

音脈分凝を活用した音楽電子透かし技術 "ゲンコーダ Mark" の開発

茂出木 敏雄[†] 千葉 誠[†]

[†]大日本印刷株式会社 C & I 事業部/情報コミュニケーション研究開発センター
〒141-8001 東京都品川区西五反田 3-5-20

あらまし 筆者らは、音源に携帯電話を向けるだけで情報が抽出できる音楽電子透かし技術の開発を進めている。そのためには、ヒトの聴感特性が鋭敏な音域を改変せざるを得ず、埋め込みノイズをいかにして抑圧するかが課題であった。これまで、筆者らは2チャンネル・ステレオ再生を応用して、データを埋め込んだL側スピーカで発生するノイズをR側スピーカで相殺する手法を提案してきた。しかし、音源から離れた位置で抽出することが難しいといった運用上の問題を抱えていた。本稿では、R側の雑音抑圧信号をL側に多重化し、更に視聴者に音脈分凝を誘導することにより埋め込みノイズを低減させ、モノラル再生で運用を可能にした改良手法について提案する。

Development of Audio Watermark Technique Applying Auditory Stream Segregation Phenomenon, "G-encoder Mark"

Toshio MODEGI[†] Makoto CHIBA[†]

[†] Communication and Information Operations, Dai Nippon Printing Co., Ltd.
3-5-20, Nishigotanda, Shinagawa-ku, Tokyo 141-8001, Japan
E-mail: [†] {Modegi-T, Chiba-M10}@mail.dnp.co.jp

Abstract We are developing audio watermarking techniques which enable extraction of embedded data by cell phones. For that we have to embed data onto frequency ranges, where our auditory response is prominent, therefore data embedding will cause much auditory noises. Previously we have proposed an applying two-channel stereo play-back feature, where noises generated by a data embedded left-channel signal will be reduced by the other right-channel signal. However, this proposal has practical problems of restricting extracting terminal location. In this paper, we propose an improved method, which supports monaural play-back operations. This proposes multiplexing the noise reducing right-channel signal with the left-signal and reduces noises completely by generating an auditory stream segregation phenomenon to users.

1. まえがき

筆者らは空間に分布する音源の情報を携帯端末等で非接触に抽出し、音源に関連するサイバー空間と連動できるユビキタス音響空間を提案している[5]。それを実現する方法として、音楽電子透かし手法に注目し、携帯電話を音源に近づけるだけで、音響信号に埋め込まれた情報を抽出できる新規な音楽電子透かし手法の開発を検討した。既提案の音楽著作権保護を目的とした電子透かし手法[1]では、埋め込みノイズが聞こえないようにヒトの聴感特性が鈍感な音域を改変し、音質を余り劣化させない程度の各種圧縮・変調処理などに対する耐性を備えていれば良かった。

これに対して、現行の3G携帯電話では電話回線帯域と同様に4kHzを越える音は収録できず、更に録音された音声データは3GPP準拠圧縮によ

り振幅側も顕著な歪みが発生する。一方、電子透かしが埋め込まれたソース音源は、アナログ/デジタル放送やストリーミングで配信される場合もあり、同様に4kHzを超える周波数成分は変調や圧縮による劣化が加わることも考慮する必要がある。従って、開発すべき電子透かし技術において、4kHz未満の聴感特性が最も鋭敏な音域に、振幅歪みが加わっても抽出可能なように顕著な改変を加え、かつ再生時にノイズが聴取されないようにするという矛盾する課題を解決する必要がある。

そこで、筆者らは2チャンネル・ステレオ再生を応用して、データを埋め込んだL側スピーカで顕著に発生するノイズをR側スピーカで相殺する手法を提案した[3]。空間分割法、周波数分割法、時分割法の3つの基本方式を順次提案し、特に時

分割法を用いれば、携帯電話による非接触抽出が可能であることを確認した[3]。本手法は、既提案の電子透かし手法と比較すると、10 [bps]というレートで多くの情報を埋め込めるという優位性もあったが、携帯電話以外の高性能な音楽録音機能をもつ端末を用いれば、周波数分割法と時分割法を組み合わせるにより、400[Hz]以下の周波数帯域に埋め込む条件で、最大 61.5[bps]のレートで埋め込めることも確認できた[4]。更に、その後の実験により、埋め込み周波数帯を電話回線帯域の高域側(1.7kHz~3.4kHz)に変更すれば、最大 80 [bps]のレートで埋め込めることも確認している[5]。

しかし、これらの既提案手法はモノラル再生やヘッドフォン再生環境ではノイズが目立ち運用が難しいこと、受信端末を再生音源のL側スピーカから遠ざけたり、室内残響が豊かであるとR側信号の影響を受けやすく、抽出精度が低下するという問題を抱えていた。

本稿では、データを埋め込んだL側信号にノイズを抑圧するためのR側信号成分を多重化して埋め込み、L側信号だけのモノラルで運用できるようにする改良手法を提案する。本提案では、多重化された信号パターンが、視聴者に対して音脈分凝[2]を引き起こすように工夫されており、マイクロフォンではデータ埋め込みノイズ成分を聴取できるが、ヒトには聴取されにくいという特徴をもつ。以下改良方式の概要と既提案方式との抽出精度に関する比較実験結果について述べる。

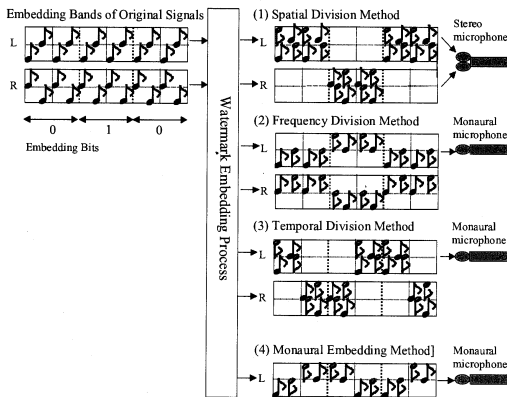


図1 既提案3方式および新規提案の音楽電子透かし埋め込み方式の概念

2. 先提案の電子透かし埋め込み方式

図1(1)~(3)は筆者らがこれまで提案してきた2チャンネルステレオ音楽信号を対象とした電子透かし埋め込み方式の概念図で、(4)は本稿で新規に提案するモノラル信号にも適用可能な電子透かし埋め込み方式の概念を示したものである。

図中の各音符は埋め込み周波数帯(例えば、1.7kHz~3.4kHz)の代表的な信号成分を示しており、同図左の2つの枠は2チャンネルの原信号より周波数分離した埋め込み周波数帯の信号成分を時系列に示し、横軸は時間で縦軸は周波数である。1フレーム(例えば、約100 [msec]、サンプリング周波数 44.1kHzで4096サンプル)を時間方向および周波数方向に各々2分割し、全3フレームで構成される原信号の各フレームごとに4つに分割された領域に信号成分が1つずつ均等に分布していると仮定する。

同図右側の(1)~(4)は3つのフレームに1ビットづつ、「0, 1, 0」というビット列を埋め込んだ結果を示す。図中の音符の向きが上下反転しているものは、変更された信号成分であることを示し、(1)~(3)においては、時間軸上および周波数軸上の位置を変更させずに、L側とR側との間で相互に移動させた信号成分である。従って、(1)~(3)においては、埋め込み前後で音符の増減は無く、L側とR側という空間的な定位のみが変更されているだけであるため、理論的に品質上の劣化は無い。

実験の結果、どのような周波数帯に埋め込んでもステレオスピーカで再生している限り、定位変動による品質の劣化には気づきにくい。ヘッドフォン視聴では周波数帯によりノイズが聴取される。ただし、埋め込み周波数帯を200 [Hz]以下に設定すれば、低音部におけるヒトの音源定位感覚が鈍いため、ヘッドフォン視聴しても再生品質に殆ど影響を与えない。ところが、300 [Hz]未満の周波数帯は電話回線帯域外であるため、携帯電話による受信は困難で、スピーカの再生能力も低く、抽出精度も安定しないため実用には適さない。筆者らが探索した最適な埋め込む周波数帯は、1.7kHz~3.4kHzで、この帯域は携帯電話による受信を含め、最も安定した抽出が行なえ最も埋め込み容量を確保できる。フレーム長をサンプリング周波数 44.1kHzで512サンプル(25 msec)まで縮小することができ、最大 80 [bps]のレートで埋め込むことができる。

図1右(1)は、最初に筆者らが提案した手法で、ステレオの左右チャンネル間で信号成分を一方のチャンネルに偏らせるものである[3]。フレームにビット0を埋め込むときは、R側の全成分をL側に移動し、ビット1を埋め込むときはその逆の操作を行なう。データを抽出する際は各フレームごとに左右の音圧を比較すれば行なえる。本手法は抽出感度の点でベストであるが、抽出時にステレオマイクを用いて、左右のスピーカの音を独立して収録しなければならず実用的ではない。

これに対して、モノラルマイクで抽出できるようにしたのが、図1右(2)である。L側にデータを埋め込み、R側はL側で欠損する信号成分を補うもので、再生時のノイズを抑圧する役目を果たす。各フレームを周波数分割し、ビット0を埋め込む場合はL側の信号成分が低音側に偏り、ビット1を埋め込む場合は高音側に偏るように、L側とR側の信号成分を相互に移動させる。データを抽出する際は、L側の各フレームごとに周波数分割された上下の領域間で音圧を比較すれば行なえる。本手法はスピーカやマイクロフォンの周波数特性の影響を受け、特に携帯電話による運用は困難であることが判明し、同様に実用的ではない。

この周波数特性の影響を受けない手法として提案したのが、図1右(3)である。L側にデータを埋め込み、R側はL側で欠損する信号成分を補う構成は上記(2)と同様である。各フレームを時分割し、ビット0を埋め込む場合はL側の信号成分が前側に偏り、ビット1を埋め込む場合は後側に偏るように、L側とR側の信号成分を相互に移動させる。データを抽出する際は、L側の各フレームごとに時分割された前後の領域間で音圧を比較すれば行なえる。本手法は時間的な位置ずれの影響を受けやすいという欠点があったが、フレームの時分割において非対称窓関数を提案することにより、携帯電話でも抽出可能で上記(1)と(2)に比べ最も実用性があることが確認できた[3]。

上記(1)~(3)のいずれの手法においても、埋め込まれたデータはアナログ音響空間に放射されても消失することは無く、強力なアナログ耐性をもつ。また、埋め込み周波数帯を電話回線帯域以下に設定すれば、MP3等の各種音楽圧縮や3GPP携帯電話で採用されている低ビットレートの音声圧縮を施しても、データが埋め込まれている周波数帯域は殆ど影響を受けないため、各種データ圧縮に対するロバスト性もある。

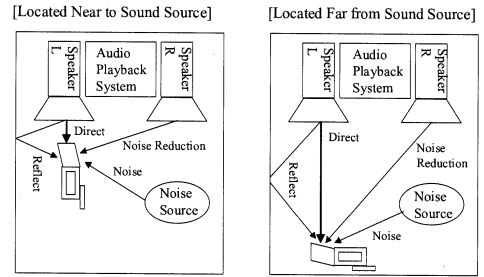


図2 先提案手法による電子透かし抽出時における端末位置の制約

3. 改良型の電子透かし埋め込み方式の概念

3.1. 先提案の問題点

図2は先に提案した図1(3)に基づく手法により、室内で携帯電話による電子透かし抽出を運用する上での、音源からの直接音以外の影響を示している。端末のマイクロフォンに入力される音源としては、左側スピーカからの直接音以外に右側スピーカからのノイズ抑圧信号音と室内外の騒音がある。これに加えて、室内の壁・天井・床を介した反射音および残響がある。図2左のように音源と端末間の距離が短ければ、直接音以外の音圧は低いため、抽出精度にあまり影響を与えないが、図2右のように距離が長くなると、無視できなくなり、ある程度以上離れると抽出不能になる。

この中で、第1の問題は、右側スピーカからのノイズ抑圧信号音で、これは左側スピーカに埋め込まれているデータパターンを打ち消すように設計されている。左右のスピーカが充分離れていても、室内残響が豊かであると、右側スピーカの影響が無視できなくなる。

従って、先提案のステレオ特性を用いた電子透かし埋め込み手法では音源と端末間の距離を離すことが基本的に難しい。これに加え、モノラル音源やヘッドフォン音源では運用できないという問題も抱える。ちなみに、イヤフォンやヘッドフォンでもマイクロフォンを密着させることにより、電子透かしを抽出することができ、図2に示されるような環境要因が存在しないため、高精度な抽出を実現できる。

仮に第1の問題が解決でき、左スピーカだけのモノラル音源による運用が実現できても、第2の問題として、左スピーカを音源とする室内反射音がある。これは、図1(3)に基づく時分割手法の欠

点で、位相遅れの音波が重なることにより抽出時に誤判断が発生しやすい。

本稿では、以上2つの問題を解決する新規な方式を提案する。

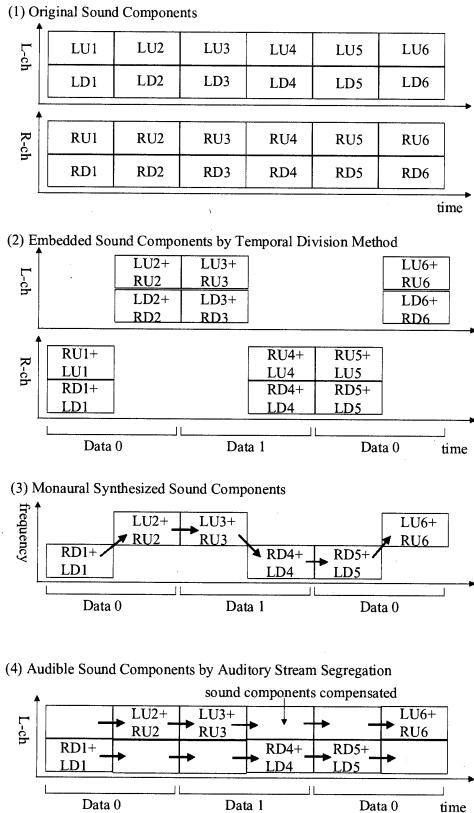


図3 ノイズ抑圧信号のモノラル信号への多重化プロセスと音脈分凝の発生

3.2. ノイズ抑圧信号の多重化

図3(1)(2)は図1(3)の時分割埋め込み方式により3ビットのデータを埋め込んだことを示している。尚、図3では各フレームを時間軸および周波数軸に2分割した4分割領域の信号成分を記号で表記している。図3(3)は図3(2)のR側成分をL側に周波数分割により多重化したものである。図3(3)の各フレームの上側はL側成分を使い、下側はR側成分に差し替えている。

図3(3)においては、図3(2)における信号成分の半分は欠落しており、かつ上下いずれかの信号成

分が0になっているため、このモノラル信号を再生するとノイズが発生する。しかし、図3(2)のL側だけをモノラル再生する音と実際に聴き比べると、ノイズレベルが予想以上に低減されることがわかる。

この顕著なノイズ低減効果は音響物理の範疇だけでは説明がつかず、聴覚心理学の分野で知られている音脈分凝[2]の効果が加わったものと考えられる。図3(3)では上下いずれか一方の音が鳴る単旋律の流れをもつが、音脈分凝により図3(4)に示されるように上下2つの複旋律の流れとしてヒトには聞こえていと推察される。即ち、データを埋め込むために故意に作成した空白の信号成分の箇所、ヒトは適当な音成分を前後関係から補間して聞いてくれていると考えられる。

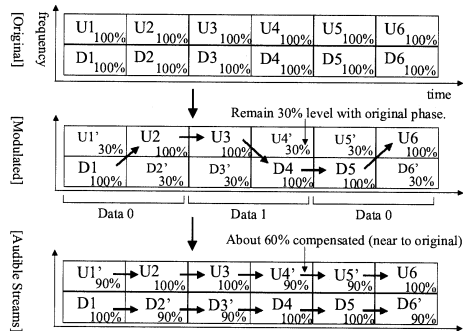


図4 音脈分凝を活用したモノラル信号への電子透かし埋め込みの概念

3.3. 音脈分凝によるノイズの完全抑圧

図3(4)の段階で、ノイズが若干残る原因としては、空白信号成分を完全に埋めることができないという理由のほかに、信号成分0の領域では位相情報をもたないため位相が不連続になるという理由も考えられる。そこで、図3(3)の信号成分0の箇所を完全に0にせず、原信号成分の位相を維持しながら若干残留させるようにすれば、ほぼ完全に補間して聞こえるようになると推察できる。

図4はソースをモノラル音響信号として、図3(3)のように各フレームを時間軸方向および周波数方向に各々2分割し、対角領域のレベルを増減させながらビットパターンを埋めるようにしたものである。ビット0を埋め込む場合はU1とD2のレベルを30%に減らし、ビット1を埋め込む

場合はD1とU2のレベルを30%に減らす。そうすると、減らされた成分は原音レベルまで補間されてヒトの耳に聞こえるためノイズは殆ど聴取されず、マイクロフォンには減らされた成分がそのまま収録されることが実験により確認できた。

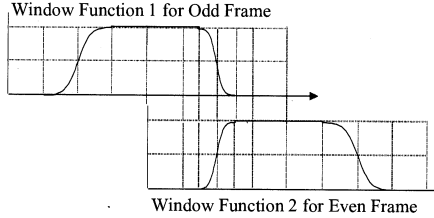


図5 提案する2つの窓関数

4. 改良型電子透かし埋め込み方式の詳細

4.1. 電子透かし埋め込みアルゴリズム

4.1.1. 音響信号の周波数次元変換

本節では以下、ソースがモノラル信号であることを前提に説明するが、ステレオ信号の場合は各チャンネルごとに同様な処理を行えばよい。サンプリング周波数 f_s のソース音響信号から窓幅 N のフレーム $X(j)$ ($j=0, \dots, N-1$) を $N/2$ 間隔で抽出し、奇数フレームと偶数フレームをセットにして1ビットを埋め込むものとする。 $f_s=44.1$ [kHz] のとき、 N は 4096 (10bps), 2048 (20bps), 1024 (40bps), 512 (80 bps) のいずれかを与える。図5に示される2種類の窓関数 $W_k(j)$ ($k=1, 2$) を用意し、奇数フレームに対して、第1番目の窓関数 $W_1(j)$ を適用し、偶数フレームに対しては第2番目の窓関数 $W_2(j)$ を用いて離散フーリエ変換を施し、変換後データの実部を $A_k(j)$ ($j=0, \dots, N-1$, $k=1, 2$), 虚部を $B_k(j)$ ($j=0, \dots, N-1$, $k=1, 2$) として、以下のように与えられる。

$$\begin{aligned} A_k(j) &= \sum_{i=0, N-1} W_k(j) X(i) \cos(2\pi ij/M), \\ B_k(j) &= \sum_{i=0, N-1} W_k(j) X(i) \sin(2\pi ij/M), \\ & (j=0, \dots, N-1, k=1, 2) \quad (1) \end{aligned}$$

ここで、図5に示される2種の窓関数 $W_k(j)$ ($k=1, 2$) は以下の通り、互いに $N/2$ ずらして重ねると相補的な関係になるように、ハニング窓関数 $W(i)=0.5-0.5\cos(2\pi i/M)$ を時間軸で非対称に変形させたものである。

$$\begin{aligned} W_1(i) &= 0.5-0.5\cos(4\pi i/N) \quad (i \leq N/4), \\ W_1(i) &= 1.0 \quad (N/4 < i \leq 11N/16). \end{aligned}$$

$$\begin{aligned} W_1(i) &= 0.5-0.5\cos(8\pi(i-9N/16)/N) \\ & \quad (11N/16 < i \leq 13N/16). \end{aligned}$$

$$W_1(i) = 0.0 \quad (13N/16 < i).$$

$$W_2(i) = 0.0 \quad (i \leq 3N/16).$$

$$W_2(i) = 0.5-0.5\cos(8\pi(i-3N/16)/N)$$

$$(3N/16 < i \leq 5N/16).$$

$$W_2(i) = 1.0 \quad (5N/16 < i \leq 3N/4).$$

$$W_2(i) = 0.5-0.5\cos(4\pi(i-N/2)/N) \quad (3N/4 < i). \quad (2)$$

4.1.2. 最適埋め込み割合の決定

奇数フレームおよび偶数フレームにおける周波数2分割バンドのエネルギー総和値 $E_{1\phi}$, $E_{1\psi}$, $E_{2\phi}$, $E_{2\psi}$ を計算する。 m と M を埋め込む周波数領域の下限値と上限値とし、 p を2分割バンド幅 $p=(M-m)/2$ とする。例えば、 $N=4096$ の場合、 $m=160$ (1.7kHz) に設定し、 p は2分割バンド幅で $p=(M-m)/2$, $p=160$ (1.7kHz) に設定する。

$$\begin{aligned} E_{1\phi} &= \sum_{j=m, m+p-1} \{A_1(j)^2 + B_1(j)^2\}, \\ E_{2\phi} &= \sum_{j=m, m+p-1} \{A_2(j)^2 + B_2(j)^2\}, \\ E_{1\psi} &= \sum_{j=m+p, m+2p-1} \{A_1(j)^2 + B_1(j)^2\}, \\ E_{2\psi} &= \sum_{j=m+p, m+2p-1} \{A_2(j)^2 + B_2(j)^2\}. \quad (3) \end{aligned}$$

続いて $\gamma = (E_{1\phi} + E_{2\psi}) / (E_{1\psi} + E_{2\phi})$ を算出し、埋め込みビットが1の場合は、 $\gamma = 1/\gamma$ に変換し、埋め込み割合 α , β の初期値を0.7, 0.3として、 γ が1.0未満の場合、

$$\alpha' = \alpha \cdot \gamma^{-1/2}, \quad \beta' = \beta \cdot \gamma^{1/2} \quad (4)$$

に補正する。ただし、パルス性ノイズの発生を抑えるため、 $\alpha' \leq 1.4$, $\beta' \geq 0.15$ に制限する。

4.1.3. 周波数成分の変換

上記算出した埋め込み割合 α , β を用いて、 $A_j(j)$, $A_2(j)$, $B_j(j)$, $B_2(j)$ の各周波数成分に対して、以下(5)式に基づき改変を行なう。

$$\begin{aligned} A_k'(j) &= A_k(j) \delta / \{A_k(j)^2 + B_k(j)^2\}^{1/2}, \\ B_k'(j) &= B_k(j) \delta / \{A_k(j)^2 + B_k(j)^2\}^{1/2}. \quad (5) \end{aligned}$$

1) ビット0を埋め込む場合:

$j=m, \dots, m+p-1$ の範囲で、 $k=1$, $\delta=\alpha$ として(5)式を実行し、 $j=m, \dots, m+p-1$ の範囲で、 $k=2$, $\delta=\beta$ として(5)式を実行し、

$j=m+p, \dots, m+2p-1$ の範囲で、 $k=1$, $\delta=\beta$ として(5)式を実行し、 $j=m+p, \dots, m+2p-1$ の範囲で、 $k=2$, $\delta=\alpha$ として(5)式を実行する。

2) ビット1を埋め込む場合:

$j=m, \dots, m+P-1$ の範囲で、 $k=1, \delta=\beta$ として (5) 式を実行し、 $j=m, \dots, m+P-1$ の範囲で、 $k=2, \delta=\alpha$ として (5) 式を実行し、

$j=m+P, \dots, m+2P-1$ の範囲で、 $k=1, \delta=\alpha$ として (5) 式を実行し、 $j=m+P, \dots, m+2P-1$ の範囲で、 $k=2, \delta=\beta$ として (5) 式を実行する。

4.1.4. 音響信号の時間次元逆変換

更新されるフレームのデータ $X'(i)$ は、直前の処理で逆変換されたフレームのデータ $X^p(i)$ に対して $N/2$ だけオフセットして加算される。奇数フレームの場合 $k=1$ 、偶数フレームの場合 $k=2$ として以下式を実行する。

$$X'(i) = 1/N \left\{ \sum_{j=0, N-1} A_k'(j) \cos(2\pi ij/M) - \sum_{j=0, N-1} B_k'(j) \sin(2\pi ij/M) + [1 - W_k(i)] X^p(i+N/2) \right\} \quad (6)$$

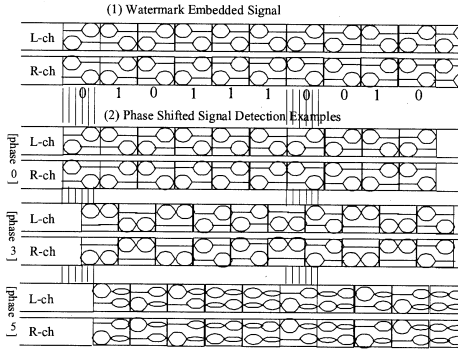


図6 電子透かし抽出時におけるフレーム位相補正処理

4.2. 電子透かし抽出アルゴリズム

4.2.1. 音響信号の周波数次元変換

電子透かしが埋め込まれた音響信号に対して、音源がステレオの場合はそれらを合成した信号に対して、モノラル信号として周波数 f_s でサンプリングを行ない、4.1.1 節と同様に、窓幅 N のフレーム $X(i)$ ($i=0, \dots, N-1$) を $N/2$ 間隔で抽出する。この時、埋め込み時の奇数フレームと偶数フレームに対応する区間が未知であるため、図6に示されるように位相 h を H 通り (上記例では $H=6$) にずらしながら、 H 通りの奇数フレーム ($k=1$) および偶数フレーム ($k=2$) のセットを用意する。そして (1) と同様に 2 つの窓関数 $W_k(i)$ ($k=1, 2$) を用いて、

各位相 h 別に離散フーリエ変換を施し、変換後データの实部を $A_k(j, h)$ 、虚部を $B_k(j, h)$ ($j=0, \dots, N-1, k=1, 2, h=0, \dots, H-1$) とする。

$$A_k(j, h) = \sum_{i=0, N-1} W_k(i) X(i+hN/H) \cos(2\pi ij/N) \\ B_k(j, h) = \sum_{i=0, N-1} W_k(i) X(i+hN/H) \sin(2\pi ij/N) \\ (j=0, \dots, N-1, k=1, 2, h=0, \dots, H-1) \quad (7)$$

4.2.2. 残響補正とビットデータの判定

前節で算出した各変換後データには残響成分として 1 フレーム前の各変換後データ $A_k^p(j, h)$ 、 $B_k^p(j, h)$ が割合 q ($0 \leq q \leq 1$) だけ含まれていると仮定し、以下のように補正を加えながらエネルギー成分を算出する。尚、 q の目安としては一般的な室内残響で運用する場合を $q=0.06 \cdot 4096/N$ とし、窓幅に反比例させた値を与える。

$$E_k(j, h) = A_k(j, h)^2 + B_k(j, h)^2 - q \{ A_k^p(j, h)^2 + B_k^p(j, h)^2 \} \quad (8)$$

次に、上記算出したエネルギー成分を奇数フレームと偶数フレームとの間で、周波数ごとに対応する要素を時間軸上で相互に大小比較し、大きい要素のカウント値として $C_1(h)$ 、 $C_2(h)$ 、 $C_3(h)$ 、 $C_4(h)$ を以下の通り算出する。

$j=m, \dots, m+P-1$ の範囲で、 $E_1(j, h) > E_2(j, h)$ を満たす周波数要素の個数を $C_1(h)$ とし、 $C_2(h) = p - C_1(h)$ とする。

$j=m+P, \dots, m+2P-1$ の範囲で、 $E_1(j, h) > E_2(j, h)$ を満たす周波数要素の個数を $C_3(h)$ とし、 $C_4(h) = p - C_3(h)$ とする。

また、 $E_1(j, h)$ 、 $E_2(j, h)$ の 2 分割周波数帯別の総和値を以下の通り算出する。

$$E_1(h) = \sum_{j=m, \dots, m+P-1} E_1(j, h) \\ E_2(h) = \sum_{j=m, \dots, m+P-1} E_2(j, h) \\ E_3(h) = \sum_{j=m+P, \dots, m+2P-1} E_1(j, h) \\ E_4(h) = \sum_{j=m+P, \dots, m+2P-1} E_2(j, h) \quad (9)$$

位相 h に対応する符号判定テーブルを $B(h)$ 、位相判定テーブルを $S(h)$ (初期値を全て 0 に設定) とすると、以下の規則で符号の仮判定を行なう。

- 1) $C_1(h) + C_4(h) > C_2(h) + C_3(h)$ の場合
 $B(h) = 0$
 $S(h) \leftarrow S(h) + E_1(h) + E_4(h) - E_2(h) - E_3(h)$
- 2) $C_1(h) + C_4(h) \leq C_2(h) + C_3(h)$ の場合
 $B(h) = 1$
 $S(h) \leftarrow S(h) + E_2(h) + E_3(h) - E_1(h) - E_4(h)$

全ての位相 h において上記仮符号を判定したら、初期フレームから積算される位相判定テーブル $S(h)$ の値が最大となる h_{max} を探索し、これを最適位相と判断し、そのときの候補符号 $B(h_{max})$ を本フレームにおける符号と確定する。

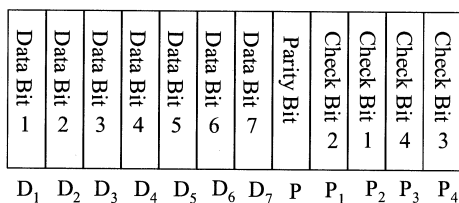


図7 バイトデータの埋め込みシーケンス

4.3. バイトデータの埋め込みシーケンス

前節までは、フレーム単位にビットデータを埋め込み、抽出するアルゴリズムを説明してきたが、本節では実用頻度の高い7ビットのASCII文字を誤り検出符号を付加して一連の信号に埋め込む手法について述べる。

通信の分野で一般に使われる誤り検出符号は、予測不能な種々の外乱によるビット誤りを対象としたもので、本稿のような電子透かし抽出においては図2に示されるようなものが該当する。しかし、本稿で提案する電子埋め込み手法においては、原信号により4.1.2節で述べた最適埋め込みを実現できない場合があり、例えば原信号レベルが0に近い区間ではデータを埋め込むことができない。これに対して、文献[3]~[5]では3値符号で埋め込む手法を提案したが、抽出時のロジックが複雑になりかえって誤判断が増大するという問題が露見した。そこで、本稿では4.1節で述べた通り、原信号の状態を問わず2値符号を埋め込むように修正した。そのため、原信号の状態により、適切な埋め込みが行なわれていないフレームが存在し、その箇所では抽出時にビットエラーになる可能性が高い。

また、シリアル通信においてはバイトデータの先頭ビットを示す同期符号を挿入する方法がとられることがあり、文献[4][5]の通り種々な埋め込みシーケンスを検討した。その結果、同期符号を挿入せず、誤り検出符号で代用する方法が最良と判断した。

このように、外乱によるビット誤り以外に、原信号レベルに起因する埋め込み処理の失敗、パイ

ト先頭ビットの識別という多機能をもたせた誤り検出符号を検討した結果、図7に示されるような7ビットのデータに5ビットの検査符号を付加した全12ビットで符号化する方法を採用した。

$D_1 \sim D_7$, および $P, P_1 \sim P_4$ は0または1の二進符号とし、 \oplus を排他的論理和とすると、 $P, P_1 \sim P_4$ は以下で定義される。 P はパリティ符号で、 $P_1 \sim P_4$ は4ビットのハミング符号である。

$$\begin{aligned}
 P &= D_1 \oplus D_2 \oplus D_3 \oplus D_4 \oplus D_5 \oplus D_6 \oplus D_7, \\
 P_1 &= D_1 \oplus D_2 \oplus D_3 \oplus D_7, \quad P_2 = D_1 \oplus D_4 \oplus D_5 \oplus D_7 \oplus P, \\
 P_3 &= D_2 \oplus D_4 \oplus D_6 \oplus D_7 \oplus P, \quad P_4 = D_3 \oplus D_5 \oplus D_6 \oplus P \quad (10)
 \end{aligned}$$

P が不合格で、 P_1 と P_2 , P_1 と P_3 , P_1 と P_4 , P_2 と P_3 , P_3 と P_4 , P_3 と P_4 のいずれか2ビットが不合格の場合、または P_1 と P_2 と P_3 , P_2 と P_3 と P_4 のいずれか3ビットが不合格の場合、各々 $D_1, D_2, D_3, D_4, D_5, D_6, D_7, P$ の1ビット誤りと特定でき、自動訂正可能である。

一度5ビットの検査符号が全て合格になれば、バイトデータの同期はとれていると判断できるため、それ以降は1ビットの自動訂正機能を働かせることができる。

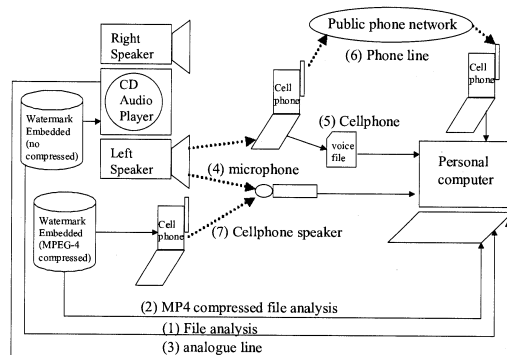


図8 各種録音形態による電子透かし抽出実験の構成

5. 提案手法の評価実験

文献[5]と同様に2分のCD音楽素材(44.1kHz/16bits/2-ch)に対して図1(3)の既提案の時分割埋め込み手法と図1(4)の本提案の手法を用いて、各々埋め込みフレーム幅 N を4096(10bps), 2048(20bps), 1024(40bps), 512(80bps)の4通りに変化させながら全8通りの手法で電子透かしを埋め込んだ。いずれにおいても、埋め込み周波数帯は1.7kHz~3.4kHzとし、80バイトのASCII

文字列を 4.3 節で述べた検査符号を付加して各々 12 ビットに符号化して繰り返し埋め込んだ。これら 8 種の音楽素材に対して、図 8 の実験系で以下で述べる全 7 通りの方法で抽出実験を行なった。

- (1) 埋め込み非圧縮音楽ファイルを直接デジタル解析して抽出する方法。
- (2) MPEG-4/AAC128[kbps] に圧縮された埋め込み音楽ファイルをデジタル解析して抽出する方法。
- (3) 再生している CD プレーヤからのライン出力を PC に接続し、アナログライン経由で抽出する方法。
- (4) 再生しているスピーカ (Sony SRS-Z510) の近くにマイクロフォン (Sony ECM-MS957) を設置して非接触に抽出する方法。
- (5) (4) と同様な再生系で携帯電話 (FOMA F901iS) のボイスレコーダ機能で録音を行ない、3GPP/AMR-NB 圧縮音声ファイルをパソコンで解析する方法。
- (6) (4) と同様な再生系で、2 台の携帯電話を通話状態にした状態で、スピーカで再生されている音を公衆電話回線経由でパソコン側でアナログキャプチャする方法。
- (7) (2) と同じ MPEG-4/AAC ファイルを携帯電話で再生し、携帯電話のスピーカ再生音をマイクロフォンで抽出する方法。

表 1 電子透かし抽出実験結果

Method	1	1	1	1	2	2	2	2
Frame Length	4096	2048	1024	512	4096	2048	1024	512
Embed Rate[bps]	10	20	40	80	10	20	40	80
Source	111	223	447	894	111	223	447	894
(1) File Analysis	106	221	447	894	102	220	440	874
(2) MP4 compressed	106	221	446	894	105	202	439	867
(3) Analogue line	111	218	444	884	101	215	437	841
(4) Microphone	105	214	440	881	104	215	439	849
(5) Cellphone	105	186	168	52	69	113	39	14
(6) Phone line	109	211	329	400	74	142	171	71
(7) Cellphone-SP	110	218	442	887	95	209	417	785

表 1 はその抽出結果で各数値は誤り検出に合格し、かつ原理埋め込み文字列と照合の上、誤りなく抽出できた文字数を示す。表 1 の Method-1 は既提案の時分割埋め込み手法で、Method-2 は本提案手法である。全体的に本提案の方が既提案より抽出精度が劣るが、携帯電話や電話回線経由といった

抽出項目(5)(6)を除けば顕著な差ではない。ただし、本稿で主目的にしている携帯電話による検出精度は既提案に比べ顕著な低下が見られるため、今後より一層の改善が求められる。

6. あとがき

本提案の改良型電子透かし埋め込み方式により、モノラル信号だけで聴取されるノイズを抑制しながら、従来と同等の最大 80 [bps] のレートでデータを埋め込むことが可能になった。これにより既存の手法に比べ、モノラル音源で運用でき、ステレオ音源の場合においても音源と受信端末間の距離を拡大でき、より自由度の高い運用が実現可能になった。

今後は本改良方式によるホストのコンテンツ品質の問題を詳細に評価し、特に携帯電話による抽出を想定したロバスト性を強化する方針である。具体的には、データ抽出可能な音源と端末間の距離、音源と端末間における外乱ノイズの影響、ビットエラーの特性等について評価を進め、最適なエラー訂正符号の設計を進める予定である。

また、ユビキタス音響空間を実現するため、抽出ソフトウェアを携帯電話等に実装し、民生用の安価なスピーカで再生中の音楽に関連する情報サイト URL 等を抽出できるようにする予定である。

文 献

- [1] 小野東：“第 4 章. 音に入る電子透かし”，「電子透かしとコンテンツ保護」，オーム社、第 1 版，pp.139-168, 2001.
- [2] 九州大学・中島祥好「聴覚心理学」研究室 <http://www.design.kyushu-u.ac.jp/~ynhome/JPN/index.html>
- [3] 茂出木敏雄：“携帯電話で非接触抽出可能な音楽への電子透かし埋め込み技術の開発”，電気学会・電子情報システム部門誌，Vol.126-C, No.7, pp.825-831, July, 2006.
- [4] Toshio Modegi and Makoto Chiba：“Nearly Lossless Audio Watermark Embedding Techniques to be Extracted Contactlessly by Cell Phone,” Proceedings of 3rd International Conference on Mobile Computing and Ubiquitous Networking, ICMU2006, pp.242-247, October, 2006.
- [5] 茂出木敏雄：“音響空間のユビキタス化に向けた電子透かし埋め込み容量の拡大技術”，電気学会・電子情報システム部門誌，Vol.127-C, No.7, July, 2007(掲載予定).