聴覚特性に基づいた音響情報ハイディング技術

Audio Information Hiding Techniques Based on Human Auditory Characteristics

鵜木祐史 Masashi UNOKI

アブストラクト 音響情報ハイディング技術は、音響コンテンツの情報保護技術だけでなく、音声改ざん検出といったセキュリティ技術として注目されている。筆者の研究グループでは、音響情報ハイディングの研究に取り掛かり、ちょうど10年が経過した。本稿では、音響情報ハイディングの基礎並びにその実現に役立つ聴覚特性、代表的な音響情報ハイディング法を紹介する。また、蝸牛遅延に基づく方法、適応位相変調に基づく方法、特異スペクトル分析に基づく方法など、筆者の研究グループが提案した方法も紹介する。最近の取組みとして、音声セキュリティとしての音声改ざん検出の話題も提供する。

キーワード 音響情報ハイディング, 蝸牛遅延, 適応位相変調, 特異スペクトル分析

Abstract For audio content security, such as audio content protection, as well as speech tampering detection, audio information hiding (AIH) techniques are attracting considerable attention. Ten years have passed since our research group started attempting to resolve many issues in AIH projects. In this paper, we introduce the basis of AIH, important knowledge of human hearing characteristics effective in realigning AIH, and typical methods of AIH. Then, AIH methods based on cochlear delay characteristics, based on adaptive phase modulation, and singular spectrum analysis proposed by the author's research group are also introduced. In addition, we present the research topic of speech tampering detection for secure speech communications, which is one of the current issues in speech technology.

Key words Audio information hiding, cochlear delay, adaptive phase modulation, singular spectrum analysis

1. はじめに

「透かし (watermark)」とは、光の当て具合によって印刷物に識別可能な画像を浮き上がらせる技術のことである。例えば、透かしが紙幣の偽造防止に利用されていることは、誰もが周知の事実である。「電子透かし (digital watermarking)」とは、名前から類推できるように「透かし」のデジタル版のことである。この技術は、紙幣の偽造防止に倣って、音や画像といったマルチメディアコンテンツの偽造防止 (違法コピーや違法配布, 改ざん) に利用されることを目的に開発されてきた。また、この技術はマルチメディア・情報通信技術の普及とともに1990年代から活発に研究されてきた。そのため、当時からデジタルコンテンツの新しい著作権保護 (DRM: Digital Rights Management) 技術として注目を浴びて

きた。

しかし、2000年後半に入ってからでは DRM フリーの時代に突入したことも手伝って、電子透かしを DRM のみに利用することはなくなった。現在では埋め込む情報やその利用法によって呼称も分類され、広い意味でマルチメディア情報ハイディング (MIH: Multimedia Information Hiding) 技術あるいはデータハイディング (data hiding) 技術と呼ばれている。MIH 技術は、データ保護だけでなくデータに付加価値を与えるような応用に利用されている。

本稿では、MIH の基盤技術として、デジタル音響信号に何らかの付加情報 (例えば文字情報) を聴こえないように隠し、それを検出する技術である音響情報ハイディング (AIH: Audio Information Hiding) 法^{(1)~(3)}を紹介する。**2.**では、音響情報ハイディングで利用する用語や原理を紹介する。**3.**では音響情報ハイディングに役立つ聴覚特性を紹介し、**4.**では1990年代~2000年代初頭に提案された代表的な方法を概説する。**5.**では、聴覚特性に基づいた音響情報ハイディング法、特に、筆者の研究グループが提案した聴覚特性に基づいた音響情報ハイディング法を概説する。**6.**では、音響情報ハイディングの応用例として改ざん検出の事例を紹介する。最後に、本稿のまとめを述べる。

鵜木祐史 正員: シニア会員 北陸先端科学技術大学院大学情報科学系 ヒューマンライフデザイン領域
E-mail unoki@jaist.ac.jp
Masashi UNOKI, Senior member, School of Information Science, Japan Advanced Institute of Science and Technology
電子情報通信学会 基礎・境界ソサイエティ
Fundamentals Review Vol.13 No.4 pp.284-293 2020年4月
©電子情報通信学会 2020

図1 音響情報ハイディング技術の概要

2. 音響情報ハイディングの概要

2.1 音響情報ハイディング法の原理

図1に一般的な音響情報ハイディング技術の概要を示す。まず、原信号（ホスト信号）と埋め込む情報（付加情報）を用意する。ホスト信号とは、情報を埋め込む前の信号のことであり、キャリア信号とも言う。ここでは、付加情報として2値マップ画像（{0,1}の2値情報から成る画像）を利用者に気付かれないように（聴こえないように）ホスト信号に埋め込むことを考える。

次に、埋込処理にて、ホスト信号に何らかの手法を利用して付加情報を埋め込む（情報秘匿される）。具体的な埋込方法については後半で説明するが、例えば、ホスト信号そのものに直接、付加情報を重畳することで埋め込んだり、フーリエ変換・ウェーブレット変換など特定の領域に変換したところで付加情報を埋め込んだりする。

次に、付加情報が埋め込まれた信号（情報秘匿された信号）は、様々な経路を通して、そのまま情報機器端末等で利用者に提示される。メディアがオーディオCDであれば、情報秘匿された信号はPCM信号を再生して聴くことになるし、MP3プレーヤーであればMPEG情報圧縮が掛かった形で聴くことになる。また携帯電話やVoIP（Skypeなど）を利用した情報通信では、音声符号化が施された形で情報秘匿された信号を聴くことになる。この際、利用者にとって情報秘匿された信号を聴いても、音響信号にひずみがなく、更には付加情報が埋め込まれているとは全く気付かないことが重要になる。

次に、情報秘匿された信号から付加情報を検出することを考える。検出処理は埋込方法に従って行われるのが一般的である。図2に検出処理の種類を示す。原信号を利用して検出するノンブラインド法（図2(a)）と原信号を利用せずに検出するブラインド法（図2(c)）に大別される。原信号を利用しないが、埋込処理に使った別の重要な情報（例えば、秘密鍵情報）を利用して検出するセミブラインド法（図2(b)）

図2 音響情報ハイディングでの検出法の分類

もある。

著作権保護のような応用の場合は、原信号を利用できないため、秘密鍵を利用したセミブラインド検出法か完全なブラインド検出法を利用しなければならない。一方で、データ照合・データ識別のように原信号を利用できる場合は、ブラインド検出法だけでなくノンブラインド検出法を利用することもある。そのため、音響情報ハイディングの応用範囲を考えると、検出処理は一般にブラインド法であることが望ましいと言える。

検出処理では、通常、情報秘匿された信号をそのまま利用して付加情報を検出する。しかし、利用状況によっては、前述したようにMPEG情報圧縮や音声符号化のような処理が施されたものから付加情報を検出しなければならない。また、悪意ある利用者からの攻撃に耐え、正確に付加情報を検出しなければならない。これらは頑健性（あるいはロバスト性）と呼ばれ、音響情報ハイディングの重要な評価項目の一つである。一方で、場合によっては情報秘匿した付加情報が攻撃によって壊れることで、攻撃を受けたと判断することも可能である。このようなものをぜい弱性（あるいは脆弱性）といい、音響改ざん検出の技術に使われることがあ

る。

最後に、検出された付加情報は、著作権保護の場合は、そのデータ保護に、データ照合・データ識別の場合はその真正性に利用されることになる。上述したように、付加情報の検出に対する頑健性やぜい弱性は、その応用先の用途によって捉え方が変わることになる。

2.2 音響情報ハイディングの評価項目

音響情報ハイディングの技術基盤を整備するためには、次の三つの要求項目を深く検討する必要がある。

- (1) 知覚不可能性 (inaudibility) …埋め込んだ付加情報が利用者に知覚されないこと。また、埋込みによる知覚可能なひずみが生じないこと。
- (2) 頑健性 (robustness) …様々な信号変換処理を施した後や悪意のある攻撃を受けた後でも、埋め込んだ付加情報を頑健でかつ正確に検出できること。
- (3) 秘匿性 (security) …情報が埋め込まれていること自体を気付かせないこと。また、気付かれたとしてもその情報が容易に検出されないこと。

知覚不可能性では、人を対象に音質評価を行う主観評価実験や定量的な評価指標を利用した客観評価実験を行う必要がある。前者の場合、隠れ基準信号を利用した3刺激2重盲検法 (double-blind triple-stimulus with hidden reference) (ITU-R BS. 1116-1)⁽⁴⁾と MUSHRA (multi stimulus test with hidden reference and anchor) (ITU-R BS. 1534-1)⁽⁵⁾がある。後者の場合、音響信号向けに PEAQ (Perceptual Evaluation of Audio Quality) (ITU-R BS. 1387-1)⁽⁶⁾が、音声向けに PESQ (Perceptual Evaluation of Speech Quality) (ITU-R P. 862)⁽⁷⁾がある。いずれも評価音源を用意して検討しなければならない。

頑健性では、付加情報の検出に関するビット誤差率 (BER: Bit Error Rate) を評価尺度として様々な外乱・攻撃等に対する耐性を評価しなければならない。基本的な頑健性評価では、雑音付加、量子化、エコー加算、非可逆符号化 (MP3 や AAC 符号化など) を付与した状態での BER が検討される。これに対し、高度な頑健性評価では、クリッピングやピッチスケールリングといった非線形処理やサンプル置換、A-D/D-A 変換といった処理を施した状態での BER が検討される。悪意ある攻撃に関しては想定しにくいものであるが、代表的なものとして、結託攻撃 (同一のホスト信号に異なる付加情報が付与されたサンプルを集め、その平均化処理により情報秘匿された付加情報の検出を無効にする攻撃) に対する評価がある。

秘匿性では、付加情報が秘匿されていることに気付かれないことを評価しなければならないが、上記二つのように明確な評価方法は定められていない。秘匿性は知覚不可能性とも密接に関わっているため、多くは知覚不可能性を担保することで秘匿性を間接的に評価している。評価の代わりにどのような手法で情報ハイディングしたのかを伏せることで秘匿性を担保するものもある。

上述の三つの評価項目のほかに、付加情報の検出において原信号を利用するかどうか (ブラインド性)、埋込容量 (ペイロード) の上限はどれだけあるか (情報秘匿量)、特定の処理に過敏に反応して付加情報が壊れるかどうか (ぜい弱性)、可逆的に情報を埋込み・検出・削除できるかどうかなどの項目もある。いずれも音響情報ハイディングの応用に依存して決まるため、その応用ごとに議論する必要がある。可逆性に関しては、埋込情報の編集作業が可能であり、応用の幅が広がることになる。なお、標準的な評価方法は、IHC (information hiding and its criteria for evaluation) 基準が整備されている⁽⁸⁾。

3. 聴覚特性

音響情報ハイディングでは、利用者に知覚されないように音響データに付加情報を埋め込むことを狙いとしている。そのため、人の聴覚特性を知ることが知覚不可能な音響情報ハイディング法を実現する上で重要なことである。本章では、そのための基礎知識を提供する。

3.1 可聴域

我々人間の可聴域 (audible range) は、図3に示すように、周波数範囲で 20 Hz から 20 kHz、音圧レベルで 0 dB からおよそ 120 dB まで (音圧で 20 μ Pa から 20 Pa まで) である^{(9),(10)}。そのため、周波数で 10^3 倍、音圧で 10^6 倍のダイナミックレンジを持つことになる。そのため、可聴域ではないところ (加齢に伴い高域側から増えてくる難可聴域も含む) に付加情報を埋め込むことが重要な考え方の一つになる。

3.2 音の知覚の三属性

音の知覚には三つの属性がある。一つ目は音の大きさ (ラウドネス (loudness)) であり、音の強さに関係する。二つ目

図3 人間の可聴領域 (周波数・音圧レベルの範囲)

は音の高さ（ピッチ（pitch））であり、音の基本周波数あるいは周期性・調波性に関係する。三つ目は音色（timbreあるいはtone color）であり、最初の二つと異なり、様々な物理量と関係し、多次元的な性質を持つ。

音の大きさは、音の強さあるいは音圧のべき乗に比例し、音圧レベルの増加とともに増加する。これはラウドネスの成長則と呼ばれる。周波数が1 kHzで音圧レベルが40 dBの音の大きさは1 soneであり、そこを基準に音圧レベルが10 dB増加するごとに2 sone（2倍）、4 sone（4倍）、8 sone（8倍）と増えていく。音の強さの差を弁別できる最小の刺激差異は丁度可知差異（JND: Just Noticeable Difference）と呼ばれ、純音が帯域音かで異なるがおおよそ1 dB程度と言われている。

音の高さは、基本周波数（信号を構成する周期の逆数）に対応する。そのため、基本周波数が増加すると音を高く知覚することになる。音の高さの違いが分かる周波数差は基本周波数が高いほど大きくなり、この差を弁別できる丁度可知差異は、最も感度が良い0.5~2.0 kHzでおおよそ5 cent ($(2^{5/1200}-1) \times 100 \approx 3\%$)と言われている。

音色は、最も複雑な性質を表すものであり、多次元的な表現がなされている。また、それに伴う物理量も多種多様であり、識別的側面と印象的側面を持つ。例えば、音の高さ知覚には、音楽的高さ（トーンクロマ）と音色的高さ（トーンハイト）の二面性があり、オクターブ類似性（octave similarity）にも関係する。これは、音の周波数全体が上昇すると、音階は無限階段のように循環するが音色的には高さが上昇する現象のことである。

以上から、音の強さと周波数を操作し、それらの差が丁度可知差異未満になるように付加情報をホスト信号に埋め込むことが知覚不可能な音響情報ハイディングを実現することにつながる。

3.3 マスキング特性

目的音（信号）を聴取する際、それを妨害する音（マスカ）に影響を受け、聴き取り難くなる、あるいは聴き取りできなくなる。このように、目的音の聴覚検知しきいが上昇する現象あるいはそのしきい値の上昇量をマスキング（masking）という。マスキングには、時間配置、周波数配置、興奮・抑圧の効果による見方がある⁽¹¹⁾。

図4に時間配置における分類を示す。図4(a)のように信号とマスカが同じ時間に存在したときに生じるマスキングのことを同時マスキング（simultaneous masking）と言い、逆の場合のマスキングのことを非同時マスキング（non-simultaneous masking）と言う。これは更に、図4(b)のようにマスカの後に信号が続くときに生じるマスキングのことを順向性マスキング（forward masking）と言い、図4(c)のように逆の提示順で生じるマスキングのことを逆向性マスキング（backward masking）と言う。マスキング量はマスカが信号と同時にあるときに高く、信号から離れるにつれ低くなる。

図4 同時・非同時マスキング

図5 心理物理的同調曲線

図6 聴覚フィルタによるマスキングパターンの推定

図5に周波数配置によるマスキングとして、心理物理的同調曲線（PTC: Psychoacoustical Tuning Curve）を示す。これは、純音を信号としたときにマスカを掃引しながらマスキングしきい値を調べたものであり、マスカの周波数が信号の周波数に近いときに最も低いマスカレベルでマスキングできることを示している。

マスキングは一般的に興奮性マスキング（excitatory masking）と抑圧性マスキング（suppressive masking）に分類される。図6に興奮性マスキングに関係したマスキングパターンを示す。例えば、1 kHzで音圧レベルが20, 40, 60, 80 dBのマスカがあったとすると、低周波側よりも高周波側に興奮が広がり、その広がり音圧レベルの増加に伴い大きくなる。これは、マスキングの上方への広がり（upward spread of masking）と呼ばれる現象である。一般に低い周波数の音が高い音をマスクしやすいことを指している。

一方、抑圧性マスキングは、そのマスカが存在することにより興奮が抑えられることに起因する。図5(b)のように同

時マスキングと非同時マスキングの同調曲線の差が抑圧性マスキングによる影響量を示す。これは二音抑圧 (two-tone suppression) としても知られるものである。

音響情報ハイディングでは、ホスト信号をマスカ、付加情報を信号と見たときのマスキング現象を考えることに等しい。そのため、付加情報がホスト信号によっていつでもマスクされるように情報秘匿することが知覚不可な音響情報ハイディング法を実現することにつながる。

3.4 聴覚情景解析

聴覚情景解析 (ASA: Auditory Scene Analysis) とは、音を通じて環境を把握するときどのようなメカニズムで音のイベントを知覚しているかを議論したものである⁽¹²⁾。1990年に Bregman の著書が発刊された後、それを要約した書籍論文にて、音脈分凝 (auditory segregation) に関する四つの発見的規則がまとめられた⁽¹³⁾。これらの四つの発見的規則の概要を図7に示す。

①音の立上りと立下りの同期性 (図7(a))

図7 聴覚情景分析から得られた四つの発見的規則

②音の倍音構造 (調波構造を持つこと) (図7(b))

③音の変化の滑らかさ (図7(c))

④音の時間的な振幅包絡線の類似性 (図7(d))

四つの発見的規則は、一つの音源から一つの音が出ていると考えたときに音が時間・周波数領域において、どのような性質を持つのかを表したものである。これらの規則は、音を一つの音脈と見たときの二つの音脈に分離する条件あるいは、音を二つの音脈の集まりと見たときの一つの音脈に融合する条件とも解釈できる。

音響情報ハイディングでは、ホスト信号を聴いているときに新しいイベント (付加情報) を知覚しないように、四つの発見的規則を利用して付加情報をホスト信号に埋め込むことが知覚不可能な音響情報ハイディング法を実現することにつながる。

ここまでは、片耳受聴の聴覚特性を説明してきたが、人は二つの耳を利用して音を聴くため、両耳受聴についても検討する必要がある。誌面の都合のためここでは詳細を省略するが、例えば最小可聴角度 (MAA: Minimum Audible Angle) や空間知覚におけるマスキング解除 (SRM: Spatial Release of Masking) は知覚不可能な音響情報ハイディングを検討するための、最も重要なバイノーラル特性である。

4. 代表的な音響情報ハイディング法

本章では、1990年代から2000年初頭に提案された音響情報ハイディング法を紹介する。全てを網羅して説明できないが、その中でも代表的なものを表1に示す。

4.1 量子化ビットを利用した方法

古典的ではあるが、基礎的なものとして量子化ビット置換法⁽²⁾が知られる。この方法ではホスト信号の量子化ビットの一部を付加情報で置換することで音響情報ハイディングを実現している。図8(a)では、付加情報の2値情報 $\{0,1\}$ をホ

表1 代表的な音響情報ハイディング法

方法	処理領域	処理概要	文献
ビット置換法 (LSB 法)	量子化ビット表現 (LSB や符号ビット)	音の大きさに影響を与え難い量子化ビット表現の下位ビットや符号ビットに情報を埋め込む方法	(2)
オクターブ類似性に基づく方法	スペクトル領域 (調波性/オクターブ類似)	音楽的な高さあるいは音色的な高さに影響を与えない形でスペクトル領域の調波性に情報を埋め込む方法	(14), (15)
スペクトル拡散法	スペクトル領域	原信号の広範なスペクトル領域に情報を埋め込む方法	(2)
周波数マスキング法	周波数領域	マスキング特性を利用して聴こえないように埋め込む方法	(16) ~ (18)
エコーハイディング法 (ECHO 法)	時間領域	継時マスキングあるいはエコー知覚特性を利用して聴こえないように情報を埋め込む方法	(20), (21)
振幅変調法	振幅変調領域	振幅変調マスキングを利用して情報を埋め込む方法	(23)
周期的位相変調法	位相領域	周期的位相変調を利用して情報を埋め込む方法	(26)
蝸牛遅延を利用する方法	時間領域 (群遅延)	蝸牛遅延特性を考慮して群遅延に情報を埋め込む方法	(30) ~ (32)
群遅延を利用する方法	時間領域 (群遅延)	位相特性を考慮して群遅延に情報を埋め込む方法	(35) ~ (37)
特異スペクトル分析法 (SSA 法)	特異スペクトル	マスキング特性を考慮した特異スペクトルに情報を埋め込む方法	(38) ~ (40)

図8 ビット置換法 (a) LSB 置換と (b) ゼロ置換

図9 オクターブ類似性に基づく音響情報ハイディング

スト信号の LSB (Least Significant Bits) に直接埋め込む。これは、LSB の操作による音の強さの違いが丁度可知差異未満になるため、その違いを検知できず、知覚不可能な音響情報ハイディングを実現している。

図 8(b) では、1 の補数表現から +0 と -0 の符号ビットの冗長性に着目して、ホスト信号が 0 の値のとき符号ビットに直接付加情報を埋め込む。音の大きさの違いの丁度可知差異で見るとその差を検知することができない。

これらの方法は、知覚不可能性の点で非常に優れた方法であるが、雑音付与など外乱の影響にぜい弱であり、頑健性が低いことが大きな問題である。

4.2 オクターブ類似性を利用した方法

音色的な高さと音楽的な高さの二面性に着目した方法として、オクターブ類似性に基づく音響情報ハイディング法がある⁽¹⁴⁾。この方法では、図 9(a) に示すように、ある特定のスペクトルピーク $X(k_0)$ (例えば 400 Hz) に着目し、付加情報の 2 値情報 $\{0, 1\}$ に合わせて、その周波数の 2 倍 (800 Hz) ないし 1/2 倍 (200 Hz) のスペクトル ($X(2k_0)$ あるいは $X(k_0/2)$) を 0 にすることで音響情報ハイディングを実現する。この結果、音色的高さは変わるが、両者では音楽的な高さは変わらない。

他の実現法として、図 9(b) のように、調波性を満たす

ペクトル $X(k_0)$ を k_0 の 2 倍、4 倍、8 倍というように高域側へシフトすることで付加情報をホスト信号に埋め込む方法もある (図中では 2 倍の高域シフトを赤の点線で示す⁽¹⁵⁾)。この場合も上述のものと同様に、音色的高さは変わるが音楽的高さは変わらない。

オクターブ類似性に基づく方法は、Bregman の発見的規則②と密接な関係がある。付加情報の周波数構造が、ホスト信号の周波数構造と調波関係にあるように調整することでオクターブ類似性と同様の効果を持たせることも可能である。

4.3 スペクトル拡散を利用した方法

デジタル変調方式の一つにスペクトル拡張通信がある。この方法を付加情報に直接利用してホスト信号に変調することで実現した音響情報ハイディング法が数多く提案されている。例えば、付加情報を位相偏移変調 (PSK: Phase Shift Keying) や周波数偏移変調 (FSK: Frequency Shift Keying)、周波数ホッピング (FH: Frequency Hopping) することでホスト信号のスペクトルに広範囲に拡散することで情報秘匿できる^{(1)~(3)}。

拡散符号系を直接利用する直接拡散方式 (DSS: Direct-Sequence Spread Spectrum) では、音響情報ハイディングを実現している。例えば、メッセージ信号 $m(n)$ を擬似乱数 (PN: Pseud-random Noise) 系列 $c(n)$ でスペクトル拡散変調した付加情報 $w(n)$ を作成し、それを a 倍してホスト信号 $x(n)$ に加算することで情報秘匿された信号 $y(n) = x(n) + am(n)c(n)$ を得る。付加情報の検出では、情報秘匿に利用した PN 系列 $c(n)$ を $y(n)$ に乗じ、その平均値の符号を見ることでメッセージを取り出すことができる ($c^2(n) = 1$)。しかし、検出時に異なる PN 系列 $c'(n)$ を $y(n)$ に乗じると逆拡散できずメッセージを正しく取り出すことができない。

DSS 法を利用した音響情報ハイディング法は、このようにデジタル通信方式をそのまま採用できるため、簡便で、検出性能やその頑健性に期待できるが、知覚不可能性を担保するために様々な工夫が必要になる。

4.4 心理音響モデルを利用した方法

マスキング特性で紹介したように、低い周波数成分は高い周波数成分をマスクしやすい。図 10 に示すように、周波数マスキングを考慮した音響情報ハイディング法がいくつか提案されている^{(1)~(3)}。代表的なものとして、心理音響モデルを利用したものがある。例えば、先の DSS 法における係数 a を音の大きさの差を弁別できないように調整する方法^{(16), (17)} や心理音響モデルからホスト信号によるマスキングパターンを推定して、マスクされる場所に付加情報を埋め込む方法⁽¹⁸⁾ などがある。

これらの方法は、聴覚特性をうまく活用する方法として高く期待されたが、同様の特性を使った情報圧縮等に頑健でな

図 10 周波数マスキングを利用した情報ハイディング

いことが大きな弱点になっている。

5. 聴覚特性を考慮した音響情報ハイディング法

5.1 エコー知覚を利用した方法

エコー知覚 (echo perception) とは、図 11 のように、音源から発生した音 (直接音) に対し、壁などで反射した単純な音 (第一反射音) に対する知覚を指す。エコー検知限 (echo-detection threshold) では、二音の音圧レベルが等しく、二音の時間のずれが短いとき (約 0.6~1 ms) に一つの音に知覚され、ずれが長いとき (25 ms 以上) に二つの音に分離して知覚されることが知られている⁽¹⁹⁾。

一方、直接音をマスカ、反射音を信号音としたとき、エコー知覚は、継時マスキングとみなすこともできる。信号音がマスカに近いときは継時マスキングとしてマスクされ、マスカから 50 ms 以上離れるとマスキング量が相当減り、信号音を聴き取れることになる。また、信号音がマスカから 25 ms 以上離れたときにエコー知覚の意味でも二音に分離されることになる。

これら二つの知見に対し、Bregman の発見的規則①を利用することでもう一つの解釈も生まれる。音の立上りと立下りの同期性の観点から、直接音と反射音の立上りの差が 50 ms 未満のときは一つの音として知覚され、差がそれ以上開いたときは二つの音として知覚される。

これらの結果に基づき、図 11 のように、エコー時間とエコーの振幅を調整して、単一エコーを利用することで、音響情報ハイディング法を実現することができる⁽²⁰⁾。この方法は、知覚不可能性を満たすが、自己相関法などで容易に付加情報を検出できるため、秘匿性に問題がある。この問題を解

図 11 エコー知覚に基づいた音響情報ハイディング

決するものとして、エコー成分を PN 系列で構成するエコー拡散法も提案されている⁽²¹⁾。

5.2 振幅変調の知覚特性を利用した方法

3.3 で紹介した聴覚マスキングは、聴覚末梢系の周波数選択性を調査するためのものである。これと同様のアナロジーから、振幅変調マスキングの検知限を調べることで変調周波数に関する周波数選択性を調査した研究がある⁽²²⁾。この検知限を変調周波数の関数として調べたものが、時間的変調伝達関数 (TMTF: Temporal Modulation Transfer Function) であり、変調検知干渉 (MDI: Modulation Detection Interference) や共変調マスキング解除 (CMR: Co-modulation Masking Release) といった現象が報告されている。特に CMR に関しては、Bregman の発見的規則④で説明されるように、振幅変調の手掛かりを利用して音の融合・分凝が説明されている。

このような知見に基づくと、常に、ホスト信号の振幅変調成分 (妨害変調) が付加情報の振幅変調 (目的変調) を振幅変調マスキングすることで、知覚不可能な音響情報ハイディングを実現できる⁽²³⁾。この方法では、ホスト信号 $x(n)$ をフィルタバンクにより帯域分割し、各帯域でのチャンネル信号 $x_k(n)$ を対象に、付加情報を振幅変調する (4 値の PSK を持つ振幅変調信号を利用) ことで情報秘匿する。検出時は、特定のチャンネル信号をリファレンスとして、情報秘匿されたチャンネルと合わせて振幅復調後に振幅包絡線間の位相偏移差を検出することで付加情報を抽出する。この方法は、音響情報ハイディング法の評価基準 IHC⁽⁸⁾ を満たした方法であり、頑健性や秘匿性に優れる方法である。

5.3 周期的位相変調を利用した方法

古くから人の聴覚は位相知覚に関して鈍感であると言われている。しかし近年では、位相に鈍感であるということを否定するような研究報告も少なくない⁽²⁴⁾。例えば、二成分複合音を利用して位相変化に伴う音色知覚を調査した結果、二つの成分音の相対的な位相関係の変化が音色知覚に影響を与えるという報告もある⁽²⁵⁾。

この研究成果に基づいた音響情報ハイディング法が提案されている⁽²⁶⁾。この方法では、位相変調に関する音色知覚の弁別結果に基づき、二次 IIR (Infinite Impulse Response) 全域通過フィルタを利用して、付加情報に対応する 2 値情報 $\{0,1\}$ に合わせて、周期的に位相変調をホスト信号にかけることで情報秘匿を実現させた。

$$H(z) = \frac{a + bz^{-1} + z^{-2}}{1 + bz^{-1} + az^{-2}} \quad (1)$$

フィルタ係数 a と b は特定の中心周波数で群遅延が変化するように設計され、周期的にゆっくりとした速度で変化する

図 14 特異スペクトル分析に基づく方法

のある特異スペクトルを知覚不可能な形で操作し、秘匿情報の埋込にも成功している。

一般に、付加情報の埋込みによる音質の劣化と埋込情報の検出の正確さ・頑健さにはトレードオフの関係がある。ここでは、このトレードオフの関係性のある評価関数で定義し、非線形最適化手法により、その最適性を得ることで提案法の全てのパラメータが得られている。その結果、従来から指摘されている音響情報ハイディングの要求項目（知覚不可能性と頑健性）を同時に満たす方法として実現されている。現在、この方法は、音声信号にも活用することが可能であり、音声改ざん検出にも応用されている⁽⁴⁰⁾。

6. 応用例：音声改ざん検出

筆者の研究グループで検討した音声改ざん検出⁽⁴¹⁾は、音響情報ハイディングの埋込部、検出部、識別部の三つの処理ブロックから成る。音声収録側では原信号 $x(n)$ と秘密データ $s(k)$ をペアとして保存しておき、音響情報ハイディング法を利用して、 $s(k)$ を $x(n)$ に埋め込み、外部へ公開するものとする。一方、ユーザ側は許可を得て流通した信号（埋込信号 $y(n)$ ）を取り扱うものとする。ここで、悪意のある第三者が公開された信号 $y(n)$ の内容そのものを、音声分析合成技術を利用して改ざんし、更にそれを原信号だと偽って公開したものとする。検出部では、 $y(n)$ から埋込情報 $s(k)$ を検出する。識別部では収録側から $s(k)$ を受け取り、 $s(k)$ が $s(k)$ と同一であるか判断する。もし同一でない場合、第三者によって $y(n)$ が改ざんを受けた可能性を知ることができ、同時にその壊れ具合のパターンからどのような改ざんを受けたのかその様態を知ることできる。

例えば、図 15 (a) のようなパターンを音声信号 $x(n)$ に埋め込んだ場合の結果の一例を図 15 (b) ~ (h) に示す。改ざんがない場合は図 15 (b) のように完全に同じパターンを得る。また、音声符号化等かけた場合は、図 15 (c) のようにほぼ同じパターンを得る。しかし、雑音や残響といった外乱を与えると、検出パターンは図 15 (d) (e) のように全体的に壊れ、特定の音声の素片合成を行ったり、スペクトルパターン、時間パターンを加工したりすると図 15 (f) ~ (h) のように当該箇所に対応して検出パターンで壊れることが分かる。これらのことから、検出された透かし情報の欠落箇所からどこが改

図 15 音声改ざん検出の例 (a) 埋込みデータ, (b) 改ざんなし, (c) PCM 符号化・復号後, (d) 白色雑音, (e) 残響, (f) 波形編集型音声合成による改ざん, (g) STFT によるスペクトルの変形, (h) GTFB による時間情報の変形。

ざんされたか（改ざんの場所）、どのような改ざんを受けたのか（改ざんの様態）まで検出可能である。

7. まとめ

本稿では、音響情報ハイディングの基礎とその実現に役立つ聴覚特性を紹介した。また 1990 年代から始まる代表的な音響情報ハイディング法を含め、筆者の研究グループで提案した聴覚特性に基づく音響情報ハイディング法（蝸牛遅延を利用した方法、群遅延特性を利用した方法、特異スペクトル分析を利用した方法）を概説した。また、一つの応用例として音声改ざん検出の事例も紹介した。音響情報ハイディング技術は著作権保護の目的で開発されてきたが、今後は、付加情報のエンリッチメントやデータの真正性の証明、音声信号の改ざん検出・なりすまし検出、更にはデジタルフォレンジクスへの応用へと高く期待されるものと考えている。

文 献

- (1) 鷗木祐史, 西村竜一, 伊藤彰則, 西村 明, 近藤和弘, 瀧田光太郎, 音響情報ハイディング, コロナ社, 東京, 2018.
- (2) N. Cvejic and T. Seppänen, Digital audio watermarking techniques and technologies, IGI, 2007.
- (3) Kondo, Multimedia Information Hiding Technologies and Methodologies for Controlling Data, IGI, 2012.
- (4) ITU, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," Recommendation ITU-R BS. 1116-1, 1997.
- (5) ITU, "Method for the subjective assessment of intermediate quality level of coding systems," Recommendation ITU-R BS. 1534-1, 2003.
- (6) ITU, "Method for objective measurement of perceived audio quality," Recommendation ITU-R BS. 1387-1, 2001.
- (7) J.G. Beerends, A.P. Hekstra, A.W. Rix, and M.P. Hollier, "Perceptual evaluation of speech quality (pesq) the new its standard for end-to-end speech quality assessment part ii: psychoacoustic model," Journal of the Audio Engineering Society, vol. 50, no. 10, pp. 65-778, 2002.
- (8) <https://www.ieice.org/iss/emmm/ihc/> (2019 年 12 月 16 日アクセス確認済み)

- (9) 鈴木陽一ら, 音響学入門, 日本音響学会 (編), 音響学入門シリーズ A1, コロナ社, 東京, 2011.
- (10) 聴覚モデル, 日本音響学会 (編), 森周 司, 香田 徹 (編著), 音響サイエンスシリーズ 3, コロナ社, 東京, 2016.
- (11) 音響キーワードブック, 日本音響学会 (編), コロナ社, 東京, 2016.
- (12) A.S. Bregman, Auditory Scene Analysis : The Perceptual Organization of Sound, MIT Press, 1990.
- (13) A.S. Bregman, "Auditory scene analysis : hearing in complex environments," in Thinking in Sounds, S. McAdams and E. Bigand, eds., pp. 10-36, Oxford University Press, 1993.
- (14) 村松 敏, 荒川 薫, "オクターブ類似性に基づくオーディオ信号への電子透かし," 信学論 (A), vol. J87-A, no. 6, pp. 787-796, June 2004.
- (15) 杉山清一, 渡部英二, "オクターブ類似性に基づく電子透かしの埋込法の改良," 信学論 (A), vol. J91-A, no. 8, pp. 828-831, Aug. 2008.
- (16) L. Boney, A.H. Tewfik, and K.N. Hamdy, "Digital watermarking for audio signals," IEEE Int. Conf. on Multimedia Computing and Systems, pp. 473-480, 1996.
- (17) L. Boney, A.H. Tewfik, and K.N. Hamdy, "Digital watermarking for audio signals," EUSIPCO1996, 1996.
- (18) 中山 彰, 陸 金林, 中村 哲, 鹿野清宏, "心理音響モデルに基づいたオーディオ信号の電子透かし," 信学論 (D-II), vol. J83-D-II, no. 11, pp. 2255-2263, Nov. 2000.
- (19) 播磨敏雄, 安倍幸治, 高根昭一, 曾根敏夫, "音像定位における先行音効果とエコー知覚の限界に関する考察," 信学技報, HIP2004-75, pp. 13-18, Dec. 2004.
- (20) D. Gruhl, A. Lu, and W. Bender, "Echo hiding," Proc. Information Hiding, 1st Workshop, pp. 295-315, 1996.
- (21) B. -S. Ko, R. Nishimura, and Y. Suzuki, "Time-spread echo method for digital audio watermarking," IEEE Trans. Multimed., vol. 7, no. 2, pp. 212-221, 2005.
- (22) T. Houtgast, "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am., vol. 85, no. 4, pp. 1676-1680, 1989.
- (23) A. Nishimura, "Audio watermarking based on subband amplitude modulation," Acoust. Sci. & Tech., vol. 31, no. 5, pp. 328-336, 2010.
- (24) 赤木正人, "位相と知覚一人間ははたして位相か? -," 音響論集(秋), no. 1-2-2, pp. 193-196, 1997.
- (25) K. Ozawa, Y. Suzuki, and T. Sone, "Monaural phase effects on timbre of two-tone signals," J. Acoust. Soc. Am., vol. 93, no. 2, pp. 1007-1011, 1993.
- (26) 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, vol. 60, no. 5, pp. 268-272, 2004.
- (27) T. Dau, O. Wegner, V. Mallert, and B. Kollmeier, "Auditory brainstem responses (ABR) with optimized chirp signals compensating basilar membrane dispersion," J. Acoust. Soc. Am., vol. 107, no. 3, pp. 1530-1540, 2000.
- (28) E. Aiba and M. Tsuzaki, "Perceptual judgement in synchronization of two complex tones : Relation to the cochlear delays," Acoust. Sci. Technol., vol. 28, no. 5, pp. 357-359, 2007.
- (29) 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, vol. 50, no. 4, pp. 204-213, 2008.
- (30) M. Unoki and D. Hamada, "Method of digital-audio watermarking based on cochlear delay characteristics," International Journal of Innovative Computing, Information and Control, vol. 6 no. 3(B), pp. 1325-1346, 2010.
- (31) 鵜木祐史, 宮内良太, "蝸牛遅延特性に基づいた音響電子透かし," 音響誌, vol. 71, no. 1, pp. 15-22, 2015.
- (32) M. Unoki, and R. Miyauchi, "Robust, blindly-detectable, and semi-reversible technique of audio watermarking based on cochlear delay," IEICE Trans. Inf. & Syst., vol. E98-D, no. 1, pp. 38-48, Jan. 2015.
- (33) N. M. Ngo, M. Unoki, R. Miyauchi and Y. Suzuki "Data hiding scheme for amplitude modulation radio broadcasting systems," Journal of Information Hiding and Multimedia Signal Processing, vol. 5 no. 3, pp. 324-341, 2014.
- (34) N. M. Ngo and M. Unoki, "Watermarking for digital audio based on adaptive phase modulation," Lect. Note Comput. Sci, IWDW2014, pp. 105-119, Springer, 2015.
- (35) N. M. Ngo and M. Unoki, "Robust and reliable audio watermarking based on phase coding," Proc. ICASSP2015, pp. 345-349, Brisbane, April 2015.
- (36) C. O. Mawalim and M. Unoki, "Audio information hiding based on Cochlear delay characteristics with optimized segment selection," Proc. 3rd International Conference on Security with Intelligent Computing and Big-data Services (SICBS 2019), New Taipei, Dec. 2019.
- (37) N. M. Ngo, B. M. Kurkoski, and M. Unoki, "Robust and reliable audio watermarking based on dynamic phase coding and error control coding," Proc. EUSIPCO2015, pp. 2316-2320, Niece, 2015.
- (38) J. Karnjana, M. Unoki, P. Aimmanee, and C. Wutiwiwatchai, "Singular-spectrum analysis for digital audio watermarking with automatic parameterization and parameter estimation," IEICE Trans. Inf. & Syst., vol. E99-D, no. 8, pp. 2109-2120, Aug. 2016.
- (39) J. Karnjana, M. Unoki, P. Aimmanee, and C. Wutiwiwatchai, "Audio watermarking scheme based on singular-spectrum Analysis and psychoacoustic model with self synchronization," Journal of Electrical and Computer Engineering, vol. 2016, Article ID 5067313, pp. 1-15, 2016.
- (40) K. Galajit, J. Karnjana, M. Unoki, and P. Aimmanee, "Semi-fragile speech watermarking based on singular spectrum analysis with CNN-based parameter estimation for tampering detection," APSIPA Trans. on Signal and Information Processing, vol. 8, e11, pp. 1-13, 2019.
- (41) M. Unoki and R. Miyauchi, "Detection of tampering in speech signals with inaudible watermarking technique," Proc. IHHMSP2012, pp. 118-121, 2012.

(EA 研究会提案, 2019 年 12 月 18 日受付, 2020 年 1 月 8 日再受付)

鵜木祐史 (正員: シニア会員)

1994 職業能力開発大学校・情報卒. 1996 北陸先端大情報科学研究科博士前期課程了. 1999 同博士後期課程了. 博士 (情報科学). 同年 ATR 人間情報通信研究所第一研究室客員研究員, 2000 英国ケンブリッジ大・生理・CNBH 客員研究員, 2001 北陸先端大情報科学研究科助手を経て, 2005 同助教授 (2007 に准教授), 2016 同教授. 現在に至る.

1998~2001 の間, 日本学術振興会特別研究員 (DC2, PD の 2 期) を兼任. 主に, 聴覚機能のモデル化とそれに基づく信号処理並びに音声信号処理 (残響音声回復, 骨導音声回復, 音響電子透かし) の研究に従事. 日本音響学会佐藤論文賞 (1999 年度, 2010 年度, 2013 年度), 山下太郎学術奨励賞 (2005 年度), 本会論文賞 (2016 年度) 各受賞. 信号処理学会, 日本音響学会, アメリカ音響学会, IEEE, ISCA 各会員.