

非可聴域の音信号を用いる通信システムの開発

銭 顯^{†1} 森勢 将雅^{†2} 新妻 雅弘^{†3} 山下 洋一^{†3}

概要: 本論文では、有用な情報を音信号の非可聴域である高周波帯域に埋め込み、音による通信を行うシステムの開発について述べる。本システムを利用することにより、専用の送受信デバイスを必要とせず、汎用のスピーカー、およびモバイル端末に備え付けのマイクロホンによって通信が実現できる。本論文では、非可聴域へ情報を埋め込む手法と情報を取り出す解析手法を提案し、実験により性能評価を行った。また、タブレット端末上でアプリケーションを試作し、駅でのアナウンスにアナウンス内容のテキストを埋め込んだ音声、および歌詞を埋め込んだ楽曲のデコードを実装した。

Development of a Communication System Using Sound Signals in Inaudible Frequency Bands

HAO QIAN^{†1} MASANORI MORISE^{†2}
MASAHIRO NIITSUMA^{†3} YOICHI YAMASHITA^{†3}

Abstract: This paper describes the development of a communication system using sound signals in which useful information are embedded in inaudible high frequency bands. The proposed system can realize communication to a mobile terminal device just using a built-in microphone and a general speaker, and does not require any special devices for sending and receiving information. This paper describes a method of encoding information in inaudible bands and decoding embedded signals to extract information, and evaluation results. We implemented a prototype system on a tablet PC, which involves decoding the station announce message embedded in speech and lyrics in music sound.

1. はじめに

スマートフォンやタブレットを代表とするモバイル端末の普及が進み、どこでも情報が簡単に得られる時代になった。一方で、広域ネットワークに接続出来ない環境において情報を得るために、ネットワークの接続がなくても応用できる近距離通信も広く注目されている。その中、RFID (Radio Frequency Identifier) を基づく NFC や Bluetooth などの技術が開発されており、各領域で使われている。しかし、これらの近距離通信技術を使うには端末に対応する機能チップの搭載が必要であり、全てのモバイル端末で使えるわけではない。

画像や音などのメディア情報に対して、人が知覚できないように情報を埋め込むステガノグラフィーが広く研究されており、著作権保護や改ざん防止などを主な目的とした電子透かしなどがその例として挙げられる。一方で、メディア情報に有用な付加情報を埋め込むことによって、情報ハイディングを通信手段として利用する試みも始まっている。このような通信方式では、画像であればカメラ、音で

あればマイクを使って信号を受信し、埋め込まれた情報を抽出することができるため、一般的なモバイル端末で情報を受け取ることができる。本研究でも、音信号に対して情報を埋め込む通信方式の開発を行う。

音信号への情報埋め込みの手法としては、エコーハイディング法 [7], エコー拡散法 [2], スペクトル拡散法 [10], 心理音響モデルに基づく方法 [5], 振幅変調マスキング [4], 周期的位相変調法 [3], サブバンド位相シフト法 [8], 聴覚特性を利用する手法 [9] などの様々な手法が提案されているが、これらは音の歪みや変化に気づいてしまうことや埋め込み情報の検出の容易さ、通信速度の欠如などの問題がある。これらの可聴域で行われている手法に対して、近年では非可聴域の音を用いた通信方式が提案されている [1]。非可聴域での情報埋め込みは可聴域での従来法と違い、音質の確保が可能となり、より高い通信速度を利点とし、埋め込み情報の検出の容易さと頑健さの向上も可能になる。本報告では、モバイル端末でも利用できる非可聴域を用いた近距離通信システムを試作し、評価実験を行った結果について述べる。

2. 高周波音を用いた情報通信

提案システムはエンコードとデコードの2つの処理から構成される。図1のように、送信するテキストを付加情報

^{†1} 立命館大学大学院情報理工学研究科
Graduate School of Information Science and Engineering, Ritsumeikan University.

^{†2} 山梨大学大学院医学工学総合研究部
Graduate School of Medical and Engineering Science Department of Research, Yamanashi University

^{†3} 立命館大学情報理工学部
College of Information Science and Engineering, Ritsumeikan University

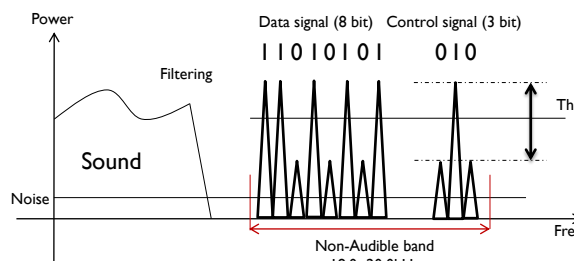
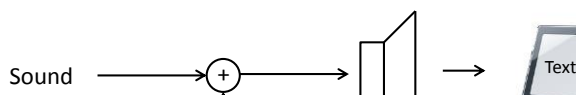


図 2 高周波数帯域における情報の埋め込み

Figure2 Information hiding in high-frequency bands

として高周波数帯域に埋め込んだ音信号をスピーカから放射し、モバイル端末で収録した音信号から付加情報を復号し表示する。

2.1 エンコードの概要

提案法では、図 2 のように、高周波帯域に設定した複合正弦波によりビット情報を送信する。各周波数の正弦波が 1 ビットを表現し、パワーが閾値を超えるか否かで 1, 0 を判断する。通信用信号で用いる帯域は、聴取が困難な高周波帯域であることと、モバイル端末で収録可能であることから、18 – 20kHz に設定する。

計算機における文字が通常 8bit あるいは 16bit で表現されることが多いため、同時に送信する付加情報のビット数を 8bit に設定する。

2.1.1 パケットの定義

図 2 に示されるように、提案法では 8 ビットのデータ信号に 3 ビットの制御信号を加え、11 ビットで通信用信号を構成して送信する。この通信用信号を 1 パケットと定義する。

2.1.2 制御ビット

提案法では、連結されたパケットを区別するために、また、文の開始と終了を表すために制御信号を用いる。図 3 は、24bit (8bit×3) の付加情報の通信を例とした制御信号の役割、及び、パケット列の時間周波数表現を示す。ここで制御の 3 ビットに周波数の昇順により Trigger1, 2 と Condition とする。本方式では、Trigger1, 2 のどちらか一方が 1 の場合、その信号にはパケットが存在することを示す。また、パケット列の開始・終了を識別するため、Condition が 1 となる開始パケット・終了パケットを通信パケット列の前後に付加する。通信パケット列ではデータ信号が 8 ビットの付加情報となっているが、開始・終了パケットの場

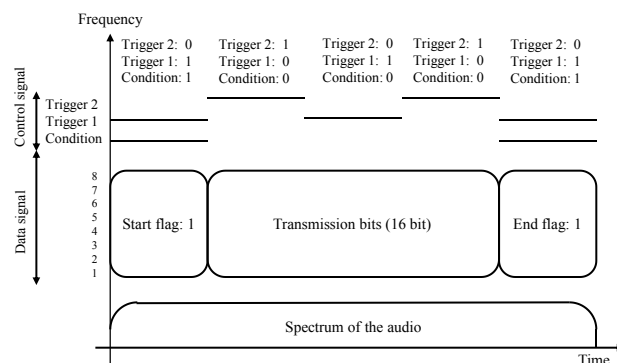


図 3 パケット列中のビット割り当て

Figure 3 Bit assignment in a packet sequence

合、データ信号は付加情報ではなく、そのパケットが開始、終了のどちらかであることを識別するフラグから構成される。パケットの種類を識別するビットとして、パケットの開始・終了を示す Start flag と End flag の 2 ビットを予約語としているが、それ以外の 6 ビットを用いてパケットの種類を識別することができる。本システムでは、パケットが切り替わるタイミングで Trigger1, 2 の 0, 1 を反転させることで異なるパケットを区別する。

2.1.3 通信用信号の位相

提案法は複合正弦波における各周波数のパワーが閾値を超えているか否かで判定するため、位相を任意に設定することができる。本システムでは、Schroeder 位相 [6] を用い、複合正弦波に時間的にエネルギーを分散することで聴覚不可能という特性が守れる。エネルギーが時間的に分散することは時間軸での振幅の最小値と最大値の差が小さくなることと等価となる。ここで使う Schroeder 位相は、波形における振幅の最小値・最大値の幅を最小化する位相として提案されており、本研究では複合正弦波の位相を Schroeder 位相で与えることとする。

2.2 デコードの概要

デコードは、マイクロホンで音を収録したモバイル端末側で行われる。基本的には窓関数により切り出した波形をフーリエ変換し、各ビットに対応する周波数のパワーが閾値を超えているかを判定してビット列を復元する。

2.2.1 開始パケットの抽出

収録を開始すると、デコード側では窓関数により切り出し、パワースペクトルの計算を行う。制御信号により閾値を設定し、開始パケットを検出するまでループする。Trigger1, 2 は片方が 0 となるともう片方が必ず 1 であるため、Trigger1 と 2 のパワーの中間に信号検出の閾値を設ける。開始パケットの検出は、Condition が 1 であり、データ信号の Start flag が 1 である条件を満たすことで行われる。

2.2.2 パケット列からテキスト情報のデコード

開始パケットを検出した後は、Condition が 1, End flag が 1 となる終了パケットを検出するまでパケット列から付

表 1 実装されたシステムの条件

Table 1 Conditions of the implemented system

Sampling frequency	44.1 kHz
Windows function	Hamming window
Packet length	1536 sample
Frame length	1024 sample
Shift length	512 sample
FFT size	1024 sample
Frequency spacing	172.26 Hz
Frequency band	17.98–19.78 kHz

加情報を復号し続ける。本システムでは日本語を表示するため Shift-JIS の文字コードをデータ信号として送っている。終了パケットを検出した場合は、それまで読み込んだ付加情報を画面に出力し、初期状態に戻る。

2.3 システム実装に関する各条件の設定と実装上の工夫

本システムでは、用いる窓関数の種類、1 パケット長、分析フレーム長、シフト長、FFT 長、各ビットの周波数間隔と用いる周波数帯域を設定する必要がある。本論文では、表 1 に示す条件でシステムを実装した。

本実装における通信速度は、8 ビットの情報を表現するパケットの長さが 1536 サンプルであることから 229.7 (44100 / 1536 × 8) bps となる。

窓関数により切り出す時刻によっては 2 つのパケットが混ざったフレームを切り出す可能性があることから、表 1 に示されるフレーム長、シフト長の条件下において確実に 1 パケットのみを含むフレームが存在する条件を設定している。

本実装では、2 つのパケットが混ざったフレームを使わず 1 パケットのみ含まれるフレームを常に選択する処理を実装している。具体的には、Trigger が切り替わったタイミングから、次に Trigger が切り替わったタイミングまでのすべてのフレームの Trigger のパワーを計算しておき、Trigger のパワーが最大となるフレームを選択するようにしている。

3. 評価

駅アナウンス音声にアナウンス内容のテキスト情報を文字コード (Shift-JIS) として埋め込み、情報を取り出す評価実験を行った。音データの収録はスピーカーとモバイル端末との距離を 0.5 メートルから 3.5 メートルまで変化させ、簡易防音室で行った。埋め込んだテキストは図 3 に示す 4 つの発話で、合計 116 個の文字から成り、3 回繰り返して収録した。収録した音声を PC 側で分析し、正しくデコードできた文字の数を集計した。

各距離において 3 つの音声から集計したデータを平均し、距離により埋め込んだ 11 ビット正解率と 10 ビット正解率を図 4 に示す。パケットの定義により 1 パケットが 11 ビット (情報 8 ビットと制御 3 ビット) を含めている。そのた

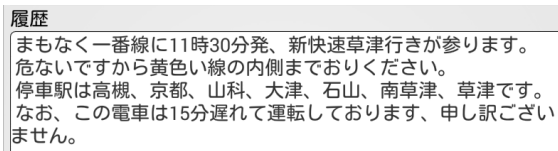


図 3 駅アナウンスシステムで埋め込んだ情報

Figure 3 Embedded message of the station announce system

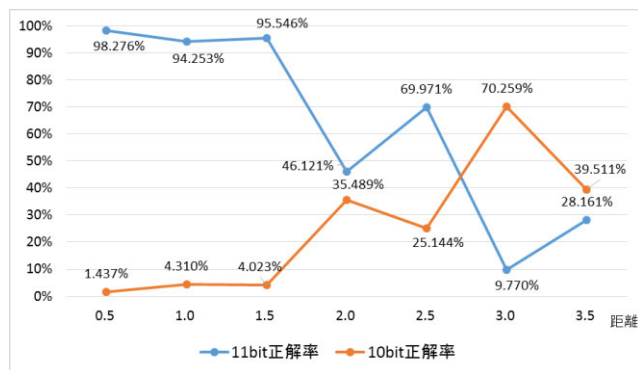


図 4 デコードされた文字の正解率

Figure 4 Correct Rate of decoded characters

め、11 ビット正解率を、音声に埋め込んだ全パケットの中で 11 ビット全部正しいパケットの比率と定義し、10 ビット正解率を 10 ビットが正しい、すなわち 1 ビット誤りがあるパケットの比率と定義している。本システムでは 8 ビット情報部と 3 ビット制御部から構成される 11 ビットのコードを通信しているため、1 ビット誤り訂正能力のある誤り訂正符号を導入することが可能である。実際に誤り訂正システムを用いると、1 ビット誤りの部分は理論上で全て修正することができるため、実験結果には 10 ビット正解率も集計することとした。

結果により、0.5 メートルから 1.5 メートルまではほぼ 100% の正解率で、2 メートルになると、正解率が下がり、さらに 3 メートルになると非常に低くなる。実験により、1.5 メートル以内の距離では信頼できる誤りの少ない通信が実現できていることが分かった。一方、2 メートル以上になると正解率が半分ぐらい低くなり、1 ビットの誤り率がたくさん増えている。誤り訂正の修正を前提とすると 3.5 メートルまで信頼できると考えられる。

収録音声の分析により、閾値の設定が不適切であることから正しくデコードされない例があることが分かった。また、距離が離れることにより、信号のパワーが減衰し、雑音の影響も大きくなると考えられる。

4. 応用システムの試作

2 章で述べた情報通信技術を用い、Android 上で音信号に埋め込んだテキスト情報をデコードして表示するシステムを試作した。

図 5 は、実装されたアプリケーションのスクリーンショット

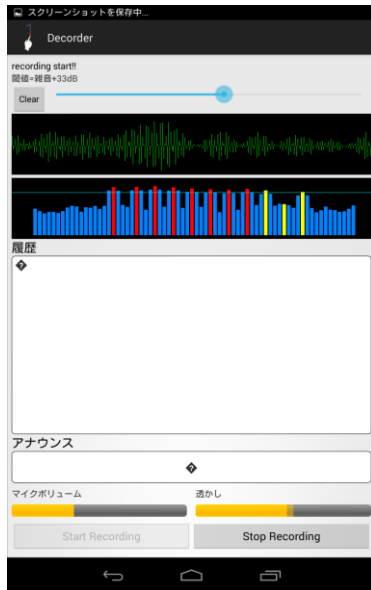


図 5 実装したアプリケーションのスナップショット

Figure 5 Snapshot of the implemented system

トを示す。アプリケーションには、操作用のボタンとして情報をリセットする「Clear」ボタン（左上）とデコードを開始・終了するボタン（左下・右下）が実装されている。波形窓（中上）はリアルタイムの音信号の波形を表示する。スペクトル窓（中）はリアルタイムの通信に用いる帯域のパワースペクトルを表示している。開始ボタンを押すと、デコードが始まり、デコードされた文字テキストが「アナウンス」窓に映され、同時に「履歴」窓に保存されておく。「履歴」枠にスクロール操作で隠れた結果を見ることがもできる。

情報を埋め込んだコンテンツとして、駅でのアナウンスにアナウンス内容のテキストを埋め込んだ音声と歌詞を埋め込んだ楽曲を作成した。

駅のような騒々しい場所で重要なアナウンス通知を聞き取れない場合があるため、駅アナウンスを埋め込むコンテンツとして実装した。駅アナウンスの内容は図 3 のような発話となっている。本実装により、音声流される直前からアナウンス内容のテキストを表示することで、喧騒な環境や短時間しか聞けない場合や外国人の場合などの原因で聞き取りにくい時においても最低限の情報を与えることができる。また、既知の音声を聞く際も、音声を聞きながら内容を確認する効果が期待される。

一方、人間で歌唱が困難とされるほど高速である「初音ミクの消失」に対して歌詞を埋め込むコンテンツとする実装を行った。デコードされた文字テキストの例を図 4 に示す。初音ミクの消失では、1 秒間に 9 文字程度の歌詞を表示する必要があることから、求められる通信速度の下限は 144bps (9×16bit) となる。本手法では、最大で 229.7bps の転送レートが得られるため、この条件を満たしている。

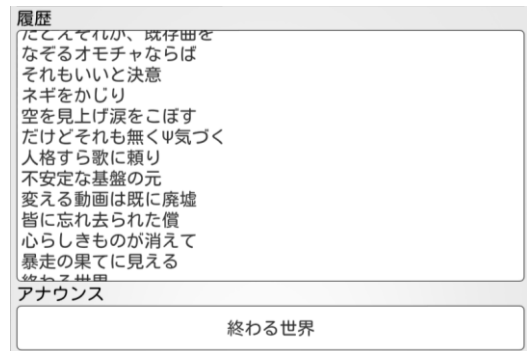


図 4 歌詞表示システムの表示例

Figure 4 An example of the lyrics display system

5. おわりに

非可聴域の高周波帯域に情報を埋め込み、Android 上で開発したアプリケーションによって、埋め込んだ情報を端末に表示するシステムを試作した。評価実験の結果、正解率はスピーカーとの距離が増加するにつれて下がるのがわかった。その原因として、閾値の設定が不適切であることや雑音の影響があると考えられる。今後の課題として、より精度の高い閾値の設定や、頑健性を向上させることを検討したい。

参考文献

- 1) 森勢将雅, 山下洋一: “モスキート音に基づく電子透かしを用いた歌詞連動型音楽試聴システムの提案”, 情報処理学会研究報告, 2012-MUS-96, 4, pp. 1-6, 2012.
- 2) 高橋弘宜, 西村竜一, 鈴木陽一: “エコー拡散法を用いた電子透かし信号への不可逆圧縮の影響”, 日本音響学会研究発表会講演論文集, 2003 (1), pp. 793-794, 2003.
- 3) 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”, 日本音響学会誌, 60-5, pp. 268-272, 2004.
- 4) 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ー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, J93-A, 2, pp. 91-99, 2010.
- 5) 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”, 電子情報通信学会論文誌 (D-II), J83-D-II, 11, pp. 2255-2263, 2000.
- 6) M.R. Schroeder, “Synthesis of low-speak-factor signals and binary sequences with low auto-correlation”, IEEE Trans. Information Theory, Vol. 16, No. 1, pp. 85-89, 1970.
- 7) Gruhl, D, Lu, A and Bender, W: “Echo Hiding”, Proc. Information Hiding 1st Workshop, pp. 295-315, 1996.
- 8) 松岡保静: “サブバンド位相シフトを用いた音響電子透かし埋め込み法”, 電子情報通信学会技術研究報告, EA, 応用音響, 106(205), 7-11, 2006-08-03.
- 9) 鶴木祐史, 羽二生篤, 小杉敏三[他], 宮内良太: “音声信号への蝸牛遅延特性を利用した情報ハイディングの検討”, 電子情報通信学会技術研究報告, EA, 応用音響 110(71), 31-36, 2010-06-03.
- 10) Boney, L, Tewfik, H. H and Hamdy, K. N: “Digital watermarks for audio signals”, Proc. ICMCS, pp. 473-480, 1996.