

特集 社会情報

原著論文

音響情報ハイディングを用いる 長距離空間伝搬音響情報伝送技術とその評価

西村 明*

要旨：空間を伝播する可聴帯域に限る音響信号によってデジタル情報を伝送する技術として、音響情報秘匿技術と音響モデム技術、それらを合わせたハイブリッド技術を、空間伝搬音響情報伝送 (aerial acoustic data transmission) 技術として概観する。これまで、空間伝搬に用いる情報秘匿や符号化の手法の開発は行われてきたが、空間伝搬を前提とした技術の評価は定量的に行われてきていなかった。それらの技術を、アナウンス音声で伝送したい内容を符号化して音響信号へ秘匿することで、言語あるいは音声バリアフリー利用する事例を挙げ、その際の技術の要求と評価方法について検討する。そして、特に長距離伝搬が必要な防災無線放送に伴う各種妨害要因を検討し、それらを模擬するシミュレーション実験環境において、従来から提案していた両側エコー拡散法の改善手法によってアナウンス音声へ情報秘匿した場合の性能を、秘匿ビット値と検出ビット値とのビットエラー率によって評価し、その有効性を確認した。

キーワード：音響情報ハイディング, 防災無線放送, 両側エコー拡散法, バリアフリー応用, 音声信号

Long-distance Aerial Acoustic Data Transmission Technology Using Audio Data Hiding and its Evaluation

Akira NISHIMURA*

Abstract: Technologies for transmitting digital information using an aerial acoustic signal limited to an audible band are reviewed as a part of the aerial acoustic data transmission technology. They include an acoustic information hiding method, an acoustic modem, and a hybrid technique combining them. Although several technologies for aerial data transmission were developed, these technologies have not been evaluated quantitatively. Application of these technologies for languages and/or barrier-free speech by embedding the encoded data of the announcement to be transmitted into the audio signal and their proper evaluation method are proposed. The improved bilateral time-spread echo hiding method applied to the speech signals of an announcement is evaluated using the proposed computer simulation. The simulation considers various factors of disturbances induced by long-distance propagation, assuming the use of a voice evacuation and mass notification system. The results of the computer simulation show that the bit error rates between the embedded and the extracted information bits are satisfactory for applications for long-distance propagation.

Keywords: Audio information hiding, Voice evacuation and mass notification system, Bi-lateral time-spread echo hiding, Barrier free application, Speech signal

1. 背景

1.1 はじめに

音響信号になんらかの規則的加工を加えることでデジタル情報を秘匿し、その情報秘匿済み音響信号から検出したデジタル情報を、何らかの用途に用いる技術を総称して、音響情報秘匿 (audio data hiding/audio information hiding) とよぶ[1]。この技術は、1990年代後半から音楽信号の著作権管理を目的とした電子透かし技術 (digital watermarking) として盛んに開発されるようになった。一方、音響信号そのものによってデジタル情報を表現 (符号化) して伝送し、受信側で検出 (復号) して活用する技術を、音響モデム (acoustic modem) とよぶ。デジタル情報を電話帯域音響信号に符号化して通信するファクシミリやコンピュータ通信用のアナログモデムは、それに相当する。また、ガスパ管や水中では、デジタル情報を音響信号に符号化した通信が、既に開発されている[2][3]。

本稿では、屋内あるいは屋外空間を伝播する可聴周波数帯域 (20 Hz ~ 20 kHz) に限る音響信号によってデジタル情報を伝送する技術を対象とする。つまり、一般用途の音響システムにおけるスピーカから情報を秘匿した音響信号を再生し、それをスマートフォンなどの携帯機器に内蔵されたマイクロホンで受音する利用形態である。よって、可聴帯域のみ再生・伝送・受音し、これを通信システムとしても用いることを前提とするために、可聴帯域を扱った音響システムのみを対象とする。その応用として、言語バリアフリーあるいは音バリアフリー的な応用事例を挙げ、特に長距離伝搬時の技術の評価方法についても言及する。そのため、音響情報秘匿技術と音響モデム技術、またそれらを合わせたハイブリッド技術を、空間伝搬音響情報伝送 (aerial acoustic data transmission) 技術として概観する。さらに、新たな空間伝搬音響情報伝送技術として、従来音楽用の電子透かし技術として開発された両側エコー拡散法を改善した手法を提案し、これを長距離伝搬 (数百メートル) を前提とした利用を行う場合の性能評価を、シミュレーション実験によって行った結果で示す。

1.2 空間伝搬音響情報伝送技術の長距離伝搬利用

本節では、空間伝搬音響情報伝送技術の長距離伝搬利用についてのべ、その利用形態に応じた空間伝搬音響情報伝送技術に対する伝送情報量や耐性などの要求を示す。ここでの長距離とは、数百メートルを上限としており、長距離伝搬時に利用できる技術は、当然数十メートルの中距離でも利用可能である。

1.2.1 音声バリアフリー

これまで音声バリアフリーを実現する研究では、音声コミュニケーションの場面において、受信者側のバリア (難聴) あるいは伝送路におけるバリア (残響や背景雑音) に起因して了解度が低い音声信号を、より了解度が高くなるよう音響変換あるいはまた伝送する方法や、音声の内容を音声認識や字幕等を用いて文字で伝える方法が考えられてきた[4]。音声認識の性能は近年各段に向上しているが、背景雑音が音声信号と同程度の強度で混入する、信号対雑音比 (SNR) 0 dB 程度の環境では、誤認識も多く、受信者側および伝送路におけるバリアを克服するのはいまだ困難である。一方、伝送したい音声の文字情報をそのまま、あるいは符号化し秘匿して音声信号と共に伝送する利用方法は、空間伝搬音響情報伝送技術によって可能になる。

事故や災害の発生情報、避難情報、復旧情報などを緊急かつ確実に伝送する必要性は、2011年3月の東日本大震災や2018年9月の北海道胆振東部地震において改めて認識された。大規模事故や災害における緊急放送メディアとしては、ラジオやテレビ放送、防災無線放送、緊急速報メールなどが挙げられる。しかし、停電や施設の損傷により、それぞれのメディアが使用不可能となる場合も想定され、特に2018年9月には地震による火力発電所の停止によって全道停電が生じ、携帯電話通信網にも混乱が生じた。よって、複数のメディアによって緊急放送の内容を相互補完することが必要になる。

特に防災無線放送は、頑強な広域無線システムと拡声放送システムを持ち、音声により直接情報を伝えるという点で、災害時における情報伝達の要であるといえる。しかし、数100 mの伝搬距離が必要なため、適切なスピーカ配置とエリア設定が必要となる上、風向きや地点によっては数秒の遅延音が重畳されることもあり、より音声伝達のバリアは大きく

なる。

防災無線放送等の緊急用途の場合、伝送すべきアナウンス内容のパターンは限られており、200～1,000パターンもあれば十分と考えられるため、8～10bitの情報を数秒～10秒程度のアナウンスで伝送できれば、実用に足りると考えられる。

1.2.2 言語バリエーション

大都市圏では、公共交通機関における表示やアナウンスにおいて、近年多言語化が進められている。しかし、アナウンス音声自体の多言語化は困難で、せいぜい英語と中国語を加えた三か国語程度が限界であろう。多言語表示やネットワークからのプッシュ配信により、海外からの旅行者向けの情報を表示するスマートフォンのアプリはあるが、いま、旅行者個人が居る場所に必要な情報を配信するコストは膨大なものになると考えられる。特に2018年9月の北海道胆振東部地震では、訪日外国人に交通機関の運行や電力供給状況などの災害時に必要な情報を、母国語で伝達する手段が不十分であったことが指摘されている[5]。

おりしも円安基調により日本を訪れる旅行者は年々増加し、2020年の東京オリンピックにおいては、日本の「おもてなし」を体現することが求められている。公共でのアナウンス内容は有限であることから、前項で示した音声バリエーション手法、つまり音声内容を符号化してアナウンス音声に秘匿して伝送する手法を実現することにより、同時に言語バリエーションも実現できる。すなわち、伝送する情報をコード化することによって、文字情報そのものを伝送するよりも伝送情報量は少なくなるのに加えて、そのコードを各言語に変換するテーブルを事前にアプリ内に用意しておけば、各言語での情報表示が可能となる。例えば、首都圏の駅構内のアナウンスで使用する場合は、目的地（最大256か所：8 bit）、発車時刻（1日1,440分：11bit）、ホーム番号（最大32ホーム：5 bit）などが符号化伝送すべき情報であり、合計24bit程度で表現できる。

利用環境としては、防災無線放送のような長距離伝送が必要な環境や、数十メートルまでの中距離伝送が必要な、建物屋内や公共交通機関構内のスピーカ放送システムによるアナウンス音声を対象として考えられる。いずれも、屋内の天井や壁、床からの強い反射音や残響音、また周囲の強い背景雑音が妨

害となる環境での使用が前提となる。

1.2.3 空間伝搬音響情報伝送技術への妨害要因

上述したような利用場面において、空間伝搬音響情報伝送技術を用いる場合、情報を音響信号に秘匿してから、受信された音響信号から秘匿された情報が復号されるまでに、次のような妨害要因が生じる。

- 主にスピーカに起因する高調波歪みや混変調歪み
- 室内や構造物による反射音や残響音
- 背景雑音
- 長距離伝搬による空気吸収に伴う主に高周波数帯域に生じる減衰

さらに、音響信号に情報を秘匿する符号化方式にも依存するが、次のような要因も考慮する必要がある。

- スピーカやマイクロホンの周波数特性や指向特性に起因する伝達特性の変化
- 再生装置のDA変換器と受音装置のAD変換機の僅かなサンプリング周波数差や、受音装置の移動によるドップラー効果に基づく周波数偏移

これらはいずれも直列的に相乘される妨害要因であり、4章にて行う長距離伝搬シミュレーションの妨害要因として採り上げる。スピーカから再生される直前の音響信号と、マイクロホンで受音された音響信号との間には上述の線形／非線形な変形が生じることを前提に、秘匿されたデジタル情報の頑強な復号ができる符号化手法を検討する必要がある。

また実用化時には、エラー訂正符号を採り入れることも必要であるが、ここではエラー訂正符号は実装せずに、秘匿したデータを検出した際のビットエラー率を示すことで、エラー訂正符号を採り入れた場合について検討する。

2. 空間伝搬音響情報伝送技術の概観

本章では、空間伝搬音響情報伝送技術として従来提案された技術を概観し、それらの技術が前節で挙げたような利用に適するかどうかを議論する。いずれの技術も、音響信号に情報を秘匿する処理は、通常のパーソナルコンピュータあるいは組み込み機器において実時間で動作可能な程度の演算処理である。また利用者端末は、スマートフォンやタブレット等の携帯情報端末を想定しており、2018年時点でのそれら端末の演算性能は、秘匿された情報を実時

間で復号するために十分な能力を保持している。

2.1 音響情報秘匿

キャリア信号を音楽や音声といった既存の音響コンテンツとし、その様々な特徴量に対して広義のデジタル変調や符号化を行うのが、音響情報秘匿である。音楽信号の著作権管理を目的とした電子透かし技術 (digital watermarking) として研究が主流であり、デジタル変調の結果として、もとの音楽や音声の品質を損なうことは、当然ながら望ましくないで、どのような特徴量を操作するのが技術の鍵ともなる。音響情報秘匿の具体的手法は参考文献[1]に解説されているので、そちらを参照されたい。

空間伝搬を前提とした用途に利用される音響情報秘匿手法としては、映画館内で再生される音響トラックに位置情報を検出するためのタイミング情報を秘匿しておき、再撮による著作権侵害に対抗するため、再撮時に映像と同時に録音された音響信号から再撮位置を特定する、という技術が報告されている[6]。しかし、利用環境は映画館という雑音や反射・残響音が少ない環境を前提としており、前節で挙げたような利用環境での妨害要因については評価されていない。

前節の空間伝搬に伴う妨害要因に対して定量的評価が行われている音響情報秘匿技術は少ない。筆者が開発した技術[7]は主に屋内での空間伝搬に関する定量的評価を行っている[8][9][10]が、数百メートルにも及ぶ長距離伝搬を前提とした評価は行っていない。

音響信号は時系列信号であるので、ある一定時間区間 (フレーム) 毎に、秘匿するデジタルデータを変える処理を行い秘匿する情報量を増やす。あるいはフレーム毎に繰り返して同じデータを秘匿し、秘匿情報量は少ないが頑強にデータを復号できる処理のいずれか、またはそれらを組み合わせた秘匿処理を行うのが一般的である。このフレームは、送信時と受信時で時間区間が一致したときに最も検出効率がよいため、これらを一致させる処理を同期処理とよぶ。

2.2 音響モデム技術

デジタル情報をアナログ信号でどう表現するか、という問題は、デジタル通信の分野で研究と実用が先行している。通常のデジタル通信では、

情報の媒体は電磁波や電気信号であり、キャリアやサブキャリアと呼ばれる時間周波数信号を様々に変調 (振幅あるいは周波数シフト符号化、位相シフト符号化、直交周波数分割多重 (OFDM)、スペクトル拡散符号化など) して、情報を符号化する。このキャリア信号を可聴周波数の音波や純音に置き換えることで実現されるのが、音響モデムである。

可聴帯域における音響モデムによる符号化信号は、ファックスや通信モデムの符号化信号を聴取すれば想像できるように、人間が聴いた場合にあまり心地よい信号であるとはいえない。Lopesら[11]は、符号化信号が、空気伝搬時に人間にとって不快にならないように、既存の環境音や音楽に似た信号となる符号化方式を複数提案している。たとえば、鳥やキリギリスの鳴く声、R2D2 (映画スターウォーズに登場するロボット) の声に聞こえるような符号化方式、全音階や五音階の音階に情報を割り当てて疑似旋律を生成する方法などである。この復号処理はパソコンやPDAで可能であり、比較的静かな事務室ではエラー率は実質ゼロで800 bps 程度の情報伝送量を実現できると報告している。しかし、部屋の残響やスピーカおよびマイクロホンの周波数特性、背景雑音の影響などは定量的に検討されていない。

音響情報秘匿技術におけるキャリアをホワイトノイズとして、音響モデムを構成するアプローチもある。エコー法を用いて4 bps の情報をホワイトノイズに秘匿し、校舎内の廊下においたスピーカより再生した音をマイクロホンで受信した復号した実環境試験[12]では、音圧レベル 39 dB 程度まで減衰した場所や、音源から 77 m 離れた廊下でも秘匿したデータの復号ができたとの報告がある。ただし、この際の暗騒音の性質やレベル、反射音や残響音の状態は明らかでは無く、定量的に伝送性能が計測されたとは言い難い。

著者は、背景雑音と残響のある空間でのスピーカ再生によるデータ伝送に頑強な音響モデム技術を開発した[13]。伝送データはその基本周波数がクロマチック音階に従った複合音に符号化された。携帯電話を用いた復号化を容易にするため、マイクロホン受信された音からのデータ復号には、CELP系音声符号化におけるピッチ検出アルゴリズムを用いた。コンピュータシミュレーション実験の結果、0.6 秒の残響のある空間での様々な環境騒音下 (SNR 5 dB)

の条件では、32 bpsで伝送されるデータから検出されるデータの平均エラー率は10%を下回った。また、符号化時に使用する音高を和音の構成音とし、和音進行を前提とすることで、情報伝送量は減少するが、より音楽的な旋律によって情報伝達を行う可能性を示した。しかし、この評価実験では長距離伝搬に伴う空気吸収や長経路エコーの影響は考慮されていない。また、本技術の評価にはエラー訂正符号は使われておらず、エラー訂正を行った場合は、さらにデータ伝送量は少なくなる。

なお、防災無線放送等の音響へ情報を秘匿する場合、即時性と緊急性を要するため、必ずしもアナウンス音声に重畳する必要はなく、サイレンなどの警報音に情報を重畳する方式[14][15]も有効である。

いずれにせよ強度の妨害要因の存在する環境において、少ない情報量でもよいので、頑強に検出できる手法の開発とその検証は十分とはいえない。

2.3 ハイブリッド方式

音響モデムによるデジタル変調信号と、音楽や音声などのコンテンツ信号とで、周波数帯域を分けて伝送するのが、ハイブリッド方式である。一般に前者のモデム信号は聴感上目立たないように高周波数帯域を利用し、後者のコンテンツ信号は音声や音楽の伝送に重要な低・中周波数帯域を利用する。

ハイブリッド方式の代表的な技術は、音響OFDM[16][17]であり、OFDMとよばれるデジタル変調を音響モデム信号に用いる。この信号帯域を6.4 kHz - 8 kHzとして、その帯域にあった原音響信号のパワースペクトルに似せたOFDM信号を生成して原音響信号と置き換え、音質劣化が顕著にならないように工夫されている。この条件下ではエラー訂正符号による符号化率も考慮した上で、約300 bpsの実データを伝送できるとしている。送受信の音響システムの性能が適合するならば、このOFDM信号の帯域をより高い周波数帯域として、より聴感上目立たなくすることも可能である。一方、スピーカからの数mの伝送距離や受音角度による伝送性能の評価、ドップラー効果への対策は行われているが、知覚符号化への耐性や、長距離伝搬や屋外伝搬時に顕著となる背景雑音や反射音・残響音による妨害に対する定量的な評価は行われていない。このような、可聴域の一部の帯域を音響モデム信号に置き換える方法は他にも提案されている

[18]。

18kHz以上に直接スペクトラム拡散による音響モデム信号を加算付加する方式が、2011年頃から実用化されている。伝送レートは最大約80bps程度とされているが、このような高い周波数帯域は、中高年齢層ではほとんど聞こえず、この帯域での環境雑音のパワーも小さいため背景雑音への耐性も高い。また、地上デジタル放送の音響符号化(MPEG2 AAC 256 kbps)と復号を経てスピーカ再生された後でも、スマートフォン等での復号が可能であるとしている[19]。多くの音響コンテンツは18 kHz以上の成分をほとんど含まないため、音響コンテンツへの加工は最小限で済み、音響コンテンツが無音の場合でも、音響モデム信号の再生だけで実施が可能である。しかし、技術内容や技術の定量的評価は公開されておらず、一般的なスピーカシステムを用いて再生した場合、高周波数帯域ゆえに遮蔽物からの回折を期待しにくいこと、再生指向特性が鋭くなること、長距離伝搬時には4.3節で論じる空気吸収の効果からも分かるように減衰が激しいこと、などの問題点が挙げられ、その設置方法に依存して受信可能なエリアや伝送距離にかなり制限を受けると考えられる。

さらに、ハイブリッド方式において重要となる、44.1 kHzあるいは48 kHz標準化時のナイキスト周波数付近でのAD変換器の遮断特性やエリアジグ歪み、18 kHz以上におけるマイクロホン感度については、受音を想定されるスマートフォンにおいて、多くの機種での検討が進んでいるとは言い難い[20]。

3. エコー拡散法

本章では、音響情報秘匿技術として従来から提案されているエコー拡散法[21][22][23]を紹介し、これに改善を加えることによって、長距離伝搬に耐性を持たせた新たな手法を説明する。

3.1 秘匿手法

エコー拡散法は、乱数系列によって正負の符号を与えられたエコーカーネルを含むインパルス応答を、音楽や音声信号に畳み込むことで、それらキャリア信号のスペクトルに情報を秘匿する電子透かし法の一つである[21][22][23]。

長さ L 、振幅が ± 1 である疑似乱数(PN)系列

$P(n)$ を用いて、ホスト信号に畳み込むエコーカーネル $k(n)$ を次式で表す。

$$k(n) = \delta(n) + \alpha P(n - d_0) \cdots (1)$$

ここで、 α はエコーを形成する PN 系列の振幅であり、秘匿強度となる。 n は離散時刻であり、 $\delta(n)$ はディラックのデルタ関数である。情報が秘匿された信号 $r(n)$ は、秘匿前の信号 $s(n)$ と $k(n)$ の畳み込みとして、畳み込み記号 $*$ を用いて次式のように表される。

$$r(n) = s(n) * k(n) \cdots (2)$$

この畳み込み演算は、予め定めたフレーム長 (F) よりオーバーラップ時間分 (T_r サンプル) だけ長い信号に対して行われ、秘匿情報が異なる前後のフレームとは $T_r/2$ サンプル分、raised cosine 関数による重み付けによって重ねる。

$P(n)$ を環状に m だけシフトした、 $P'(n)$ を PN 系列として用いることで、シフト量に応じた情報を秘匿できる。シフト量 m は、秘匿時と検出時のサンプリング周波数がわずかに異なる場合への耐性を高めるため、 0 から m' ステップで $\lfloor \frac{L}{m} \rfloor$ 段階に変化させる。このとき、秘匿データは $0 \sim \lfloor \frac{L}{m} \rfloor$ の整数値で表現され、このフレームに対する秘匿情報量は、 $\log_2(\lfloor \frac{L}{m} \rfloor + 1)$ bit となる。

$$P'(n) = \begin{cases} P(n + L - m) & 1 \leq n \leq m \\ P(n - m) & m + 1 \leq n \leq L \end{cases} \cdots (3)$$

妨害要因への耐性を高める必要があるため、エコー振幅は、最大となる $\alpha = \frac{1}{\sqrt{L}}$ とした。

3.2 両側エコー拡散法

この手法は、エコーカーネルを時間的に反転させ、時刻ゼロのデルタ信号の直前（負時間側）にも付加する[23][24]。以降両側エコー拡散法とよぶ。よって、インパルス応答の長さは、 $2(L + d_0) - 1$ となり、エコーカーネルの振幅が $\alpha = 1/(2\sqrt{L})$ のとき、正時間側のみエコーに比べて、ケプストラム上に現れるエコー遅延振幅を最大約1.7倍に強調できる。この手法の効果は、既報[25]にて既に示した。

3.3 検出手法

式(2)において、両辺に対して離散フーリエ変換(DFT)後に対数変換を行い逆DFTを行う、いわゆるケプストラム変換を行う。ここで、ケプスト

ラム変換は演算記号 \sim で表し、DFT演算を DFT、逆DFT演算を IDFT と表すとき、 $\tilde{r}(n) = \text{IDFT}(\log \text{DFT}(r(n)))$ と表される。すると、次式のように、式(2)の右辺の畳み込み演算はケプストラムの加算として表現できる。

$$\tilde{r}(n) = \tilde{s}(n) + \tilde{k}(n) \cdots (4)$$

$\tilde{k}(n)$ には、 $k(n)$ における、 $\delta(n)$ と、 $P'(n)$ に含まれる各遅延パルスの遅れ時間と振幅に対応する正負のピークが現れる。このため、情報秘匿済み信号のケプストラム $\tilde{r}(n)$ と、 $P(n)$ との相互相関関数は、埋め込み時の環状遅延時刻 (m) に最大値を持つ。実際には、相互相関は $P(n)$ の時間を反転させ、 $\tilde{r}(n)$ との環状畳み込みにより求める。なお、相互相関の最大位置が $m'/2$ を超えてずれて検出された場合でも、ビット誤りを少なくできるように、秘匿ビット情報には連続する整数値に対する変化ビットが 1 bit となるグレイ符号化を施す。

3.4 秘匿区間と検出区間の同期

エコー拡散法に関する一連の論文[21][22][23]では、フレーム同期については、全く述べられていない。秘匿時のフレームより短いフレームからでも、秘匿情報の検出は可能であるが、秘匿時のフレームと全く同じ長さの同期した検出フレームから、最も正確な情報が検出できる。

既報[24][25]の同期手法を示す。まず、秘匿フレーム長 (F) は既知なので、検出フレーム区間を $F/8$ づつずらしながら ($7F/8$ だけ重ねながら)、検出処理を繰り返す。検出された情報が正しいと見込まれるのは、秘匿フレームと検出フレームが最も重なるときであるが、その前後のフレームでも、同一の正しい情報が検出されやすい。よって、連続するフレーム (i 番目と $i+1$ 番目) から検出されたピーク位置 (検出データ) の差分の絶対値 (D_i) を調べると、8 フレームごとに極小値が現れやすいことを利用して、同期位置を算出した。

しかしこの手法では、SNRが低いときに同期したフレーム位置前後で同じデータが検出されにくく、同期精度が落ちる。また、SNRが高いときにも同じデータが連続して検出されるため、正確な同期位置とならない場合がある。よって、検出データによらない新たな同期手法を検討した。

$F/8$ だけ検出フレームをずらしながら検出処理を行う際の、PN 系列とケプストラムとの相互相関関

数のピーク値を A_i とする。この値は、フレーム同期された付近で大きい値をとるため、8フレームごとに大きい値をとる。このことを利用して、それまでに得られる N フレーム分の A_i に対して $\sum_{n=1}^{\lfloor \frac{N}{8} \rfloor} A_{(k+8n)}$ ($k = 1, 2, \dots, 8$)を求め、これが最も大きくなる k をフレーム同期時刻とした。

3.5 エコー拡散法の高耐性化：対数スペクトル抑圧

背景雑音と音声信号とのSNRが0 dBを下回るような場合、つまり雑音が信号より強い場合、音声帯域信号(200 Hz - 3 kHz)が、主にエコーカーネルによる遅延情報を保持している。この帯域はホーンスピーカが有効に伝送できる帯域でもある。よって、対数スペクトル算出後に音声帯域外を抑圧してケプストラムを算出することで、検出力を高める新たな検出法を開発した。既報[25]では、 $\log \text{DFT}(r(n))$ において、200 Hz ~ 3 kHz以外の帯域の強度を1/10し、良好な結果が得られたため、これを採用した。

4. 評価シミュレーション実験

本章では、1.2.3節で指摘した長距離伝搬に伴う妨害要因を元に、現実の屋外拡声放送における伝送環境を模したシミュレーション環境を構築し、3章で示したエコー拡散法の改善手法を評価するシミュレーション実験条件とその結果を示す。

具体的には、4.2節で示す防災無線放送を模擬した音声信号へ、秘匿情報としてランダムなビット値を、3.1節および3.2節で示した手法によって4.1節で示すパラメータ値を採用して秘匿する。

情報秘匿された音声信号には4.3節で示す長距離屋外伝搬シミュレーション環境での変形を与えた後、3.3、3.4、および3.5節で示した手法によって秘匿情報を検出する。秘匿したビット値と検出したビット値とが異なる割合であるエラー率を算出することで、性能の評価を行った結果を4.4節に示す。

4.1 秘匿パラメータ

エコー拡散法の改善手法における、秘匿時パラメータは表1のように定めた。

4.2 シミュレーション音声

音声データベースに含まれる音声信号にランダムビットからなる情報を両側エコー拡散法によって秘匿し、防災無線放送を想定した屋外スピーカ再生による伝搬と受音を前提とした様々な妨害要因を重畳した後に、秘匿情報を検出するシミュレーション実験を行う。このとき、情報秘匿量は、bpsの単位を用いて評価し、妨害要因に対する頑強性は、情報秘匿された音声区間に対して検出された情報を秘匿情報と比較して得られるビット誤り率を算出して評価する。音声信号は、日本音響学会研究用連続音声データベースVol.1に収録されている音声ファイルを2つづつ連結し、男性話者10名753ファイル(4.8~18.1秒)、女性話者12名の903ファイル(5.6~18.1秒)を用いた。

既報[24][25]は、この音声信号をそのまま情報秘匿対象音声としていたが、実際の防災無線放送で拡声されるアナウンス音声は、長経路反射音が重畳しても音声聞き取りやすいように、音節ごとに0.7~1秒程度のポーズを入れて発話される。これを模擬するために、データベースの音声信号に含まれる

表1 エコー拡散法の秘匿パラメータと秘匿情報量

パラメータ	値	
サンプリング周波数	16,000 Hz	
エコーカーネル遅延時間 (d_0)	80, 240 samples	
フレームオーバーラップ ($Tr/2$)	50 samples	
フレーム長 (F)	16,384 samples	32,768 samples
PN系列長 (L)	2,047 samples	2,047 samples
フレームあたりの秘匿量	9 bits	9 bits
秘匿ビットレート	8.8 bps	4.4 bps

-50 dBFS以下の信号が100 ms以上続く部分に対し、0.8秒の無音時間を加える条件でも実験を行う。こうして付加された無音区間は、実際の音節の区切りに入るとは限らないが、1～2秒程度の音声部分と1秒程度の無音部分の繰り返しは、実際の防災無線放送におけるアナウンス音声の時間的特徴をよく表している。図1は、上に音声データベースの音声波形、その波形にポーズを挿入して放送音声を模擬した波形を中に、下には実際に伊丹市で用いられるテスト版防災無線放送音声波形（TOA株式会社提供）を示した。中と下の音声波形は、上述のような時間的特徴において似ているといえる。なお、この音声を用いた実験結果は4.4節の最後に示した。

4.3 模擬した妨害要因と評価ブロックダイアグラム

図2に、防災無線スピーカから再生される情報秘匿済み音声を、録音機器で受信して秘匿情報を検出する評価シミュレーションにおける妨害要因と、そのブロックダイアグラムを示した。

妨害要因のパラメータは、表2に示した。前節で説明した情報秘匿済み音声信号（ポーズなし、あるいはあり）に対して、以下に示すように模擬した妨害要因を、拡声用アンプおよびスピーカの振幅非線形特性、スピーカ伝達特性（空気減衰を含む）、残響音、長経路遅延音、サンプリング周波数変換、背景雑音の順に与えて、実環境での收音の模擬を行った。

伊丹市で放送した防災無線試験放送音声の原信号

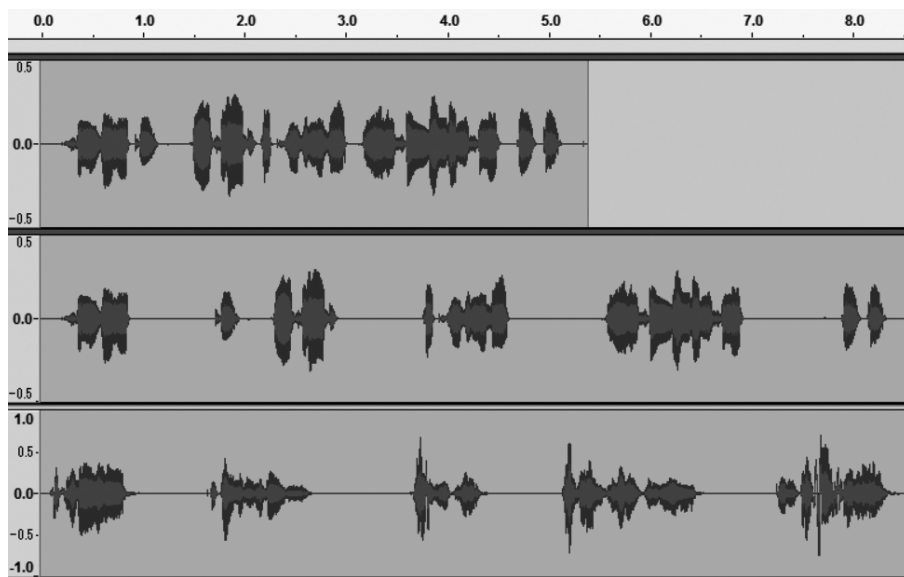


図1 上：音声データベースの音声波形、中：ポーズを挿入して放送音声を模擬した波形
下：伊丹市のテスト版防災無線放送音声波形（TOA提供）

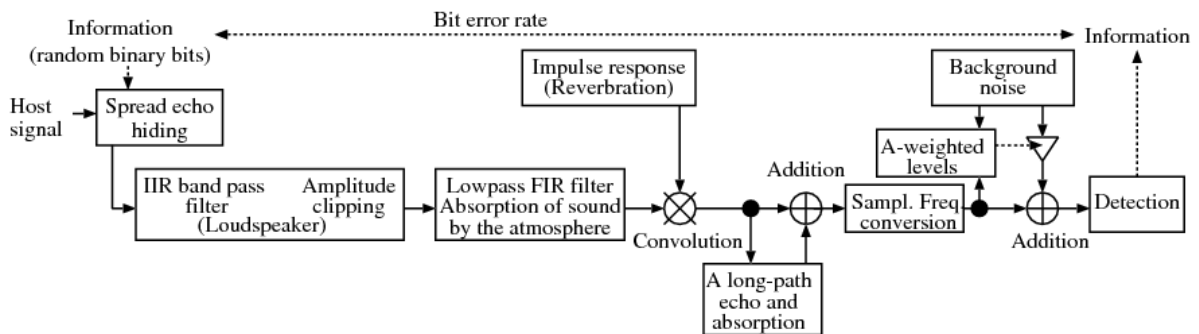


図2 評価シミュレーションのブロックダイアグラム

表2 長距離伝搬シミュレーション実験で模擬した妨害要因とその実装方法およびパラメータ値

妨害要因	実装方法	パラメータ値
スピーカ周波数特性	バンドパスIIRフィルタ	高域通過：3次バターワース300 Hz 低域通過：2次バターワース2 kHz
スピーカ非線形歪	振幅クリッピング	16bit量子化正負最大振幅の1/8で振幅制限
大気による吸収	低域通過FIRフィルタ	ISO-9613-1 気温：15° C 相対湿度：60%，大気圧：1013 hPa
建造物等による残響	指数減衰をもつガウス雑音に基づく合成インパルス応答	残響時間：1 s, 直接間接比率：4 dB STI [26]：0.6
複数スピーカ等による長経路遅延音	残響と空気吸収を含む単発遅延音	振幅：-6 dB 0.1-1.0 sの間でランダムに遅延時間を選択
ドップラーシフトによる周波数偏移	再サンプリング	+0.1%
背景雑音	DEMAND データベース[30]	A特性SNR：0, -5, +5 dB

と、ホーンアレイスピーカから直線距離で約70 m離れた位置で収録した同音声信号のスペクトルを比較することでスピーカ伝送特性を推定し、低域通過および高域通過IIRバターワースフィルタを設計して模擬した。図3にそのフィルタ特性を示した。

ホーンアレイスピーカの非線形振幅歪特性は測定データがなく、あったとしてもその厳密なコンピュータシミュレーションによる模擬は非線形がゆえに難しい。よって、16 bit直線量子化振幅において、正負最大振幅の1/8を超える振幅を制限することによる振幅クリッピングにより歪を生成した。これによって生じる歪の量は、一般的なスピーカの歪より十分に多いため、最悪の条件を模擬するために設定した。

70 mより長い伝搬距離を想定する遅延音あるい

は直接伝達音（残響音を除く）については、ISO9613-1で定められた摂氏15度、相対湿度60%での空気吸収減衰を元に32タップのFIRフィルタにより模擬した。なお、空気吸収特性は、高域で強く減衰し、かつ湿度に大きく依存する。図4に、ISO9613-1から計算される100 mあたりの減衰周波数特性を、典型的な3つの温度・湿度条件において示した。

スピーカからの直接到来音には、様々な経路による地表面や建造物からの反射音が、残響音として付加される。この残響音は、実地測定を行った文献[27][28]を元に、残響時間1秒かつインパルス応答波形から算出したSTI値[26][29]が0.6となるように直接音対残響音エネルギー比を4 dBとし、ガウス雑音波形をベキ指数で減衰させた残響をもつイン

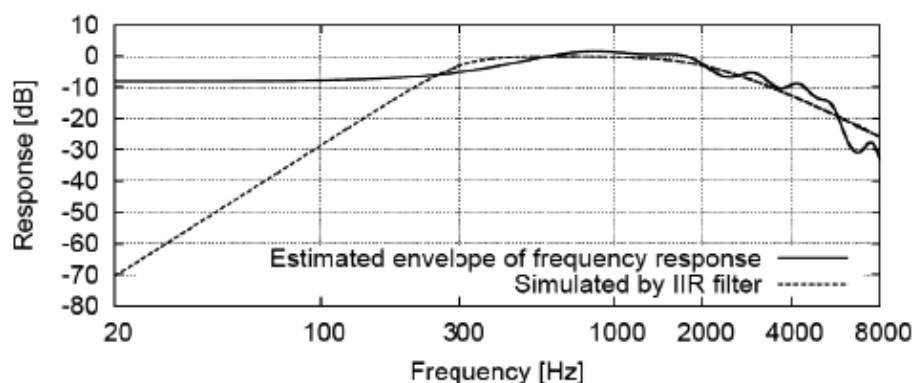


図3 ホーンスピーカ再生音から推定された70m離れたスピーカの再生周波数特性とIIRフィルタによるその模擬特性

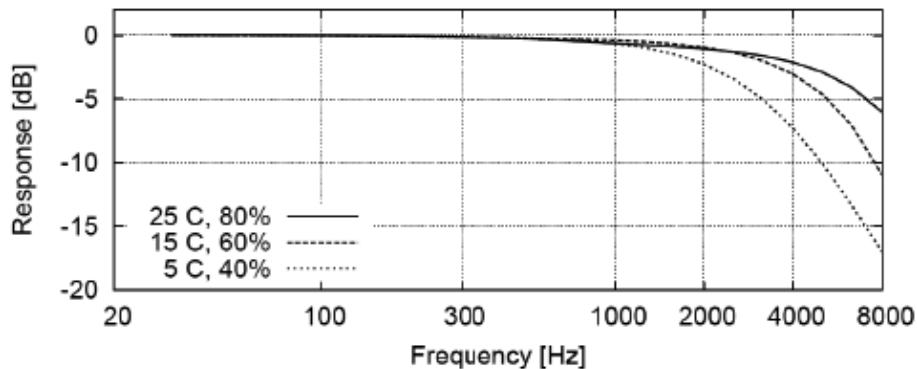


図4 ISO9613-1より計算される大気圧1013 hPa のときの、100 m あたりの減衰周波数特性

パルス応答を畳み込み模擬した。この残響音には空気吸収によるスペクトル変化は含まれていない。

長経路遅延音（ロングパスエコー）として、硬い平面的な構造をもつ建造物からの強い反射音や、別のスピーカからの再生音が遅延して到達する場合の両方が考えられる。問題を単純にするため遅延音は単発とし、その振幅は距離減衰分を除いて半分（-6 dB）に固定し、音声ファイルごとに0.1~1.0秒の範囲でランダムに決定して付加した。長経路遅延音には、スピーカ伝達特性および残響特性を直接到来音と同様に与えた後、遅延時間を伝搬距離に換算した空気減衰特性を、FIRフィルタによって模擬した。

周波数偏差は、再生時と収録時にDA/AD変換器が異なることによって生じる僅かなサンプリング周波数のずれ、および風速の変化、あるいはわずかに受信機器が移動することによって生じるドップラー効果を想定し模擬した。ここでは、サンプリング周波数が0.1%高くなるように変換した。

背景雑音は、マルチマイク集音された環境雑音データベース DEMAND[31]を16 kHzにダウンサンプリングしたものをを用いた。街中で録音された3種類の背景雑音（STRAFFIC：交通量の多い交差点、SCAFE：広場に面した屋外カフェ、SPSQUARE：旅行者の多い街中の広場）のいずれかのch01.wavファイルから音声時間長区間をランダムに抜きだして用いた。

背景雑音を加算する際のSNRは、騒音計によって測定される騒音レベル比として定義した。これは、背景雑音および音声信号のそれぞれについて、A特性フィルタを通した波形の二乗値に対して、時定数1秒の時間積分（Slow特性）を与えた後に時

間平均し、それぞれをdB値に変換し、その差として求めた。

4.4 シミュレーション結果

背景雑音の種類は、エラー率に大きく影響する。情報秘匿済み音声信号とスペクトルが類似している背景雑音は、エラー率が他の背景雑音と比較して高い。また、男声と女声での結果を比較すると、いずれの条件でも男声の方がSNRが低いときにエラー率がより低く、最大10ポイント、平均8ポイント程度エラー率が低かった。これは、男声のほうが女声よりも基本周波数が低いいため、SNRが相対的に高い低域周波数領域において、エコーによって与えられるスペクトル形状のピークディップが保持されやすいためと考えられる。

よってはじめに、よりエラー率の高い女声のみを対象とし、かつ会話雑音を含んでおりそのスペクトルが音声信号と類似しているSCAFEを背景雑音として用いた条件での結果を示す。この結果を、3.2節に示した両側エコー拡散法を用いた既報[24]、さらに3.5節に示した対数スペクトル抑圧を用いた既報[25]と比較する。これらを公正に比較するため、表1においてエコーカーネル遅延時間（ d_0 ）は80サンプルとし、表2においてスピーカ非線形歪は導入せず、スピーカ距離は70 m、音声信号はポーズなしとして、シミュレーション条件を合わせた。

図5において、エラーバーは各条件での90パーセントイル値および10パーセントイル値であり、点は中央値を示している。この結果から、3.5節で示した対数スペクトルでの抑圧は、低いSNRのときに特に有効であることが分かる。また、3.4節で示した同期手法の改善により、数ポイント程度エラー率が改善されることが分かる。

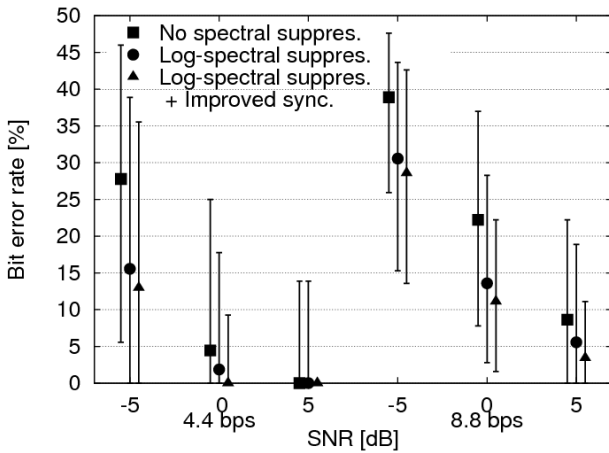


図5 対数スペクトルでの抑圧と同期手法の改善

表1においてエコーカーネル遅延時間 (d_0) は240サンプルとし、表2においてスピーカ非線形歪は導入せず、音声信号はポーズなし伝搬距離 (70 m ~ 800 m) に伴う空気吸収特性の影響を図6に示した。実験条件としてSNRは同じに保たれているものの、距離が遠くなるほど、エラー率は高くなる。これは距離による空気吸収が特に高周波数帯域の減衰 (図4参照) を引き起こすためである。つまり音声に含まれる高い周波数帯域が秘匿情報の伝送には重要であることが分かる。

これまでのシミュレーション結果は、エラー率が相対的に高い条件である、女声かつSCAFEを背景雑音とした結果を示した。図7には、スピーカ距離400 mでの、環境騒音の種類と、男声および女性によるエラー率の違いを示した。図7からは、背景雑音SCAFEにおける女声で最もエラー率が高く、男声の場合はエラー率が低いことが分かる。

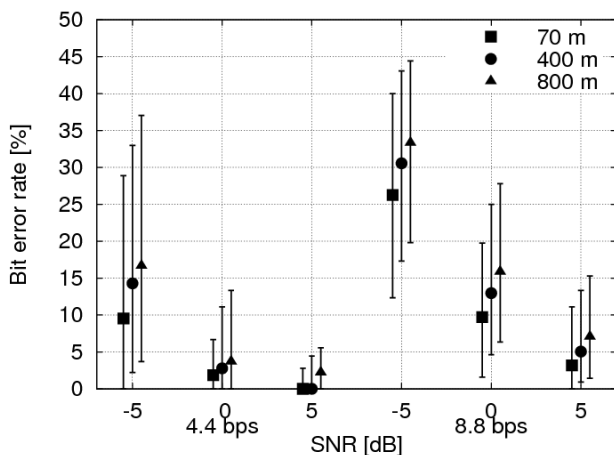


図6 伝搬距離によるエラー率の違い

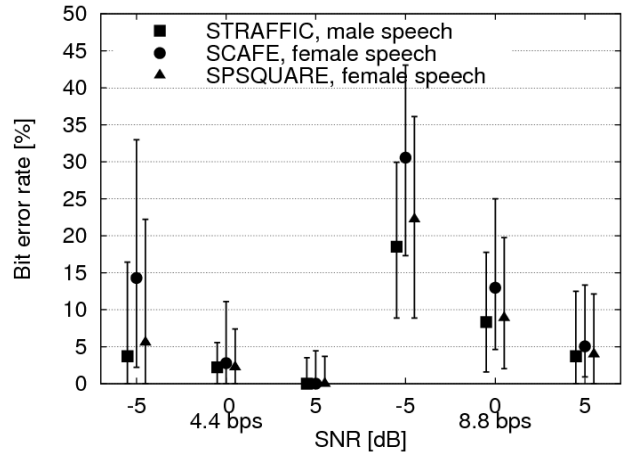


図7 環境騒音の種類、男声および女声によるエラー率の違い。スピーカ距離は400 m

図8には、防災無線放送のアナウンス音声を想定して無音区間 (ポーズ) を挿入した信号へ、情報秘匿した場合の結果を示した。ポーズを挿入することで、音声信号は平均1.564倍の長さになった。また、スピーカの非線形歪を模擬する振幅クリッピングを行った場合の結果も併せて示した。その他のシミュレーションのパラメータは、図6における距離400 mの場合と同じである。

図8を図6と比較すると、振幅クリッピングの有無は、ほとんど結果に影響を及ぼさないことが分かる。また、無音区間を挿入することで、エラー率が上がり、特に8.8 bpsでの秘匿条件の時には最大10ポイント程度上昇することが分かった。この理由は、エコー拡散法は、無音の部分には基本的に情報を秘匿できないため、ポーズがあると伝送量が低下する

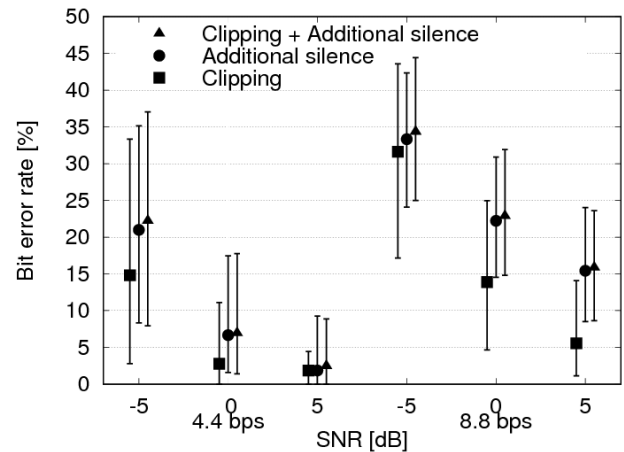


図8 無音区間の挿入および振幅クリッピングによるエラー率の違い。距離は400 m

ためである。また、音声信号と背景雑音とのSNR算出にはポーズを含んだ音声信号を元に算出されるため、ポーズを入れた場合の音声の直接音部分は平均で約2 dB弱いためである。これらを加味すると、フレーム時間長を2倍としてより長い時間区間に含まれる音声部分へ情報を秘匿する4.4 bps条件のほうが、無音区間の存在による性能劣化が少なく、利用に適していることが分かった。

シミュレーション結果をまとめると、男声より女声においてエラー率が高く、背景雑音の種類によってもエラー率は影響を受ける。女性アナウンスへ情報を秘匿し、会話音が背景雑音に含まれ、音声にポーズを挿入しない場合での、防災無線放送による情報伝達を想定してみる。スピーカ距離が400 mのとき、図6からは、SNR 0 dBのとき、4.4 bpsで秘匿を行った場合、11%以下のエラー率で90%の音声から情報を検出できることが分かる。もし、音声にポーズを入れた場合は数ポイントエラー率が高くなってしまいが、フレーム長を長くすることで、エラー率を下げることができることが図8から示唆される。また図4を見てわかるように、気温が低く湿度が低いほど、空気伝搬による高域減衰が大きく生じるので、そのような場合もエラー率はやや高くなるといえる。

5. 考 察

エコー拡散法が脆弱となる妨害要因は、風速の変化や受信装置の移動に伴うドップラー効果によって生じる周波数変移である[24]。これは、検出時のケプストラムと秘匿時に使った疑似乱数系列の長さが合致しなくなると、それらの相互相関関数のピークが小さくなるためである。これを回避するには、受信信号に±0.5%程度周波数変換を行った後に、同様に検出処理を行い、得られた相互相関のピーク振幅を比較することで、周波数変移量を推定し補正してから検出を行う方法[31]がある。しかし、演算量が過倍されるのが難点である。周波数変換後の検出処理を間引きすることなどにより、周波数変移量を効率よく求める改善が必要である。また、風速の変化やドップラー効果は、周波数を一定にシフトさせるだけでなく、音速の相対的变化や音源への接近と離遠によって、周波数を上昇あるいは下降させる周波数変動効果もあるため、この影響も調べる必要が

ある。

エコー拡散法は、ランダムなエコーをもつインパルス応答を畳み込むため、室内環境などでの実際の空間伝搬時に生じる反射音によって生じる音色の変化に似た音質変化をもたらす。よって、音声明瞭度を損ねるような音質劣化はあまりないと考えられる。しかし今後は、今回のようなシミュレーション環境で受信された情報秘匿済み音声の了解度などを主観評価することで、音声伝達に大きく影響するような音質劣化が無いことを確認する必要がある。

実際の防災無線放送音声への情報秘匿とその検出を行うシステムを構築する場合、検出ソフトを動作させたときに、エラー訂正符号を活用して正しいビット値を検出する率である正検出率 (True positive rate) を向上させ、秘匿情報が無いのに間違えて検出する誤検出率 (False positive rate) を最小化する必要がある。BCH (31, 16) 符号化を用いて、正検出率の向上と誤検出率の低下を図る例として、16 bitsのペイロードを31 bitsへ符号化し、3 bitsのエラーまでは復号する。これは4.4 bpsで情報秘匿を行った場合、7秒間の音声信号で伝送できる。このとき、 $3/31=9.7\%$ のビットエラー率までは正しい情報へ復号できる一方、ランダムなビット値が検出された場合、それを正しく復号する確率 (誤検出率) は0.152である。8 bitsの真のペイロードを二重化して、16 bitsのペイロードとする場合、 $0.152 \times 2^{-8} = 0.059\%$ まで誤検出率を下げるができる。この例では、エラー訂正符号を用いて誤検出率を下げることを検討したが、今後は秘匿に用いたPN系列以外の複数のPN系列に対して得られる相互相関関数のピーク振幅を元に、秘匿に用いたPN系列によって検出した秘匿情報の相対的信頼度を算出し、より誤検出率を下げる検討が必要である。

6. ま と め

本論文では、空間伝搬音響情報伝送技術を用いた音声バリエーションおよび言語バリエーション応用について着目した。さらにその技術を概観し、音響モデム方式やハイブリッド方式は長距離伝搬利用には適していないことを指摘した。そこで、長距離伝搬利用に適する音響情報秘匿手法である両側エコー拡散法の改善法に対して、その秘匿区間同期検出の改善

を加えた。そして、屋外長距離伝搬において考えられる各種妨害要因を検討し、それを模擬するシミュレーション実験環境を構築し、それらの妨害要因に対する性能を伝送情報ビットのエラー率として比較評価した。実験結果から、秘匿時と検出時のフレーム同期手法の改善によって数ポイントのエラー率の低減が認められた。また、女声のほうが男声よりもエラー率が大きく、背景雑音の種類によってエラー率が変わることを示した。さらに、スピーカの非線形歪はほとんどエラー率に影響を与えず、防災無線放送音声として明瞭度を増すために音節間に挿入する0.8秒の無音区間を模擬した条件では、数ポイントのエラー率の増加が見られることが分かった。

謝辞

本研究は、TOA株式会社およびエヴィクサー株式会社による援助を受けた。ここに謝意を表する。

【引用文献】

- [1] 西村明「音響信号への情報秘匿技術—電子透かしとステガノグラフィー」, 日本音響学会誌, 63(11), pp.660-667. (2007)
- [2] Sakuma, H., Nakamura, K. and Ueha, S., "Study on influence of gas flow on acoustic communication over gas pipeline," *The Journal of the Acoustical Society of Japan*, 62(8), pp.580-586. (2006)
- [3] Stojanovic, M., "Recent Advances in Underwater Acoustic Communications," *IEEE J. Oceanic Eng.*, 21(2), pp.125-136. (1996)
- [4] 上羽貞行・荒井隆行・栗栖清浩・倉片憲治・坂本真一・船場ひさお・佐藤洋「音バリアフリーの現状と課題」, 日本音響学会誌, 63(12), pp.723-730. (2007)
- [5] 『デジタル毎日新聞』, 2018.9.14, <https://mainichi.jp/articles/20180915/k00/00m/040/138000c>, (2018.10.29閲覧)
- [6] Nakashima, Y., Tachibana, Y. and Babaguchi, N., "Watermarked movie soundtrack finds the position of the camcorder in a theater," *IEEE Transactions on Multimedia*, 11(3), pp.443-454. (2009)
- [7] Nishimura, A., "Audio watermarking based on subband amplitude modulation," *Acoustical Science and Technology*, 31(5), pp.328-336. (2010)
- [8] Nishimura, A., "Data hiding for speech sounds using sub-band amplitude modulation robust against reverberations and background noise," in *Proceedings of IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp.7-10 IEEE. (2006)
- [9] 西村明・坂本真一「音響データハイディングを用いるスピーカ再生音と同期した情報呈示システム」, 電子情報通信学会論文誌A, J93-A, 2, pp.91-99, (2010)
- [10] Nishimura, A., "Audio data hiding that is robust with respect to aerial transmission and speech codecs," *International Journal of Innovative Computing Information and Control*, 6(3(B)), pp.1389-1400, (2010)
- [11] Lopes, C.V. and Aguiar, P.M.Q., "Acoustic modems for ubiquitous computing," *IEEE Pervasive Computing*, 2(3), pp.62-71. (2003)
- [12] Munekata, T., Yamaguchi, T., Handa, H., Nishimura, R. and Suzuki, R., "A portable acoustic caption decoder using IH technique for enhancing lives of the people who are deaf or hard-of-hearing-System configuration and robustness for airborne sound-," in *Proc. of IIHMSP2007*, pp.406-409. (2007)
- [13] Nishimura, A., "Aerial acoustic modem that is suitable to decode using a CELP-based speech encoder," *Proc. of IIHMSP2010*, pp.514-517. (2010)
- [14] Sonoda, K., Yoshioka, K. and Takizawa, O., "Information Hiding for Public Address Audio Signal using FH/FSK Spread-spectrum Scheme," in *Proc. of IIHMSP2007*, pp.279-282. (2007)
- [15] Nishimura, A., "Encoding data by frequency modulation of a high-low siren emitted by an emergency vehicle," in *Proc. of IIHMSP2014*, pp.225-259. (2014)
- [16] Matsuoka, H., Nakashima, Y. and T. Yoshimura, "Acoustic OFDM system and performance analysis," *IEICE transactions fundamentals*, vol.E91-A, no.7, pp.1652-1658. (2008)
- [17] 松岡保静「音響データ通信技術—音響OFDM—」, 日本音響学会誌, 68(3), pp.143-147. (2012)
- [18] Cho, K., Choi, J., Jin, Y.G. and Kim, N.S., "Quality enhancement of audio watermarking for data transmission in aerial space based on segmental SNR adjustment," *Proc. of IIHMSP2012*, pp.122-125. (2012)
- [19] YAMAHA 株式会社「『INFOSOUND™ (インフォサウンド)』を活用したスマートフォン向け2つのサービス提供を開始」, https://archive.yamaha.com/ja/news_release/2013/13050701.html. (2018.10.29閲覧)
- [20] 西村明「空間伝搬音響による情報伝送技術のバリアフリー利用とその評価法」, 聴覚研究会資料, 46(2), pp.107-112. (2016)
- [21] Ko, B.-S., Nishimura, R. and Suzuki, Y. "Robust

- watermarking based on time-spread echo method with subband decomposition,” IEICE Trans. Fundamentals, E87-A, pp.1647-1650, (2004)
- [22] Ko, B.-S., Nishimura, R. and Suzuki, Y. “Time-spread echo method for digital audio watermarking,” IEEE Trans. on Multimedia, 7, pp.212-221, (2005)
- [23] Chou, S.A. and Hsieh, S.F., “An echo-hiding watermarking technique based on bilateral symmetric time spread kernel,” Proceedings of ICASSP 2006 III, pp.1100-1103, (2006)
- [24] Nishimura, A., “Simulation of long-distance aerial transmissions for robust audio data hiding,” Advances in Intelligent Information Hiding and Multimedia Signal Processing Part I, Springer Smart Innovation Systems and Technologies, 81, pp.361-369, Springer, (2017)
- [25] Nishimura, A., “Improvement and evaluation of time-spread echo hiding technology for long-distance voice evacuation system,” Digital Forensics and Watermarking, 10431, pp.391-405, Springer LNCS, (2017)
- [26] Schroeder, M.R., “Modulation transfer functions: Definition and measurement,” Acustica, 49, pp.179-182, (1981)
- [27] 齋藤文孝・崔正烈・佐藤逸人・森本政之・荳木禎史・宇佐川毅・岩谷幸雄・坂本修一・鈴木陽一・青木雅彦・高島和博「屋外における災害情報伝達拡声系の高度化プロジェクト」, 日本音響学会秋季講演論文集, pp.725-726, (2012)
- [28] 佐藤逸人・森本政之・宮川祐輔「ロングパスエコーを考慮した屋外拡声システムの評価
その1：両耳間差の差に着目した評価」, 日本音響学会秋季講演論文集, pp.409-412, (2015)
- [29] 小椋靖夫・浜田晴夫・三浦種敏「音場における音声伝送品質のためのMTFとSTIについて」, 日本音響学会誌, 40(3), pp.181-191, (1984)
- [30] Thiemann, J., Ito, N. and Vincent, E., “DEMAND: Diverse environments multichannel acoustic noise database,” <http://parole.loria.fr/DEMAND/>, (2013)
DOI: <https://doi.org/10.5281/zenodo.1227121>
- [31] Nishimura, A., “Detection of frequency-scale modification using robust audio watermarking based on amplitude modulation,” Digital-forensics and Watermarking: 14th International Workshop; IWDW2015 Revised Selected Papers, LNCS 9569, pp.299-311, (2016)