

特異スペクトル分析に基づいた音響ゼロ電子透かし法

渡辺 瑠伊 鵜木 祐史

北陸先端科学技術大学院大学 〒923-1292 石川県能美市旭台 1-1

E-mail: {s2210191, unoki}@jaist.ac.jp

あらまし 近年の生成 AI 技術の発展は、ディープフェイクによる音声改ざんなどの新たな脅威を生み出しており、コンテンツの真正性を保証する技術が不可欠となっている。本稿では、このような背景から、スパースコーディングを用いた音響ゼロ電子透かし法を提案する。提案法では、時系列データ解析法であり、スパースコーディング法の一つである特異スペクトル分析を用いて音声信号を複数の成分に分解し、信号の本質的特徴を捉える主要な成分から 1-dimension local binary pattern によってバイナリパターンを生成する。そして、このバイナリパターンと透かし情報との排他的論理和をとることで検出鍵を生成する。性能評価の結果、提案法は MP3 や AAC 符号化といった非攻撃的処理に対して、従来法よりも高い頑健性を示すが、過度な頑健性を有するため意図的な改ざん攻撃に対して検出感度が低下することもあった。

キーワード 特異スペクトル分析, 音響ゼロ電子透かし, スパースコーディング, 音声改ざん検出

Audio zero-watermarking method based on singular spectrum analysis

Rui WATANABE and Masashi UNOKI

Japan Advanced Institute of Science and Technology 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan
E-mail: {s2210191, unoki}@jaist.ac.jp

Abstract Recent advancements in generative AI technology have created new threats such as deep-fakes and speech tampering, making it essential to have technologies that can guarantee the authenticity of content. Against this backdrop, this paper proposes an audio zero-watermarking method using sparse coding. The proposed method utilizes singular spectrum analysis (SSA) as one of sparse coding and time-series analysis methods, to decompose an audio signal into multiple components. It then generates a binary pattern from the principal components, which capture the essential features of the signal, using the 1-dimension local binary pattern. A detection key is subsequently created by performing an eXclusive OR (XOR) operation between this binary pattern and the watermark information. Evaluation results showed that the proposed method demonstrates robustness against non-attacking such as MP3 and AAC encoding, compared to conventional methods. On the other hand, it was also found that detection sensitivity to tampering attacks was insufficient due to extreme robustness against any attacks.

Key words singular spectrum analysis, audio zero-watermarking, sparse coding, speech tampering detection

1. はじめに

近年の生成系 AI の普及により、誰でも容易に動画像や音声を作成できるようになり、コンテンツ制作等の多岐に渡る分野で活用されている。その一方で、生成系 AI による情報操作やオリジナルのメディアの改ざんなど情報悪用の問題が生じている。この問題を解決するため、デジタルコンテンツの真正性や完全性を保証する技術の一つとしてゼロ電子透かし技術がある。その中でも、音響ゼロ電子透かしは音響信号に透かし情報を埋め込むことで信号の真正性や完全性を保証する技術である。

音響信号に情報を埋め込む方法として、振幅情報に直接埋め

込む方法 [1] や、聴覚の蝸牛遅延数特性を利用する方法 [2] など様々なものが提案されている。音響電子透かしでは、直接信号に情報を埋め込むため、メディアの品質劣化が生じる。

この問題を解消する方法として、音響ゼロ電子透かしがある。この方法では、音響信号に直接情報を埋め込まず、信号から得られる特徴量を用いて検出鍵を作成する。そして、検出鍵を利用することで透かし情報を抽出することができる。代表的な方法として、信号の振幅情報を利用する方法 [3] や、離散ウェーブレット変換と離散コサイン変換を利用する方法 [4] がある。

一方、聴覚的スペクトル表現に基づく音響ゼロ電子透かし法が提案されている。この方法は、信号を複数基底で表現するス

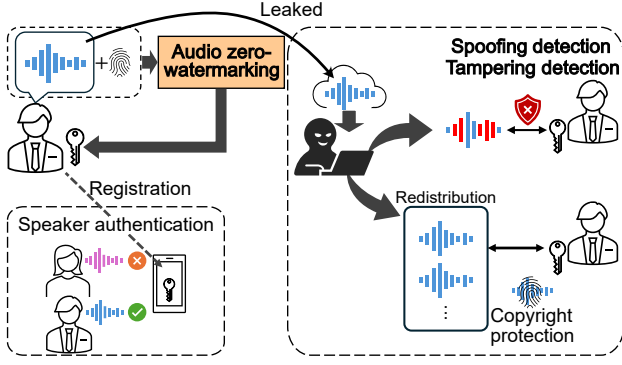


図 1: 音響ゼロ電子透かし技術の応用例

パースコーディングの一つとして Spikegram を利用した検出鍵の作成や透かし情報の検出を行うものである。この方法は、前述した音響ゼロ電子透かし法 [3], [4] と比較し頑健な透かしの情報検出が可能であるが、符号化処理の種類によって頑健であるとはいえず、音声改ざん検出の応用で問題となる。

本稿では、Spikegram を利用した音響ゼロ電子透かし法の残された課題を解決するために、他の効果的なスパースコーディングを用いた音響ゼロ電子透かし法を提案する。そして、評価シミュレーションによる性能評価を行い、既存法との比較を通じて提案法の実現可能性と今後の課題について考察する。

2. 音響ゼロ電子透かし技術の利用シーン

本稿で想定する音響ゼロ電子透かし法の応用例を図 1 に示す。この技術では、ユーザが保護したい音声データとそれを証明するための透かし情報を用いて検出鍵を作成する。音声データがインターネット上に公開され、悪意を持った第三者が改ざん処理を加えた場合、持ち主である本人が生成した鍵を用いて照合することで、なりすましや改ざんの検知が可能となる。また、データが複製され二次的に流布された場合も、検出鍵を用いて同一のデータであることを証明することで、著作権の保護が可能となる。その他の応用例として、話者認証にも利用される。予め端末に音声データと検出鍵を登録しておくことで、端末の所有者の声のみに反応するような認証技術が実現可能となる。

これまでに、Spikegram を用いた音響ゼロ電子透かし法 [5] が提案されている。Spikegram とは、聴覚の神経発火を模した Spike を横軸時間、縦軸チャネルの二次元領域に配置したものである。各 Spike の発火強度や発火タイミングや位置を特徴量として信号処理に利用することが可能である。また、Spike は反復処理によって算出され、反復初期に得られる Spike ほど強度が強く、信号にとって重要な要素である。この方法では、Spikegram からバイナリパターンを生成し、透かし情報 (画像) との排他的論理和 (XOR) を算出することで検出鍵を作成し、対象信号に対する音響ゼロ電子透かしを適用する。

この方法では、代表的な音響ゼロ電子透かし法 [3], [4] と比較して多くの音情報処理に対し頑健性が高いことや、音声改ざん検出技術に応用可能であることが実験により明らかとなった。しかし、一部の音声符号化では未だ頑健性を得られていない点やバイナリパターン生成時の二値化処理において Spike の特徴量を全て反映できていないことによって誤差なく透かし情報を

検出できない場合があることや、改ざん攻撃に対する検出精度が低い点が課題として残っている。

本稿では、これらの問題点を解決するため、音声信号のスパース性に基いた新たな音響ゼロ電子透かし法を提案する。

3. 特異スペクトル分析

特異スペクトル分析 (singular spectrum analysis: SSA) [6] は、時系列データの解析に使用されるスパースコーディング法の一つである。この方法は、時系列データからパターン構造を抽出し、トレンド、周期性、及びノイズに分解する。SSA は、主に次の四つの処理で構成される。

(1) **埋め込み** 長さ N の時系列データ $\{x_n\}_{n=1}^N$ を考え、遅延子 L ($2 \leq L \leq N/2$) を選択し、時系列を L 次元ベクトルに埋め込む。具体的には、以下のように $K = N - L + 1$ 個の遅延ベクトル $X_i \in \mathbb{R}^L$ を生成する。

$$X_i = (x_i, x_{i+1}, \dots, x_{i+L-1})^T, \quad i = 1, 2, \dots, K \quad (1)$$

これらのベクトルを持つ $L \times K$ の軌跡行列 X を構成する。

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{bmatrix} = [X_1 : X_2 : \cdots : X_K] \quad (2)$$

ここで、軌跡行列はハンケル行列 (各反斜辺上の要素が全て等しいという特徴をもつ行列) となる。

(2) **特異値分解** 軌跡行列 X に対して、次式のように特異値分解 (singular value decomposition: SVD) を適用する。

$$X = U \Sigma V^T = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T \quad (3)$$

ここで、 $U = [U_1 : U_2 : \cdots : U_d]$ は $L \times d$ の左特異ベクトルからなる行列、 $\Sigma = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_d})$ は特異値 $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \cdots \geq \sqrt{\lambda_d} > 0$ を対角成分に持つ $d \times d$ の対角行列 ($d = \text{rank}(X)$)、 $V = [V_1 : V_2 : \cdots : V_d]$ は $K \times d$ の右特異ベクトルからなる行列である。また、 $\{U_i\}$ と $\{V_i\}$ はそれぞれ正規直交基底である。

(3) **グループ化** $I = \{1, 2, \dots, d\}$ を m 個の互いに素な部分集合 $I_1, I_2, \dots, I_m$ に分割する。各部分集合 I_p に対して、対応する行列を $X_{I_p} = \sum_{i \in I_p} \sqrt{\lambda_i} U_i V_i^T$ とする。これにより、元の行列 X を m 個の行列の和に分解できる。

$$X = X_{I_1} + X_{I_2} + \cdots + X_{I_m} \quad (4)$$

この操作は、グループ化と呼ばれ、時系列の異なる成分 (トレンド、周期成分、雑音成分) に分解する。

(4) **対角平均化** 各グループの行列 X_{I_p} に対して、対角平均化を適用して、対応する成分を再構築する。 $L \times K$ 行列 Y の要素を y_{ij} とし、 $L^* = \min(L, K)$, $K^* = \max(L, K)$, $N = L + K - 1$ とする。また $y_{ij}^* = y_{ij}$ ($L < K$ の場合) $y_{ij}^* = y_{ji}$ ($L > K$ の場合) とする。対角平均化では、次式のよ

うにして行列 Y から長さ N の時系列 $\tilde{y}_k$ を生成する.

$$\tilde{y}_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k y_{m,k-m+1}^* & (1 \leq k < L^*) \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+1}^* & (L^* \leq k \leq K^*) \\ \frac{1}{N-k+1} \sum_{m=k-K^*+1}^{N-K^*+1} y_{m,k-m+1}^* & (K^* < k \leq N) \end{cases} \quad (5)$$

この対角平均化により, 各部分行列 X_{I_p} から長さ N の時系列を得る. これらを足し合わせることで, 元の時系列の近似再構成が可能となる.

3.1 SSA による音声信号の分解例

音声信号に対して SSA を適用し, 各成分に分解した例を図 2 に示す. 図の最上部のパネルは, SSA の対象となる音声信号を示し, 下部のパネルは, $L = 5$ として SSA を用いて, 特異値の大きい順に, 分解された各成分を示す. 特異値が最大となる成分 (a) は信号全体の概形 (トレンド成分) を表現している. 成分 (b) ~ (d) は, 高調波などの周期振動成分となり, 特異値が最小となる成分 (e) は残差や環境雑音といった雑音成分となる. 特異値の大きさは元信号を構成するうえでの寄与度を表しており, 寄与度の高い成分を用いることで, 元信号に近い品質で再合成が可能となる. 従って, SSA に基づいて分解された上位成分には信号の本質的な特徴が集約されると考える. 上位成分に透かし情報を埋め込むことで, 様々な音情報処理に対して頑健性を獲得できると期待される. また, 埋め込む際に, それぞれの成分に対して埋め込み処理を行うことで, 音情報処理によって埋め込み情報が崩れた成分があったとしても, 各成分を重畳することで誤りを補間でき, 頑健となることが期待できる.

3.2 スパースコーディング法による音声信号の分析・再構成

音声信号の再構成品質の観点から, 音響ゼロ電子透かしに有効な方法であるかを検討する. SSA によって, 信号を分解し得られた成分を用いて再び元信号を構成するように信号を再構成する. この時, 元信号を完全に再現できていない場合, 対象の信号に重要な特徴量を正しく抽出し成分分解できていないといえる. つまり, 音響ゼロ電子透かしに利用するにあたり, 音声信号の本質的な特徴量を捉えることができないため, 符号化処理などが施された場合, 頑健性を獲得できない. このような観点から SSA について信号の再構成品質の評価を行った.

評価指標には, 時間領域での信号間の誤差を表す信号対誤差比 (signal to error ratio: SER) [7], スペクトル領域での歪みを表す対数スペクトル歪み (log spectral distance: LSD) [8], 音声品質の客観評価指標である perceptual evaluation of speech quality (PESQ) [9] を用いた. 元信号を $x(n)$, 再構成した信号を $\hat{x}(n)$ とすると, SER は次のように算出される.

$$\text{SER} = 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} |x(n)|^2}{\sum_{n=0}^{N-1} |x(n) - \hat{x}(n)|^2} \right) \quad (6)$$

元信号の対数パワースペクトルを $S(k)$, 再構成した信号の対数パワースペクトルを $\hat{S}(k)$ とすると, LSD は次式のように算出される.

$$\text{LSD} = \sqrt{\frac{1}{K} \sum_{k=1}^K (S(k) - \hat{S}(k))^2} \quad (7)$$

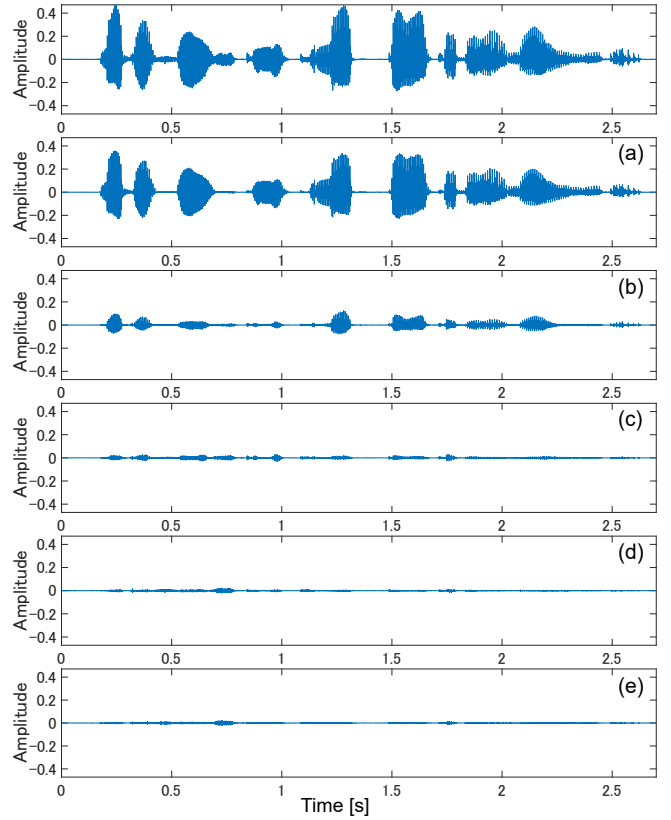


図 2: SSA による音声信号の分析

これらの音質評価指標は一般に, SER は 20 dB 以上, LSD は 0.1 dB, PESQ は 4.5 以上であれば高品質な信号の再合成であるとされている.

男女 100 名分の新聞読み上げ音声データセット [10] に対し SSA を用いて信号の再構成品質を測った. その結果, SER は 303.41 dB, LSD は 0.00 dB, PESQ は 4.63 であった. 従って, SSA では元信号の重要な成分を欠落させることなく成分分解が可能であり, これらの特徴量を音響ゼロ電子透かしに利用できるものと考えられる.

4. 提案法

本稿では, SSA を用いた音響ゼロ電子透かし法を提案する. 提案手法のフレームワークを図 3 に示す. 図 3(a) の埋め込み処理では, まず音声信号に SSA を施し複数の成分に分解する. 次に, 各成分の特異値から元信号への寄与率を算出し, 寄与率の大きい上位成分を選択する. そして, 選択された成分から 1-dimension local binary pattern (1D-LBP) を用いてバイナリパターンを生成する. 最後に, このバイナリパターンと透かし情報 (0/1 の二値画像) との XOR により検出鍵を生成する.

4.1 透かし情報の埋め込みと検出鍵生成

図 3(a) に示す検出鍵の生成法では, 次の手順で処理を行う. まず, 音声信号を $x(t)$ とし, $x(t)$ に対して SSA を施し, 各成分 $Q(L)$ を得る. SSA における分解成分数 L は任意に設定可能なパラメータである. 次に, 得られた $Q(L)$ のうち, 重要成分を選択する. 各成分の特異値に基づき元信号への寄与率を算出し, 累積寄与率が 90% に達するまでの成分を選択する. その後, 選択した各成分に対して 1D-LBP を用いてバイナリパ

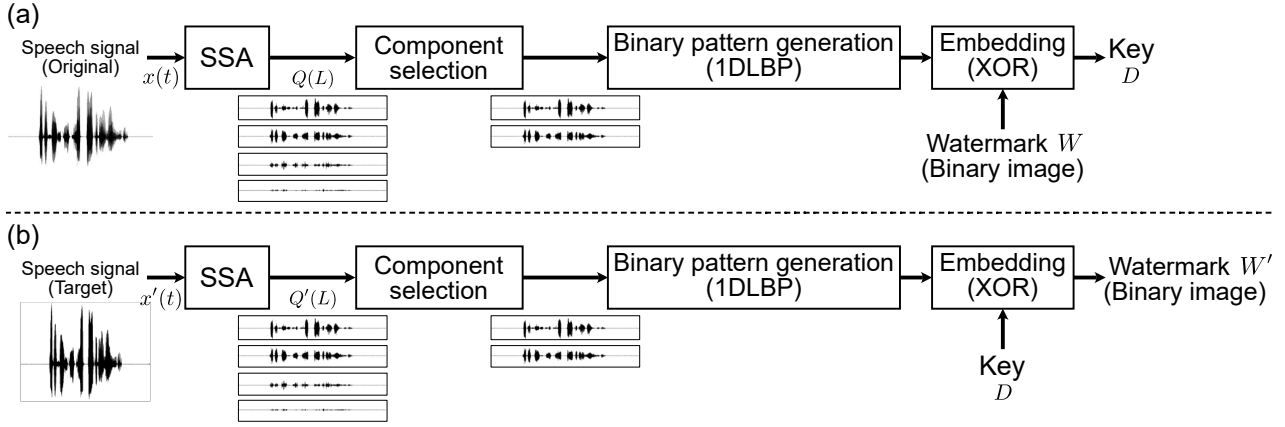


図 3: 提案法の概要図：(a) 透かし情報の埋め込み処理, (b) 透かし情報の検出処理

ターン $B(l) \in \{0, 1\}^{2^P}$ を生成する．ここで P は 1D-LBP の近傍点数である．続いて，選択した成分から得られたバイナリパターン群 $\{B(l)\}$ の各次元で多数決処理を施し，バイナリベクトル $B \in \{0, 1\}^{2^P}$ を生成する．例えば， $B(l)$ の対応する次元のバイナリパターン列が $\{1, 1, 0, 1\}$ であった場合，多数決処理によって値として 1 を得る．最後に，透かし情報であるバイナリ画像を平坦化したベクトル $W \in \{0, 1\}^{2^P}$ と，集約されたバイナリベクトル $B \in \{0, 1\}^{2^P}$ との XOR を算出することにより，検出鍵 $D \in \{0, 1\}^{2^P}$ を得る．

4.2 二値化処理方法

本稿では，バイナリパターン生成に 1D-LBP を用いる．LBP [11] は画像用テクスチャ特徴量であり，本稿では時系列信号に適用可能な 1D-LBP [12] を用いる．1D-LBP はノイズ耐性を有し，信号の時間的局所構造を捉えるため，頑健なバイナリパターン生成に適していると考えられる．

長さ N の信号 $x = \{x_0, x_1, \dots, x_{N-1}\} \in \mathbb{R}$ を対象とし，解析中心点を i_c とおく．左右 R 点の近傍（計 $P = 2R$ 点）の集合

$$\mathcal{N}(i_c) = \{i_c - R, i_c - R + 1, \dots, i_c - 1, i_c + 1, \dots, i_c + R\} \quad (8)$$

を用いる．そして， $p = 0, 1, \dots, P - 1$ に対し近傍インデックス $i_p = i_c - R + p + 1_{p \geq R}$ を導入し，以下の条件で 1 ビット符号 b_p を得る．

$$b_p = \begin{cases} 1 & (x_{i_p} \geq x_{i_c}) \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

ビット列 $\{b_p\}$ を下位ビット優先で重み付けすると，1D-LBP 値

$$\text{LBP}_{P,R}(i_c) = \sum_{p=0}^{P-1} b_p 2^p \quad (10)$$

が得られる．取り得る値は $0 \leq \text{LBP}_{P,R} \leq 2^P - 1$ であり， P ビットの局所形状を一つの整数に圧縮している．

この 1D-LBP 値に対して，ヒストグラム（度数分布）

$$H(k) = \sum_{i_c \in S} \delta(\text{LBP}_{P,R}(i_c), k), \quad k = 0, \dots, 2^P - 1 \quad (11)$$

を求める．ここで S は信号 x 中の解析可能な中心点 i_c の集合

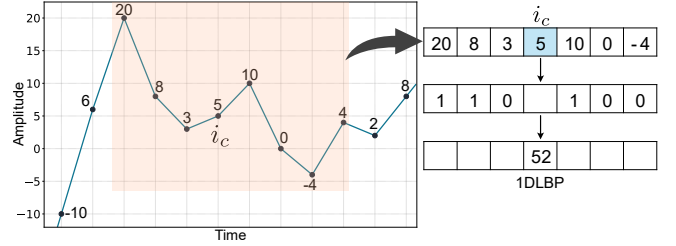


図 4: 1D-LBP の算出方法

である． $H(k)$ の全 2^P ビンにわたる出現回数の平均

$$\bar{H} = \frac{1}{2^P} \sum_{k=0}^{2^P-1} H(k) \quad (12)$$

を閾値として，最終的に成分ごとのバイナリパターン $B(l)_k = \Theta(H(k) - \bar{H})$ を得る． $\Theta(\cdot)$ はヘヴィサイド関数である．これによりベクトル $B(l) \in \{0, 1\}^{2^P}$ が定まる．

図 4 に提案法の 1D-LBP の算出法の例を示す． $R = 3$ とし，解析中心点 i_c が図 4 のような位置にあるとき，得られるビット列は $\mathbf{b} = \{1, 1, 0, 1, 0, 0\}$ となる．これを 2 進数とみなし，式 (10) のように，解析点 i_c の LBP を算出する．この処理を 1 サンプルずつずらしながら全サンプル点について計算を行う．

4.3 透かし情報の検出

図 3(b) に透かし情報の検出法を示す．ここでは，以下の処理を行う．まず，検出対象の信号を $x'(t)$ とし， $x'(t)$ に対して SSA を施し各成分 $Q'(L)$ を得る．次に，得られた $Q'(L)$ に対し，埋め込み時と同様の基準（累積寄与率 90%）で重要成分を選択する．選択した各成分に 1D-LBP を適用してバイナリパターン $B'(l) \in \{0, 1\}^{2^P}$ を生成する．これら $B'(l)$ に対し，埋め込み時と同様に多数決処理を施し，集約バイナリベクトル $B' \in \{0, 1\}^{2^P}$ を得る．最後に，この集約バイナリベクトル $B' \in \{0, 1\}^{2^P}$ と検出鍵 $D \in \{0, 1\}^{2^P}$ との XOR を算出することによって，透かし情報 $W' \in \{0, 1\}^{2^P}$ を検出する．

5. 評価シミュレーション

提案法について，非攻撃的処理に対する頑健性と攻撃的処理に対する検出性能の評価を行う．評価に用いた様々な音情報処理の種類を表 1 に示す．非攻撃的処理では，透かし情報を検出

表 1: 評価に使用する非攻撃的処理及び攻撃的処理の種類

Non-attack processing	
No processing (NOP)	
Low-pass filter (LPF)	Cut-off frequency: 8 kHz
Resampling (RSM _d)	16 kHz → 8 kHz → 16 kHz
Resampling (RSM _u)	16 kHz → 44.1 kHz → 16 kHz
Requantization (RQZ _d)	16 bit → 8 bit → 16 bit
Requantization (RQZ _u)	16 bit → 24 bit → 16 bit
MP3 coding (MP3 _n)	Bit-rate n kbps (n : 32, 64, 128, 256)
AAC coding (AAC _n)	Bit-rate n kbps, (n : 32, 64, 128, 256)
G.711 coding (G.711)	$F_s = 8$ kHz, log-PCM, μ -law
G.729 coding (G.729)	$F_s = 8$ kHz, CELP
Attack processing	
Zero interpolation (ZI)	Zero padding and interpolation
Pitch shift (PSH _{±20%})	Fundamental frequency shifted by $\pm 20\%$
Replacement (RPM)	Sample replacement w/ synthetic speech
Cropping (CRP)	Remove a specified segment

する際、なるべく誤差が少ないことが望ましい。逆に、攻撃的処理では、検出された透かし情報が元信号の透かし情報と大きく異なることが望ましい。

本稿では、Spikegram を用いた音響ゼロ電子透かし法（従来法）を比較対象として、非攻撃的処理に対しての頑健性及び攻撃的処理に対しての検出精度について評価を行う。

5.1 評価方法

評価用データセットには、新聞読み上げコーパスを用いた。これは男性及び女性の文章読み上げ音声を含み、男女各 50 名ずつ計 100 名分収録された。サンプリング周波数は 16 kHz、量子化ビット数は 16 bit のモノラル信号であった。このデータセットに対して表 1 にある音情報処理を施し、評価用信号を作成した。音声信号に SSA を施す際、分解する成分数は 30 とした。また、埋め込みを行う透かし画像のサイズは 16×16 , 32×32 , 64×64 の 3 種類を用いた。

評価指標に、Bit error rate (BER) を用いて、従来法と提案法の性能比較を行った。サイズ $m \times m$ の透かし情報 W と W' の誤差を測る際 BER (%) は以下のように算出された。

$$\text{BER} = \frac{\sum_{i,j} (w_{i,j} \oplus w'_{i,j})}{m \times m} \times 100 \quad (13)$$

$\oplus$ は排他的論理和を表し、以下のように定義される。

$$a \oplus b = (a + b) \bmod 2 \quad (14)$$

情報通信分野において、一般的に BER が 1% 以下であれば、情報の完全復元が十分に可能である。従って、本稿では BER が 1% 以下となれば頑健性を有すると評価した。

攻撃処理に対する検出精度の評価として F 値を用いた。 F 値は、二値分類の性能を示す指標であり、真陽性を TP、偽陽性を FP、偽陰性を FN とするとき、適合率 P 、再現率 R 、および F 値は次式で定義された。

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

$$F \text{ score} = \frac{2PR}{P + R} \quad (17)$$

F 値は 0 から 1 の範囲の値を持ち、1 に近いほど検出精度が優れていることを示す。つまり、攻撃の見逃し (FN) が少なく、誤検知 (FP) が少ない場合、 F 値が向上することになる。

5.2 頑健性に対する評価

非攻撃的処理に対する頑健性評価の結果を図 5 に示す。ここでは、100 名分のデータの平均 BER% を示す。また、3 種類のサイズの透かし画像を用いた結果をそれぞれ示す。元信号と同一の信号を用いた場合、BER は 0% となっていることから、提案法について正常に実装できていることが確認できた。G.711 符号化及び G.729 符号化以外の非攻撃的処理について、提案法は頑健性の指標となる $\text{BER} \leq 1\%$ を達成しており、音響ゼロ電子透かしに有効であることが示唆された。また、透かし画像のサイズが増加する場合でも、BER は 1% を超えないため、情報量が増えた場合についても透かし情報を検出できていることが示された。MP3 符号化に関して、低ビットレート (32 kbps, 64 kbps) の場合、提案法が従来法よりも高い頑健性を持つことが分かった。一方で、高ビットレート (128 kbps, 256 kbps) の場合は、提案法の方が頑健性を有することがわかった。AAC 符号化については全種類のビットレートにおいて提案法の方が高い頑健性を持つことが確認された。G.711 符号化については従来法と比較して提案法の BER が高く、BER が 1% を超える結果も確認された。G.729 符号化では、従来法に及ばない性能ではあるが、その差は僅かであった。

5.3 改ざん検出に対する評価

攻撃処理を施した後の音声信号から得られた透かし画像の平均 BER の結果を図 6(a) に示す。この結果から、従来法では、全攻撃処理で BER が 1% を大幅に超えており、改ざんを高確率で検出可能であることがわかった。一方、提案法では、BER が 0.8% から 1.7% にとどまり、BER の増加による改ざん検出を行う際に十分な値に達していないこともわかった。

音声信号を改ざんあり、改ざん無しに分類した際の F 値の結果を図 6(b) に示す。提案法は、従来法と比較し検出精度 (F 値) が低く、誤検知及び改ざんデータの見逃しが多いことが示される。従って、攻撃的処理に対しても頑健となっており、透かし画像の破壊を通じて改ざん検出を行う場合、実用性に欠けることが明らかとなった。

5.4 全体考察

非攻撃的処理では、提案法は従来法と同様に殆どの条件で $\text{BER} \leq 1\%$ を維持し、特に MP3 及び AAC の低ビットレート符号化では従来法より BER を低減した。従って、非攻撃的処理に対して頑健性を有することが確認され特に音声符号化処理では従来法よりも頑健であることが明らかとなった。これは、SSA が抽出する周期主成分が音声符号化後も保存されやすく、1D-LBP による同期ずれに強い二値化処理が寄与していると考えられる。

攻撃的処理ではクロッピング、ゼロ補間、ピッチシフト、サンプル置換など局所改ざんを施しても、提案法の BER は 0.8% から 1.7% にとどまり、改ざん検出には十分な上昇が得られな

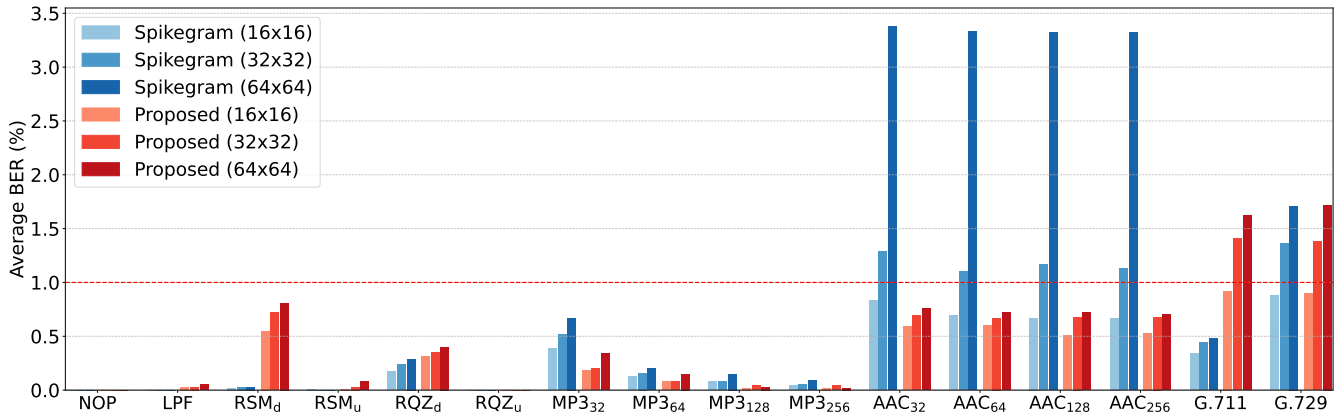
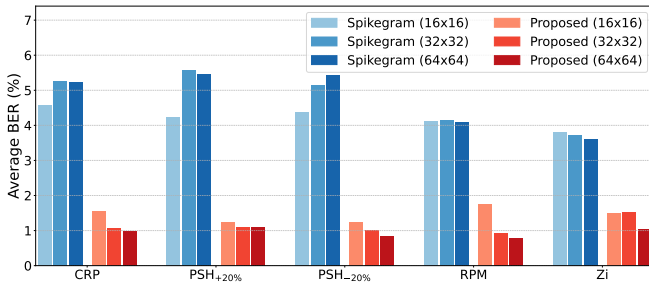
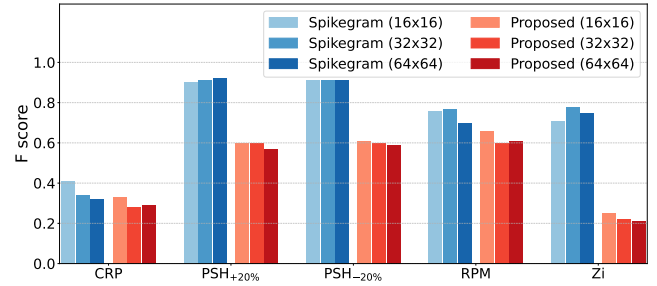


図 5: 非攻撃の処理に対する頑健性評価



(a) 各攻撃に対する BER



(b) 各攻撃に対する F 値

図 6: 攻撃の処理に対する評価

かった。SSA が大域周期構造を鍵とするため攻撃処理適用後も特徴量が保持されやすく、さらに 1D-LBP が隣接サンプルの大小関係のみを符号化するため、短時間の置換や位相ずれではパターンが変化しにくいことが主因と考えられる。結果として F 値も 0.21 から 0.66 と低く、従来法に大きく劣った。

以上より、提案法では著作権保護などの透かしの検出では様々な音情報処理に頑健である。しかし、透かし破壊による改ざん検出では感度が不足しており、実用可能な性能には達していないといえる。

6. ま と め

本稿では、スパースコーディング法である特異スペクトル分析を用いた音響ゼロ電子透かし法を提案した。頑健性に対する評価では、多くの非攻撃の処理に対して $BER \leq 1\%$ を達成し、特に MP3 及び AAC 符号化処理に対しては従来法よりも高い頑健性を示した。また、透かし画像のサイズを増加させた場合においてもその性能は維持された。一方で、攻撃の処理に対する評価では、BER 及び F 値の結果から、過度な頑健性を有し、改ざん検出の感度が低下することがわかった。

今後は、現在の頑健性を保ちつつ、改ざん攻撃では透かし情報が破壊されるように改良することで、実用可能な方法に確立できると考えられる。そのため、SSA で得られた成分について埋め込み対象を再考することや、特徴量の情報を損なわない二値化法を依然検討する必要がある。

謝辞 本研究は、SCAT 研究助成ならびに科研費 (23K18491) の支援を受けて行われた。また、本研究開発は、大学発新産業

創出基金事業 スタートアップ・エコシステム共創プログラム JPMJSF2318 の支援を受けた。

文 献

- [1] Cvejic N. and Seppanen T., “Increasing robustness of LSB audio steganography using a novel embedding method,” in *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004.*, vol. 2, pp. 533–537, 2004.
- [2] Unoki M. and Miyauchi R., “Robust, blindly-detectable, and semi-reversible technique of audio watermarking based on cochlear delay characteristics,” *IEICE TRANSACTIONS on Information and Systems*, vol. 98, no. 1, pp. 38–48, 2015.
- [3] Ali Z., Imran M., Alsulaiman M., Zia T., and Shoaib M., “A zero-watermarking algorithm for privacy protection in biomedical signals,” *Future Generation Computer Systems*, vol. 82, pp. 290–303, May 2018.
- [4] Panda J., Choudhary S., Nath K., and Kumar S., “Audio zero watermarking scheme based on sub band mean energy comparison using DWT-DCT,” in *2016 International Conference on Signal Processing and Communication (ICSC)*, pp. 352–357, 2016.
- [5] 市川 敦暉, 鶴木 祐史, “聴覚的スペクトル表現に基づく音響ゼロ電子透かし法,” *信学技報*, vol. 122, no. 368, EMM2022-65, pp. 20–25, 2023.
- [6] Broomhead D. S. and King G. P., “Extracting qualitative dynamics from experimental data,” *Physica D: Nonlinear Phenomena*, vol. 20, no. 2–3, pp. 217–236, 1986.
- [7] Vincent E., Sawada H., Bofill P., Makino S., and Rosca J. P., “First stereo audio source separation evaluation campaign: data, algorithms and results,” in *International Conference on Independent Component Analysis and Signal Separation*, pp. 552–559, 2007.
- [8] 福井勝宏, 島内末廣, 日岡裕輔, 中川朗, 羽田陽一, 大室仲, 片岡章俊, “雑音抑圧のための雑音対信号比率に基づく雑音パワー推定,” *Journal of Signal Processing*, vol. 18, no. 1, pp. 17–28, 2014.
- [9] Rix A. W., Beerends J. G., Hollier M. P., and Hekstra A. P., “Perceptual evaluation of speech quality (PESQ)—a new method for speech quality assessment of telephone networks and codecs,” in *2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings*, vol. 2, pp. 749–752, 2001.
- [10] 日本音響学会, 毎日新聞社, “日本音響学会新聞記事読み上げ音声コーパス (JNAS).” 2006.
- [11] Ojala T., Pietikainen M., and Maenpaa T., “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [12] Kaya Y., Uyar M., Tekin R., and Yildirim S., “1D-local binary pattern based feature extraction for classification of epileptic EEG signals,” *Applied Mathematics and Computation*, vol. 243, pp. 209–219, 2014.