係わる知覚特性に基づく方法として、エコーハイディング法(ECHO法[7]; 図2(a)の群遅延操作を行う方法)や、周期的位相変調(PPM: Periodical Phase Modulation)法(図2(b)の群遅延操作を行う方法)[8]がある。

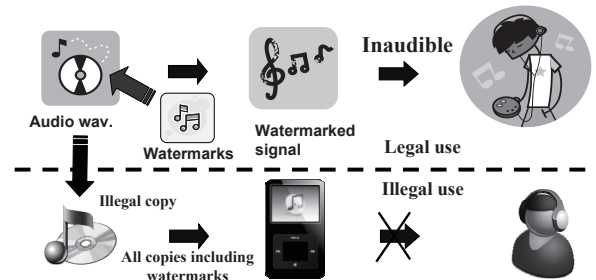


図1: 電子音響透かし法の概要

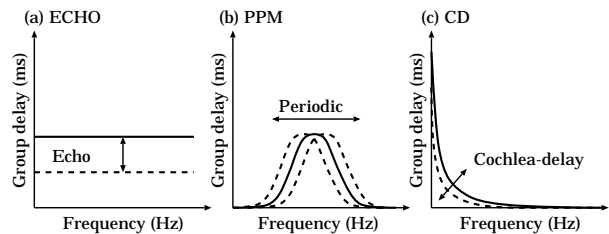


図2: 埋め込みに利用する主な群遅延特性。(a) ECHO法、(b) PPM法、(c) CD法の場合。

LSB法は、振幅情報に大きく影響を与えない下位ビットに情報を埋め込むため、知覚不可能性を満たすが、ビット変化に敏感なため頑健性に問題が残る。DSS法はスペクトル全体に情報を埋め込むため、信号変形処理には頑健であるが、埋め込まれた情報を容易に知覚できるため知覚不可能性に問題が残る。ECHO法は、エコー時間や1次反射音の振幅を調整することで歪みがなく、知覚不可能な埋め込みを実現できるが、自己相関法やケプストラム処理を利用することで透かし情報を容易に検出・除去できるため、これまでの方法の中でもっとも頑健性・秘匿性に欠ける。PPM法は、周期的な位相変調が比較的知覚され難いという聴覚特性に基づいているが、位相変調が高い周波数成分の位相スペクトルをランダムに歪ませるため、知覚不可能性に問題が残る。最後に、ここで紹介されたすべての方法は、埋め込みにどのようなアルゴリズムを利用したのかを伏せることで秘匿性を担保している。将来的には、アルゴリズムを公開しても埋め込み情報を検知されないようにすることが理想である。

一方、著者らはこれまでに、蝸牛遅延(Cochlear Delay: CD)に着目した新しい電子音響透かし法を提案した[9, 10]。以後、この方法をCD法と呼ぶ。CD法は、

\*北陸先端科学技術大学院大学 情報科学研究科

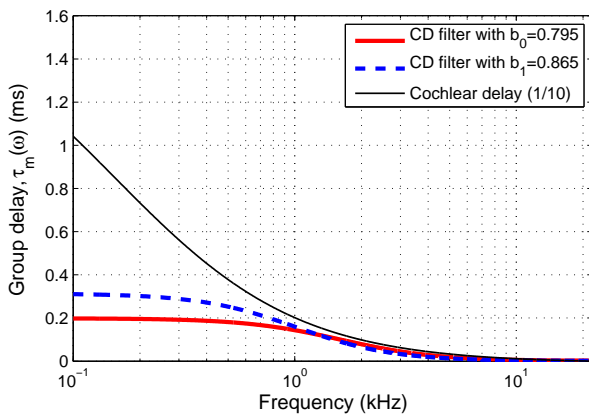


図3: 蝸牛遅延特性と蝸牛遅延フィルタの群遅延特性

三つの条件の中でも特に知覚不可能性と頑健性を満たすことに重点が置かれ、深く検討されてきた。秘匿性に関しては、他の方法と同様にアルゴリズムを非公開とすることで担保している。先の報告 [11] では、様々な客観評価実験 / 主観評価実験 / ビット検出実験 / 耐性試験を通じて、上記にあげた代表的な方法と比較検討し、知覚不可能性と頑健性に関して CD 法の有効性を示すことができた。また、別の報告 [12, 13] では、CD 法の埋め込み限界を検討し、電子音響透かし法を構成するフィルタ構成を並列型や縦続型とすることで、埋め込み限界を更に 2~3 倍まで拡張することができた。しかし、CD 法は、透かし情報を検出するために原信号の参照を必要とする。そのため、現状では応用範囲が限定され、原信号の参照を必要としないブラインド検出法の確立が急務となっている [14]。

本稿では、CD 法のデータ埋め込みならびにデータ検出を再考し、ブラインド検出法の実現可能性を探る。本稿の構成は次のとおりである。まず、第 2 節で CD 法の原理（透かし情報の埋め込み・ノンブラインド検出）を説明し、第 3 節でブラインド検出法を提案する。次に第 4 節で提案法の評価をするために、CD 法ならびに代表的な方法を利用したときの評価結果を示し、これらと比較して、ブラインド検出法を組み込んだ場合の CD 法（提案法）の特性を評価する。最後に、第 5 節で結論を述べる。

## 2. 蝸牛遅延に基づいた電子音響透かし法

蝸牛遅延とは、蝸牛内の基底膜振動で見られる、信号の周波数成分に依存した進行波の伝搬の遅延のことである。この蝸牛遅延と音の同時性判断に関する心理物理学的な検討結果から、蝸牛遅延に似た遅延パターンをもった音と原信号との弁別が非常に難しく、人間の聴覚系はこういった遅延に対して鈍感なシステムであることが示唆されている [15]。著者らは、この特性に着目し、透かし情報の二値データ（“0” と “1”）に対応する二種類の異なる蝸牛遅延に似た遅延パターンを原信号に付与することで、電子音響透かしを実現する

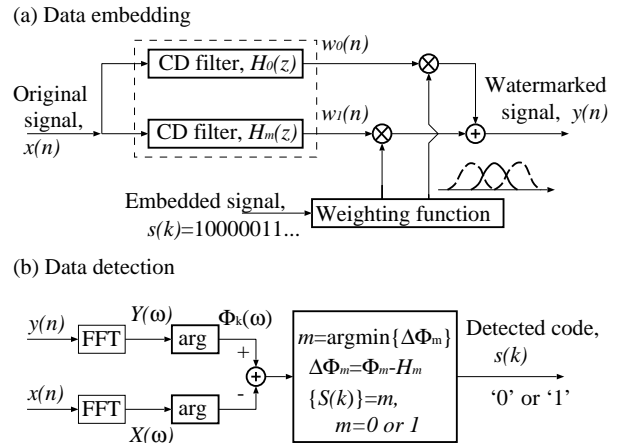


図4: 提案法におけるデータ埋め込み処理 / データ検出のブロックダイアグラム。

ことを考えた [10]。ここでは、1 次の IIR 全域通過フィルタ  $H_m(z)$  とその群遅延  $\tau_m(\omega)$

$$H_m(z) = \frac{-b_m + z^{-1}}{1 - b_m z^{-1}}, \quad m = 0, 1 \quad (1)$$

$$\tau_m(\omega) = -\frac{d \arg(H_m(e^{j\omega}))}{d\omega} \quad (2)$$

を利用して、群遅延  $\tau_m(\omega)$  が蝸牛遅延特性（図 3 の破線はこの 1/10 の特性を示す）となるように、二種類の蝸牛遅延をもつフィルタ  $H_0(z)$  ( $b_0 = 0.795$ ) と  $H_1(z)$  ( $b_1 = 0.825$ ) を設計し、透かし情報の埋め込み・ノンブラインド検出法を実現した [9]。図 3 の蝸牛遅延特性に群遅延  $\tau_m(\omega)$  を適合し、これらの間の自乗誤差が最小になるようにパラメータ  $b_m$  の最適値を求めたところ、 $b_m = 0.795$  であった [9]。以後、この値をベースに  $b_m$  の値を利用する。また、スペクトルの計算には、 $H_m(e^{j\omega}) = H_m(z)|_{z=e^{j\omega}}$  を利用する。

### 2.1. データ埋め込み処理

図 4(a) は、提案法のデータ埋め込み処理のブロックダイアグラムを示す。ここでは透かし情報を次の手順で埋め込む。(i) 二つの蝸牛遅延フィルタ  $H_0(z)$  と  $H_1(z)$  を設計する。式 (1) のパラメータ  $b_m$  は、 $b_0 = 0.795$  と  $b_1 = 0.865$  とする。(ii) 式 (3) と式 (4) を用いて、原信号  $x(n)$  に蝸牛遅延パターンを付与し、中間信号  $w_0(n)$  と  $w_1(n)$  を得る。(iii) 埋め込みデータ  $s(k)$  は、著作権情報に関する二値情報（例えば、“10000011...”）をもつものとする。(iv) 埋め込みデータ  $s(k)$  (“0”あるいは“1”)の論理値に沿って、各フレーム毎に中間信号  $w_0(n)$  と  $w_1(n)$  を選択・統合することで、式 (5) の透かし入り信号  $y(n)$  を得る。

$$w_0(n) = -b_0 x(n) + x(n-1) + b_0 w_0(n-1) \quad (3)$$

$$w_1(n) = -b_1 x(n) + x(n-1) + b_1 w_1(n-1) \quad (4)$$

$$y(n) = \begin{cases} w_0(n), & s(k) = 0 \\ w_1(n), & s(k) = 1 \end{cases} \quad (5)$$

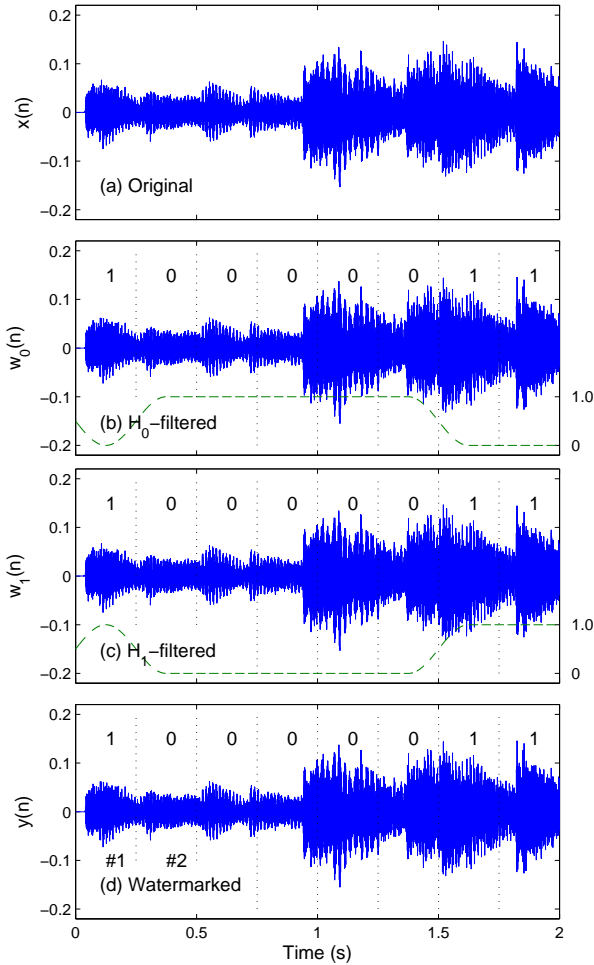


図 5: CD 法のデータ埋め込みの例 .

但し,  $(k-1)\Delta W < n \leq k\Delta W$ ,  $k = 1, 2, \dots$  である .  
 ここで,  $n$  はサンプル値,  $k$  はフレーム番号,  $\Delta W = f_s/N_{\text{bit}}$  はフレーム長 (半分のフレーム長で重複),  $f_s$  は原信号のサンプリング周波数,  $N_{\text{bit}}$  は 1 秒間の割当てビット数 (ビットレート) である . 最後に, 二値データ系列  $s(k)$  に応じて, 各中間出力の荷重和を取ること  
 で出力信号  $y(n)$  を得る . なお, ここでは, フレーム間の接続部で不連続性の発生を避けるために荷重関数 (Hanning 窓関数) を利用した .

2.2. データ検出処理

図 4(b) は, 提案法のデータ検出処理 (ノンブラインド法) のブロックダイアグラムを示す . 透かし情報 (埋め込みデータ) は次の手順で検出される . (i) 原信号  $x(n)$  と透かし入り信号  $y(n)$  の両方が検出処理で利用可能であるものとする (ノンブラインド検出) . (ii)  $x(n)$  と  $y(n)$  は, データ埋め込みで利用されたものと同じ窓関数を利用して各セグメントに分解される . (iii) 位相差  $\phi(\omega)$  は, 式 (6) を利用して, フレーム毎に計算される . ここで,  $\text{FFT}[\cdot]$  は高速フーリエ変換である . (iv) データ埋め込みで利用された群遅延特性を求めするために, 位相差  $\phi(\omega)$  に対する各フィルタの位相スペク

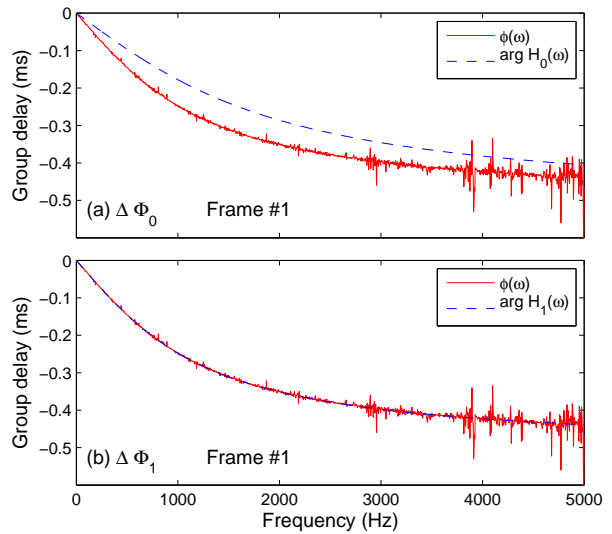


図 6: CD 法のデータ検出の例 (フレーム #1 の場合) .

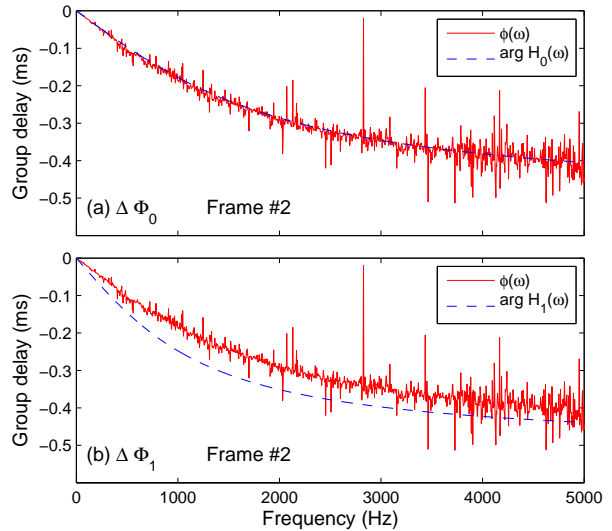


図 7: CD 法のデータ検出の例 (フレーム #2 の場合) .

トルの差 ( $\Delta\Phi_0$  と  $\Delta\Phi_1$ ) を式 (7) と式 (8) から得る .  
 (v) 式 (9) を利用して  $\hat{s}(k)$  を検出する .

$$\phi(\omega_m) = \arg(\text{FFT}[y(n)]) - \arg(\text{FFT}[x(n)]) \quad (6)$$

$$\Delta\Phi_0 = \sum_m |\phi(\omega_m) - \arg(H_0(e^{j\omega_m}))| \quad (7)$$

$$\Delta\Phi_1 = \sum_m |\phi(\omega_m) - \arg(H_1(e^{j\omega_m}))| \quad (8)$$

$$\hat{s}(k) = \begin{cases} 0, & \Delta\Phi_0 < \Delta\Phi_1 \\ 1, & \text{otherwise} \end{cases} \quad (9)$$

2.3. データ埋め込み / データ検出の事例

埋め込み・検出処理の一例を図 5 ~ 図 7 に示す . まず, 透かし情報を 8 文字 (“AIS-Lab.”, その先頭 8 ビットのみを表示すると  $s(k) = 10000011 \dots$ ) とし,  $N_{\text{bit}} = 4$  bps の埋め込みを考える . ここでフレーム長  $\Delta W$  は 250 ms,  $k$  はフレーム番号を示す . 原信号  $x(n)$  (図 5(a))

に  $H_1(z)$  と  $H_2(z)$  を利用して、それぞれの群遅延を付与して中間出力  $w_0(n)$  と  $w_1(n)$  (図5(b)と図5(c))を得る。次に、二値データ系列  $s(k)$  に応じて、各中間出力の荷重和 (Hanning窓, 図5(b)と図5(c)の破線) を取ることによって出力信号  $y(n)$  (図5(d)) を得る。ここで図中の縦の点線はフレームの境界を示す。

透かし情報  $s(k)$  の検出は、 $x(n)$  と  $y(n)$  の位相スペクトルの差分を求め、透かし情報の埋め込み時に使用した  $H_0(z)$  と  $H_1(z)$  の群遅延特性をマッチングすることで得られる。例えば、フレーム#1と#2では、それぞれ“1”と“0”が埋め込まれている。図6はフレーム#1での位相スペクトルの差を、図7はフレーム#2での位相スペクトルの差を示す。図6(b)と図7(a)から、各位相スペクトルの差と蝸牛遅延フィルタの遅延量  $\arg H_m(\omega)$  の一致を確認できるが、これらの一致は式(6)~式(9)の関係を表していることから、 $s(k)$  を正確に検出できる。ここでは、図6(b)の結果から  $s(1) = 1$  を検出し、図7(a)の結果から  $s(2) = 0$  を検出している。

### 3. ブラインド検出法

#### 3.1. 蝸牛遅延フィルタの特性

現状の遅延の埋め込み情報の検出方法(図4(b))は、原信号  $x(n)$  を利用したノンブラインド法である。この方法は、原信号と遅延情報を埋め込んだ信号との位相特性の差分から群遅延特性を逆推定することで、遅延の埋め込み情報(“0”か“1”)を検出している。現状の大きな課題は、観測した信号  $y(n)$  のみから遅延の埋め込み情報(“0”か“1”)を検出することである。これは、一般にブラインド検出法と呼ばれており、何らかの前提条件などがなければ、その実現は非常に難しい。

遅延情報の付与は、式(1)で定義されるように、パラメータ  $b_m$  ( $b_0$  と  $b_1$ ) で構成される二つの蝸牛遅延フィルタを利用して行われる。このとき、それぞれのフィルタの極・零点の配置は、図8のようになる。蝸牛遅延フィルタは前述するように、1次IIR型全域通過フィルタであり、その特徴として極( $\times$ )と零点( $\circ$ )は中心点から単位円に向かって垂線を描いたときに交差する半径とその逆数( $b_m$  と  $1/b_m$ )の関係にある。一般に  $b_m$  の値が減少するに従い、極は中心点に近付き、零点は単位円から外側に向かって離れていく。反対に  $b_m$  の値が増加するに従い、極と零点は互いに単位円に向かって近付いていく。このとき群遅延量は、図3に示すように、 $b_m$  の値の増加とともに増加する。

$y(n)$  は、このような遅延情報が埋め込まれた信号として観測される。そのため、ブラインド検出の本質は、 $y(n)$  から遅延情報、つまり遅延情報の付与に利用されたフィルタの極・零点の配置を正確に推定できるかということである。ここで、原信号  $x(n)$  自体も数列の特性として極・零点をもつため(音源が有界であるとして、その信号の減衰に関係する極など)、観測信号  $y(n)$  から仮に極・零点の配置を推定できたとしても、それ

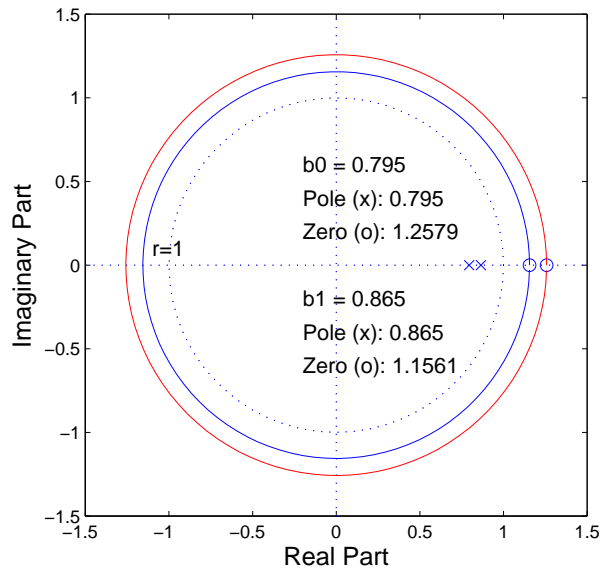


図8: 蝸牛遅延フィルタの極( $\times$ )・零点( $\circ$ )の配置。

はIIR型全域通過フィルタによって付与されたものか、対象信号そのものがもつものであるか、正確に見極めなければならない。

#### 3.2. チャープz変換

チャープz変換(CZT)は、周波数スペクトルのフレキシブルな分析を可能とする手法[16, 17]として知られ、例えば、高速Fourier変換(FFT)の実装にも活用されている。チャープz変換は、離散Fourier変換(DFT)と比較して、周波数分解能や周波数応答のダイナミックレンジを自由に換えられる特徴を有している。また、z平面上で任意のM点でのz変換を効率よく求めることもできるという特徴も有している。

一般に、z変換は  $z = r \exp(j\omega_n)$  でN点のDFTと結ばれる(大きさ  $r = 1$  で正規化周波数  $\omega_n = 2\pi n/N$  のとき単位円周上のDFTと等価である)関係にある。ここで、チャープz変換は次式で表される。

$$Y(z_k) = \sum_{n=0}^{N-1} y(n) A^n W^{nk}, \quad k = 0, 1, \dots, M-1, \quad (10)$$

ただし、 $A = A_0 \exp(j2\pi\theta_0)$ 、 $W = W_0 \exp(j2\pi\phi_0)$  である。ここで、 $\theta_0$  と  $\phi_0$  は初期位相である。前述したように、 $A = 1$ 、 $M = N$ 、 $W = \exp(-j2\pi/N)$  のときCZTはDFTに一致する。

#### 3.3. フィルタの極・零点の位置の推定

本稿では、式(1)の蝸牛遅延フィルタの零点  $r = 1/b_m$  を通るように  $r$  を選択して、原信号  $x(n)$  ならびに遅延情報を埋め込んだ信号  $y(n)$  をチャープz変換 ( $A = r$ 、 $M = N$ 、 $W = \exp(-j2\pi/N)$ ) して周波数分析してみる。予想されることは、 $x(n)$  には蝸牛遅延フィルタの零点を含まないため、零点に対応する周波数近傍では、 $x(n)$  の周波数スペクトルには零点の影響が見られない

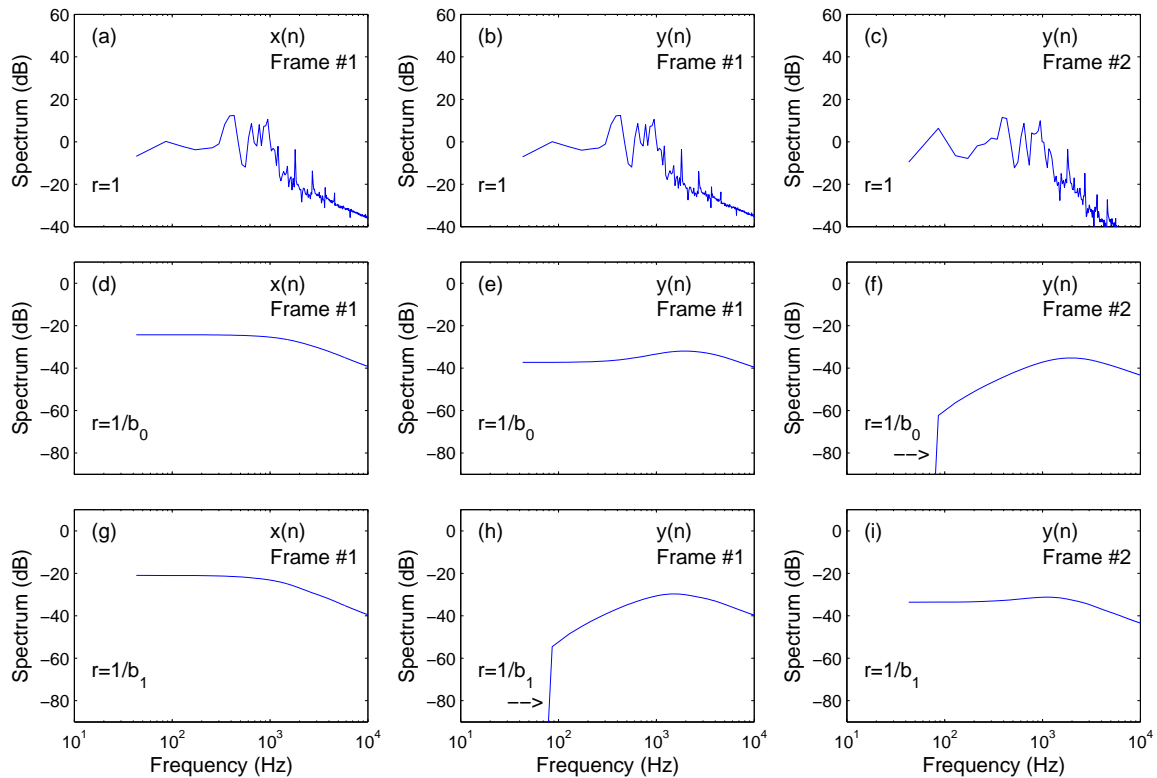


図9: 図5の楽器音 ( $x(n)$  と  $y(n)$ ) の先頭2フレームにおける分析結果. 埋め込み情報は同じく“AIS-Lab.”. この場合, 先頭2フレームでは“10”が埋め込みされている. (a)-(i) 原信号  $x(n)$  と観測信号  $y(n)$  (フレーム#1と#2) のチャープ  $z$  変換による分析結果. 図中の  $\dashrightarrow$  は, 遅延情報の埋め込み箇所とその検出位置を指している.

が,  $y(n)$  の周波数スペクトルには見られるはずである.

図5で利用した楽器音の事例を同じく利用して調べる. 原信号  $x(n)$  として, 蝸牛遅延フィルタを利用して“AIS-Lab.”の透かし情報を埋め込んだ信号  $y(n)$  とする. ここでは, 図4(a)の埋め込み方式を取らず, フレーム単位に該当する蝸牛遅延フィルタ処理のみを施し, フィルタの選択処理や Hanning 関数の荷重和をとる処理を入れない. 蝸牛遅延フィルタはいずれも直流成分のところにも極ならびに零点を配置しており,  $r = 1/b_0$  あるいは  $r = 1/b_1$  としたチャープ  $z$  変換の周波数分析を考える. サンプル周波数は 44.1 kHz, ビットレートは  $N_{\text{bit}} = 4$  bps として, 1フレーム (250 ms) に1ビット相当の遅延情報を埋め込むものとする.

図9に, 上記の分析結果を示す. 図9(a)-(i)は, 左から右にフレーム#1での  $x(n)$ , フレーム#1での  $y(n)$ , フレーム#2での  $y(n)$  の周波数スペクトルを, 上から下に  $r = 1$ ,  $r = 1/b_0$ ,  $r = 1/b_1$  でのチャープ  $z$  変換により分析した結果を示す. ここでは, 図6, 図7と同様にビット検出を考察するため, 先頭のフレーム#1とフレーム#2での  $y(n)$  のみ掲載している. まず下一番左の  $x(n)$  に関する分析結果では, 極・零点配置の周波数付近でのスペクトルには特段変化がみられない. しかし, フレーム#1の  $y(n)$  では  $r = 1/b_1$  でのチャープ  $z$  変換の結果 (図9(h)) で, フレーム#2の  $y(n)$  では

$r = 1/b_0$  でのチャープ  $z$  変換の結果 (図9(f)) で, 最も低い周波数 (直流成分から低周波数域までの範囲; 例えば図3の遅延が見られる周波数帯) のところでスペクトル成分が劇的に減少していることがわかる (図中の  $\dashrightarrow$  を参照). これは, 零点の影響によるディップに対応しているため, 原理的にはその大きさは  $-\infty$  dB になる. それ以外の分析 ( $r = 1$  と  $r = 1/b_1$  (フレーム#1のとき) あるいは  $r = 1/b_0$  (フレーム#2のとき)) では, 最も低い周波数のところでスペクトル成分の変化はほとんどみられない (あるいは  $-\infty$  dB (線形で0) に近付かない) ことがわかる. この結果に関しては, 他のフレームや他の対象信号でも同様のことが起こることを確認している.

まとめると, 図9の結果から, 対象信号に係わらず, IIR フィルタの零点を交差するように  $z$  平面上の軌跡に沿って, チャープ  $z$  変換することにより, 観測信号  $y(n)$  から IIR フィルタの零点配置を推定できたといえる. 原理的には,  $r$  を零点ではなく極の値にしてチャープ  $z$  変換を利用することも可能である (極の場合は  $\infty$  dB のスペクトルピークを得ることになる) が, 計算機上でのダイナミックレンジ内のオーバーフローを検出しなければならないため, あまり有効な手法ではない. 零点を利用する場合は, ダイナミックレンジ内の0を探せばよいため, 群遅延の付与に利用した IIR フィル

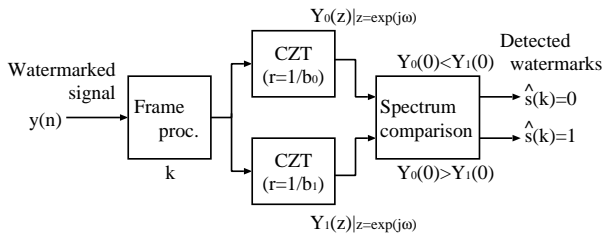


図 10: ブラインド検出法のブロックダイアグラム。

タの零点検出がより現実的で容易な手法である。

### 3.4. ブラインド検出法の実装

上記の検討から、遅延情報の埋め込みに利用する IIR 型全域通過フィルタの零点の情報、ビットレート ( $N_{bit}$  bps 等) の情報を事前に検出側が知っているという前提条件のもとであれば、零点を  $z$  平面上の軌跡が通るようなチャープ  $z$  変換を利用することで、容易にフィルタの零点情報をブラインド検出可能である。

上記のコンセプトに基づき、シンプルなデータ埋め込み法とビット検出法を実装した。本稿では、CD 法の基本的な諸特性 (知覚不可能性とビット検出の精度) のみ検討する。ブラインド検出法のブロックダイアグラムを図 10 に示す。これは、ビットレート ( $N_{bit}$  bps) の情報に基づき、フレーム長を固定し、セグメント毎に 2 種類のチャープ  $z$  変換 (CZT,  $r = 1/b_0$  と  $r = 1/b_1$  の  $z$  平面上の軌跡に沿ったチャープ  $z$  変換) を施したあと、得られた二つの周波数スペクトルに対し、最も低い周波数でのスペクトルの値が小さくなる時の蝸牛遅延フィルタの零点 ( $1/b_0$  か  $1/b_1$ ) を正しい検出結果として、“0” ないし “1” と判定するものである。なお、透かし情報  $s(k)$  の埋め込みには、 $b_m$  に対する時変的な処理や Hanning 関数による荷重和の処理を入れない。これは、観測信号  $y(n)$  からの安定したフィルタの零点を検出するための処理である。しかし、現状のままではフレームの接続点に不連続点が生じて (スペクトル拡散の原因でもある) 知覚不可能性に影響を与えるため、接続部前のフレームの後ろ数点 (1 ms 程度) を Spline 補間でなめらかにすることで不連続点の発生を抑えている。

### 4. 提案法の評価

CD 法によって埋め込まれた透かしの知覚不可能性とその透かし情報に対するビット検出の正確性に関して、客観評価実験により検討する。

知覚不可能性の検討として音質評価に関する代表的な客観評価実験を行う。客観評価実験では、RWC 音楽データベース [18] の全 102 曲を評価用の原信号 (サンプリング周波数 44.1 kHz, 16 ビット量子化) として利用する。ここでは、冒頭 10 秒間を元曲として、8 文字の情報 (“AIS-Lab.”) を透かし情報として各原信号に埋め込む。また、 $N_{bit} = 4$  bps をベースとし、12 条件の  $N_{bit}$  ( $N_{bit} = 4, 8, 16, 32, 64, 128, 256, 512, 1024,$

2048, 4096, 8192 bps) で、透かし情報を原信号の両チャンネルに埋め込み、その特性評価を行う。音質評価に関しては、Lin & Abdulla の報告 [19] に基づき、オーディオ信号に対する知覚評価尺度 (PEAQ) [20] と対数スペクトル歪尺度 (LSD) を利用する。また、比較評価では、代表的な電子音響透かし法として、LSB 法、DSS 法、ECHO 法、PPM 法を利用する。ここで、PPM 法を除き、すべてブラインド検出法である。蝸牛遅延に基づく電子音響透かし法では、著者らによって既に提案された CD 法 (ノンブラインド検出法を利用したもの、CD (UH08)) と今回の提案法 (ブラインド検出法を利用したもの、CD (Blind)) を比較する。

#### 4.1. 評価結果

はじめに、PEAQ による客観評価の結果を図 11(a) に示す。図中の結果は 102 曲の平均値である。PEAQ の ODG (objective difference grade) 値は 0 (知覚不可能) ~ -4 (非常に耳障りである) であるため、-1 (知覚される可能性があるが耳障りではない) を知覚不可能性の閾値と定めた。代表的な方法のなかでは、DSS 法が一番悪く、次いで ECHO 法もビットレートが 8 bps 以降から先で急激に悪くなった。PPM 法は一般的に ODG で -2 程度であることがわかった。LSB 法は、今回の 12 条件で全く問題ない結果となった。一方、CD (UH08) 法では、ビットレートが 4 bps では ODG 値に全く問題がないが、 $N_{bit} \geq 128$  bps あたりから ODG 値が減少し、閾値は、1024 bps 程度となった。CD (Blind) 法では、64 bps の時点で既に -1.0 付近にあり、bps の増加とともに -3.0 付近まで低下した。

次に、LSD の結果を図 11(b) に示す。一般に LSD は 1 dB 内の歪みであれば、音質がよいといわれているため、LSD の閾値を 1 dB に設定した。LSB 法は予想どおり、bps を変えても埋め込みによる歪みの影響を受けおらず、評価閾値のかなり下にあることがわかった。DSS 法も予想どおり、bps の増加にかかわらず評価閾値の上であり、音質評価では問題があることがわかった。残る ECHO 法と PPM 法では、いずれも評価閾値内にあり、特に音質に関して問題があるようには見えなかった。一方、CD (UH08) 法は、すべての  $N_{bit}$  で閾値内にあるが、256 bps まで 0.5 dB 以内を維持していた。CD (Blind) 法は、bps の増加に対して単調増加しており、 $N_{bit} < 1024$  bps までは閾値以下 (-1 dB) にあるが、CD (UH08) 法と比較すると若干大きな値となった。しかし、4 ~ 64 bps の付近では、CD (Blind) 法での LSD が CD (UH08) 法のものよりも若干小さい値になった。また、LSD での差は図 11(a) の PEAQ ほど大きくはならなかった。これは、単純なスペクトル歪みでみるのとは異なり、聴覚的印象に基づく尺度では、違いとして現れてきているためと考えられる。

最後に、埋め込みデータのビット検出の結果を図 11(c) に示す。LSB 法を除き、いずれの手法とも  $N_{bit}$  の増加に伴いビット検出率の低下が見られた。CD (UH08) 法

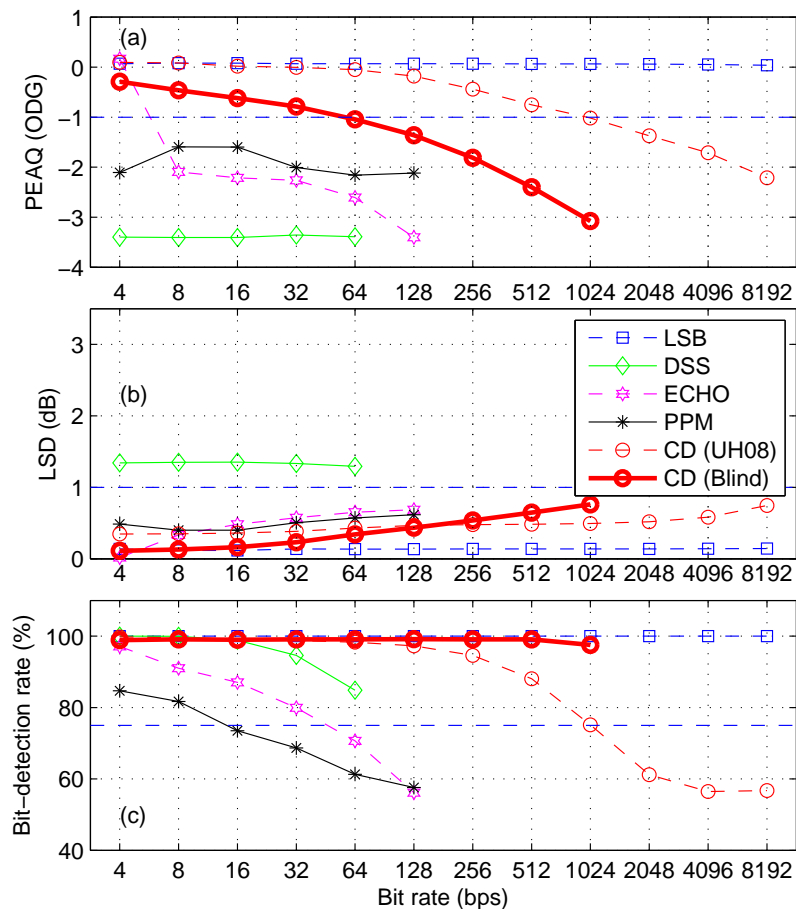


図 11: 客観評価実験の結果：(a) PEAQ, (b) LSD, (c) ビット検出率 .

は  $N_{\text{bit}} = 1024$  bps 程度で 75% 閾値を切るが、他の手法はもっと低い  $N_{\text{bit}}$  で閾値を切った。一方、CD (Blind) 法では、CD (UH08) 法と比較して、大幅な改善が確認できた。特に、 $N_{\text{bit}} \leq 512$  まではほぼ 100% であり、1024 bps で 98% となることがわかった。これは、フィルタの零点検出精度に必要なフレーム長に起因するものと考えられる。しかしながら、ブラインド検出であるという点で非常に大きな利点を持ち、高精度にビット検出できることは提案法の大きな特徴であるといえる。

以上の客観評価から、CD 法 (CD (UH08)) ならびに提案法 (ブラインド検出法, CD (Blind)) が他の手法に比べて優れていることが確認できた。LSB 法は知覚不可能性ならびにビット検出率について何も問題ないが、埋め込みされた信号  $y(n)$  が少しでも改変された場合に検出できないため、頑健性に大きな問題があることが既に指摘されている [10, 11]。なお、知覚不可能性に関する評価では、CD (UH08) の結果に対して、CD (Blind) の結果が bps の増加とともに更に減少する傾向にあったが、これに関しては以下の考察で議論する。

#### 4.2. 考察

図 11 の結果は、Unoki & Hamada によって提案された蝸牛遅延に基づく電子音響透かし法 (CD 法, CD

(UH08)) の有効性を客観的に示すものである。これに対し、CD 法で残されていた課題として、本稿ではブラインド検出法 (CD (Blind)) を提案した。図 11(a), (b) の知覚不可能性に関する PEAQ と LSD の評価では、CD (Blind) 法は、CD (UH08) 法のものよりも低下してしまっていたが、ビット検出率の評価ではそれを十分に上回るものとなった。これは、チャープ  $z$  変換を利用して、正確に零点の位置を検出するために、埋め込み処理で利用する Hanning 窓関数による荷重処理を利用しないことに起因していると考えられる。この処理ではフレーム間に位相の急激な変化が生じ、これによって音が歪んだためと考えられる。

従来からの報告にあるように、知覚不可能性と頑健性の間にはトレードオフの関係があるといわれている。今回の評価は、もしかすると頑健なビット検出のために強めに透かし情報を埋め込んでいることと等価な状況であるかもしれない。つまり、フィルタの零点の検出精度の正確性を若干弱めることで、知覚不可能性に関する PEAQ や LSD の評価値をさらに向上させることができるかもしれない。

先の論文 [10, 11] で報告したように、主観評価実験から CD (UH08) 法が知覚不可能性を十分に満たすこ

とが確認されている。この実験では、 $N_{\text{bit}} = 4$  bps で PEAQ の ODG が最良なもの (+0.18) と最悪なもの (-0.27) の刺激を利用したが、いずれの場合でも知覚不可能なことが実証されている。一概に PEAQ の ODG 値だけで議論することは難しいかもしれないが、提案法 (CD (Blind)) の結果をみると、4 bps での ODG の平均値が -0.29 であることを考えると、おそらく現状のものでも主観評価実験をすることで CD (Blind) 法の知覚不可能性を実証できるかもしれない。

一方、先の報告 [11] で行った耐性試験により、CD (UH08) 法が、ビット変換やリサンプリング、MP3 情報圧縮の信号変換のほか、StiarMark ベンチマークテスト [21] といった攻撃にも頑健性があることが実証済みである。これらの結果についてはまだ予想を立てられる状況にないが、あくまでビット検出率だけをみると CD (Blind) 法でも同程度の頑健性があるのではないかと予想できる。いずれにしても、CD (Blind) 法では、知覚不可能ならびに頑健性について、先の検討を踏襲して議論する必要がある。これらは今後の課題である。

## 5. まとめ

本稿では、著者らによって提案された蝸牛遅延に基づく電子音響透かし法を再考し、チャープ  $z$  変換を利用した埋め込み情報のブラインド検出法の可能性について検討した。これは、群遅延情報の付与に利用した IIR 型全域通過フィルタの零点を観測信号  $y(n)$  から推定する方法である。これまでの我々の考察から既にわかっているように、知覚不可能性 (PEAQ/LSD) とビット検出率にはトレードオフの関係がある。今回の評価結果から、提案法の知覚不可能ならびにビット検出率について総合的に検討すると、知覚不可能性に多少影響を与えているが、高いビット検出率でブラインド検出を実現することができた。今後は、トレードオフの関係からどの程度の知覚不可能性の評価閾値を緩めるか、あるいはビット検出率の評価閾値を緩めるかについても検討し、検出法を改良する予定である。

## 謝辞

本研究は、科学研究費補助金・挑戦的萌芽研究 (No. No. 21650035) ならびに基盤研究 B (No. 23300070)、H22 年度 JST つなぐしくみの援助を受けて行われた。ブラインド検出に関して議論くださった NICT 西村竜一博士に感謝する。また、本文の推敲にあたって適切なアドバイスをくださった秘匿査読者に感謝する。

## 参考文献

- [1] 小特集 マルチメディア情報ハイディング技術とその応用, 信学論 A, J93-A(2), 41, 2010.
- [2] 竹居智久, Phil Keys, “「コピーに自由を」生まれ変わる DRM,” 日経エレクトロニクス, 51-74, 2008.
- [3] STEP2001. “プレスリリース (2001.10.19), 音楽電子透かし 4 社を技術認定,” 社団法人 日本音楽著作権協会 (JASRAC), [http://www.jasrac.or.jp/release/01/10\\_2.html](http://www.jasrac.or.jp/release/01/10_2.html).
- [4] N. Cvejic and T. Seppänen, *Digital audio watermarking techniques and technologies*, IGI Global, 2007.
- [5] 西村 明, “音響信号への情報秘匿技術 -電子音響透かしとステガノグラフィ-,” 音響誌, 63(11), 660-667, 2007.
- [6] Boney, L., Tewfik, H. H., and Hamdy, K. N., “Digital watermarks for audio signals,” *Proc. ICMCS*, 473-480, 1996.
- [7] Gruhl, D., Lu, A. and Bender, W., “Echo Hiding,” *Proc. Information Hiding 1st Workshop*, 295-315, 1996.
- [8] 西村 竜一, 鈴木 陽一, “周期的位相変調に基づく音響電子透かし,” 音響誌, 60(5), 268-272, 2004.
- [9] Unoki, M. and Hamada, D. “Audio watermarking method based on the cochlear delay characteristics,” *Proc. IHMSP2008*, 616-619, 2008.
- [10] Unoki, M. and Hamada, D. “Method of digital-audio watermarking based on cochlear delay characteristics,” *Int. J. Innov. Comp., Inf. Cont.*, 6(3(B)), 1325-1346, 2010.
- [11] Unoki, M., Imabeppu, K., Hamada, D., Haniu, A., and Miyauchi, R. “Embedding limitations with digital-audio watermarking method based on cochlear delay characteristics,” *J. Information Hiding and Multimedia Signal Processing*, 2(1), 1-23, 2011.
- [12] Imabeppu, K., Hamada, D., and Unoki, M. “Embedding limitations with audio-watermarking method based on cochlear delay characteristics,” *Proc. IHMSP2009*, 82-85, 2009.
- [13] Unoki, M., Kosugi, T., Haniu, A., and Miyauchi, R. “Design of IIR all-pass filter based on cochlear delay to reduce embedding limitations,” *Prof. IHMSP2010*, 526-529, 2010.
- [14] 羽二生篤, 小杉敏三, 宮内良太, 鶴木祐史, “蝸牛遅延特性を利用した電子音響透かしのブラインド検出に向けた基礎検討,” マルチメディア情報ハイディング研究会資料, 13-18, 2010.
- [15] E. Aiba, M. Tsuzaki, S. Tanaka, and M. Unoki, “Judgment of perceptual synchrony between two pulses and verification of its relation to cochlear delay by an auditory model,” *Japan Psychological Research 2008*, 50(4), 204-213, 2008.
- [16] Rabiner, L. R., Schafer, R. W., and Rader, C. M., “The chirp  $z$ -transform algorithm,” *IEEE Trans. Audio and Electroacoustics*, AU-17(2), 86-92, 1969.
- [17] Wang, T. T. “The segmented chirp  $z$ -transform and its application in spectrum analysis,” *IEEE Trans. Instrumentation and measurement*, 39(2), 318-323, 1990.
- [18] 後藤, 橋口, 西村, 岡, “RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース,” 情報学 研究報告, 2002-MUS-45-4, 19-26, 2002.
- [19] Y. Lin and W. H. Abdulla, “Perceptual evaluation of audio watermarking using objective quality measure,” *Proc. ICASSP2008*, 1745-1748, 2008.
- [20] P. Kabal, “An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality,” *TSP Lab. Technical Report*, Dept. Electrical & Computer Engineering, McGill Univ. 2002.
- [21] M. Steinebach et al. “StirMark Benchmark: Audio watermarking attacks,” *Proc. ITCC*, 49-54, 2001.