

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

森勢 将雅^{1,a)} 山下 洋一¹

概要: スマートフォン等のマイクロフォンを備えた携帯端末は、携帯電話としての機能だけではなく、音楽再生、音収録等様々な機能をユーザに提供しつつある。本研究では、スマートフォンを用いた音楽試聴スタイルの1つとして、タイトルやアーティスト等の情報、および歌詞が音楽に同期して表示される音楽試聴システムの実現に取り組む。本システムを実現するため、音楽に知覚的な影響を与えず情報を埋め込み、埋め込まれた情報をリアルタイムで抽出する技術を提案する。知覚に影響せず情報を埋め込むための技術として電子透かしが一般的に用いられるが、雑音に対する頑健性や通信速度の問題があり、実時間での情報伝達手段に用いることは困難であった。本研究では、モスキート音の考え方に基づくシンプルな方法を提案することで、目的とするシステムへの要求を満たす技術の確立を目指す。本稿では、モスキート音を用いた電子透かしの概要、および情報を埋め込むための仕様について説明する。次いでスマートフォン上に実装されたプロトタイプシステム、およびいくつかの音楽を用いた動作確認テストにより本システムの有効性について議論する。

キーワード: 音楽鑑賞, 電子透かし, モスキート音, スマートフォン

Proposal on a lyric-synchronized music listening system using a watermark based on mosquito sound

MORISE MASANORI^{1,a)} YAMASHITA YOICHI¹

Abstract: Smartphone that has a microphone has been used as a good tool for dairy life, and we can easily record a sound. In this research, a lyric-synchronized music listening system with smartphone is introduced to provide users with a new music listening style. The system can display the information such as title, lyric, and so on, extracted from the recorded music. The information is hidden in the music signal to avoid deteriorating the sound quality. Digital watermarking is one of the technology to hide the information, and several watermarking technologies for audio signal have been proposed. However, it was difficult for them to realize the good noise tolerance and transmission performance. In this paper, we focus on the mosquito sound to develop the digital watermarking to solve these problems. This paper shows the method to hide the information and the specification about the transmission of the information. The implemented system is introduced and tested in a real environment, and its effectiveness is also discussed.

Keywords: Music listening, watermarking, mosquito sound, smartphone

1. はじめに

スマートフォンを代表とする携帯端末はすでに世の中に

広く普及しており、特に、QRコードによる情報抽出は、パンフレットや雑誌等で幅広く利用されている。一般的に携帯端末で情報を入力する場合 URL 等の情報を手動で打ち込む手間がかかるため、QRコードのように手間をかけずに情報を抽出する技術には情報入力の手間を省く大きな利点がある。近年では音データに情報を埋め込む音響電子透

¹ 立命館大学
Ritsumeikan University
^{a)} morise@fc.ritsumei.ac.jp

しかし、QRコードと同様のことが可能な技術として着目されつつある [1]。音楽に情報を埋め込み、それを携帯端末で録音し QRコードと同様に情報を抽出することが可能になれば、街頭で何気なく聞こえた音楽から、タイトルやアーティスト等の情報を容易に抽出するサービス等が実現可能となる。

このサービスを実現する基礎技術として、音信号に対する電子透かしが提案されている。電子透かしは、画像や音楽等のデジタルデータに情報を埋め込むために用いられる技術であり、不正コピーの抑止による著作権の保護を目的に発展してきた。スピーカ再生音に対する電子透かし技術は、反射や残響、スピーカやマイク特性等様々な外乱が存在するため、デジタルデータを対象とした電子透かしとは異なる問題を扱うこととなる。

スピーカ再生音を対象とした電子透かしには研究事例があり [2], [3], 振幅変調を用いたもの [4], [5], エコー拡散法 [6], サブバンド位相シフト法 [7] 等が代表的な手法として挙げられる。これらの方法は、通信速度の面で実用的とは言い難く、また、反射や雑音に対する頑健性に問題がある可能性がある。2006年に提案された音響 OFDM (Orthogonal Frequency-Division Multiplexing) [8] は、音楽信号に情報を埋め込むための技術として提案されており、通信速度は上述の方法を遥かに上回る。一方、方法の特性上特に反射や雑音に対する頑健性と音楽信号の品質劣化が課題となる。

本研究では、情報通信速度、雑音への頑健性、および音楽信号の品質向上を満たす方法として、モスキート音に基づく音響電子透かしを提案する。本稿では、提案する電子透かしの詳細について示し、プロトタイプとして実装した音楽試聴システム「Clairvoyant」について述べる。

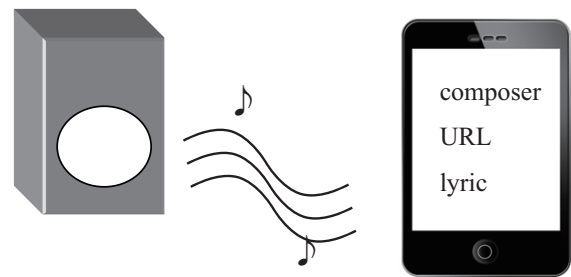
以下、2章では提案する音楽試聴システムの概要と音楽信号に情報を埋め込む技術への要求事項について述べる。3章で、要求事項を満たす情報埋め込み技術として、モスキート音に着目した方法を提案する。4章では、実装されたシステム Clairvoyant について、インタフェースと仕様の詳細について述べる。5章で、提案法および実装システムの有効性について議論し、6章で本稿のまとめを述べる。

2. 提案する音楽試聴システムへの要求事項

ここでは、提案する音楽試聴システムのコンセプトについて述べる。また、関連する技術について述べ、提案法の特徴や位置付けについて議論する。

2.1 本システムを実現するための要求事項

図 1 に示されるように、本研究では、音楽に埋め込まれた情報をスマートフォンでリアルタイムに抽出しディスプレイに表示するシステムの構築を目指す。このシステムを実現するためには、電子透かしのように情報を埋め込む必



Music including information

図 1 提案法のコンセプト

Fig. 1 Concept of the proposed system

要があるが、音楽試聴システムの目的を達成するためには以下の条件を全て満たす必要がある。

- 埋め込まれる情報が音楽の品質に影響しないこと
- スマートフォンで収録可能な帯域のみ用いること
- 室内の反射や残響、屋外の場合雑音に対し頑健であること
- 音楽に連動して歌詞を表示するための通信速度を満たすこと

特に、雑音に対する頑健性と通信速度を両立することが大きな課題といえる。

2.2 要求事項に対する従来法の問題点と提案法のコンセプト

従来法の多くは通信速度の面で問題があり、音楽に合わせて歌詞をリアルタイムで表示することは困難であった。音響 OFDM は、本研究での要求を満たす通信速度を達成できるが、音楽信号の高域のスペクトル包絡の位相を制御する特徴がある。音楽信号は高域よりも低域に強いエネルギーを有するため、SNR の低下が問題となる。また、ほぼすべての方法が可聴域を対象に音楽を処理するため主に位相を制御するが、位相はパワーよりも雑音に対する影響を受けやすく、反射や雑音に対する影響、および耐雑音性の向上が課題となる。

本研究では、従来法よりも高い耐雑音性を達成しつつ、歌詞表示をリアルタイムで可能な程度の通信速度を達成する方法の実現を目指す。本稿では、若年者のみ聴取が可能な信号であるモスキート音に着目し、位相を使わない情報埋め込み法を提案する。提案法は、電子透かしのように音楽信号に情報を埋め込む考え方ではなく、モスキート音という聴取が困難な信号に基づいて送信用信号を生成し、帯域制限した音楽信号に重畳するのみで埋め込める利点がある。

3. モスキート音に基づく音響電子透かし

モスキート音は、夜間の公園やコンビニでたむろする若者を音で撃退することを目的とし、一般的に 17.5 kHz 付近の周波数を有する正弦波と定義される。この周波数は、

表 1 3つの周波数帯域の定義

Table 1 Definition of three frequency bands

	Frequency band
Audible sound	20-20,000 Hz
Mosquito sound	17,400-21,000 Hz
Ultrasound	20,000- Hz

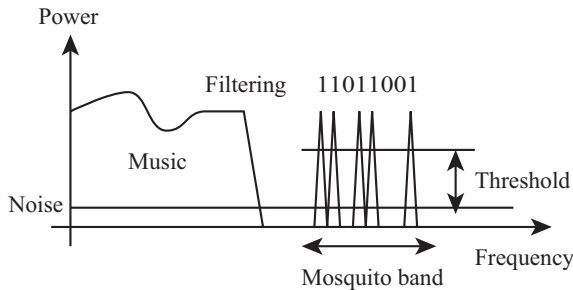


図 2 提案法の概要

Fig. 2 Overview of the proposed system

加齢とともに聴取能力が衰える高齢者は聴取出来ず、目的とする若年者にのみ聞こえる特徴がある。なお、初めて用いられたモスキート音は 17.4 kHz の正弦波であるとされるが、日本では 17.6 kHz が用いられたこともあるため、本稿では、可聴域、超音波、モスキート音の帯域 (以下ではモスキート帯域と呼称する) を表 1 のように定義する。モスキート帯域の上限を 21 kHz としているが、これは、本実験で用いるスマートフォンで収録可能な音の最大周波数に基づいて決定している。

3.1 提案法の概要

提案法の概要を図 2 に示す。まず、提案法ではローパスフィルタにより音楽に含まれるモスキート帯域の成分を除去する。その後、モスキート帯域に複数の正弦波を重畳する。1つの正弦波が1つのビット情報を表すため、 N 個の正弦波で N bit の情報を埋め込むことが出来る。また、各ビットの 0, 1 判定は、正弦波が生ずる周波数にエネルギーが存在するか否かにより行われる。エネルギーの有無を判定する閾値は、周辺帯域の雑音のエネルギーから計算される。

提案法の基本的な考え方は音響 OFDM に近いが、音響 OFDM は、各正弦波のエネルギーで音楽のスペクトル包絡を近似し品質劣化を避ける点、および隣接する正弦波の位相によりビット情報を制御する点で性質が異なる。また、従来法は、モスキート帯域より低い帯域も利用するため、パワーのみを制御することは音楽信号の品質劣化を引き起こす。音楽の劣化を避けるためには聴覚的に鈍感な位相の制御が一般的とされてきたが、提案法は元々聴取が困難なモスキート帯域を用いるため、パワーのみ制御した場合においても品質劣化が生じにくい利点がある。

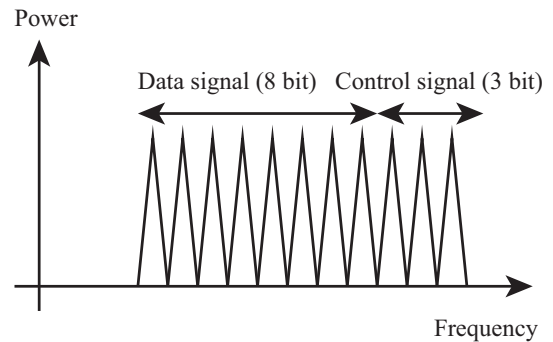


図 3 ビット列と、各ビットに関連付けられた情報の定義

Fig. 3 Bit sequence and the definition of the information associated with each bits

3.2 パケットの定義

提案法は単に複数のビット列を送信するだけである。提案法を情報通信に利用するためには、ビット列の存在を検出する信号や、情報の種類を識別する条件を示す信号を与える必要がある。ここでは、情報通信を行う際の仕様について述べる。

本システムでは、図 3 に示されるように、11 bit の情報をまとめて送信し、そのうち 8 bit をデータ信号、3 bit を制御信号として用いる。また、この 11 bit の情報を含む信号をパケットとする。3 bit の制御信号は、以下の目的で用いられる。

- パケットの有無の検出
- パケット列の開始・終了時刻の検出
- 通信速度の向上

3.3 パケットの制御信号

図 4 は、制御信号として用いられる 3 bit の種類と、あるパケット列を送信する時間周波数表現の例を示す。Trigger1, 2 は、デコードの際切り出された波形にパケットが含まれるか識別するために用いられる。パケットは、Trigger1, 2 のどちらか一方が 1 の場合のみ存在するものとする。Condition は、パケットの種類を示すフラグであり、パケット列の先端と終端で 1 を示す。Condition が 1 の場合は、情報通信の 8 bit でパケット列を識別する情報を送信する。本実装では、Condition が 1 の場合、情報通信の 8 bit でパケットの開始・終了を識別、および送信される情報が歌詞かタイトルや URL かを識別するフラグを与えることとした。特にパケットの開始・終了のフラグは、音楽信号を途中から収録した場合の検出、読み出された情報を画面に出力するタイミングを与えるために用いられる。

図 4 のパケット列の時間周波数表現では、まず、開始を示すパケット (以下では開始パケットとする) が、Trigger1 が 1 の状態で送信される。開始パケットを受信後は、終了を示すパケット (終了パケット) を受信するまで 8 bit ずつ情報を送信する。終了パケットを受信した場合、これまで

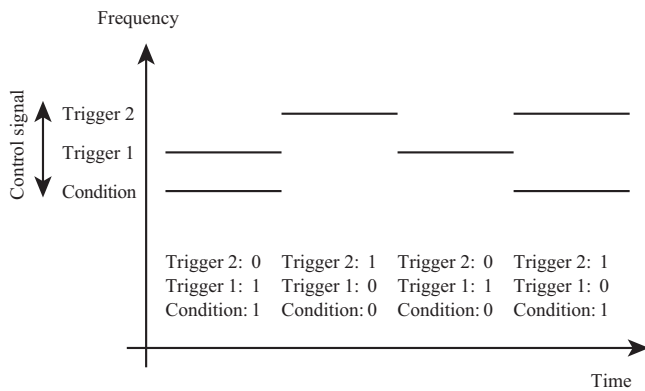


図 4 制御信号の種類と、パケット列を示す時間周波数表現の例
 Fig. 4 The types of the control signal and the example of a time-frequency representation of a packet sequence

受信してきた情報を、開始パケットで与えられた条件に基づいて画面に出力する。

検出トリガーが 1 つの場合、連続する 2 つのパケットで同一の情報を送信する際パケット間の継ぎ目を検出することが困難である。提案法では、この問題に対応するため、パケットが切り替わるタイミングでパケットの有無を検出するためのトリガーが切り替える。トリガーはパケット毎に切り替わり、パケットは一方のトリガーが 1 の場合のみ検出されることとする。2 つのパケットの中央を切り出し 2 つのトリガーが同時に 1、あるいは 0 となった場合は、そのフレームを無視することとする。

終了パケットが受信できなかった場合は次のパケットを受け取ることもできないため、2 つのトリガーが 2 フレーム以上 0 になった場合、全情報を初期化し開始パケットの待ち状態とする。また、開始パケットを抽出されずに他のパケットを検出した場合は、検出されたパケットをそのまま捨てることとする。

3.4 パケット信号の位相

提案法はモスキート帯域におけるパワーの有無のみで情報を埋め込むことが出来るため、任意の位相を付与することが出来る。ここでは、適切な位相を与える方法について議論する。

パケットとなる波形に与える要求事項として、時間的にエネルギーが分散し切り出し時刻に依存せずパケットを検出できることが望まれる。この要求は、同じエネルギーの信号の場合、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の与え方として、Schroeder 位相 [9] を採用する。

Schroeder 位相 $\theta(n)$ は、入力波形 $x(t)$ が基本周波数を f_c Hz とする周期信号の場合、以下の式により与えられる。

$$x(t) = \sum_{n=1}^N s(n) \sin(2\pi n f_c t + \theta(n)), \quad (1)$$

$$\theta(n) = -2\pi \sum_{l=1}^{n-1} (n-l)p(l) \quad (2)$$

$$p(n) = \frac{s^2(n)}{\sum_{l=1}^N s^2(l)} \quad (3)$$

$s(n)$ は、各調波の振幅を示す。Schroeder 位相を計算するため、 $s(n)$ は $n = 1$ から値が与えられるが、提案法で用いる 11 bit の区間以外の n において $s(n) = 0$ となる。なお、スピーカやマイクロフォンの特性によりモスキート帯域の正弦波が均一の音圧で出力されない場合、 $s(n)$ を 1 以外の値に設定することで、スピーカ・マイク特性を補正することもできる。

図 5 に、同じパワースペクトルを有する信号に、ゼロ位相、ランダム位相、Schroeder 位相を与えて得られた波形を示す。各図の上部には、最大振幅と最小振幅の差を示す数値が記載されている。ゼロ位相の場合、周期的なピークが観測され、その結果大きい振幅差が生じることとなる。ランダム位相は、ゼロ位相と比較して振幅差が大きく低減できていることが分かる。Schroeder 位相は、ランダム位相と比較してさらに振幅差が低減可能である。これは、Schroeder 位相を用いることで、相対的に大きいエネルギーが放射できることを意味する。

4. 歌詞連動型音楽試聴システム「Clairvoyant」

ここでは、3 章で述べた情報通信技術を実装した音楽試聴システム「Clairvoyant *1」について述べる。本システムは、情報を埋め込まれた音楽信号から埋め込まれた情報を抽出し表示する機能を有する。

4.1 インタフェースの説明

図 6 は、iPod touch 上に実装された Clairvoyant の実行画面を示す。Clairvoyant には、操作用のボタンとして情報をリセットする Reset ボタン (左上)、収録を開始・終了するボタン (左下) が実装されている。音楽に埋め込む歌詞以外の情報には、楽曲を試聴した際重要になる情報として、「曲のタイトル」「アーティスト」に加え、Web サイトへリンクするための「URL」を採用した。曲のイントロ部分や間奏部分では歌詞を表示する必要が無いため、これらの空白の時間に上記 3 つの情報を埋め込むことが出来る。タイトル、アーティスト、URL はそれぞれ抽出後の情報が表示されている。URL に関しては、クリックすることでブラウザが起動しその URL の Web ページが表示される。

図 6 の下部に表示されている 2 つのスライダーは、左が観測信号波形の振幅絶対値の最大値、右がモスキート帯域の音圧をそれぞれ表示している。音信号を収録する場

*1 Clairvoyant は透視能力者の意味である。

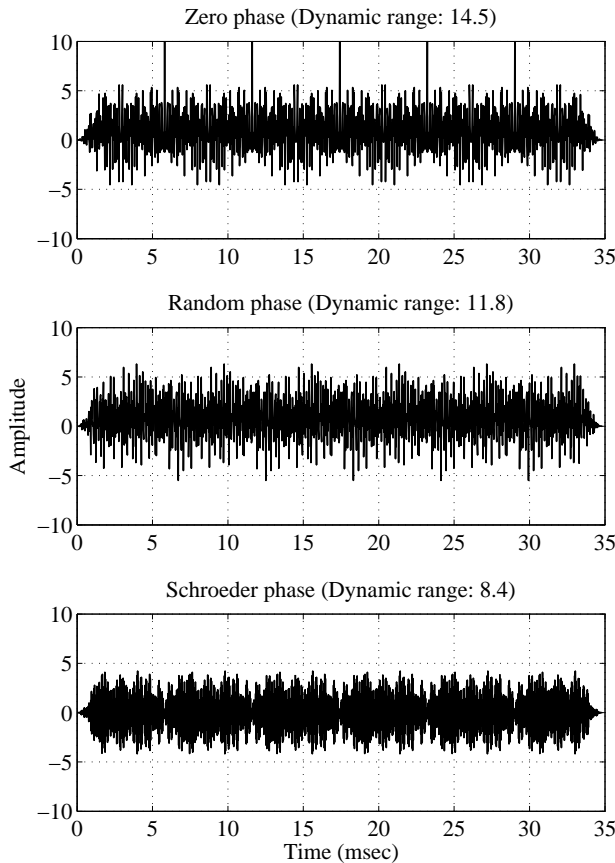


図 5 Schroeder 位相の効果
 Fig. 5 Effectiveness of Schroeder phase

合、クリッピングによる歪はモスキート帯域にも影響を及ぼすため、マイクボリュームの-slider は、適切な音圧で音信号が収録されているか確認する指標となる。なお、iPod touch はマイクゲインをコントロールできないため、ボリューム制御は音源とマイクフォン間の距離を物理的に制御する必要がある。

4.2 リアルタイムの歌詞表示機能

図 6 の中央に表示されているピッカーは、開始パケットにより示された通信情報が歌詞の場合、終了パケットを検出した直後に回転し読み込んだ歌詞を表示する。本実装では、任意の時刻でピッカーを回転させたい場合、終了パケットの時刻とピッカーを回転させたい時刻とが一致するようにパケット列の開始時刻を決定する必要がある。なお、本システムでは、2 つのパケットが時間的にオーバーラップすると正しく検出できないため、2 つの長い歌詞が短時間に近接する場合、提案法を用いて埋め込むことができない。

4.3 信号処理に関する実装の詳細

ここでは、本システムに用いた提案法の詳細なパラメタ



図 6 Clairvoyant のスナップショット
 Fig. 6 Snapshot of the implemented system named Clairvoyant

設定について説明する。本システムは、CD 相当の品質で動作することを想定しており、サンプリング周波数を 44.1 kHz、量子化ビット数を 16 bit に設定した。波形の切り出しには 1024 サンプルの Hamming 窓を利用し、フレームシフト幅は 512 サンプルとした。パケット長は、フレーム長 1024、シフト幅 512 サンプルの条件に対し、切り出し時刻の初期値に依存せず 1 つのフレームが 1 つのパケットのみを確実に含む条件を満たす最小の長さとして 1536 サンプルとした。

各ビットに対応する正弦波の周波数は、1024 点で切り出した際の離散周波数と対応させるため 44100/1024 Hz の整数倍とし、440 点目 (18,949 Hz) から 480 点目 (20,672 Hz) まで 4 点刻み (172.3 Hz) とした。なお、この上限は、iPod touch のマイク特性を計算し読み取れる上限周波数からやや低い値として決定している。

本実装では、1 パケットが 1536 サンプルで 11 bit の情報を送信できるため、通信速度は 315.8 bit/sec となる。ただし、3 ビットは制御信号として用いるため、実際に送信されるビット数で計算すると 229.7 bit/sec となる。全角 1 文字を 2 バイトで計算した場合 1 秒につき 14.4 文字送信できるが、開始・終了を伝えるパケットを前後に付与するため、歌詞情報を転送する通信速度はこの数値よりやや低い値となる。この通信速度の妥当性については、次章で議論する。

5. 考察と今後の展望

本章では、上述の信号処理技術、および実装された音楽試聴システムの有効性について議論する。なお、ここでは詳細な評価までは行っておらず、定性的に行われたシステ

ムアンケート調査を議論の対象とする。

5.1 信号処理技術の有効性

モスキート音を用いた情報通信技術は、位相を用いた従来法と比較して、実環境で混入する雑音に対して頑健であることが期待される。実装されたシステムを雑音や雑音、あるいは別の音楽が再生されている環境で利用したところ、これらの雑音に影響されずに情報を通信できていた。また、1秒当たりの歌詞の表示に関して、Vocaloid 楽曲で極めて高速に歌詞が表示される「初音ミクの消失 [10]」を用いてテストしたところ、概ね実時間で動作させることが出来ていた。

本実装では 1,723 Hz の帯域幅を情報埋め込みに用いているが、この刻み幅は、実装に用いた窓関数の種類と長さから、隣接する周波数からの漏れを理論上十分低減するため広く設定されている。実用上十分な刻み幅については今後検討する必要がある。

5.2 提案システムの有効性

大学生約 30 名を対象に提案システムを体験してもらったところ、多くの学生が歌詞の実時間表示、および表示される情報に対し好意的な回答を寄せた。一方、iPod touch のマイクロフォンをスピーカ方向に向けない場合、精度が著しく低下することも指摘された。これは、高周波であるモスキート音を用いるため音の回折が生じにくいこと、および iPod touch に搭載されたマイクロフォンの設置位置に影響されたことが考えられる。また、読み込みされた情報が誤っていた場合文字化けすることについても指摘された。この問題については、誤り訂正符号を導入することで改善できると考えられる。

5.3 課題と今後の展望

実装されたシステムを利用した結果、実環境の雑音に対しても頑健であり、早いテンポの楽曲を対象にしても実時間で情報を埋め込めることが確認された。特に、iPod touch に備え付けられたマイクロフォンでも正しく情報を読み出せることから、本技術はスマートフォンに対して情報を送るために有効であると考えられる。

本システムは、埋め込まれた情報を読み込むデコーダとしての機能しか有さないため、情報を音楽に埋め込むためのソフトウェアが別途必要となる。現時点での実装では、歌詞情報と表示するタイミングをコマンドラインで与えることでパケット列を生成する方法を採用しているが、システムを利用する観点からは、GUI を用いて容易に情報の埋め込みが可能ないように改善する必要がある。

本稿では、音楽試聴システムを実現を目的としているが、提案する情報通信技術は、様々な状況での応用が考えられる。具体的には、駅のホームのアナウンスにおける聴覚障

害者のサポートや、BGM として再生される様々な音楽信号に情報を埋め込むことで、必要な情報を容易に取得することが考えられる。また、従来の音響電子透かし技術は、情報を埋め込む媒体となる音楽信号を要求するため無音での通信は不可能であるが、提案法は聴取が困難なモスキート音のみで情報通信することが可能である。美術館における展示品の説明や、図書館でのアナウンスの場合、音楽を放射することは好ましくないため、静かな環境における非可聴情報通信は、提案法と相性の良い応用例といえる。

6. おわりに

本稿では、スピーカ再生音を対象として音信号に情報を埋め込む信号処理技術、および埋め込まれた情報をリアルタイムで抽出し画面に表示する音楽試聴システムを提案した。提案したシステムは iPod touch 上で動作し、音楽信号に埋め込まれた曲のタイトル、アーティスト、音楽紹介用 Web サイトにリンクする URL、歌詞をリアルタイムで表示する機能を有する。提案した情報埋め込み技術は、雑音や反射が存在する環境でも情報通信を行うことが可能であり、早いテンポの楽曲に対しても動作可能な通信速度が達成可能であることが示された。今後は、通信性能を向上させるための誤り訂正符号の導入や、通信速度の改善を目指した信号処理に関するパラメタの最適化を行う予定である。

参考文献

- [1] 平川学, 飯島淳一: 音声データに対する電子透かし技術の実装とモバイルサービスに関する一考察, 経営情報学会 2009 年春季全国研究発表大会, B4-3, pp. 113-116 (2009).
- [2] Tachibana, R.: Improving audio watermark robustness using stretched patterns against geometric distortion, in Proc. PCM'02, pp. 647-654 (2002).
- [3] Kirovski, D., Malvar, H.: Robust spread-spectrum audio watermarking, in Proc. ICASSP'01, Vol. 3, pp. 1345-1348 (2001).
- [4] 西村明, 坂本真一: 音響透かしを用いたカラオケ歌詞表示システムの性能評価, 信学技報, EA2008-59, pp. 83-88 (2008).
- [5] Nishimura, A.: Audio watermarking based on sinusoidal amplitude modulation, in Proc. ICASSP 2006, Vol. IV, pp. 797-800 (2006).
- [6] Ko, B., Nishimura, R., and Suzuki, Y.: Robust watermarking based on time-spread echo method with sub-band decomposition, IEICE Trans. Fundamentals, Vol. E87-A, No. 6, pp. 1647-1650 (2004).
- [7] 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, 信学技報, EA2006-43, pp. 7-11 (2006).
- [8] 松岡保静, 中島悠輔, 吉村健: 可聴帯域における音波情報伝送技術—音響 OFDM—, 信学技報, EA2006-24, pp. 25-29 (2006).
- [9] Schroeder, M.R.: Synthesis of low-peak-factor signals and binary sequences with low auto-correlation, IEEE Trans. Information Theory, Vol. 16, No. 1, pp. 85-89 (1970).
- [10] 初音ミクの消失 (LONG VERSION), <http://www.nicovideo.jp/watch/sm2937784>