

## 音脈分凝を活用した携帯電話で抽出可能な音楽電子透かし技術

### "ゲンコーダ Mark"

茂出木 敏雄<sup>†</sup>

<sup>†</sup> 大日本印刷株式会社 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。はじめに、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た。その後、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手法を提案した。本稿では、この音脈分凝を二重に誘導させることにより、電話回線全帯域に対してデータを埋め込むことを可能にし、3G携帯電話で安定した抽出を実現できる改良手法について提案する。

## Audio Watermark Embedding Technique Applying Auditory Stream Segregation: "G-encoder Mark," Able to Be Extracted by Mobile Phone

Toshio MODEGI<sup>†</sup>

<sup>†</sup> Media Technology Research Center, Dai Nippon Printing Co., Ltd.  
3-5-20, Nishigotanda, Shinagawa-ku, Tokyo 141-8001, Japan  
E-mail: <sup>†</sup> Modegi-T@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. Then we have applied an auditory stream segregation phenomenon to reducing noises during monaural play-back operations. In this paper, we propose an improved method causing dual auditory stream segregation phenomena, which enables data embedding on whole public phone frequency ranges and stable extractions with 3-G mobile phones.

### 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ことにより、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[5]。更に、左右双方のスピーカからも抽出可能な抽出自由度の高い埋め込み方式も考案し、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[6]。

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。そこで、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した[4]。本提案では、多重化された信号パターンが、

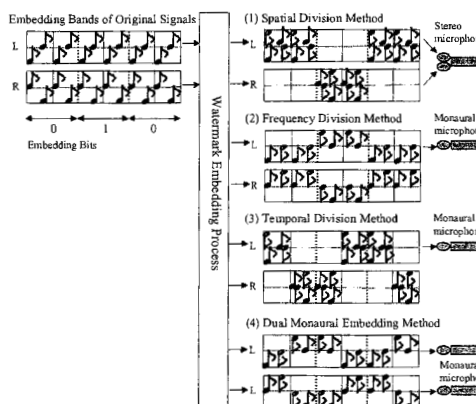


図1 既提案の4種の音楽電子透かし埋め込み方式の概念

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。しかし、抽出精度についてはそれまでの既提案方式に比べ劣るという問題があった。特に、本研究の主目的である携帯電話による抽出成績が顕著に劣るという問題があった。

本稿では、この音脈分凝が二重に引き起こされるようにして再生ノイズを抑圧しながら、埋め込み周波数帯域を電話回線帯域(300Hz~3.4kHz)の全体まで拡大させ、抽出感度を向上させた改良方式について提案する。

## 2. 先提案のステレオ音源に対する電子透かし埋め込み方式

図1(1)~(4)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ある[3][5][6]。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ビットずつ、「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、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、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、ヘッドホ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 (1)~(4)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。

### 3. 音脈分凝を活用したモノラル音源に対する電子透かし埋め込み方式

#### 3.1. 先提案の概要と問題点

図 1 右(2)~(4)方式ではデータの抽出にあたり左右一方のモノラル音源しか必要としないが、モノラルで再生すると顕著なノイズが聴取される。その中で、両側から抽出を可能にした図 1(4)については、モノラル再生時のノイズレベルが顕著に小さくなることが判明した[6]。特に図 1(3)と(4)の

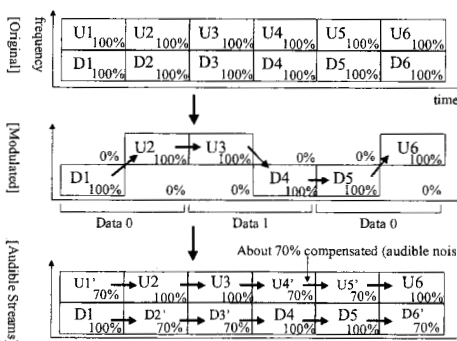


図 2 音脈分凝による電子透かし埋め込みノイズ抑圧の概念

L 側だけのモノラル再生音を比較すると、その差が著しい。

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 2 は図 1(4)の L 側チャンネル成分だけを書き直したもので、図 2 中段において U1, D2, D3, U4, U5, D6 の成分を 0 にするような改変を加えている。そうすると、図 2 中段に示されるように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 2 下段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ると推察される。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 (U1, D2, D3, U4, U5, D6) の箇所に、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 2 下段の段階で、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 2 中段の信号成分 0 の箇所を完全に 0 にせずに、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いう推察できる。

図 3 はそのように変更したもので、図 3 中段において、U1, D2, D3, U4, U5, D6 の信号レベルを 30% 程度残すような改変を加えている。そうすると、残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取

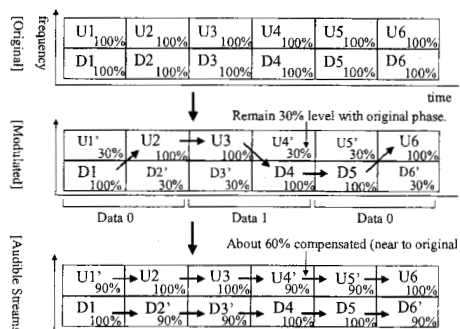


図 3 音脈分凝を活用したモノラル信号への電子透かし埋め込みの概念

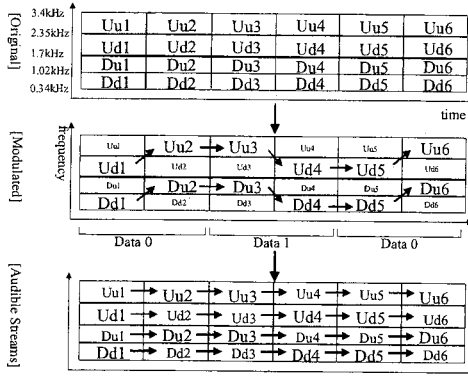


図4 二重音脈分凝を活用した広帯域電子透かし埋め込みの概念

録されることが実験により確認できた。ただし、変更された信号レベルの差が小さくなることにより抽出精度も悪くなり、文献[4]で報告の通り、図1(3)方式に比べ抽出成績が顕著に下がるという問題を抱えていた。

### 3.2. 改良方式の概要

前節で述べた方式を基本にして抽出感度を向上させるためには、埋め込み周波数帯を従来の1.7kHz～3.4kHzの全域に拡大すればよい。しかし、単に帯域を拡大すると埋め込みノイズも増大するため、本稿では低域側にもう1セットのデータ埋め込み領域を設定し、二重に音脈分凝が働くように、埋め込み周波数帯を電話回線帯域の全域に拡大する方法を提案する。

図4はその概念図で従来の1.7kHz～3.4kHzの帯域の下にほぼ半分の帯域幅をもつ埋め込み領域0.34kHz～1.7kHzを設け、上下で同一ビットパターンを重複して埋め込むようにする。実験の結果、このように上下パターンを同調させると、埋め込みノイズの発生は起こらなかった。本稿では以下本改良方式について詳細に述べる。

## 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  $X(i)$  ( $i=0, \dots, N-1$ ) 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 (80bps) のいずれかを与える。図5に示される2種類の窓関数  $W_k(i)$  ( $k=1, 2$ ) を用意し、奇数フレームに対して、第1番目の窓関数  $W_1(i)$  を適用し、偶数フレームに対しては第2番目の窓関数  $W_2(i)$  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。

$$A_k(j) = \sum_{i=0, N-1} W_k(i) \{X(i)+R(i)\} \cos(2\pi ij/M),$$

$$B_k(j) = \sum_{i=0, N-1} W_k(i) \{X(i)+R(i)\} \sin(2\pi ij/M),$$

$$(j=0, \dots, N-1, k=1, 2) \quad (1)$$

ここで、 $R(i)$  はソース音響信号に付加する白色雑音成分で、ソース音響信号が音声信号の場合など無音区間が多く存在する場合、データの埋め込みが適切に行えないため、必要に応じて付加する。ソース音響信号が16ビットの場合、 $R(i)$  として  $\pm 32$  以内の範囲で正負に変化する一様乱数を与えれば、音質にあまり影響を与えず、無音部にも適切にデータを埋め込むことができる。

2種の窓関数  $W_k(i)$  ( $k=1, 2$ ) 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。

$$W_1(i)=0.5-0.5\cos(4\pi i/N) \quad (i \leq N/4),$$

$$W_1(i)=1.0 \quad (N/4 < i \leq 11N/16),$$

$$W_1(i)=0.5-0.5\cos(8\pi(i-9N/16)/N) \quad (11N/16 < i \leq 13N/16),$$

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

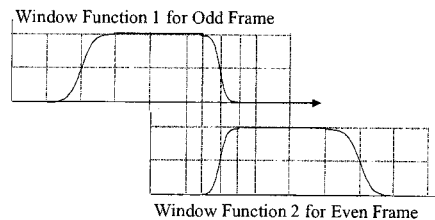


図5 提案する2つの窓関数

$$\begin{aligned}
W_2(i) &= 0.0 \quad (i \leq 3N/16). \\
W_2(i) &= 0.5 - 0.5 \cos(8\pi(i-3N/16)/N) \\
&\quad (3N/16 < i \leq 5N/16). \\
W_2(i) &= 1.0 \quad (5N/16 < i \leq 3N/4). \\
W_2(i) &= 0.5 - 0.5 \cos(4\pi(i-N/2)/N) \quad (3N/4 < i). \quad (2)
\end{aligned}$$

#### 4.1.2. 最適埋め込み割合の決定

奇数フレーム ( $k=1$ ) および偶数フレーム ( $k=2$ ) における周波数 4 分割バンドのエネルギー総和値  $E_{k1}, E_{k2}, E_{k3}, E_{k4}$  を計算する。 $m1, M1$  と  $mh, Mh$  を下位と上位の埋め込み周波数領域の各々下限値と上限値とし、 $p1, ph$  を各々の領域の 2 分割バンドとする。例えば、 $N=4096$  の場合、 $m1=32$  (0.34kHz)、 $M1=160$  (1.7kHz)、 $Mh=320$  (3.4kHz) に設定し、 $p1=(M1-m1)/2=64$  (0.68kHz)、 $ph=(Mh-mh)/2=160$  (1.7kHz) に設定する。

$$\begin{aligned}
E_{k1} &= \sum_{j=m1, m1+p1-1} \{A_k(j)^2 + B_k(j)^2\}. \\
E_{k2} &= \sum_{j=m1+p1, m1+2p1-1} \{A_k(j)^2 + B_k(j)^2\}. \\
E_{k3} &= \sum_{j=mh, mh+ph-1} \{A_k(j)^2 + B_k(j)^2\}. \\
E_{k4} &= \sum_{j=mh+ph, mh+2ph-1} \{A_k(j)^2 + B_k(j)^2\}. \quad (k=1, 2) \quad (3)
\end{aligned}$$

続いて  $\gamma = (E_{11}E_{22}E_{13}E_{24}) / (E_{12}E_{21}E_{14}E_{23})$  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  $\alpha, \beta$  の初期値を 0.7, 0.3 とし、 $\gamma$  が 1.0 未満の場合、

$$\alpha = 0.7 \gamma^{-1/4}, \quad \beta = 0.3 \gamma^{1/4} \quad (4)$$

に設定する。ただし、パルス性ノイズの発生を抑えるため、 $\alpha \leq 7.0, \beta \geq 0.03$  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  $\alpha, \beta$  を用いて、 $A_1(j), A_2(j), B_1(j), B_2(j)$  の各周波数成分に対して、以下 (5) 式に基づき変換を行なう。

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

1) ビット 0 を埋め込む場合：

$j=m1, \dots, m1+P1-1$  および  $j=mh, \dots, mh+Ph-1$  の範囲で、 $k=1, \delta=\alpha$  として (5) 式を実行し、 $k=2, \delta=\beta$  として (5) 式を実行し、

$j=m1+P1, \dots, m1+2P1-1$  および  $j=mh+Ph, \dots, mh+2Ph-1$  の範囲で、 $k=1, \delta=\beta$  として (5) 式を実行し、 $k=2, \delta=\alpha$  として (5) 式を実行する。

2) ビット 1 を埋め込む場合：

$j=m1, \dots, m1+P1-1$  および  $j=mh, \dots, mh+Ph-1$  の範囲で、 $k=1, \delta=\beta$  として (5) 式を実行し、 $k=2, \delta=\alpha$  として (5) 式を実行し、

$j=m1+P1, \dots, m1+2P1-1$  および  $j=mh+Ph, \dots, mh+2Ph-1$  の範囲で、 $k=1, \delta=\alpha$  として (5) 式を実行し、 $k=2, \delta=\beta$  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  $X'(j)$  は、直前の処理で逆変換されたフレームのデータ  $X^p(j)$  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。

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

### 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  $X(j)$  ( $j=0, \dots, N-1$ ) 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 (図 6 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) 式と同様に 2 つの窓関数  $W_k(j)$  ( $k=1, 2$ ) を用いて、各位相  $h$  別に離散フーリエ変換を施し、変換

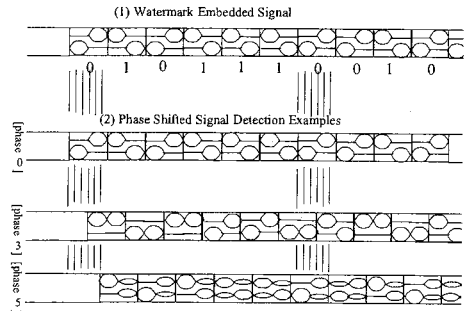


図 6 電子透かし抽出時におけるフレーム位相補正処理

後データの実部を  $A_k(j, h)$ 、虚部を  $B_k(j, h)$  ( $j=0, \dots, N-1, k=1, 2, h=0, \dots, H-1$ ) とする。

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

#### 4.2.2. 音響信号の仮想振幅変換

後述するビットデータの判定において隣接フレーム間での振幅大小比較を行なうにあたり、原信号の振幅変動成分を除外するため平坦化する前処理を行う。 $H$  通りの奇数フレームおよび偶数フレームにおけるエネルギー総和値を計算し、各々が既定値  $S_0$  になるようにスケール変換倍率  $S_1(h)$  と  $S_2(h)$  を決定する。

$$\begin{aligned} S_k(h) &= S_0 / [\sum_{j=m, M-1} \{A_k(j, h)^2 + B_k(j, h)^2\}]^{1/2}, \\ &\quad (k=1, 2, h=0, \dots, H-1) \end{aligned} \quad (8)$$

$j=m, \dots, M-1$  の範囲で、 $k=1, 2$  について  $A_k(j, h)$ ,  $B_k(j, h)$  の各要素に  $S_k(h)$  を乗じるスケール変換を施す。以下(9)式を用いて、各分割バンドごとに、スケール変換後の周波数成分  $A_k(j)$ ,  $B_k(j)$  に対するエネルギー総和値  $E_{k1}(h)$ ,  $E_{k2}(h)$ ,  $E_{k3}(h)$ ,  $E_{k4}(h)$  を計算する。

$$\begin{aligned} E_{k1}(h) &= \sum_{j=ml, ml+pl-1} \{A_k(j, h)^2 + B_k(j, h)^2\}, \\ E_{k2}(h) &= \sum_{j=ml+pl, ml+2pl-1} \{A_k(j, h)^2 + B_k(j, h)^2\}, \\ E_{k3}(h) &= \sum_{j=mh, mh+ph-1} \{A_k(j, h)^2 + B_k(j, h)^2\}, \\ E_{k4}(h) &= \sum_{j=mh+ph, mh+2ph-1} \{A_k(j, h)^2 + B_k(j, h)^2\}. \end{aligned} \quad (k=1, 2) \quad (9)$$

#### 4.2.3. 残響補正とビットデータの判定

前節で算出したエネルギー総和値には残響成分として1フレーム前の各エネルギー総和値  $E_{k1}^p(h)$ ,  $E_{k2}^p(h)$ ,  $E_{k3}^p(h)$ ,  $E_{k4}^p(h)$  が所定割合  $q$  ( $0 \leq q \leq 1$ ) 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。ただし、 $E_{k1}'(h) \geq 0$ ,  $E_{k2}'(h) \geq 0$ ,  $E_{k3}'(h) \geq 0$ ,  $E_{k4}'(h) \geq 0$  とする。

$$\begin{aligned} E_{11}'(h) &= E_{11}(h) - qE_{21}^p(h), & E_{21}'(h) &= E_{21}(h) - qE_{11}^p(h), \\ E_{12}'(h) &= E_{12}(h) - qE_{22}^p(h), & E_{22}'(h) &= E_{22}(h) - qE_{12}^p(h), \\ E_{13}'(h) &= E_{13}(h) - qE_{23}^p(h), & E_{23}'(h) &= E_{23}(h) - qE_{13}^p(h), \\ E_{14}'(h) &= E_{14}(h) - qE_{24}^p(h), \\ E_{24}'(h) &= E_{24}(h) - qE_{14}^p(h). \end{aligned} \quad (10)$$

本残響補正値を使用するか否かについてはフレームごとに位相適合性を含めて判断するようにし、以下位相適合性の評価式  $D(h)$ ,  $D'(h)$  を算出する。

$$\begin{aligned} D(h) &= [\{E_{11}(h)E_{22}(h)E_{13}(h)E_{24}(h)\}^{1/4} \\ &\quad - \{E_{12}(h)E_{21}(h)E_{14}(h)E_{23}(h)\}^{1/4}] \\ &\quad / [\{E_{11}(h)E_{22}(h)E_{13}(h)E_{24}(h)\}^{1/4} \\ &\quad + \{E_{12}(h)E_{21}(h)E_{14}(h)E_{23}(h)\}^{1/4}], \\ D'(h) &= [\{E_{11}'(h)E_{22}'(h)E_{13}'(h)E_{24}'(h)\}^{1/4} \\ &\quad - \{E_{12}'(h)E_{21}'(h)E_{14}'(h)E_{23}'(h)\}^{1/4}] \\ &\quad / [\{E_{11}'(h)E_{22}'(h)E_{13}'(h)E_{24}'(h)\}^{1/4} \\ &\quad + \{E_{12}'(h)E_{21}'(h)E_{14}'(h)E_{23}'(h)\}^{1/4}]. \end{aligned} \quad (11)$$

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  $T(h)$  (初期値を全て0に設定) とすると、以下の優先順位で符号の仮判定を行なう。

- 1)  $|D(h)| > |D'(h)|$  かつ  $E_{11}(h)E_{13}(h) > E_{21}(h)E_{23}(h)$  かつ  $E_{22}(h)E_{24}(h) > E_{12}(h)E_{14}(h)$  の場合：  
 $B(h)=0, T(h) \leftarrow T(h)+D(h)$ .
- 2)  $|D(h)| > |D'(h)|$  かつ  $E_{11}(h)E_{13}(h) < E_{21}(h)E_{23}(h)$  かつ  $E_{22}(h)E_{24}(h) < E_{12}(h)E_{14}(h)$  の場合：  
 $B(h)=1, T(h) \leftarrow T(h)-D(h)$ .
- 3)  $|D'(h)| > |D(h)|$  かつ  $E_{11}'(h)E_{13}'(h) > E_{21}'(h)E_{23}'(h)$  かつ  $E_{22}'(h)E_{24}'(h) > E_{12}'(h)E_{14}'(h)$  の場合：  
 $B(h)=0, T(h) \leftarrow T(h)+D'(h)$ .
- 4)  $|D'(h)| > |D(h)|$  かつ  $E_{11}'(h)E_{13}'(h) < E_{21}'(h)E_{23}'(h)$  かつ  $E_{22}'(h)E_{24}'(h) < E_{12}'(h)E_{14}'(h)$  の場合：  
 $B(h)=1, T(h) \leftarrow T(h)-D'(h)$ .
- 5)  $|D(h)| > |D'(h)|$  かつ  $E_{11}(h)E_{13}(h)E_{22}(h)E_{24}(h) > E_{21}(h)E_{23}(h)E_{12}(h)E_{14}(h)$  の場合：  
 $B(h)=0, T(h) \leftarrow T(h)+D(h)$ .
- 6)  $|D(h)| > |D'(h)|$  かつ  $E_{11}(h)E_{13}(h)E_{22}(h)E_{24}(h) \leq E_{21}(h)E_{23}(h)E_{12}(h)E_{14}(h)$  の場合：  
 $B(h)=1, T(h) \leftarrow T(h)-D(h)$ .
- 7)  $|D'(h)| > |D(h)|$  かつ  $E_{11}'(h)E_{13}'(h)E_{22}'(h)E_{24}'(h) > E_{21}'(h)E_{23}'(h)E_{12}'(h)E_{14}'(h)$  の場合：  
 $B(h)=0, T(h) \leftarrow T(h)+D'(h)$ .
- 8)  $|D'(h)| > |D(h)|$  かつ  $E_{11}'(h)E_{13}'(h)E_{22}'(h)E_{24}'(h) \leq E_{21}'(h)E_{23}'(h)E_{12}'(h)E_{14}'(h)$  の場合：  
 $B(h)=1, T(h) \leftarrow T(h)-D'(h)$ .

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  $T(h)$  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。

### 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 文字に誤り検出符号を付加した一連のバイトデータを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それ以外の要因でもビット誤りが発生する。例えば、原信号の状態により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4.1.1節で述べた白色雑音成分の加算が品質上行なえない場合、原信号レベルが0に近い区間ではデータを適切に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、本稿では文献[6]で提案した2値符号で埋め込む手法を採用した。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、非同期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が、文献[5]で報告の通り種々の埋め込みシーケンスを検討した。その結果、文献[6]のような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、バ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、文献[6]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。しかし、文献[6]のような組織的ハミング符号では符号間のハミング距離が最小3になり、誤り検出能力が弱いため、本稿では最小ハミング距離4を確保した非組織的ハミング符号を採用した。

表1に示される12ビットの各符号間は4以上のハミング距離を確保しており、7ビットの ASCII 符号を表1に従って12ビットに変換して埋め込むと、3ビットまでのエラー検出と1ビットのエラー自動訂正が行なえる。従って、一度12ビットのデータがエラー無しに抽出できれば、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。

表1 [12,7]ビット非組織的ハミング符号表

元符号	12ビット・ハミング符号							
00-07	1	14	50	61	84	91	103	104
08-0F	151	152	164	171	194	205	241	254
10-17	770	781	817	830	855	856	868	875
18-1F	916	923	935	936	961	974	1010	1021
20-27	1284	1291	1335	1336	1361	1374	1378	1389
28-2F	1426	1437	1441	1454	1479	1480	1524	1531
30-37	1543	1544	1588	1595	1618	1629	1633	1646
38-3F	1681	1694	1698	1709	1732	1739	1783	1784
40-47	2311	2312	2356	2363	2386	2397	2401	2414
48-4F	2449	2462	2466	2477	2500	2507	2551	2552
50-57	2564	2571	2615	2616	2641	2654	2658	2669
58-5F	2706	2717	2721	2734	2759	2760	2804	2811
60-67	3074	3085	3121	3134	3159	3160	3172	3179
68-6F	3220	3227	3239	3240	3265	3278	3314	3325
70-77	3841	3854	3890	3901	3924	3931	3943	3944
78-7F	3991	3992	4004	4011	4034	4045	4081	4094

### 5. 提案手法の評価実験

文献[4]と同様に2分のCD音楽素材(44.1kHz/16bits/2-ch)に対して文献[4]で提案した音脈分解を活用した埋め込み手法と本稿で提案した手法を用いて、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。埋め込み周波数帯は既提案法で 1.7kHz~3.4kHz、本提案法で 0.34kHz~3.4kHz とし、80バイトの ASCII 文字列を既提案および本提案の12ビット・ハミング符号に変換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7の実験系で以下で述べる全8通りの方法で抽出実験を行なった。実験は、当社 W:3×D:5×H:3 [m]の会議室で行い、スピーカとマイクロフォンとの距離は10cm離して抽出した。また、本提案手法については、音源よりマイクロフォンを離して抽出することが可能になるため、以下(4)と同一の装置を

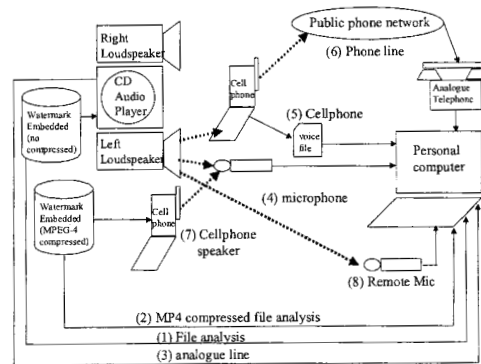


図7 各種録音形態による電子透かし抽出実験の構成

用いて、マイクロフォンをスピーカより3m離して同様な抽出を試みた。

(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(F901s)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、音源側に携帯電話、パソコン側にアナログ固定電話を設置し、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。

(8)(4)と同様な再生系で、マイクロフォンをスピーカより3m離して設置し遠方より抽出する方法。

表2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は文献[4]で提案した埋め込み手法で、Method-2は本提案の改良手法である。全体的に本提案は既提案より抽出精度が顕著に改善されている。特に本稿で主目的にしている携帯電話の検出精度についてその差が著しく、本提案手法では40bpsでも安定して抽出が行なえるという結果が得られている。また、3m離れた遠隔抽出においては、(6)の電話回線経由の抽出とほぼ同等な抽出精度が得られている。

## 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、モノラル再生音で聴取されるノイズを抑圧しながら、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でき、40[bps]以下のレートであれば携帯電話でも安定して抽出できることが確認できた。また、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歪の問題を詳細に評価し、特に携帯電話による抽出を想

表2 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	102	220	440	874	110	220	442	884
(2) MP4 compressed	105	202	439	867	110	220	442	884
(3) Analogue line	101	215	437	841	106	215	428	838
(4) Microphone	104	215	439	849	109	220	440	879
(5) Cellphone	69	113	39	14	109	220	424	570
(6) Phone line	74	142	171	71	108	219	392	426
(7) Cellphone-SP	95	209	417	785	109	217	422	828
(8) Remote Mic					108	209	349	277

定したロバスト性を強化する方針である。具体的には、データ抽出可能な最小録音時間および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、最適な埋め込み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、携帯端末を音源に近づけなくても、できれば衣服のポケットやカバンに入れた状態で、周囲に流れているBGMやアナウンスに埋め込まれた情報をリアルタイムに受信できるようにすることを目標に開発を進めて行く予定である。

## 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] 茂出木敏雄・千葉誠：“音脈分凝を活用した音楽電子透かし技術“ゲンコードMark”の開発，”情報処理学会・音楽情報科学研究報告，Vol.2007，No.37，pp.89-96，May，2007.
- [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，pp.1013-1021，July，2007.
- [6] Toshio Modegi：“Construction of Ubiquitous Acoustic Spaces Using Audio Watermark Technology and Mobile Terminals，”IEEJ Transactions on Electrical and Electronic Engineering，Vol.2，No.6，November，2007 (to be published).