

振幅変調に基づくオーディオデータハイディング手法の空中伝搬と携帯電話音声符号化への耐性*

西村 明 (東京情報大)

1 はじめに

近年、データを埋め込んだ音楽や音声などの音響信号をスピーカから再生し、携帯電話やPDAなどのマイクロホンで受信してデータを検出する応用が考えられている。利用方法としては、商品やサービスに関する情報を、CM音楽に埋め込んでの聴取者に与えて広告宣伝効果を高めるもの [1]、聴覚障害者に対して案内音声などにその内容と関連の深い情報を埋め込む福祉用途 [2] などが挙げられる。

このような用途に適合するデータハイディング技術には、スピーカおよびマイクにおける伝送周波数特性の歪、反射音や残響音、背景雑音、マイク受信時の入力過大による振幅クリッピング歪などを経ても、十分に検出が可能であることが求められる。従来いくつかの技術が提案されているが、マイクロホンがスピーカに近接することを要求したり、埋め込み情報量が少なかったりする。著者の考案した振幅変調に基づく音響信号へのデータハイディング手法は、上記条件に耐性をもつことが、多くの音楽信号と音声信号に対して確認されている。

さらに、携帯電話を利用者の端末とする場合は、公衆電話音声ネットワークを通じて、サーバコンピュータへ情報埋め込み済み音響信号を送信し、検出する利用も考えられる。よって、携帯電話音声コーデックによる符号化と復号化に対する耐性が必要となる。

本研究では、振幅変調に基づくデータハイディングが、残響、付加雑音、音声コーデックに対してどの程度の耐性を持つかを調べ、データ埋め込み済み信号の携帯電話音声による伝送の可能性について検討を行なう。また、データ埋め込みに伴う音声品質の劣化度合を評価するために、客観的音声品質評価手法のひとつである PESQ を用いる。

2 振幅変調に基づく音響信号へのデータハイディング

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。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。

データ埋め込み区間同期検出と、埋め込みデータ検出手法については、文献 [3] を参照いただきたい。

3 耐性評価実験

音声信号として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を、話者ごとに連結して、1話者あたり36秒分の音声信号を22名分(男性10名、女性12名)作成して用いた。これらの信号はサンプリング周波数16kHzであったが、8kHzに変換して用いた。

3.1 スピーカ放射に伴う残響と背景雑音のシミュレーション

音声信号の4kHz以下の帯域にランダムなデータを埋め込み、残響や背景雑音を付加した後、AMR(Adaptive Multi Rate) コーデックによる符号化および復号化を経た後の波形に対して、検出処理を行なった。データ埋め込みビットレートは8bpsとした。埋め込み強度である、振幅変調度は0.4としたが、これは4節での客観的音質劣化度合の評価を元に決定した。

半分の条件では、残響のある室内においてスピーカ再生され、マイクロホン受信されることを模擬するために、RWCP実環境音声・音響データベースより選んだ、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。その後、室内の背景雑音に似たスペクトルを持つHothノイズを、SNR 10, 20, 30 dBのいずれかで付加した。なお、残響と背景雑音の無い条件も加えた。

AMR符号化および復号化ソフトウェアは、3GPP TS 26.073に付属しているANSI-Cコードをコンパイルして用いた。シミュレーション時のAMRコーデックのビットレートは、4.75~12.2kbpsのビットレートが2フレーム毎に28フレーム(0.56秒)の周期で連続的に変化する条件、6.7kbps、12.2kbpsの3種類とした。

3.2 実環境での実験

約410m³の正方教室において、前方中央壁面より1m、高さ1mの位置においた12cmフルレンジスピーカより、前節の情報埋め込み音声信号と、オーバーオールSNRを一定としたHothノイズを合わせて再生した。スピーカから5.4mの距離、高さ1.3mの位置に携帯電話の背面をスピーカに向けてマイクスタンドで固定し、ボイスレコーダ機能(AMR 12.2kbps)にて録音した。携帯電話横の騒音計での等価騒音レベルは、音声が65dBと55dBの2条件、Hothノイズは45dB、暗騒音は31dBであった。Log-TSP信号で測定した残響時間は約1秒であった。

3.3 結果

結果は埋め込んだビット値に対して得られた正しいビット値の割合を検出率とした。図1には、22名の音声信号での平均検出率を示した。エラーバーは、

*Robustness of audio data hiding based on amplitude modulation against air transmission and speech codec of cellphone network
Akira Nishimura (Tokyo University of Information Sciences)

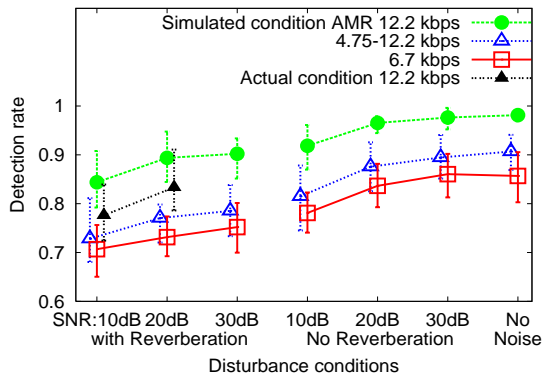


Fig. 1 Averaged detection rate for speech signals. Error bars denote the 10th through 90th percentile of 22 conditions.

22 条件中の 10 から 90 パーセントの範囲を示している。

これらの結果より、残響が無い場合は、SNR 20dB 以下では、6.7 kbps の AMR コーデックを経ても 80% 以上の検出率が得られた。残響が付加されると、検出率は大きく落ち、6.7 kbps の AMR コーデックでは検出率が 80% を下回ることが多く、事実上利用できないに等しい。しかし、12.2 kbps の AMR コーデックであれば、85%以上の検出率が得られることが分かった。また、現実の室内と機器を用いた環境では、SNR が 20dB 以上でないと、80% を上回る検出率を得るのは困難であることが分かった。

4 音質劣化の客観的評価

ITU-T 勧告 P.862 PESQ(Perceptual Evaluation of Speech Quality) は、電話帯域音声や音声コーデックの客観的品質を測定するためのアルゴリズムである。PESQ は原信号と符号化信号を比較し、心理音響特性に基づいた信号の内的表現の差分から、人間を被験者として測定した主観的な劣化度合評価値である MOS (Mean Opinion Score) に対応する評価値を算出する。

データ埋め込み後の音声信号の音質は、音声符号化を経た音質と似ているので、その客観的品質を PESQ により測定した。日本音響学会研究用連続音声データベース Vol.1 より、22 名の話者による 1200 の音素バランス文を 2 つづつ繋げて 600 文の 6~12 秒前後の音声信号とし、PESQ プログラムに入力した。

結果は、図 2 に示した。音質劣化の比較参考のため、AMR コーデックによる 4.75、6.7 および 12.2 kbps にて符号化および復号化した後の音声信号についても、同様に PESQ スコアを算出して示した。この結果から、データを埋め込んだ音声信号の品質劣化は、携帯電話品質よりやや音質が悪い程度であろうと予測される。また、同じ埋め込み強度(振幅変調度 0.4)にてデータ埋め込みを行なった前報 [4] では、VCV 音節の明瞭度試験を行っており、この場合 SNR 10 dB の条件でも平均で 93% の正答率となった。このことから、データ埋め込み済み音声信号の品質劣化は明らかではあるが、音声情報を伝えるには十分であろうと考えられる。

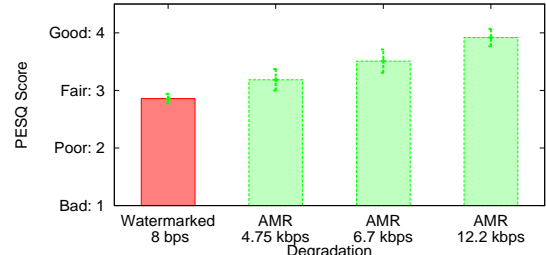


Fig. 2 PESQ score

5 考察

8 秒分の埋め込みデータ (64bit) に対して、埋め込み時に BCH エラー訂正符号を用い、復号時には軟判定復号法を併用してエラー訂正限界を 2bit 分向上させると、80% のビット検出率が得られれば、16bit 分の情報伝送が可能となり、実効データ伝送量は、2 bps となる。今回の残響のあるシミュレーションおよび実環境において、半分以上の信号条件で、80% のビット検出率を満たすのは、12.2 kbps で符号化したときのみであった。AMR 符号化を行わない場合 [4] には、より検出率が高く埋め込み情報量も多いことを考えると、AMR 符号化に適した埋め込み方法を検討する必要がある。

6 まとめ

様々な音声信号に 8 bps で振幅変調に基づくデータ埋め込みを行なった音響信号が、残響と背景雑音そして AMR 音声コーデックに対して耐性をもつかを調べた。その結果、背景雑音のみが重畳される場合は、6.7kbps 以上の AMR ビットレートにおいて、80% 以上のビット検出率が得られた。さらに残響が重畳する場合は、12.2 kbps の AMR ビットレートにおいて 80% 以上のビット検出率が得られた。また、データ埋め込みに伴う客観的音質劣化度合を、PESQ を用いて調べた。結果から、残響や背景雑音が存在しても、AMR コーデックのビットレートが高ければ、品質を大きく劣化させずにデータハイディングによる情報伝送が可能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。

参考文献

- [1] 松岡保静, 中島悠輔, 吉村健, “音響 OFDM - 音のパークード-,” 第 1 回 マルチメディア情報ハイディング研究会資料, 61—64, (2007).
- [2] 東北大学電気通信研究所, 国立特殊教育総合研究所, “空気伝搬音響透かしの福祉工学への応用,” 第 1 回 マルチメディア情報ハイディング研究会資料, 103—106, (2007).
- [3] 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).
- [4] Akira Nishimura, “Data hiding for speech sounds using subband amplitude modulation robust against reverberations and background noise,” in *Proceedings of IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 7–10 IEEE, (2006).