Audio data hiding based on amplitude modulation and its application

西村, 明
Faculty of Informatics, Tokyo University of Information Sciences

https://doi.org/10.15017/18879

第4章　情報秘匿信号の空間伝搬および携帯電話音声符号化耐性

4.1　まえがき

第3章では，振幅変調に基づく情報秘匿手法を提案し，様々な音楽信号に対する音響電子透かしとしての性能を評価した．その結果，残響や付加雑音に対して頑強であることが示された．この特徴は，この手法が音響電子透かしとしての用途だけでなく，必要な情報を音響信号に秘匿して送信/保存し，受信/読み込んで秘匿情報を取り出して利用するステガノグラフィ用途として，特に，ステゴ信号をスピーカ再生し，マイクロホンで受音する，空間伝搬条件でのステガノグラフィ用途に利用できることを示唆している．

近年，このような空間伝搬を前提とたステガノグラフィ技術がいくつか考えられている．この場合の受信器としては，携帯電話やPDAなどが挙げられる．埋め込むデータと利用方法をいくつか挙げてみると，まず，商品やサービスに関する情報を，音響信号 (例えばCM音楽) の聴取者に与えて広告宣伝効果を高めるものがある [58]．また，聴覚によって音響信号 (例えば公共空間でのアナウンス音声) の内容を把握できない聴覚障害者に対して，音響信号の内容と関連の深い情報を与える福祉用途 [12] がある．さらに，博物館等の展示情報へのポインタをスピーカ再生の音楽に埋め込んで，閲覧者のもつPDAで復号化し，ポインタの示す情報をあらかじめPDAに蓄積しておいて呈示したり，受音位置に応じた情報を呈示する方法 [59] も試行されている．

このような用途のためは，従来の音響信号に対する電子透かしやステガノグラフィにおいて用いられていた技術に無かった要求も現れてくる．つまり，スピーカ再生しマイクロホン収音することを経るなかで，ステゴ信号は以下に挙げる変形に耐性を持つ必要がある．

1. スピーカおよびマイクロホンにおける伝送周波数特性の歪
2. スピーカからマイクロホンまでの音響経路で生じる反射音や残響

3. マイクロホン受音時に付加される背景雑音

4. マイクロホン受音時の入力超過による振幅クリッピング歪

その一方で，従来の音響電子透かし用途において重要視されていた，知覚符号化 (MP3 や MPEG2AAC など) や，悪意ある攻撃 (ピッチ変換，時間長変換，データ切り取りなど) への耐性については，結果として耐性が高くなることも十分あるとは言え，あまり考慮する必要はない．また，ステゴ信号の音質劣化を用途によって許容される程度まで認めつつ，必要な耐性を確保し埋め込みデータ量を高める必要がある．

これらの要求を，主に音響電子透かしを目的として開発されたものが多い従来の音響情報秘匿技術が満たしているかを検討してみる．パッチワーク法 [39] は，強度変化を与える時間幅が 100 ms 程度と短いので，受音点がスピーカから遠ざかった場合に反射音や残響の影響を受けやすい．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 [11] は，反対側チャンネルの影響が少なくなるよう，情報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．エコー拡散法 [26] は，反射音や残響の影響を軽減するためには，第 5 章にみるように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．また，スペクトル拡散法 [60] は，付加雑音には強いが，埋め込みデータ量は数 bps と十分でない．音響 OFDM(Orthogonal Frequency-Division Multiplexing; 直交周波数分割多重方式)[61] は埋め込みデータ量は 1 kbps 程度と格段に多いが，情報を埋め込む帯域を高域に限定した上で，音楽信号のスペクトルに合わせて埋め込みデータ信号の振幅を調整するため，短時間 (100 ms 程度) の振幅の小さいデータフレームに反射音や残響音，雑音が重畳されると，埋め込み情報の検出に大きく影響を与えると考えられる．

これらをまとめると，一般的に数 100 ms 以下の短い時間波形を 1 つのデータフレームとして情報を埋め込む従来の音響情報秘匿技術は，そのフレームの信号強度が弱い場合に，先行する信号強度の強い部分の反射音や残響成分，時間変動する背景雑音などによる影響を受けやすく，スピーカから離れた位置での秘匿情報検出が十分でなくなる恐れがある．これは，もともと著作権管理を目的とする電子透かし技術が想定しているステゴ信号への変形は，MP3 などの知覚符号化圧縮や一定の付加雑音，時間伸縮やピッチ変換 [17] であり，空間伝搬に起因する残響や変動する雑音は想定されていないことも一因である．

空間伝搬用途に使用する情報秘匿手法もいくつか提案されているが，マイクロホンがスピーカに近接することを要求したり [11]，埋め込み情報量が少なかったり [60, 12]，埋め

込み情報量は多くても，空間伝搬に起因する上記のような変形に対する耐性が定量的に調べられていなかったりする [62]．また，多くの音響情報秘匿手法においてステゴ信号への変形に対する耐性や埋め込みデータ量は，音響信号の特徴に依存することが知られているが，音声から音楽までの様々な音響信号に対して耐性やデータ埋め込み量を定量的に検証した例はほとんど無い．

本章の目的は，二つある．第一に振幅変調に基づく音響信号への情報秘匿手法が，音声信号に対して空間伝搬条件においてどの程度の品質劣化と耐性を持つかを調べることである．具体的には，空間伝搬を前提とした，前述の 1 ~ 3 の条件を満たすことを，音声信号に対して確認する．なお前述 4 の振幅クリッピング歪に対する耐性条件に関しては，第 6 章において音楽信号を対象に検討する．情報秘匿に伴う音声品質の劣化に関しては，VCV 音節識別実験と，符号化音声信号の客観的音質評価手法のひとつである ITU-T P.862 PESQ (Perceptual Evaluation of Speech Quality) を用いる．ここで音声信号を取り上げた理由としては，駅や空港などの公共空間におけるアナウンス音声に対して，事前にあるいはリアルタイムに情報を埋め込み，利用者の手元の機器で検出して埋め込みデータを利用する用途を前提とするからである．これは，先に応用例として挙げた，難聴者へ健聴者と同じ音声信号から同等の情報提供を行うサービスおよび，国外からの旅行者等の音声アナウンス言語を理解しない人々へ，アナウンス内容と同等の情報を提供するサービスにおける利用を意図している．

第二に，音楽あるいは音声信号に情報を埋め込んだステゴ信号が，スピーカ再生されて空間伝搬した後，携帯電話のマイクロホンで収音され，音声通話による音声符号化を経て携帯電話ネットワークによって伝送され，受信先で復号化した音響信号より秘匿情報を検出する利用について検討を行なうことである．つまり，空間伝搬かつ音声符号化に対する耐性を調べることである．

利用者の端末へより多くの情報を呈示したい場合に，埋め込んだ情報のみを検出して呈示していたのでは，呈示できる情報量，つまり埋め込み容量が十分でない．よって，伝送したいデータをそのまま埋め込まずに，データに対するインデックス情報を埋め込み，伝送情報量を削減することが当面必要である．そのためには，伝送/表示したい情報を表示端末に事前に蓄積しておき，検出したインデックスに対応する情報を表示する方法と，埋め込まれた情報を表示端末では検出せずに，マイクロホンで受信した音響信号を音声通話にてサーバコンピュータに伝送し，サーバコンピュータ側で埋め込みインデックス情報の検出と，インデックスに対応する表示情報を端末へ電子メール等により伝送する方法が

考えられる (図 4.1 参照)．前者の利用形態での音響情報秘匿手法の有効性は，第一の目的において検証される．しかし後者のほうが，新たなデータ検出ソフトウェアを利用者端末にインストールする必要がない点で適している．この場合，マイクロホンで収音されたステゴ信号は，携帯電話の音声コーデックによって符号化の後，公衆電話ネットワークを経由した伝送先で復号化され，サーバコンピュータの検出プログラムに入力される．このため，ステゴ信号の携帯電話による音声符号化と復号化に対する耐性が必要となる．

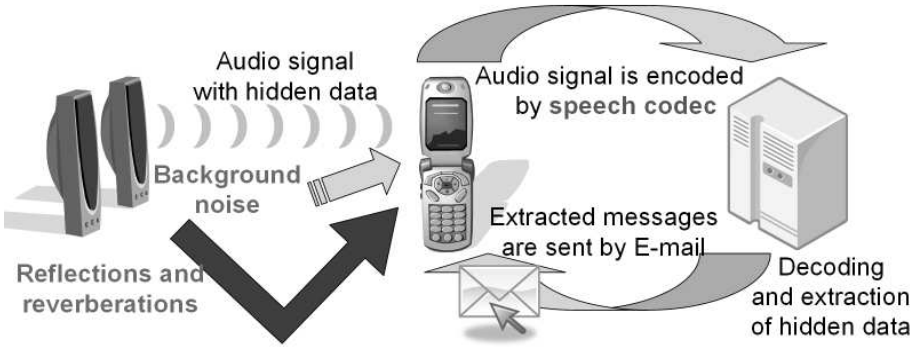


図 4.1: 携帯電話の音声通話ネットワークを通じた秘匿情報の復号化と利用者端末での情報呈示．

4.2 音声信号への情報秘匿と空間伝搬耐性

4.2.1 実験条件

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 に収録されているサンプリング周波数 16 kHz の音声を 22.05 kHz に変換して，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．埋め込み強度である振幅変調度は，リアルタイムに埋め込み処理を行うことを想定して，ホスト信号によらず一定の変調度とすることとし，0.3~ 0.7 の範囲で 0.1 ステップで設定した．埋め込む帯域は，6034Hz 以下の帯域とした．秘匿情報のビットレートは，埋め込みフレーム時間長を 3 秒としたとき 64 bps，4 秒としたとき 48 bps であり，ランダムなビット値を埋め込んだ．埋め込み時のパラメータ値は，表 4.1 に示した．

残響のある環境を想定して，ステゴ音声信号に，RWCP 実環境音声・音響データベース [63] 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を図 4.2 に示した．この波形の絶対

表 4.1: 埋め込み条件．

Parameters	Values	
bit rate	64 bps	48 bps
sampl. freq.	22050 Hz	←
freq. region	≤ 6034 Hz	←
bandwidth	21.5 Hz	←
subband pairs	140	←
subband groups	25	←
pairs per group	5 — 6	←
frame period	3 s	4 s
mod. freq. [Hz]	1, 1.67, 2.33, 3	1, 1.5, 2, 2.5

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，そのパ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振幅伝達特性を求めて，図 4.3 に示した．これにより，スピーカやマイクロホンのフラットでない伝送特性も模擬できることが分かる．なお，このインパルス応答の実際の残響時間を，シュレーダ積分法 [64] によって，–5dB から–25dB まで減衰する時間を 3 倍して求めたところ，約 1.1 秒であった．このインパルス応答の収録環境は，次に述べる環境騒音の収録環境とは異なるが，このインパルス応答は，他研究者が容易に入手可能かつ実環境で測定したものであるため用いた．

その後，背景雑音として 4 種類の環境騒音 (収録場所: 駅のホーム，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，秘匿情報を検出する処理を行った．4 つの環境騒音は，TARGET ENTERTAINMENT 製作，リッスンジャパン (http://listen.jp/store/) 販売の「効果音ライブラリ・環境音」から選び，44.1 kHz サンプリング/128 kbps の MP3 ファイルを WAV ファイルに変換して，冒頭の左チャンネルを 22.05 kHz にダウンサンプリングして用いた．5 種類の背景雑音は，オーバーオールの音声信号パワーに対して，信号対雑音比 (SNR) は 10 dB と 20 dB の 2 通りとした．全ての音響信号は，サンプリング周波数 22.05 kHz に変換後，処理を行った．5 つの背景雑音のそれぞれの平均スペクトルを，図 4.4 に示した．

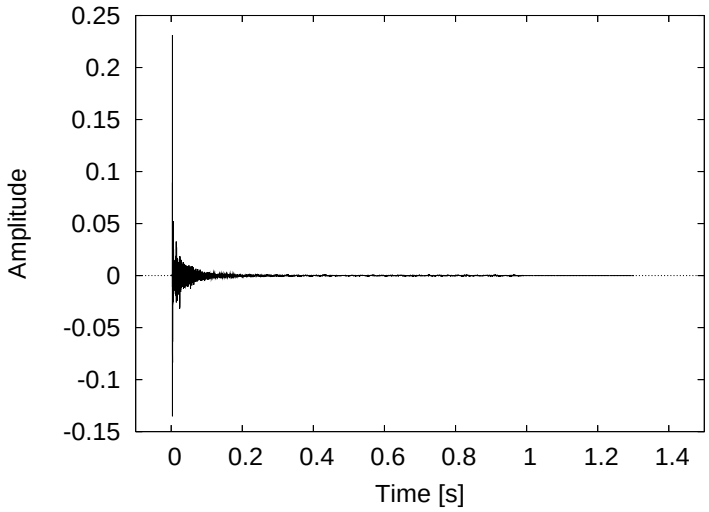


図 4.2: RWCP **実環境データベースより選んで耐性シミュレーション実験で用いた残響付加のためのインパルス応答** (ir130.dat) **波形．**

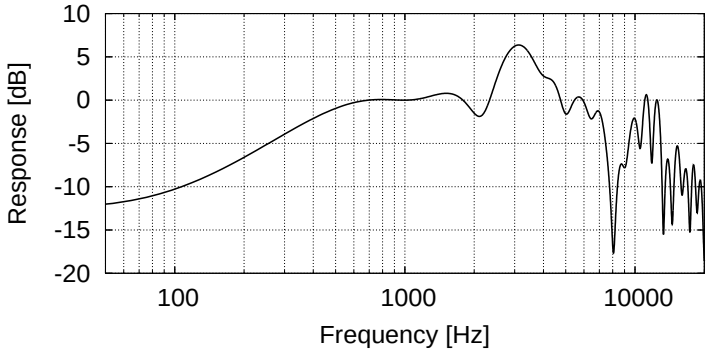


図 4.3: **インパルス応答における直接音成分のパワースペクトル**.

22 **種のステゴ音声信号と** 5 **種の背景雑音を組み合わせた** 110 **条件が，** 5 **段階の変調度，** 2 **段階のノイズ強度，** 2 **つの埋め込みビットレート条件の組み合わせに対してシミュレーションされた．**

4.2.2 結果

図 4.5 **は，**SNR **およびビットレート毎に，埋め込んだビット値と同じビット値が検出された率を，** 110 **埋め込み条件の中央値と，誤差棒により** 10~ 90 **パーセンタイル値で示した．この結果から，例えば振幅変調度** 0.4 **かつ** 48 bps **で埋め込みを行えば，**90%**の模擬条件において，**84%**以上のビット検出率が得られることが分かった．**

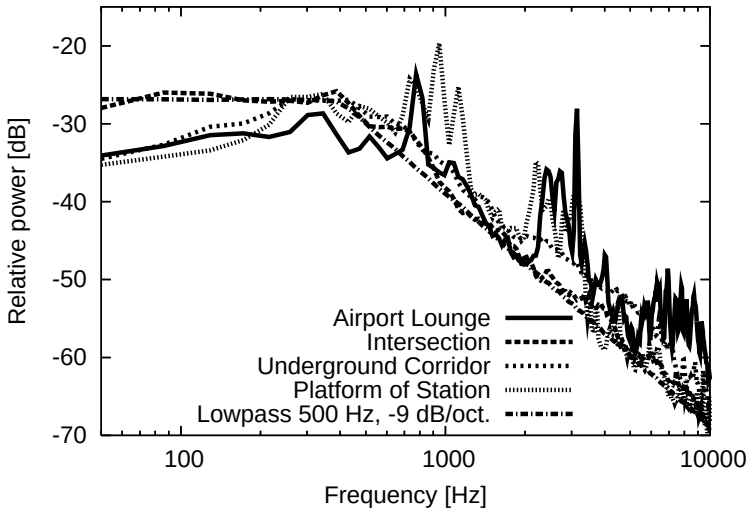


図 4.4: **ステゴ信号に付加した，**5 **つの付加雑音の平均パワースペクトル．**

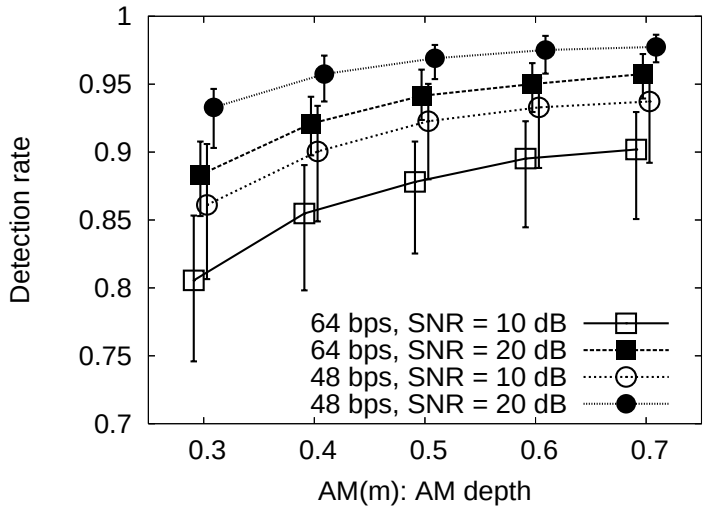


図 4.5: **ビット検出率の中央値．誤差棒は** 110 **条件中の** 10 **から** 90 **パーセンタイル検出率を示す**.

4.3 VCV 音節明瞭度試験

4.3.1 実験条件

音響電子透かしとしての利用においては，音質劣化の少ないことは重要であるが，第 4.1 節で述べたような応用場面にて音声信号に情報を埋め込む場合，ステゴ音声信号の音声了解度が十分であれば，多少の音質劣化は許容される．そこでまず，情報秘匿が音声了解度にどの程度影響を与えるのかを調べる基礎として，情報秘匿済み VCV 音節の明瞭度試験を行った．

VCV 音節は，a,i,u,e,o の先行 5 母音と 25 の日本語子音 (b, by, ch, d, g, gy, h, hy, j, k, ky, m, my, n, ny, p, py, r, ry, s, sy, t, w, y, z)，後続母音 a によって構成される 125 種とした．埋め込み条件は前節のシミュレーションと同じで，埋め込み強度は振幅変調度 0.4，0.6 および埋め込み無しとした．雑音付加時の明瞭度も調べるために，第 4.2 節で用いたローパスノイズを SNR 10 dB で付加する条件も加えた．被験者は聴力レベル 10 dB 以下の 5 名であり，防音室内でヘッドホン両耳聴 (片耳あたり実効音圧 72 dB) にて聴取した音声に対して，聞き取った音節をパソコンに入力し回答した．音節，SNR 条件，埋め込み強度条件の組み合わせをランダムに 1 巡するセットを 2 回異なる日に繰り返した．

4.3.2 結果

図 4.6 に，被験者間の平均正答率を示した．誤差棒は被験者間での最小と最大正答率を示している．この結果より，情報秘匿を行うと雑音環境下で音節明瞭度が低下することが分かる．しかし，振幅変調度 0.6 で埋め込みを行った場合でも，雑音環境下で最低でも 83%以上の正答率が得られている．文章了解度は音節明瞭度より一般的に高いゆえ，ステゴ音声信号に対しても充分な文章了解度が得られるだろうと考えられる．

異聴表を用いた分析では，どの条件でも先行五母音と第二母音には誤答がほとんど見られず，先行母音に依存した子音の異聴が目立った．これは情報秘匿のため与える振幅変調が，先行母音と子音との調音結合部に影響を与えるためと考えられる．

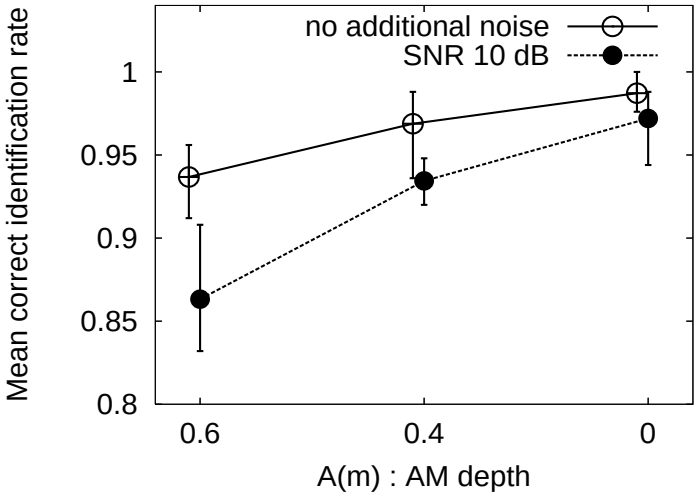


図 4.6: 125 VCV **音節に対する音節明瞭度**. **誤差棒は** 5 **名の被験者中の最大値と最小値を示す**.

4.4 携帯電話音声符号化への耐性

音楽あるいは音声信号に情報を埋め込んだステゴ信号が，スピーカ再生されて空間伝搬した後，携帯電話のマイクロホンで収音される．そして，音声通話による音声符号化を経て，受信先で復号化した音響信号より秘匿情報を検出する状況を前提に，携帯電話音声符号化に対する耐性を調べる．

4.4.1 携帯電話音声符号化方式

第三世代 3GPP(3rd Generation Partnership Project) **携帯電話においては，ディジタル音声信号符号化方式として，**CELP(Code-Excited Linear Prediction, **符号励振線形予測**) **系の最新のコーデックである** AMR(Advanced Multi-Rate) **方式が多く用いられている** [65]**．この音声符号化は，音声生成の要である声帯振動** (**励振源**) **と声道共鳴** (**フィルタ**) **を，それぞれ表現するパラメータ値として符号化することで情報圧縮を実現する．**AMR **方式では，あらかじめ単位振幅のパルスの取りうる位置と極性をお互いに少数に限定して決めておき，それら数本のベクトルの和で励振源を表現する．そして各パルスの位置の最適な組み合わせを歪の評価で選択する．伝送されるパラメータ値は，**LSP(Line Spectral Pair)**，ピッチ，コードベクトルとゲインである．**

復号時には，コードベクトルをそれぞれのゲインで調整した後加算されて生成した励振源を，線形予測フィルタに通して音声信号を合成する．よって，ベクトル符号帳の情報量

は少なく，効率のよい情報圧縮が可能である一方，入力波形に存在する微細な時間波形情報はこのような分析合成によって失われる．よって，エコー法やスペクトル拡散法によって情報を埋め込んだ音響信号に対して音声符号化を行った場合，秘匿情報の検出は困難となる．

また，AMR 方式の特徴としては，8000 Hz サンプリングかつ 13 bit 直線量子化 (8bit A-law あるいは μ-law 圧縮) された音声波形に対して，短い時間フレーム (160 サンプル，0.02 秒に相当) 毎に 4.75~ 12.2 kbps の広い範囲でビットレートを可変して伝送ができる点である．さらに，有音無音検出機能，背景雑音生成機能，フレームデータ誤り隠ぺい機能などがあるが，これらの機能はここでは扱わない．

4.4.2 実験方法

音声及び音楽信号の 4 kHz 以下の帯域にデータを埋め込み，残響や背景雑音を付加した後，AMR 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 8 bps とし，パラメータ値は，表 4.2 に従い，ランダムなビット値を埋め込んだ．埋め込み強度である振幅変調度は，0.4 で固定としたが，これは第 4.3 節における VCV 音節明瞭度実験および第 4.5 節での客観的音質劣化度合の評価を元に決定した．また実環境でも，スピーカ再生と携帯電話の AMR 符号化方式による音声録音機能を用いて，ステゴ音声信号の空間伝搬と AMR 符号化が重畳する条件での検出率を調べた．なお，同実環境においてコンデンサマイクロホンを用いた PCM 録音も同時に行うことで，携帯電話受信と音声符号化の影響を調べた．

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16 kHz であったが，8 kHz に変換して用いた．

音楽信号としては，RWC 研究用音楽ジャンルデータベース RWC-MDB-G-2001 [41] に収録された様々な音楽ジャンルの 100 曲の左チャンネル冒頭60 秒を用いた．これらの信号はデータ埋め込み時にはサンプリング周波数44.1 kHz であり，表 4.2 に示したパラメータによりランダムビットデータの埋め込んだ後，サンプリング周波数 8 kHz に変換して用いた．

半分のシミュレーション条件では，残響のある室内において，データ埋め込み済み信号

表 4.2: 埋め込みパラメータ値．

Parameters	Values
AM depth	0.4
bit rate	8 bps
embedding region	below 4000 Hz
bandwidth	31.3 Hz
subband pairs	63
subband groups	7
pairs per group	9
frame period	6 s
mod. freq. [Hz]	1.17, 1.67, 2.17, 2.67

がスピーカ再生され，マイクロホン受音されることを模擬するために，第 4.2 節で用いたものと同じ，RWCP 実環境音声・音響データベース [63] に収録されている，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，サンプリング周波数 8 kHz に変換してから畳み込んだ．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 [66] 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 [67] 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~ 12.2 kbps 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，6.7 kbps，12.2 kbps の 3 種類とした．

実環境における空間伝搬と AMR 符号化の影響を調べる実験では，容積約 410m^3，一辺 12 m の変形正方形教室において，対角線前方中央壁面より 1 m，高さ 1 m の位置に置いた 12 cm フルレンジスピーカ (Panasonic WS-X66) より，情報埋め込み済み音声信号と，オーバーオールの SNR を一定とした Hoth ノイズを混合して再生した．スピーカから 5.4 m の距離，高さ 1.3 m の位置に携帯電話 (Panasonic Mobile Communication 820P) 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 12.2 kbps) にて録音した．携帯電話機真横においた騒音計で測定した等価騒音レベルは，音声が 65 dB と 55 dB の 2 条件，Hoth ノイズは 45 dB で一定，室内の暗騒音は 31 dB であった．録音時に

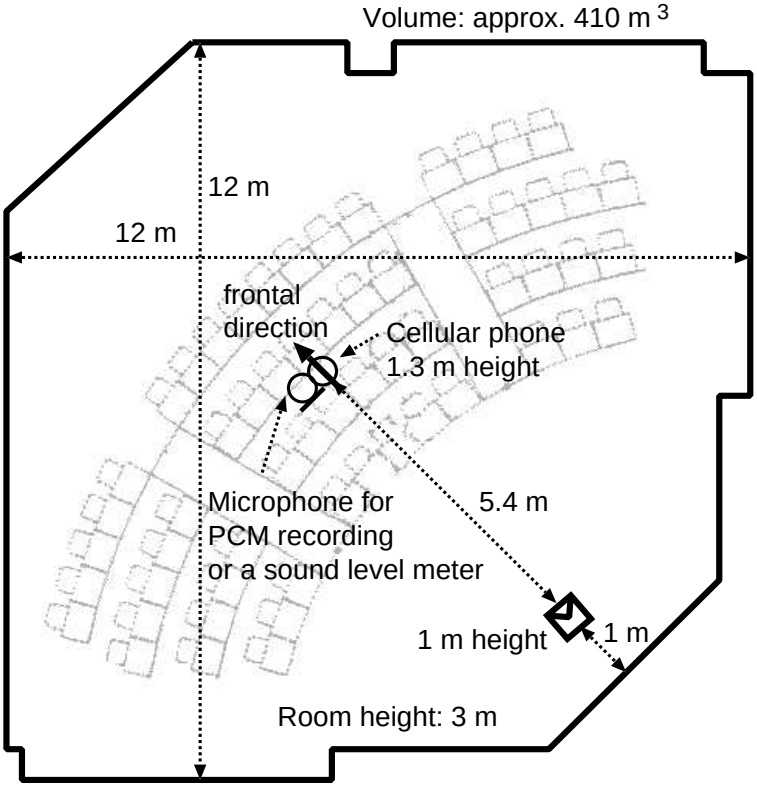


図 4.7: **実験に用いた部屋の見取図．**

は，上述の携帯電話に加えてその真横に無指向性コンデンサマイクロホン (audio-technica ATS410) **を置き，**USB **オーディオユニット** (EDIROL UA-5) **を用いて** 48 kHz, 16-bit **の直線量子化にて録音を行った．**

Log-TSP **信号をスピーカ再生して，コンデンサマイクロホンによって録音したインパルス応答からシュレーダ法** [64] **により計算した残響時間は** 1.0 **秒であった．これら部屋と使用機器の配置図は，図** 4.7 **に示した．**

4.4.3 実験結果

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 4.8 **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，**22 **条件中の** 10 **から** 90 **パーセンタイルの範囲を示している．図** 4.9 **には，**100 **曲の音楽信号に対する平均検出率を示した．エラーバーは，**100 **条件中の** 10 **から** 90 **パーセンタイルの範囲を示している．**

音声信号に対する結果の図 4.8 **より，残響が無い場合は，**SNR 20 dB **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 **符号化では検出率が** 80% **を下回ることが多く，事実**

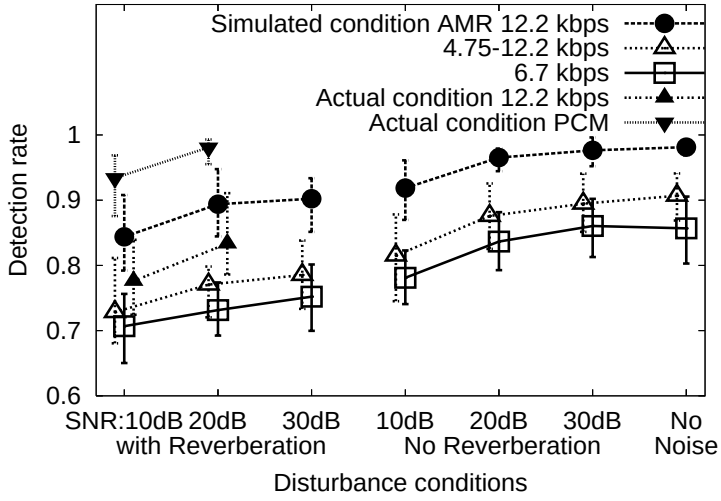


図 4.8: 22 名の音声信号に埋め込まれたデータの平均ビット検出率．エラーバーは 10 から 90 パーセンタイルの範囲を示す．

上利用できないに等しい．しかし，12.2 kbps の AMR 符号化であれば，90%以上の音声信号で 85%以上の検出率が得られることが分かった．また図 4.9 より，音楽信号の方が，検出率が 90% を下回る条件での検出率が，音声信号の場合より 5 ポイント程度高いことも確認できた．これは，音声信号の成分が時間周波数的に粗い分布をしているのに対し，音楽信号は密に分布しているため，埋め込みの効率が高く変形に対して頑強になるのが理由である．また，現実の室内環境では，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．

実環境において携帯電話にて録音した条件では，AMR ビットレートは 12.2 kbps であったが，シミュレーションでの同じビットレートよりも検出率が 0.06 程度低下した．これは音源に向かって携帯電話の画面を見ながらかざすという現実の使用環境を模して，携帯電話の画面と受話マイクロホンがスピーカと反対側を向くように配置したため，スピーカからの直接音が収録されにくく，周波数特性の劣化が大きかったためではないかと考えられる．無指向性マイクロホンにより PCM 録音した音から復号化した場合は，SNR 10 dB においても，90%以上の音声で，87%以上の検出率が得られたこと，またシミュレーション条件において AMR ビットレートによって検出率が大きく下がったことから，残響と背景雑音のある環境において音声符号化耐性を求めることは，かなり困難な要求であることが分かった．

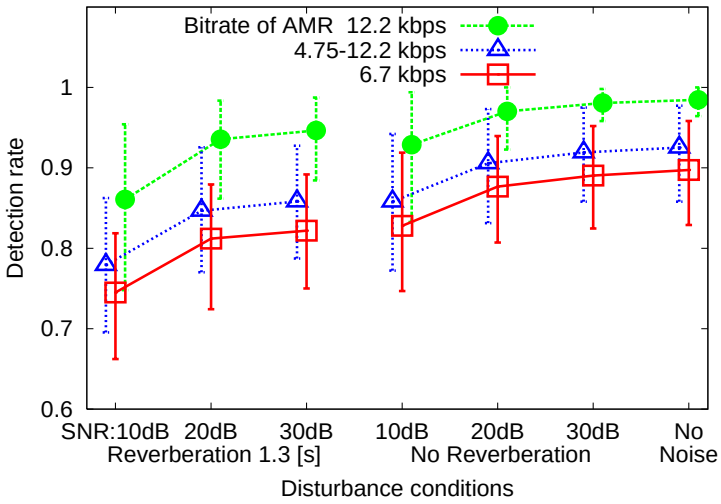


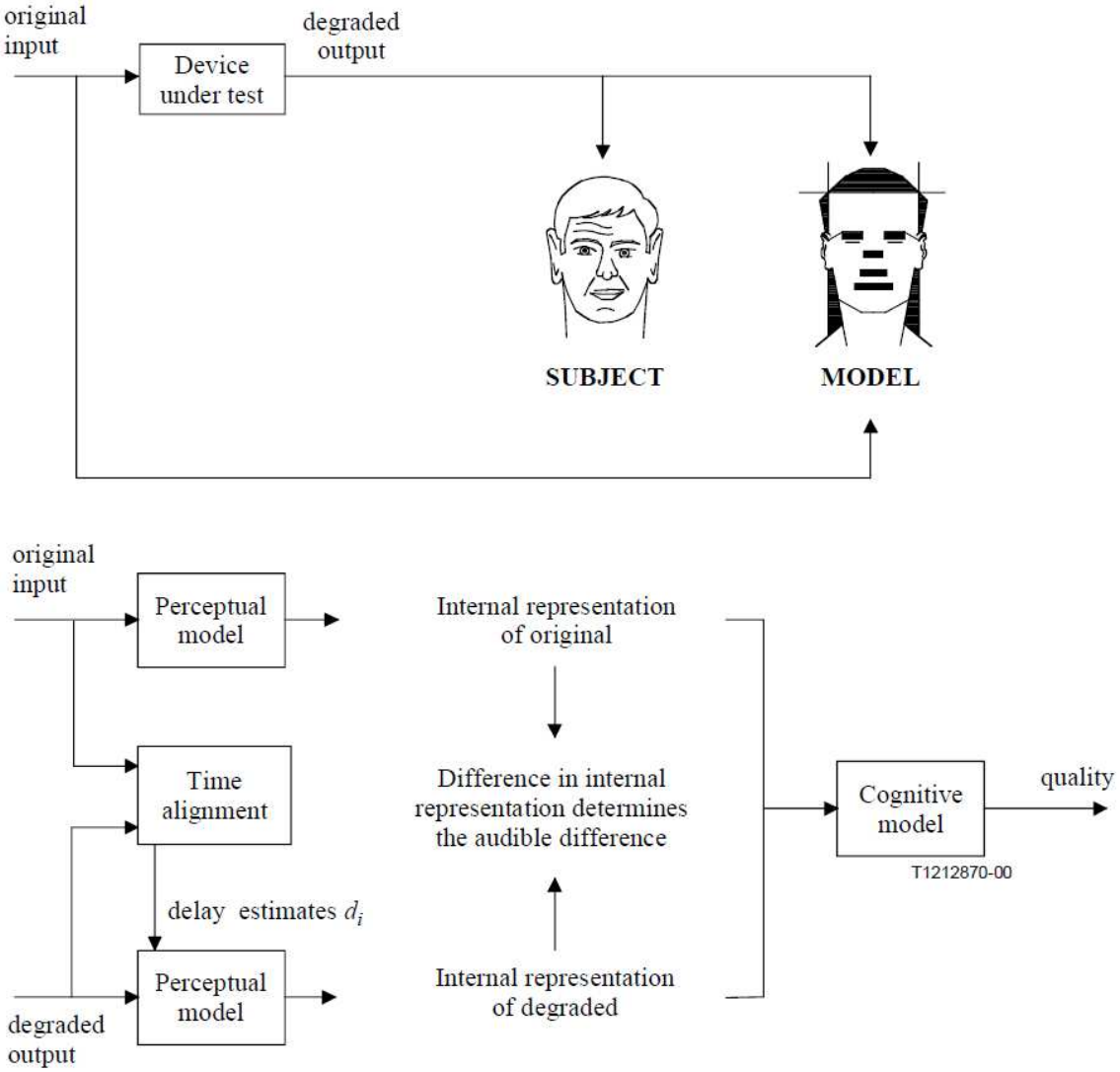
図 4.9: RWC-MDB-G-2001 100 **曲に埋め込まれたデータの平均ビット検出率．エラーバーはそれぞれ** 100 **曲に対する検出率の** 10 **から** 90 **パーセンタイルの範囲を示す．**

4.5 客観的品質劣化度合評価

4.5.1 PESQ による広帯域音声品質劣化度合評価

ITU-T P.862 PESQ **は，電話帯域音声や音声コーデックの品質劣化度合を測定するためのアルゴリズムである** [68]．PESQ **は原信号と音声コーデックを経た後の信号を比較し，心理音響特性に基づいた信号の内的表現の差分を，音質劣化度合として報告する．図** 4.10 **に，**P.862 **規格文書より，**Figure 1 **を抜粋して** PESQ **の概略図について示した．**PESQ **の結果は，**MOS-LQO (Mean Opinion Score, Listening Quality Objective) **とよばれ，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) **にほぼ対応する．**MOS-LQO **の値は，**1.02 **から** 4.56 **までが得られ，それぞれの値は，**1: Bad (**悪い**)**，**2: Poor(劣っている)，**3: Fair(まあよい)，4: Good(よい) **という評価に対応する．**

ここでは第 4.2 **節にて行った，広帯域音声信号への情報秘匿に起因する音声品質の劣化を評価する．音声信号への情報秘匿後の音質は，音声符号化を経た音質と似ているので，その劣化度合を，サンプリング周波数** 16kHz **に拡張された** Wideband PESQ **を勧告している** ITU-T P.862.2 **に基づいた** ITU-T **提供ソフトウェアにより測定した．**Wideband PESQ **への入力レベルは，**16 bit **量子化における最大振幅の純音を** 0 dB **とした実効レベルを表す** dBov **を用いて，**–26 dBov **とした．日本音響学会研究用連続音声データベース** Vol. 1 **より，**22 **名の話者による各** 50 **合計** 1100 **の音素バランス文を** 2 **つづつ繋げて作成した** 550 **文の** 8 **秒前後の音声信号を対象とした．データ埋め込みパラメータは，サンプ**



NOTE – A computer model of the subject, consisting of a perceptual and a cognitive model, is used to compare the output of the device under test with the input, using alignment information as derived from the time signals in the time alignment module.

図 4.10: PESQ **の概要図．**ITU-T P.862 Figure 1 **より抜粋．**

リング周波数が 16 kHz である以外は第 4.2 節のシミュレーションと同じとし，48 bps でデータを埋め込んだ．

結果は，図 4.11 に，CELP 系の音声符号化方式である AMR 符号化方式を広帯域 (16 kHz サンプリング) に拡張した符号化方式である，AMR-WB 方式において符号化した音声の MOS-LQO 値も併せて示した．この図から，振幅変調度 0.4 にてデータを埋め込んだ音声信号の品質劣化は，低ビットレート広帯域 CELP 系符号化音声より，やや音質が悪く，” 劣っている” 程度であろうと予測される．一方，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第 4.3 節では，VCV 音節の明瞭度試験を行なっており，この場合 SNR 10 dB の条件でも平均で 93% の正答率となった．これらより，データ埋め込み済み音声信号の品質劣化は明らかなものの，音声情報を伝えるには十分であろうと考えられる．

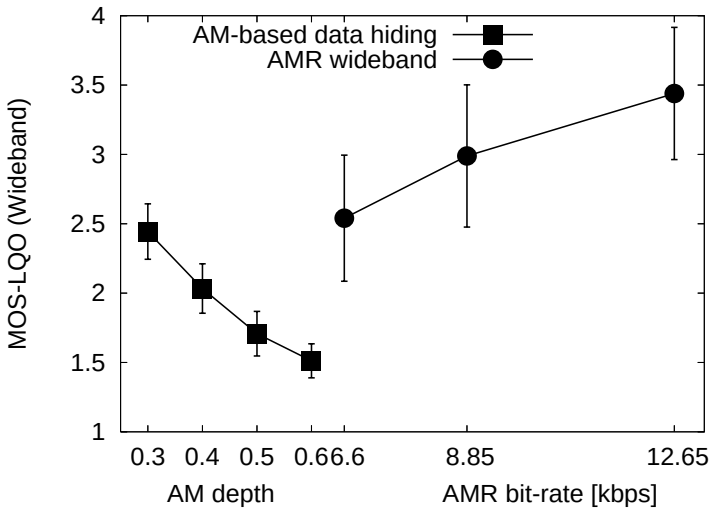


図 4.11: Wideband PESQ による広帯域音声客観的劣化評価値 (MOS-LQO). 誤差棒は 550 条件の ± 1 標準偏差を示す. 左側は振幅変調に基づく情報秘匿に起因する劣化を変調度毎に，右側は AMR-WB 音声符号化による劣化を AMR ビットレート毎に表している．

4.5.2 PESQ による狭帯域音声品質劣化度合評価

第 4.4 節の耐性シミュレーション条件下での狭帯域 (8 kHz サンプリング) 音声品質の劣化度合を，ITU-T より提供されるソースコードをコンパイルした PESQ ソフトウェアを用いて測定した．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 1100 の音素バランス文を 2 つづつ繋げて 550 文の 8 秒前後の音声信号とし，16 kHz

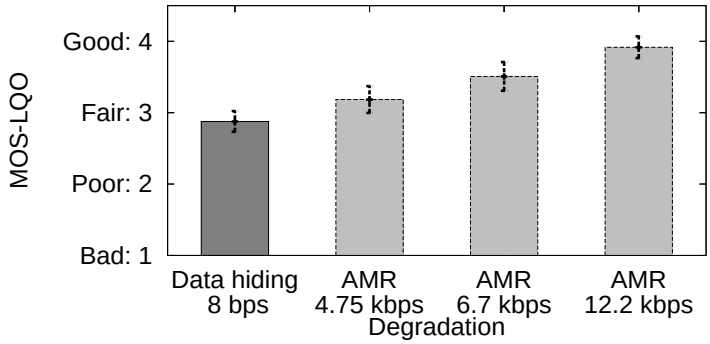


図 4.12: **狭帯域音声信号に対する** AMR **音声符号化と情報秘匿に伴う音質劣化の比較．誤差棒は** ± 1 **標準偏差とした．**

サンプリングのファイルを 8 kHz **にダウンサンプリングした後，用いた．音声信号への情報秘匿条件は，表** 4.2 **に従い，ランダムビット値を埋め込んだ．**AMR **符号化器への音声信号の入力レベルは** −26 dBov **とした．**

結果は，図 4.12 **に示した．音質劣化の比較参考のため，**AMR **狭帯域音声コーデックによる** 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，**MOS-LQO **を測定してその平均と，誤差棒にて標準偏差を示した．この結果から，データを埋め込んだ狭帯域音声信号の品質劣化は，携帯電話よりやや音質が悪い程度であろうと予測される．**

4.5.3 PEAQ による音楽音質劣化度合評価

ここでは，Kabal[14] **による** PEAQ **の基本バージョンの実装を用いて，第** 4.4 **節で用いた情報秘匿済み音楽の音質劣化度合を測定した．音楽データは** RWC-MDB-G-2001 **の** 100 **曲左チャンネル冒頭** 1 **分間とした．使用環境が，環境騒音下のスピーカ再生を前提としており，かつ理想的なステレオ聴取環境を前提としないため，左チャンネルのみのモノラル信号を評価に用いた．**44.1 kHz **サンプリングの波形データに対し，**4 kHz **以下に対して表** 4.2 **の条件にてランダムビット値の埋め込みを行った．**

図 4.13 **には，情報秘匿音楽の劣化度合と，比較対象として，**MP3 **の** 48 kbps/ch (96 kbps)**，**64 kbps/ch (128 kbps) **で符号化し復号化した音楽信号についての，音質劣化度合の平均値と** $\pm$ 1 **標準偏差の値をプロットした．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「劣化がやや気になる」よりやや悪い程度であることが分かった．また，**MP3 **によって符号化した音楽信号の劣化度合と比較すれば，**48 kbps/ch (96 kbps) **程**

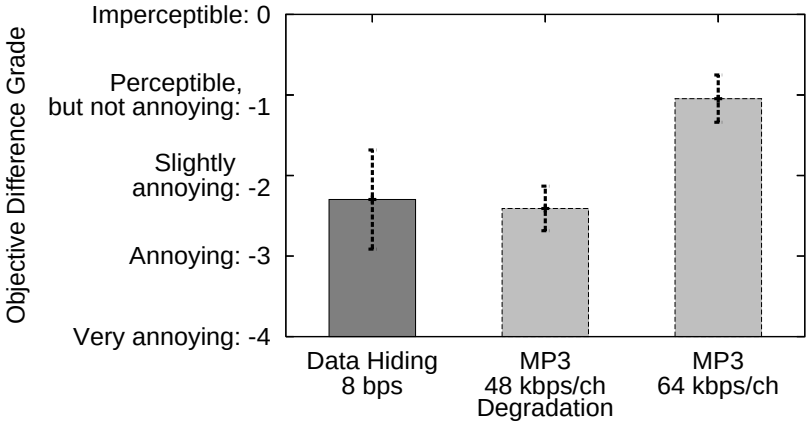


図 4.13: PEAQ **による音楽信号の客観的音質劣化度合．誤差棒は** ± 1 **標準偏差．**

度であることが分かった．情報秘匿済み音楽はスピーカ再生され，そこに背景雑音が加わることが必至である使用条件を前提とすると，データ埋め込みに伴う音質劣化は問題の無い程度であろうと考えられる．

4.6 考察

4.6.1 実時間処理埋め込みおよび検出処理

現在の情報埋め込みプログラムは，逐次フレーム処理を採用しているが，Octave **プログラミング環境で作成しているため，ハードウェアの** AD/DA **デバイスを直接制御し，入力音響信号に対して実時間処理によりステゴ音響信号を出力することはできない．そのような実時間処理ソフトウェアを実装することにより，ライブコンサート** PA **やアナウンス音声，**BGM **への情報秘匿などに対して，幅広く技術の実施が可能となる．この点は，今後の課題である．**

この場合，埋め込み強度は振幅変調度を固定して設定することとなり，理論的な最低遅延時間は，オーディオデバイスのバッファリング時間を除けば，フィルタバンク処理による遅延時間となる．表 4.1 **の埋め込み条件では，**2048 **点** FFT **を用いた** FIR **フィルタにより実装しており，この場合の遅延は約** 100 ms **となる．表** 4.2 **の場合は，**1024 **点** FFT **を用いた** FIR **フィルタによって実装しているため，遅延は** 128 ms **となる．ライブコンサート** PA **の場合には，より低遅延が求められるため，フィルタバンク処理を改善する必要がある．一方，それ以外の場合には，遅延量は問題にならず，処理負荷はクロック** 1GHz **程度のパーソナルコンピュータであれば充分であるので，実時間で情報秘匿が可能となる．**

また，現在の秘匿情報の検出ソフトウェアでは，第 6.2.2 節において検討するように，FFT を多用している．よって，現在の検出アルゴリズムは，一般的な PDA やスマートフォン (例えば PXA270 Processor 520MHz) の処理能力の約 2 倍程度の演算能力を必要としている．多くの人々が携帯する機器への検出ソフトウェアの実装は，技術を実証し，その改善および普及のために必要と考えており，今後の課題である．

フィルタバンク処理を低演算量化するにあたっては，階層型 CIC(Cascaded Integrated Comb) フィルタ [69] を用いるのが有効であろう．また，検出時のフレーム同期のため演算を，フレーム周期を予測してその近傍の時刻 u のみ (3.13) 式の演算を行うこと，および埋め込み時の変調周期をオーバラップ FFT 周期の整数倍とすることで式 (3.14) の変調周波数における強度算出を，変調周期毎の波形同期加算によって実施し FFT 処理を無くすなどの改良が考えられる．

4.6.2 携帯電話ネットワークにおけるパケットロスの影響

AMR コーデックは，伝送経路におけるパケットロスを隠蔽するような機構を必須要件として含んでいる．この主な仕組みは，伝送エラーが起きたフレームの前後のフレームのパラメータ値から補間を行なって，エラーフレームのパラメータを推定し，復号化するものである．

今回はこのパケットロス隠蔽のアルゴリズムは用いなかったが，AMR 符号化後のデータに対して，1 フレーム (20 ms) 単位でのフレームデータ抜き取り，ゼロデータフレームとの置換，ゼロデータフレームの挿入の 3 種類のパケットロスを等確率でランダムな時刻に起こすような変形を加えた後，復号化を行うシミュレーション実験も実施した．

その結果，3%程度のパケットロスでは，いずれの条件でも 1~ 2%程度の検出率の低下しか見られず，単純なパケットロスに対しては，ある程度の耐性を持つことが分かった．AMR コーデックによるパケットロスの隠蔽が行なわれた場合は，さらに検出率の低下は起こりにくくなるものと考えられる．

4.6.3 実効データ伝送量

第 4.4 節のシミュレーションでは，スピーカ伝送周波数特性，残響，背景雑音，音声コーデックの 4 つの妨害要因に対しての耐性を持たせるため，データ埋め込み量を 8 bps と少なくした．ここではエラー訂正符号を埋め込み時に用いることは無かったが，このよう

な空間伝搬情報秘匿技術を実用化する際には，なんらかのエラー訂正符号を用いる必要がある．

ここで，8 秒分のデータ (64bit) に対して，BCH(63,36,5) 符号化と軟判定復号法 [45] を併用したとすると，エラー訂正限界は 9bit 程度となる．この場合，86%のビット検出率が得られれば，36bit 分の情報伝送が可能となり，実効データ伝送量は，4.5 bps となる．図 4.8 に示した音声信号へ残響を重畳したシミュレーションにおいて，約 90%以上の信号条件で，86%のビット検出率を満たすのは，SNR 20 dB 以上での 12.2 kbps のみであった．残響が無い場合は，6.7 kbps の AMR コーデックによって，約 90%以上の信号条件で 80%のビット検出率を満たすことができたが，この場合はさらにエラー訂正限界を向上させる必要があり，実効データ伝送量は，上記の半分の 2~ 3 bps が妥当な線であろう．

データ埋め込み強度である振幅変調度を，今回用いた 0.4 から 0.6 に上げることによって，検出率は軒並 5~ 6 ポイント程度向上するが，狭帯域音声の場合，平均 MOS-LQO は 2.88 から 2.15 まで低下する．一方，振幅変調度 0.6 で埋め込みを行なっても，音節明瞭度としては，SNR 10 dB の条件において平均で 86%程度を得ており，文章了解度にはほとんど問題ないと思われる．よって，より困難な使用環境において，さらなる音質の劣化が許容できる場合には，振幅変調度を上げて頑強度を高めることも考えられる．

4.7 あとがき

データ埋め込み済み音響信号をスピーカから再生し，マイクロホンで受音してデータを検出し利用するという，空間伝搬利用を前提に，振幅変調に基づく音響情報秘匿技術の性能を検証した．

最初に，スピーカ拡声されるアナウンス音声にデータを埋め込み，利用者の手元の機器で復号化と表示を行う利用を前提とした．男女合計 22 名の広帯域日本語音声信号に対して 48 あるいは 64 bps にてデータを埋め込み，残響および背景雑音下での検出率をシミュレーション実験により調べた．その結果，48 bps のデータを振幅変調度 0.4 で埋め込むと，90%の条件で 84% 以上の検出率が得られることが分かった．また，情報秘匿に伴う音声品質劣化は，広帯域 PESQ による客観評価の結果，振幅変調度 0.4 で埋め込みを行うと平均的に “ Poor (劣っている)” 程度に劣化することが分かったが，VCV 音節識別実験からは，SNR 10 dB の環境でも 95 % 以上の明瞭度が得られることが分かり，発話内容を伝達するには問題ないことも明らかになった．

次に，スピーカ再生される音声や音楽にデータを埋め込んで，携帯電話の音声通話先にあるサーバコンピュータでデータを復号化し，利用者の携帯電話に情報を伝送する利用を前提とした．様々な狭帯域音声信号と広帯域音楽信号に 8 bps にて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 狭帯域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，音声信号に対しては，背景雑音のみが重畳される場合は，6.7 kbps 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．広帯域音楽信号に埋め込んだ場合には，音声信号より 5 ポイント程度検出率が高いことが分かった．これらの結果から，残響や背景雑音が存在しても，AMR 狭帯域音声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振幅変調に基づく情報秘匿による携帯電話ネットワークを通じた情報伝送が可能であることが分かった．

また，この利用形態におけるデータ埋め込みに伴う客観的音質劣化度合を，サンプリング周波数 8kHz の電話帯域音声信号に対しては PESQ を用いて，広帯域音楽信号に対しては PEAQ を用いて調べた．その結果，音声信号は，「まあよい (fair)」よりやや悪く，音楽信号は「劣化がわずかに気になる (slightly annoying)」よりやや悪い程度であった．