

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

西村 明[†] 坂本 真一^{††}

[†] 東京情報大学 情報文化学科

〒 265-8501 千葉県若葉区御成台 4—1

^{††} 株式会社 オトデザイナーズ

〒 351-0104 埼玉県和光市南 1—27—65

E-mail: [†]akira@rsch.tuis.ac.jp, ^{††}sakamoto@otodesigners.com

あらまし 振幅変調に基づく音響透かし技術を用いて、カラオケの伴奏に歌詞表示情報を埋め込んでスピーカ再生し、マイク受音して伴奏に同期して歌詞を表示するシステムを作成した。システムの性能評価として、データ埋め込み済み音楽信号に対してPEAQを用いた音質劣化の客観評価を行った結果、劣化は平均的に「やや気になる」以下であることが分かった。また、残響および背景雑音付加、入力過大による振幅制限、歌唱音の重畳による検出力への影響をコンピュータシミュレーションによって調べた。その結果、それらの変形を経た伴奏音からも、十分埋め込みデータの検出は可能であり、歌詞表示の時間制御の基となるデータフレーム境界時刻の検出精度も十分であることが分かった。
キーワード ステガノグラフィ、残響、歌唱音、耐性、客観的音質評価

Performance evaluation of the system for displaying Karaoke lyrics using audio watermarking

Akira NISHIMURA[†] and Shinichi SAKAMOTO^{††}

[†] Tokyo University of Information Sciences, Department of Media and Cultural Studies

4—1, Onarida, Wakaba-ku, Chiba 265-8501

^{††} Otodesigners co ltd.

1—27—65, Minami, Wako, Saitama 351-0104

E-mail: [†]akira@rsch.tuis.ac.jp, ^{††}sakamoto@otodesigners.com

Abstract A watermarking technique based on amplitude modulation was applied to a system that displays Karaoke lyrics synchronously with the watermarked audio signal reproduced by loudspeakers. The quality degradation of the embedded music signals measured by PEAQ algorithm corresponded to below ‘slightly annoying’. The performance of the system was evaluated by the computer simulation in terms of the robustness of detection against reverberations, additive noise, amplitude clipping, and singing voices. The results showed that the detection performance was sufficient despite such disturbances. The temporal precision of detected data frame boundary for lyric synchronization was also sufficient.

Key words Steganography, Reverberation, Singing voice, Robustness, Objective quality evaluation

1. はじめに

音響信号へのデータハイディング技術は、音響信号自体に気づかれないようにデータを埋め込み、必要なときに検出して利用する技術である。従来は、著作権管理を目的として、様々な変形を経ても埋め込みデータを検出できる音響電子透かし技術の開発が進められてきた。近年、スピーカから再生されたデータ埋め込み済み音響信号を、ユーザの手元の機器で受信して復

号化し、埋め込まれた情報を利用する、といった利用形態を想定した技術が開発されている [1]– [4]。

著者らは、こうしたデータ埋め込み済み音響信号の空間伝搬と受信を前提とした利用形態として、データ埋め込み済み音響信号に同期してユーザへ情報呈示する技術を開発した [5]。また、その具体的な応用として、カラオケ伴奏音楽に歌詞の呈示情報を埋め込み、マイクで受信した信号より埋め込み情報をリアルタイムに検出し、歌詞の呈示タイミングに合わせて歌詞の

表示を行うシステムを開発した。このシステムは、カラオケだけでなく、映画における字幕情報の提示など、音響信号と同期した情報の提示/活用が必要な場面において有効である。

本稿は、このシステムをカラオケ歌詞表示に利用する際に生じるであろう、情報埋め込み済み伴奏音を收音する際に混入する残響音、反射音および歌唱声によるシステム性能の低下を、定量的に評価することを目的とする。

2. 空間伝搬耐性のある音響透かし

データ埋め込み済み音響信号の空間伝搬と受信を前提とした場合、音響信号に加わる変形としては、

- 背景雑音の重畳
- 反射音や残響音の重畳
- 発信側の DA 変換器と受信側の AD 変換器のサンプリング周波数の相違
- スピーカやマイクロホンによる周波数帯域の制限や変形
- スピーカにおける歪や、AD 変換時の過大入力によるクリッピング歪

などが挙げられ、これらの変形に対する耐性を備えた技術を用いることが重要である。また、伝送情報量はなるべく多くとりながら、音質劣化は最小限に抑える必要もある。

これらの要求を、従来の音響透かし技術が満たすか検討してみる。パッチワーク法 [6] は強度変化を与える時間幅が 100ms 程度と短いので、反射音や残響の影響を受けやすい。エコー拡散法 [7] は、反射音や残響の影響を軽減するためには、埋め込み時間フレームを長く設定する必要があり、埋め込みデータ量は限られてしまう [8]。2 チャンネル伝送を前提として片チャンネル毎に強度変調を与える手法 [4] は、反対側チャンネルの影響が少なくなるよう、透かし検出時にはマイクロホンを一方のスピーカに近づける必要がある。また、スペクトル拡散法 [1] は、付加雑音には強いが、埋め込みデータ量は十分でない。音響 OFDM [9] は埋め込みデータ量は格段に多いが、反射音や残響音の付加による伝送位相特性の変化は、埋め込みデータ検出に大きく影響を与えると考えられる。

3. 振幅変調に基づく音響透かし

著者が考案した振幅変調に基づく音響透かし [3], [10], [11] は、空間伝搬を前提とする透かし技術に対する前述の要求を、ほぼ満たしている。ここでは、その埋め込みと埋め込み区間検出手法について簡単に説明する。埋め込まれたデータの検出法については、文献を参照されたい。

3.1 埋め込み方法

本方式では、2 つの隣接する周波数帯域に分割された信号同士にそれぞれ逆位相の正弦振幅変調を与える。透かし埋め込み帯域を全て帯域分割し、このペアとなる隣接帯域を複数含む 2 つ以上のグループに各帯域を分類し、そのグループ間の変調位相差にすかし情報を埋め込む (図 1 参照)。埋め込みデータフレーム毎に、基本となるパイロット帯域グループの変調位相を反転させることによって、検出時に埋め込み区間の同期検出を可能とする。さらに、すべての帯域ペア間には、埋め込み時の

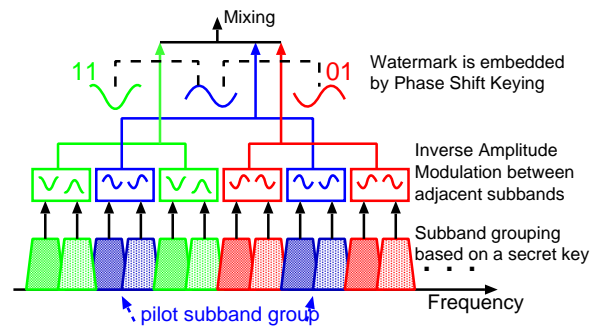


図 1 埋め込み処理の概要

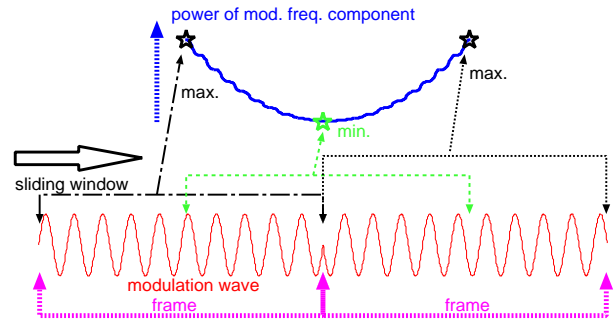


図 2 埋め込み区間境界検出の概要

鍵によってランダムに決定された初期変調位相差があらかじめ与えられる。透かしデータの符号化は、位相差 $\pi/2$ 毎に値を割り当てる 4 値の PSK 方式をとる。

3.2 埋め込み強度決定法

埋め込み時に与える振幅変調が確実に検出できるよう、音楽信号に含まれる固有の振幅変動成分を基準とした埋め込み変調強度設定方法を用いる。まず透かしを入れる前の信号において、ペアとなる 2 つの隣接帯域より得られた一定時間の振幅包絡波形に対して透かし検出演算、つまりそれらを除算し対数をとる演算を行う。得られた変動波形を周波数分析し、埋め込み変調周波数での変動振幅を A とする。このとき、透かし検出時に振幅 A の変動を与えるための帯域ペアへの振幅変調度を x とおくと、 $x = (\exp(A) - 1) / (\exp(A) + 1)$ となる。この変調度を基準 (0dB) として、埋め込み時に与える振幅変調度 X ($20 \log_{10}(X/x)$ dB) を定める [12]。

3.3 埋め込み区間同期検出方法

透かし入り信号に対して、FFT をオーバーラップさせながら実行し、絶対値をとることによって振幅変動波形を得る。その後、パイロット帯域グループの変動波形に対して、埋め込み周期おきに加算と減算と繰り返して 1 周期分の累積変動波形を得る。これは埋め込み周期ごとに位相が反転しているため、反転分を補正して同期加算することに相当する。そのフレーム開始時刻を、埋め込み周期分だけずらしながら累積変動波形を求め、埋め込み変調周波数において最大変動パワーが得られたフレーム開始時刻を埋め込み区間の先頭として特定する (図 11 参照)。

4. 透かし埋め込みに伴う音質劣化の評価

音響電子透かしをスピーカ再生とマイクロホン受信環境にお

表 1 データ埋め込み条件

Parameters	Normal	Robust
sampl. freq.	44100 Hz	←
freq. region	≤ 11025 Hz	≤ 22050 Hz
subband pairs	64	127
subband groups	17	9
frame period	3 s	←
mod. freq. [Hz]	1.67, 2.0, 2.33, 3.0	←
watermarking intensity	+12 dB	+15 dB
payload [bits/frame]	128	64
error correction	BCH(127,29,21)	BCH(63,24,7)
bit rate [bps]	9.7	8

いて利用する場合、より頑強なデータ検出のため、埋め込み強度を強めることになり、音質劣化も知覚可能な程度となる。

音質劣化度を客観的に測定する手法として、ITU-R 勧告 BS.1387-1、PEAQ(Perceptual Evaluation of Audio Quality) がある。これは原音(劣化なし音)と加工音(劣化あり音)をそれぞれ、聴覚フィルタを模したフィルタ群で帯域分割した上で、絶対閾値、周波数マスキングや時間マスキングを考慮した興奮パターン上での相違の度を計算し、主観評価実験(ITU-R BS.1116-1)で得られる劣化度を予測する手法である。

ここでは、Kabal[13]によるPEAQの基本バージョンの実装を用いて、データ埋め込み済み音楽の音質劣化度を測定した。対象音楽は、RWCポピュラー音楽データベース(RWC-MDB-P2001)[14]に含まれる100曲の左チャンネル冒頭60秒とした。また比較対象として、MP3符号化後に復号化した場合も、同様に測定を行なった。

表1には2つの埋め込み条件が示されている。Normalは、一般的な雑音や残響のある環境で利用する場合であり、BCHエラー訂正符号を用いることによって、9.7bpsの情報を検出することができる。Robustは、6.節で示す、カラオケでの歌詞表示利用を前提として残響や歌声が重畳する環境で利用する場合の埋め込みパラメータ値である。

図3には、2つの埋め込み条件の結果と、MP3で圧縮した音楽についての、音質劣化度の平均値と±1標準偏差の値を示した。この結果から、データ埋め込みに伴う音質劣化は、平均的には「やや気になる」以下程度であることが分かった。また、MP3と比較すれば、48kbps/chと64kbps/chの中間程度の音質であることが分かった。なお、Normal条件の結果は、前報[5]より劣化度が小さいが、これは前報において埋め込み済み信号を処理する際に一部の楽曲で振幅がクリップしており、今回はクリップしないようゲイン調節を行った結果である。

PEAQはMP3などの知覚符合化圧縮に伴う音質劣化を予測することを前提としているため、ここで示した振幅変調に伴う音質劣化の予測結果は、あくまで参考値である。振幅変調に伴う音質劣化も知覚符合化と同等に予測できるかどうかについては、4名の被験者に対して、4曲各4条件の音質劣化について、PEAQによる結果とITU-R BS.1116 準拠による主観評価結果との相関を調べた結果、振幅変調に伴う音質劣化の相関係数は有意ではあるが、MP3の音質劣化の相関係数より低いこと

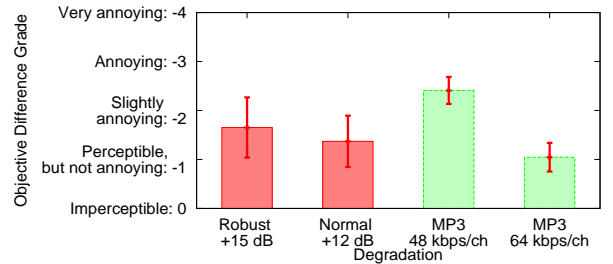


図 3 PEAQ 測定によって得られた平均客観品質劣化度合と ± 1 標準偏差。素材は RWC-MDB-P2001 に収録された 100 曲。

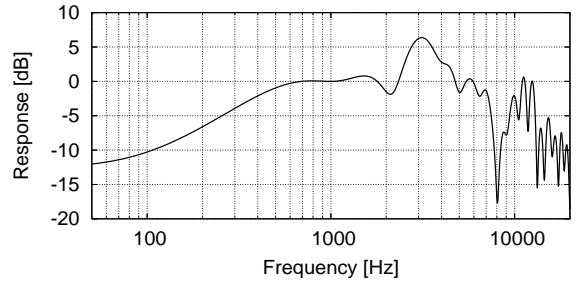


図 4 インパルス応答における直接音成分のワースペクトル。

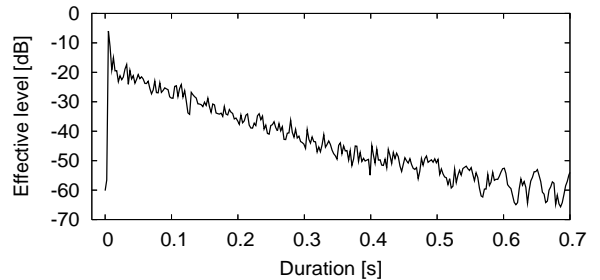


図 5 インパルス応答におけるエネルギー減衰。

が分かっている[15]。今後は、主観的劣化と客観的劣化指標との対応についてさらなる検証を進めるべきであろう。

5. 振幅変調に基づく音響透かしの空間伝搬耐性

上述の手法が、残響あるいは入力過大、また雑音に対する耐性を持つことをシミュレーションによって確認した[5]。埋め込み条件は、表1におけるNormal条件を用いた。

5.1 シミュレーション条件

対象楽曲はRWC-MDB-P2001に含まれる100曲の左チャンネル冒頭60秒とした。これに、RWC実環境音声・音響データベース[16]より選んだ、残響時間1.3秒の可変残響室で収録されたインパルス応答(ファイル名: ir130.dat)を畳み込んだ。このインパルス応答波形の絶対値ピークを中心とした128サンプルにハニング窓掛けを行って直接音成分を取り出し、ワースペクトルを求めることにより、模擬されるスピーカからマイクへの伝達特性を求めて、図4に示した。これにより、スピーカやマイクのフラットでない伝送特性も模擬できることが分かる。また、図5には、インパルス応答波形について110サンプル毎に求めた実効レベルの時間減衰を示した。ここで定義した直接音に対する間接音の比率は-0.4dBとなる。

その後、背景雑音として4種類の環境騒音(収録場所: 駅の

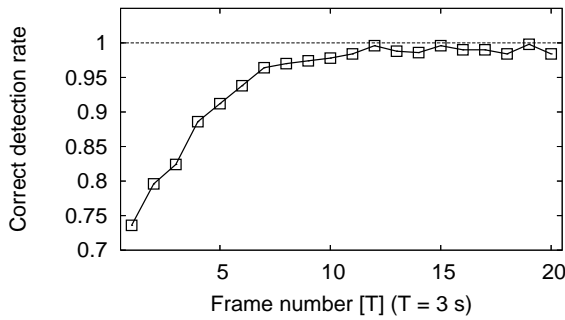


図6 残響と背景雑音加わったときの、エラー訂正限界以内に収まったデータフレームの割合。それぞれの点は500条件のシミュレーションから得られた。

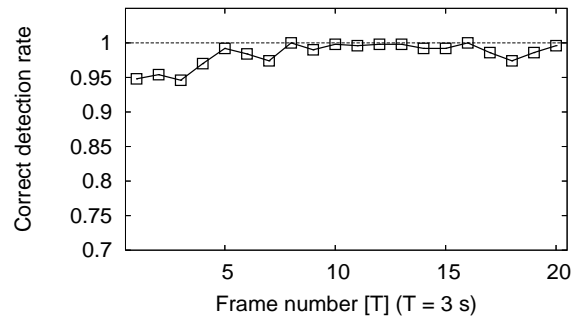


図7 背景雑音と振幅クリッピング加わったときの、エラー訂正限界以内に収まったデータフレームの割合。それぞれの点は500条件のシミュレーションから得られた。

ホーム、地下連絡通路、空港ロビー、混雑した交差点)、あるいはローパスノイズ(カットオフ 500 Hz、 -9 dB/oct. : 他の環境騒音の平均的スペクトルに近い)のいずれかを付加した後、透かし情報を検出する処理を行った。5種類の背景雑音は、オーバーオール音楽信号パワーに対して、信号対雑音比(SNR)は15 dBとした。

また、データ検出用のマイクロホンスピーカに近づけた場合は、入力過大により振幅がクリッピングする事態も考えられる。このような状況での耐性を調べるため、透かし入り音楽信号の最大振幅の0.125倍以上の振幅を制限する変形(+18 dBの入力過大)も模擬した。この振幅制限のシミュレーションでは、残響は付加せず、環境騒音のSNRは30 dBとして、振幅制限の直前に加えた。

100種の透かし入り音楽と5種の背景雑音を組み合わせて500条件において、上記のシミュレーションを行った。

5.2 フレームデータ検出率

データ検出率の指標としては、3秒間のデータフレームに埋め込まれた128ビットのうち、誤り訂正限界である21ビット以内にエラービット数が収まったフレーム数を、全体のフレーム数で割った正検出割合とした。

図6に残響と背景雑音付加の結果を、図7に背景雑音と振幅制限を与えた結果を示した。データ検出率は、先頭からのフレーム番号に依存するため、横軸をフレーム番号として示している。残響が加わった場合は、楽曲の冒頭において音楽信号レベルが相対的に小さいため、検出率が低くなっているが、第5フレーム(15秒)以降では、90%以上の検出率となっている。振幅制限の場合には、より良好に検出ができることが分かった。

6. カラオケ歌詞表示システム

ここでは、データ埋め込み済み音響信号に同期してユーザへ情報表示するシステムとして、カラオケ伴奏音楽に同期して歌詞を呈示するシステム[5]について説明する。音楽信号に埋め込まれるデータは、主に表示開始時刻と表示終了時刻、そして表示される情報へのポインタとなる情報であり、表示される情報自体は事前に表示システム側で保持する。表示システムとしては、最終的には携帯電話やPDAが望ましいが、現時点ではWindowsパソコン上のMATLABで実装している。

表2 カラオケ歌詞表示のために埋め込むレコード

相対開始フレーム	開始時刻	相対終了フレーム	終了時刻	表示歌詞のインデックス
6 bit	4 bit	4 bit	4 bit	6 bit

伴奏音楽に埋め込まない歌詞情報を表示システムへ取り込む方法としては、インターネットを経由した楽曲の販売時に付帯的なダウンロードとして実現するか、カラオケトラック入り音楽CD販売時に、ダウンロードのキーを同梱することなどが考えられるが、この仕組みは現時点でのシステムには含まれない。

以下に、歌詞表示システムの仕様を説明する。

6.1 埋め込むデータの構造

表1中のRobust条件にて埋め込みを行い、BCHエラー訂正符号を用いて符号化された63ビットの情報を、3秒間のデータフレーム毎に埋め込む。ここでは、BCH(63,24,7)を用いて、63ビットあたり7ビットまでの符号誤りを訂正できる。一方、伝送される情報量は24ビットであり、これを表2のようにビットを割り当て、データに分割する。

表2の相対開始フレームは、レコードを埋め込んだフレームに対する、表示を開始するフレームの相対的な位置を $1 \sim 64$ (6 bit)の整数で表現する。つまり、フレーム時間長3秒のときには、表示開始は最大で192秒先まで可能になる。開始時刻は、フレーム中の表示開始時刻は、フレーム時間長に対して $0/16 \sim 15/16$ で設定する。よって、フレーム時間長3秒の時は、最小の表示時間分解能は0.19秒(テンポ160での8分音符)となる。表示終了時刻については、表示開始フレームからの相対フレーム数を4bit、そのフレーム内での終了時刻を4bitで表現する。表示歌詞は、一回に表示する文にインデックスを1つ割り当てることにより、64文の歌詞を指定し、同じ歌詞には同じインデックスを割り当てることにする。このようなレコード定義は、カラオケの歌詞表示に限ったものであり、他の用途では、異なったビット割り当てやレコード長を用いるべきであろう。

復号化されたレコードデータはバッファメモリに格納され、AD変換器からデータを取り込む度に、各レコードの表示開始時刻データと終了時刻データを走査して、表示開始と表示終了処理を行なう。

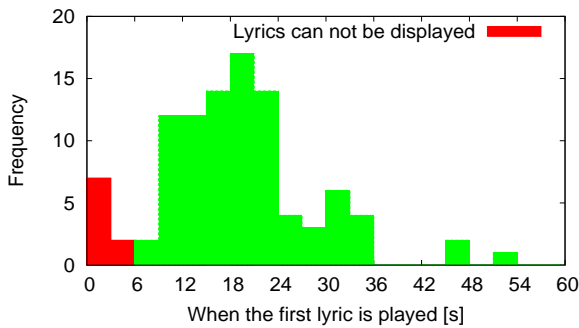


図 8 RWC-MDB-P2001 収録の 100 曲の歌詞の開始時刻と度数。

6.2 データ埋め込みと同期

1つの歌詞表示情報を埋め込む最小時間間隔はデータフレーム時間長であるが、演奏には歌の無い部分もあるため、フレーム時間長を3秒としても、実際にはフレームの数は表示する歌詞の数より数倍程度多いことが一般的である。途中から伴奏を再生したときにも歌詞が表示できるように、歌詞表示の直前の複数フレームにデータを埋め込み、冗長性を確保するのが良いだろう。実用的には、伴奏のどこに歌詞表示データを埋め込むかを最適に設計する埋め込みシステムの構築が望ましい。

データフレームの検出には、2フレーム(6秒)分の伴奏時間が最低必要である。3フレーム目以降では、検出されたフレーム境界を遡って、最初のフレームからデータの復号化と表示を行う。よって3フレーム目前に歌詞の表示は不可能であるが、このような楽曲がどの程度の割合で存在するかは明らかではない。図8に、RWC-MDB-P2001収録の100曲について、曲の開始から何秒後に歌詞が唄われるかを示した。この結果から、1割程度は冒頭の歌詞が表示できない場合があると思われる。

7. 歌詞表示システムの性能評価

歌詞表示システムの性能評価として、室内の反射や残響および利用者の歌唱音声による影響を、コンピュータシミュレーションによって調べた。性能評価の対象楽曲はRWC-MDB-P2001に含まれる100曲のMIDIファイルを、MIDI音源(timidity++)によって演奏した左チャンネル冒頭60秒とした。データ埋め込み時のパラメータは表1のRobust条件とした。

7.1 使用環境シミュレーション

残響のある室内を想定して、データ埋め込み済み音楽信号に、RWC実環境音声・音響データベースに収録されているインパルス応答(ファイル名: ir130.dat)を畳み込んだ。次にVOCALOID2「鏡音リン・レン」を用いて作成した歌唱音(男声はレン、女声はリンが歌唱)を、平均SNR -5, 0, +5 dBの3条件で付加した。この平均SNR値は、100msの矩形窓フレーム毎に、A特性フィルタリング後の残響つき伴奏音の実効値(signal)と、A特性フィルタリング後の歌唱音の実効値(noise)を計算し、歌唱音が含まれるフレームのSNR(dB)値の算術平均値として定めた。

7.2 フレームデータ検出率

データ検出率の指標としては、3秒間のデータフレームに埋

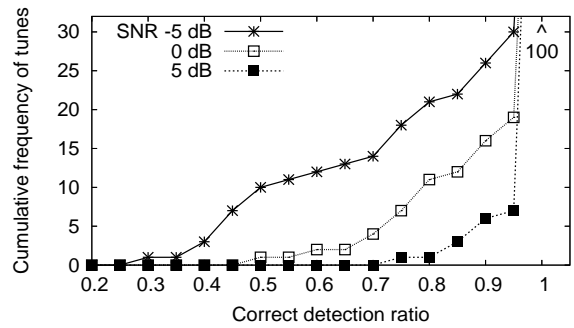


図 9 歌唱音が含まれるフレームに対するデータ検出率と累積曲数。

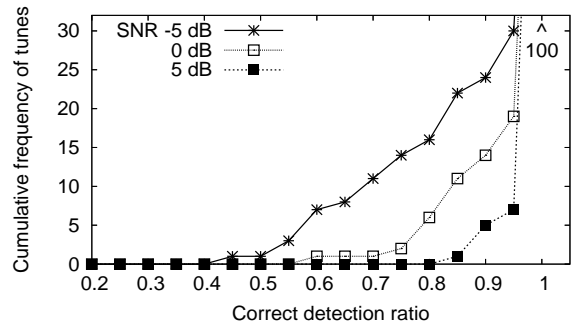


図 10 各曲の全フレームに対するデータ検出率と累積曲数。

め込まれた64ビットのうち、誤り訂正限界である7ビット以内にエラービット数が収まったフレーム数の率を調べた。歌唱音が含まれないデータフレームに対しては、全てのフレームでエラー訂正限界以内に収まった。

歌唱音がわずかでも含まれるフレームは、全2000フレーム中1369であった。歌唱音が含まれるフレーム数を母数としたデータフレーム検出率を、各曲毎に0.05刻みで調べた。図9の横軸は検出率であり、縦軸はその検出率以下となる累積曲数を示した。図10には、1曲の全フレーム数20を母数とした、データフレーム検出率を各曲毎に算出し、その検出率以下となる累積曲数を示した。

図9ではSNR -5dBの最も厳しい条件で歌唱音が重畳した場合には、100曲中18曲で歌唱音の含まれるフレームの検出率が0.75以下になることを示している。一方、歌詞表示情報の埋め込みは歌詞のあるなしに関わらず行われるため、曲全体の歌唱音のないフレームも含んだ場合の検出率を見ると、図10では、0.75以下の検出率となるのは14曲である。同じ歌詞表示情報を複数フレームに繰り返し埋め込むことを考えると、検出率はさらに高まるため、実用上は問題ないと考えられる。なお、歌唱音重畳時の検出率が低い楽曲は、伴奏がピアノのみ、あるいはギターだけの曲であった。

7.3 フレーム検出時刻の精度

データ検出時に得られるフレーム境界時刻は、伴奏音と歌詞表示の同期の基準時刻となる。このフレーム境界時刻の検出精度を、正しいフレーム境界からのずれ時間として、最も厳しい条件であるSNR -5 dBでの結果を図11に示した。

第4フレーム目(冒頭から12秒)までは、検出されたフレーム境界は平均で0.1秒程度早めとなるが、それ以降は0.1秒前

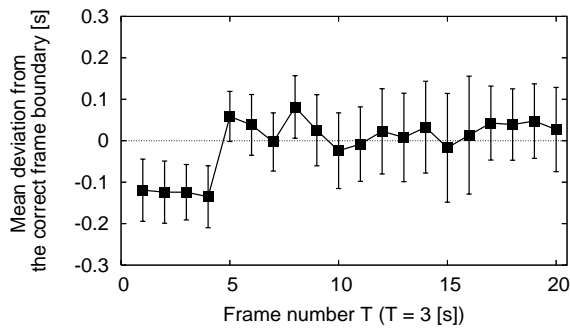


図 11 残響と歌声が加わった場合での、フレーム境界時刻からの、検出時刻のずれ。誤差棒は ± 1 標準偏差を示す。

後の標準偏差でほぼ正確にフレーム境界の検出が可能であった。また、表示における時間分解能は約 0.19 秒であるので、実用上ほぼ問題無く指定時刻に表示できるであろうことが分かった。

8. 考 察

スペクトル拡散法などの一部を除いて、音響電子透かし技術の多くは、本技術もそうであるが、音の無い部分にデータを埋め込むことはできない。本システムの用途としては、カラオケ歌詞表示以外にも、映画のサウンドトラックに字幕表示情報を埋め込んで、セリフに同期して字幕を表示するシステムが考えられる。この場合、映画サウンドトラックには、無音部分が多く含まれるという問題がある。これには、データ埋め込みフレーム時間を長くとする対応が考えられる。本技術では、データフレーム時間の一部にしか信号が含まれていなくても、その部分を有効に活用してフレーム同期とデータ検出を行うことができる。このことは、ポーズを含んだ音声信号でも、本技術は有効にデータ検出が可能であることから分かる [3]。

音響電子透かし技術は、短い時間区間 (1 秒以下) 毎に埋め込み対象の音響信号のスペクトルやマスキング特性に合わせて強度を調整しながら、データを埋め込む方法 [2], [7], [17] が多い。このような方式を用いて埋め込みを行った信号をスピーカ再生した場合、雑音の混入により、信号強度の弱い区間では検出エラーが多く生じる。本技術は、比較的長いデータフレームを用いているため、このような問題にも耐性が高いと言える。

9. ま と め

振幅変調に基づく音響透かし技術を用いて、カラオケの伴奏に歌詞表示情報を埋め込んでスピーカ再生し、マイク受音して伴奏に同期して歌詞を表示するシステムを作成した。システムの性能評価として、データ埋め込み済み音楽信号に対して PEAQ を用いた音質劣化の客観評価を行った結果、データ埋め込みに伴う音質劣化は MP3 48kbps/ch と 64kbps/ch で圧縮された音楽の中間程度であることが分かった。データ埋め込み済み音楽信号への残響および背景雑音付加、あるいは入力過大による振幅制限のシミュレーションの結果、音楽信号に埋め込まれたデータは十分検出可能であることが分かった。さらに、カラオケ歌詞表示システムとして利用時の、埋め込みデータ検

出への残響や歌唱音の影響も調べた。その結果、SNR0dB で歌唱音が付加されても、90%以上の楽曲で 80%以上のデータ検出が可能なが分かった。また、歌詞表示の時間制御の基となるデータフレーム境界時刻の検出精度も十分であることが分かった。

謝辞

本研究の一部には平成 20 年度東京情報大学共同研究 1，および科研費 (20560365) による補助を受けた。

文 献

- [1] 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, EA2006-43, 529–533 (2006).
- [2] 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,” 電子情報通信学会技術研究報告, EA2006-24, 25–29 (2006).
- [3] Akira Nishimura, “Data hiding for speech sounds using sub-band amplitude modulation robust against reverberations and background noise,” in *Proceedings of IHMSP2006*, 7–10 IEEE, (2006).
- [4] 茂出木敏雄, “音響空間のユビキタス化に向けた電子透かし埋め込み容量の拡大技術,” 電気学会論文誌 C, 127, 1013–1021 (2007).
- [5] 西村明, 坂本真一, “スピーカ再生音に同期した音響電子透かしを用いた情報提示・カラオケ歌詞表示システム,” 情報処理学会 音楽情報科学研究報告, MUS-71-2, 7–12 (2007).
- [6] Ryuki Tachibana, “Improving audio watermarking robustness using stretched patterns against geometric distortion,” in *Advances in Multimedia Information Processing, PCM2002, (LNCS) 2532*, 647–654 Springer, Hsinchu, Taiwan, (2002).
- [7] Byeong-Seob KO, Ryuichi Nishimura, and Yoiti Suzuki, “Robust Watermarking Based on Time-spread Echo Method with Subband Decomposition,” *IEICE Trans. Fundamentals*, E87-A, 1647–1650 (2004).
- [8] 西村明, “音響電子透かしの耐性評価シミュレーション,” 東京情報大学研究論集, 10, No. 2, 113–141.
- [9] 悠輔中島, 保静松岡, 健吉村, “音響 OFDM の空中伝送特性,” 電子情報通信学会技術研究報告. RCS, 無線通信システム, 106, No. 168, 139–144 (2006).
- [10] 西村明, “帯域分割と振幅変調に基づく音響電子透かし,” 暗号と情報セキュリティシンポジウム 2006, No. 3F4-2 電子情報通信学会, (2006).
- [11] Akira Nishimura, “Audio watermarking based on sinusoidal amplitude modulation,” in *Proceedings of ICASSP 2006, IV*, 797–800 IEEE, (2006).
- [12] 西村明, “振幅変調に基づく電子音響透かしによる音質劣化の評価,” 日本音響学会講演論文集, 秋 2–1–10, 455–456 (2006).
- [13] P. Kabal, “An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality,” TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University (2002).
- [14] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, “RWC Music Database: Popular, Classical, and Jazz Music Databases,” in *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, 287–288, (2002).
- [15] 西村明, “振幅変調に基づく音響電子透かしによる音質劣化の主観的評価と客観的評価,” 日本音響学会講演論文集, 春 1–7–7, 525–527 (2008).
- [16] 中村, 比屋根, 浅野, 遠藤, “実環境における音響シーンデータベースの構築,” 日本音響学会講演論文集, 秋, 621–622 (1998).
- [17] N. Cvejic and T. Seppänen, “Improving audio watermarking scheme using psychoacoustic watermark filtering,” in *Proc. 1st IEEE International Symposium on Signal Processing and Information Technology*, 169–172, (2001).