

NAIST-IS-DD0561206

博士論文

情報をさりげなく重畳した音メディアの設計
とその応用に関する研究

中山 彰

2007年 1月 31日

奈良先端科学技術大学院大学
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学) 授与の要件として提出した博士論文である。

中山 彰

審査委員：

鹿野 清宏 教授 (主指導教員)

小笠原 司 教授 (副指導教員)

猿渡 洋 助教授 (副指導教員)

情報をさりげなく重畳した音メディアの設計 とその応用に関する研究*

中山 彰

内容梗概

本論文では、音楽や音声など可聴音に対し、伝達すべき情報をさりげなく重畳した「音」メディアについて論じた。人間にとっては普通の音に聞こえると同時に、機械にとっては内部に埋め込まれた情報（信号）が受信することが可能となるような「音」メディアの設計法、およびその応用法について提案を行った。特に、そのような「音」メディアの著作権管理への分野での応用を目指し、MPEG 心理音響モデルに基づいたオーディオ信号への音声電子透かし、メディア複合化分野へのそれとして、ロボット制御情報の生成・伝送・再生への応用について述べた。

最初に、本論文では、MPEG 心理音響モデルを利用したオーディオ信号への電子透かし方法について述べた。電子透かしは、聴覚的に聴こえないということが重要である。それを考慮した透かしアルゴリズムとして、Boney らの提案する MPEG 心理音響モデルを用いた電子透かし方法がある。ただしこの方法は、MPEG 符号化、DA/AD 変換に対して脆弱であり、また透かし検出に透かしを入れる前の信号を必要とするなどのいくつかの改良すべき点がある。そこで、まず MPEG 符号化に対して電子透かしの耐性を高めるために、透かし信号を埋め込む帯域および埋め込み方法について改良を行ない、さらに心理音響実験の知見を用いて継時マスクングを導入した。これらの改良により、MPEG 符号化に対してロバストな透かしが実現できた。次に DA/AD 変換に対する耐性を高めるため、白

* 奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士論文, NAIST-IS-DD0561206, 2007年1月31日.

色化相互関連法を用い，透かし信号の同期検出を実現した．これにより DA/AD 変換に耐性をもつ透かしが実現できた．また受けとった信号のみから透かしを検出する方法についても考察し，有効な検出方法を提案した．さらに，主観評価実験の結果を示し，透かしを埋め込んだことによる品質の劣化は少ないことを示した．

次に，本論文では，ロボットや物理的なエンドエフェクタなどの動きを含む物理的なメディアをモーションメディアと定義し，ロボティックコミュニケーションサービス実現のために重要な課題となるロボット動作信号（モーションメディアコンテンツ）の効率的な流通の方式として「音」メディアを使った方式を論じた．流通のためのオーバーヘッドを最小化し，モーションメディアコミュニケーションサービスの普及促進を目的とし，音信号の中にロボット制御信号を変調して重畳するという「音によるモーションメディアコンテンツ流通プラットフォーム」の概念を提案した．この概念では音信号はロボット制御信号のキャリアとして解釈される．本概念の有効性を検証するために，高周波 FM 変調と DTMF（Dual Tone Multi Frequency）信号を変調手段としたモーションメディアコンテンツ流通プラットフォームを実装し，プロトタイプコミュニケーションロボットシステムを実現した．またそれらのロボットをウェブベースサービス，電子メール，携帯電話の着信メロディなどに適用し「音」メディアによるモーションメディアコンテンツ流通の有用性を示した．

キーワード

音，重畳，電子透かし，心理音響モデル，ロボット，メディア

Design and applications of compound audio media that embed information imperceptibly*

Akira Nakayama

Abstract

In this paper, a new medium for audio communication for humans and machines is discussed. This technique is used to embed a machine control signal into an audio stream; the signal is imperceptible to humans, but can be detected and decoded by other machines. The design of compound audio signals and a new industrial application are discussed. Several interesting applications for digital watermarking and robotic control-signal transmission are described.

First, digital watermarks for audio signals based on the psychoacoustic masking model are discussed. Currently, in digital watermarking algorithms that are being studied, it is necessary and inevitable for the watermarks to be inaudible. In this regard, the Boney algorithm, which utilizes the masking characteristics, is very promising. This algorithm, however, takes only simultaneous masking into account and is not robust to MPEG encoding/decoding. Moreover, the Boney algorithm requires the original signals to detect the embedded watermarks. In this paper, the author introduces temporal masking as well as simultaneous masking and improves the Boney watermarking algorithm with regard to the approximation of the masking characteristics. The author also introduces a self-clocking mechanism to automatically detect the watermarks and proposes detection algorithms that do not require the original signals for the detection of the watermarks.

* Doctoral Dissertation, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0561206, January 31, 2007.

The Boney and our watermarking algorithms are comparatively tested by using a watermark detection algorithm and are evaluated by subjective tests using several kinds of music sources. Both evaluation results show that our watermarking algorithm has the following advantages: (1) audio signals with embedded watermarks have no degradation and the watermarks are inaudible and (2) our watermarking algorithm is robust to additive noise attack, MPEG encoding/decoding, and DA/AD conversion.

Second, MotionMedia, a physical medium that includes the motion of active entities such as robots or physical end-effectors is defined, and an efficient method of sharing the robot-motion control data (called MotionMedia content) via audio is discussed for the realization of robotic communication services. To foster the sharing of the MotionMedia content and minimize the overhead involved in the sharing, the reuse of conventional “audio” media for sharing MotionMedia is an important concept. Therefore, a conceptual idea for MotionMedia content sharing via audio is proposed. This technique embeds robot control signals in an audio stream using a modulation technique. This concept can be explained as the usage of an audio medium as a carrier to convey control data. To test this concept, the author has developed two implementations which use the frequency modulation technique and the dual-tone multi-frequency technique for modulation and also prototype communication robot systems. These robots have been applied to conventional communication media such as the World Wide Web, E-mail, and cell phones. The author has shown that the MotionMedia sharing platform via audio can be easily realized and is a promising concept.

Keywords:

Sound, Multiplexing, Digital watermarking, Psychoacoustic model, Robot, Media

目次

第1章	はじめに	1
1.1	背景	1
1.2	機械・人間向け「音」メディアの重要性	2
1.3	さりげなく情報を重畳する「音」メディアの応用例	3
1.4	「音」メディアの設計のための課題	4
1.5	本研究の目的と研究領域別の課題と方針	6
1.5.1	音声電子透かし研究の現状とその研究課題	7
1.5.2	音によるメディア複合の研究とその課題	9
1.5.3	音による秘匿通信(ステガノグラフィー)の課題と現状, 本研究テーマとの関連	13
1.6	本論文の構成	15
第2章	心理音響モデルに基づく音声電子透かし	17
2.1	はじめに	17
2.2	心理音響モデルを用いた方法(Boneyらの方法)の追試	17
2.2.1	心理音響モデルを用いた埋め込みアルゴリズム(Boneyらの方法)	17
2.2.2	Boneyらの検出アルゴリズム	18
2.2.3	透かし検出実験	21
2.2.4	音質評価実験	23
2.2.5	本節のまとめ	24
2.3	埋め込みアルゴリズムの改良	25
2.3.1	はじめに	25
2.3.2	透かしの埋め込む周波数帯域の制限	25

2.3.3	マスキングを近似するフィルタの実現方法の改善	25
2.3.4	継時マスキングの導入	27
2.3.5	白色化相互相関法による同期検出	28
2.4	従来法からの検出率改善の評価	28
2.5	攻撃への耐性評価	29
2.5.1	雑音攻撃への耐性	29
2.5.2	MPEG-1 Audio Layer III に対する耐性評価 (2つのPN系列を埋め込んだ時)	29
2.5.3	結託した攻撃への耐性評価 (多くの数のPN系列を埋め込んだ場合の評価)	32
2.5.4	DA/AD変換での攻撃に対する耐性評価	32
2.6	冗長性導入による効果	34
2.7	音質の主観評価実験	35
2.7.1	主観評価方法	35
2.8	耐性評価および音質評価のまとめ	37
2.9	原信号を必要としない検出方法	38
2.9.1	はじめに	38
2.9.2	原音を必要としない透かし検出アルゴリズム	38
2.9.3	検出アルゴリズムの有効性の検討	40
2.9.4	本節のまとめ	43
2.10	本章のまとめ	43
第3章	「音」メディアによるモーショメディアコンテンツ流通プラットフォームへの適用	45
3.1	はじめに	45
3.2	モーショメディアコンテンツ流通プラットフォームとそのプラットフォームに望まれる要件	46
3.3	「音」メディアを用いたモーショメディアコンテンツ流通の提案とその特長	48

3.4	「音」メディアを用いたモーショメディアコンテンツ流通の要件 と実現方式	51
3.4.1	CyberPerformer/Hyper	53
3.4.2	CyberPerformer/Audio	58
3.4.3	各プラットフォームの比較	65
3.4.4	簡易なモーショメディアコンテンツの生成と音への変調・ 重畳のための GUI 編集システム	66
3.4.5	本節のまとめ	66
3.5	応用例	67
3.5.1	CyberPerformer/Hyper の応用例	68
3.5.2	CyberPerformer/Audio の応用例	68
3.6	本章のまとめ	72
第 4 章	「音」メディアを用いた多自由度モータコントローラへの展開	74
4.1	はじめに	74
4.2	USB オーディオインタフェースを利用した多自由度ロボットコン トローラの開発	75
4.2.1	USB オーディオインタフェース使用のメリット	75
4.2.2	実現方式	76
4.2.3	評価	77
4.3	CyberPerformer/Hyper の発展形としての多自由度ロボットコント ローラ	79
4.3.1	概要	79
4.3.2	回路構成とその動作	79
4.3.3	時分割多重された FM 変調波の信号フォーマット	80
4.3.4	評価	82
4.4	本章のまとめ	82
第 5 章	結論	84
5.1	本論文のまとめ	84

5.2 今後の課題	85
5.2.1 電子透かし分野に関して	86
5.2.2 メディアの複合化分野に関して	86
謝辞	88
参考文献	91
付録	101
A.1 セグメンタル SNR	101
A.2 ケプストラム合成によるマスキングカーブの近似法	101
A.3 白色化相互相関の計算方法	102
A.4 電子透かし / ネットワーク・デジタル放送と著作権に関連した新 聞報道	103
研究業績	104

目次

1.1	「音」メディアの研究マップ	6
1.2	電子透かし利用の一例	8
2.1	Boney らの提案する電子透かし埋め込みアルゴリズム	19
2.2	マスキングレベルと透かし	20
2.3	ケプストラムフィルタによるマスキングレベルの近似精度改善	26
2.4	雑音攻撃に対する耐性	31
2.5	MPEG Audio Layer III に対する耐性 (S: 同時マスキングのみ考慮, T: 継時マスキングも考慮したことをあらわす)	33
2.6	長い系列の相関を取る方法の場合の検出	41
2.7	信号の統計的性質を用いる場合の検出	42
3.1	モーションメディアコンテンツ流通プラットフォーム	48
3.2	音によるモーションメディアコンテンツ流通プラットフォームの構想とそのコミュニケーションサービスへの応用	49
3.3	CyberPerformer/Hyper の処理の概要	55
3.4	CyberPerformer/Hyper のノイズへのロバスト性評価	57
3.5	CyberPerformer/Audio の処理の概要	60
3.6	DTMF コードと角度情報への対応/サーボ制御信号と DTMF 信号のタイミングチャート	61
3.7	モータ角度情報の補間	62
3.8	CyberPerformer/Audio を用いたプロトタイプコミュニケーションロボット	64
3.9	CyberPerformer/Audio のノイズへのロバスト性評価	65

3.10	試作したモーションエディタ	67
3.11	CyberPerformer/Hyper ダンシングロボットのパフォーマンス	68
3.12	モーションポータルサービスの Web コンテンツの一例	70
3.13	スマイリーに特化した，一般ユーザを対象とした擬似的にダイレ クトな動作教示が可能な動作エディタの一例	71
3.14	モーションメディアによって強調されたビデオメッセージの一例	72
4.1	USB audio によるモータコントローラの概要	77
4.2	モータコントローラのプロトタイプボード	78
4.3	音による 4ch モータコントローラボードの概要	80
4.4	音による 4ch モータコントローラプロトタイプボード	81
A.1	ケプストラム合成による最小位相フィルタの構成法	102

表 目 次

1.1	さりげなく情報を重畳する「音」メディアの応用例の一覧	5
2.1	雑音の攻撃を受けた場合の透かし検出	22
2.2	MPEG 符号化後の透かし検出	22
2.3	原音と透かし入り音楽とのセグメンタル SNR	23
2.4	対比較によるプリファレンススコア	24
2.5	埋め込みアルゴリズムの改善の効果	30
2.6	8 つの PN 系列を埋め込んだ場合の評価 (マスキングレベルより −30 dB に埋め込み)	33
2.7	DA/AD 変換に対する耐性 (マスキングレベルより −30 dB に埋め 込んだ場合)	34
2.8	MPEG 符号化時 (64 kbits/s) の冗長性導入の効果 (音源はビオラ , マスキングレベルより −30 dB に透かしの埋め込んだ場合)	35
2.9	DA/AD 変換への冗長性導入の効果 (音源はビオラ , マスキングレ ベルより −30 dB に透かしの埋め込んだ場合)	36
2.10	透かしが埋め込まれた音に対する主観評価値	37
2.11	ピアノ音源に対する主観評価値 (再実験)	38
2.12	原音なしで透かしを取り出す方法の評価 (−20 dB 以下に透かしの 埋め込んだ場合)	43
2.13	MPEG 符号化時の原音なしで透かしを取り出す方法の評価 (−20 dB 以下に透かしの埋め込み MPEG-1 Audio Layer III 64 kbits/s で符号化)	44
3.1	二つの実現手法 CyberPerformer/Hyper・Audio の比較	66

3.2	DTMF の 8 個の基準周波数と音名との対応	73
4.1	各モータチャンネル毎の FM 変調波の周波数, 遷移幅および対応するパルス幅	82

第1章 はじめに

1.1 背景

今日我々はインターネットを震源とする情報技術革命の真っ只中で暮らしている。そして情報家電、ロボット等の高度な情報機械が我々の家庭に入り込む時代はすぐそこまで来ている。このような近未来の高度情報化社会においては、複数の人間と複数の機械が家庭やオフィス内に混在している場面が想定される。我々は今まさに、これらの相互コミュニケーションの方法や、インタフェースの有り方について、真剣に議論すべき時を迎えていると言える。

人間と機械、機械・機械のコミュニケーションの切口、すなわち入出力インタフェースに関しては、古くから様々な工夫がなされてきた。人間から機械への指示に関しては、実在する物理媒体を人間が操作することにより、機械に信号が伝えられる。一方、機械から人間に対しては、人間の持つ五感に訴える物理媒体を経由して情報伝達が行なわれる。また機械と機械へのコミュニケーションにおいても、音、電気信号、電波、光などの機械の持つセンシングデバイスに訴える物理媒体を経由して情報伝達が行なわれてきた。これら物理媒体の内、音声や音楽などに代表される音は、最も人間にフレンドリーな媒体の一つと言える。この観点から本論文では音のメディアに着目し、人間と機械が混在する空間において、人間と機械の両者に有益に働きかける「音」メディアの設計とその応用例について論じたい[1]。

1.2 機械・人間向け「音」メディアの重要性

複数の人間と複数の機械が混在する場面における，音によるコミュニケーションを考えた場合，その要素と情報伝達方向から，人間相互，機械から人へ，人から機械へ，機械相互という，4つの網羅的な場面に着目する必要がある．以下ではまず，それぞれの音媒体利用の特徴についてまず考察する．

人対人：音声は，有史以来，人対人のもっとも自然なコミュニケーション手段として利用され人間社会形成に大きな役割を果たしてきた．自然言語はおそらく人類滅亡の日まで途絶えることなく，人と人との情報流通手段の主演を演じ続けるであろう．また音楽（歌）も，人と人との情緒的なコミュニケーションの重要な手段と捉えることが可能である．

機械から人：機械から人へのコミュニケーション手段としても，音は古くから利用されている．例えば，事前に特定な意味付けがされたブザー，チャイム，モールス信号など機械的な信号音やメロディー，そして最近では自然言語を機械に発生させる音声合成技術などがあげられる．前者の信号音やメロディーは，単純な情報しか伝えることしかできないので今後も用途は限定されると思われる．また音声合成技術は極めて理想的ではあるが，現状の技術は抑揚表現が画一的であったり，雑音の元で聞き取りにくいなどの点で実用的に十分とは言えず，今後の発展を期待するところである．

人から機械：人から機械への指令伝達手段としての音の利用場面では，最近著しい進歩を遂げつつある音声認識技術が重要である．自然言語を機械に認識させる音声認識技術は，キーボードやマウスなどの力学的インタフェースの必要無い非接触通信技術の一つであり，最近商用システムの提案も数多い．しかしながら，部屋の残響や，雑音などに対する耐性を考慮すると，まだまだ不十分である [2]．また音楽への理解への研究はさらに困難を極めていると考えられる．

機械対機械：機械 機械間のインタフェース手段としては，有線通信はもちろんのこと，電波や赤外線等の無線通信手段が今後も主流ではあるが，音の利用も無視できない．プッシュホン，ショートメールに使われている DTMF (Dual Tone Multi Frequency)，モデムなどに使われている変調波，SONY のペットロボットの AIBO のサウンドコマンダーはそれらの例であろう．これらの信号は「機械が

正常に動作している」あるいは「機械の間でなんからの通信を行われている」以上の情報を人間にもたらずものではないし、場合によっては大変耳障りである。

これら4つの場面を近未来における我々の生活空間の中で想像してみると、その音コミュニケーションのやり取りを「聞いている」第三者（人間または機械）の存在に気づく。この状況は、人間のみのコミュニケーションで例えば、外国人同士が自分には理解不能な外国語会話を、傍で聞いている状態に類似している。その意味をリアルタイムで翻訳してくれる通訳は、まさに外国人と自分のコミュニケーションの壁を取り除いてくれる要のインタフェースメディアである。

筆者はこのような考察から、人間にとってなじみやすい音（例えば、音声や音楽）であると同時に、機械にとっても理解し易い音の研究の重要性を認識した。

本論文では、以上の考察から、音楽や音声など可聴音に対し、伝達すべき情報をさりげなく重畳した「音」メディアについて論じる。つまり、人間にとっては普通の音に聞こえると同時に、機械にとっても内部に埋め込まれた情報（信号）が受信可能となるような、音の設計法、およびその応用法の検討を行う。

1.3 さりげなく情報を重畳する「音」メディアの応用例

音であるため、これまでの音メディアが有していた下記のような特徴も継承され、これらの特徴により、多くの応用例が創出できる。

- 簡便性：アナログ音声端子・スピーカはもっともありふれたインタフェースである。この単純なインタフェースだけで、さまざまな機器を動作させることができる。また利用者にとっても直感的でありわかりやすい。
- 既存の音資源との親和性：既存メディア、既存インフラストラクチャ、各種録音媒体、放送・通信システムでそのまま使える。

筆者が想定している「音」メディアの応用例の一覧を表1.1に示す。以下にくつかの例について紹介を行う。

- Web連動広告：音声・楽音信号に広告のURLや，Webブラウザ制御信号を重畳させて，受信側で，Webブラウザがポップアップさせるような放送システムである．
- 放送・ロボット連動システム：音楽信号に，ロボット動作コマンドを重畳させて，受信側で，音楽に合わせてまたは，放送内容に連動して，ロボットが動作することが可能にする．
- カーナビシステム：緯度，経度情報を放送局側で埋め込むことにより，場所と店のコマーシャルなどを同時に的確に放送することができる．
- 踊るロボット（玩具への応用）：ロボットにシステムを組み込み，音楽に合わせて，ダンスをするロボットを作成することも可能である．従来ダンシングフラワーなどの音の強弱に反応するロボットがあったが，それらより，より高度な表現，音楽に同期したダンスなどを表現を行うことができる可能性があり，エンタテインメント要素かつ鑑賞に値する動きを行わせることができる．CD，MDなどのメディアに動き情報を蓄積することも可能である．
- 音声電子透かし（著作権管理への応用）：著作権情報や，再生端末の限定の条件などを楽音に重畳させて，端末の動作を制御することが可能になる．
- 音声秘匿通信（ステガノグラフィー）：携帯電話などに，システムを組み込むことで容易に実現できる．電話で話しながら，簡易なメモなどを秘密裏に同時に送ることができる．

このように，「埋め込まれる音」「埋め込む（重畳される）情報」「流通メディア」をさまざまに変更することで多くの興味深い応用が生まれてくることが期待できる．

1.4 「音」メディアの設計のための課題

表1.1のような応用を考慮すると，「音」メディアの設計の課題は，利用できる計算コスト・システムの構築コスト・「音」メディアの伝送路特性・アプリケー

表 1.1 さりげなく情報を重畳する「音」メディアの応用例の一覧

流通形態	放送		ローカル					インターネット	
	音源	放送 CM	ラジオ CM	音声合成	BGM	舞踊音楽	音声	電話音声	楽音
情報	URL など	地図情報	ロボット動作命令	移動場所	踊りコマンド	段落	家電操作命令	著作権保護情報	画像・テキスト
媒体	放送一般	ラジオ	有線	空中	空中	空中	電話網	インターネット	インターネット
応用例	Web 連動広告	カーナビ	メール読み上げロボット	病院内ワゴン	踊るロボット	紙芝居	情報家電遠隔操作	電子透かし	秘匿通信

ションに合わせ、

- 頑健性（伝送路で想定される妨害に強く）
- 音声品質（人間への聞こえが心地よい、もしくは、音声の品質を落とさない）
- 通信レート（十分な機械向けの情報伝送容量を持つ）
- 安全性・秘匿性・耐改ざん性（攻撃・解読・改ざん行為に耐える）

を兼ね備えた方法を開発することであると考えられる。

さりげなく情報を重畳する「音」メディアの研究領域を整理すると、図 1.1 のようになると考えることができる。本図では横軸に秘匿性・安全性、縦軸に通信レートをとって整理を行っている。秘匿通信の場合には機械向けの信号の安全性がなによりも重要視され、可聴の人間向け音信号には、一見、重要なコンテンツが含まれていないように見せかけること（秘匿性）が必要となる。一般に信号を埋め込まれたコンテンツについての品質は重要ではない[3]。機械向けの信号を消去しようとする積極的な攻撃の行為は想定しないため、頑健性はそれほど重要視されない。また著作権管理の場合には、積極的な攻撃を想定した上で、耐改ざん性・耐攻撃性を持つこと、そして人間向けのコンテンツ品質がなによりも重視される[3]。メディアの複合化が目的の場合には、積極的な攻撃の行為を想定しなく

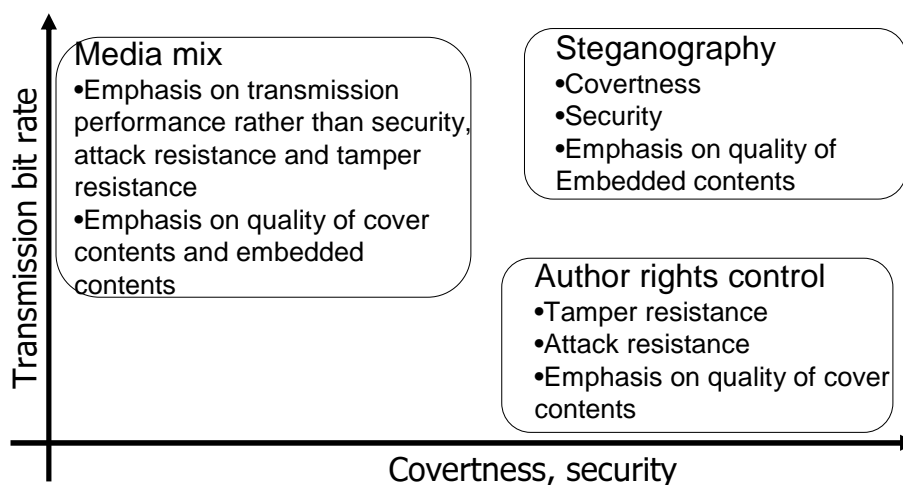


図 1.1 「音」メディアの研究マップ

てもよく，秘匿性・安全性・耐改ざん性よりも通信特性の重視が重視される．さらに，上記二つの分野と異なり，人間向けの音信号も機械向けの音信号も同等に重視される．これらの目的と技術条件の違いにより，三分野で求められる手法はお互い関連を持ちながらも，異なる技術要素を含む手法となり，個別に技術を創出していかねばならないと考えられる．これらの点については，次節，節 1.5 において研究例とともに詳述する．

1.5 本研究の目的と研究領域別の課題と方針

本研究では，前節で述べた 3 つの研究領域の内，現在強く開発が望まれている「著作権管理分野」，そして，産業用用途の規模が大きく見込まれる「メディア複合化分野」に注目しテーマを設定した．著作権管理の分野では，「心理音響モデルに基づく音声電子透かし」，またメディア複合化分野では，近年発展が著しいロボット分野に注目し，「ロボット制御信号の伝送」を主眼に取り組むこととする．以下，本論文のテーマである，電子透かしおよび，メディア複合化分野の関連研究，研究課題，本論文の研究の方向性について述べる．また音による秘匿通信の課題と現状についても，課題設定を明確にするために，紹介する．

1.5.1 音声電子透かし研究の現状とその研究課題

近年，インターネットおよび，デジタル放送の普及により，さまざまなマルチメディアの情報が，ネットワーク上を流れるようになってきた．最近，音楽業界はCDに代わるメディアとして，インターネットに注目している [4]．

しかしながら，ネットワークで流れる音楽情報は，デジタル情報である．それゆえ，オリジナルとまったく同一のコピーが容易にできる．さらに，ネットワークを介して，不特定多数の人に，音楽を配布することも可能である．これは，ネットワークによって，かつてない規模で著作権が侵されることを意味する [5]．

最近，このような行為を防ぐために，電子透かし技術が開発されてきた．電子透かしとは，そのデータ自体に，受け手に気づかれないように著作権情報などを埋め込む技術のことである．電子透かしは，コンテンツと不可分という特徴がある．余計なヘッダーやビット列を付け加えるのではなく，オーディオの場合であれば音として透かしが埋め込まれる．

これによって，著作者は，配布したコンテンツの追跡をファイルのフォーマットによらずにできるようになる．電子透かしは不正なコピー，配布に対して抑止力としての効果を持つ [6] [7]．

図 1.2 に透かし利用方法の一例を示す．この場合では，ユーザに対して音楽を配信するサーバを想定している．配信する音楽に透かしが埋め込まれているとすると，疑わしい音楽データに対して，透かし検出を行えば不正をサーバが立証することができる．また音楽を配信するときに，個人ユーザ ID などを埋め込むと，不正を行なっている本人を正確に特定することができる．

レコーダやビューワーなどの再生機器のための制御コードを埋め込み，コピーやディスク保存の可否をコンテンツ毎に指定するような用途にも用いることができる [8]．

このような目的に使われる電子透かしであるため，次のような条件を満たすことが求められている [7][9]．

[聴覚的に区別がつかない]

透かしを埋め込んだ信号と，元の信号が，聴覚的に区別がつかない．

[耐性を持つ]

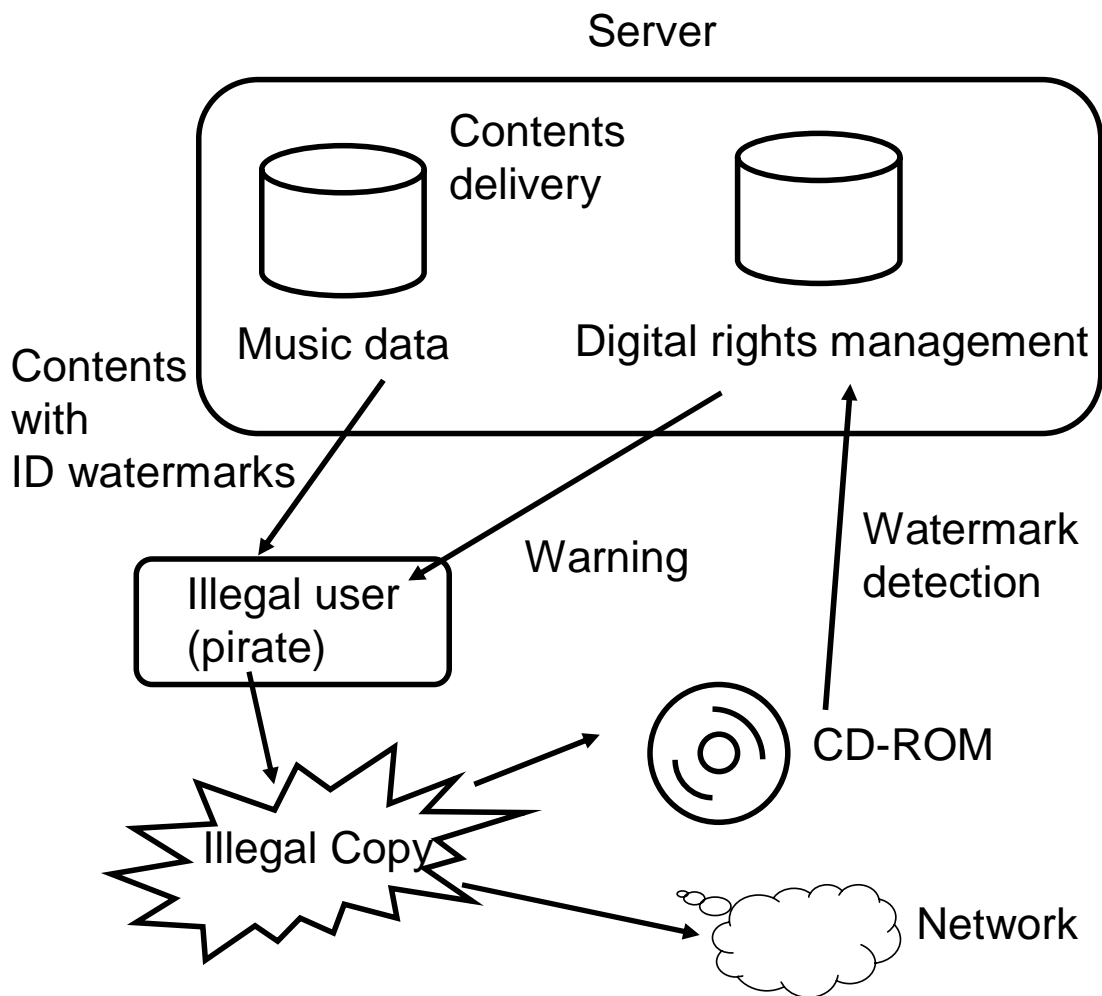


図 1.2 電子透かし利用の一例

悪意を持つ者が透かしを消そうとして透かしの入った音楽に雑音を加えたり何らかの信号処理をする，あるいは音楽データが伝送される途中で不可逆符号化処理を受けることが考えられる．その後でも透かしが取り出せる必要がある．理想的には，透かしが取り出せなくてもよいのは，透かしの埋め込まれた音楽が，観賞に耐えないほど劣化したときのみである．

[埋め込み情報を入れる余裕を持つ]

所有権や権利情報，著作権などの情報を入れるのに十分な余裕があること．

この聴覚的な歪みを小さく保ちつつ，いかにして耐性を持つ透かしを実現するかが，この分野の研究課題である [10] ．

音楽への電子透かしの方法としては，音楽の位相を変更する方法 [11] ，エコー成分を付け足す方法 [12][13] ，周波数拡散を利用する方法 [13][14] ，量子化雑音を利用する方法 [15] ，Boney らの提案する MPEG 心理音響モデルを利用した電子透かし方法 [16] [17] などの方法が提案されている．いずれの報告も，聴覚的な評価に対して不明な点が多い．

また，/聴覚的な品質/攻撃への頑健性/を同時に評価することが，透かしの研究上重要であると考えられるが，そのような報告例は少ない．

本研究では，聴覚的に聴こえないという観点から有望だと考えられる MPEG 心理音響モデルを利用した，Boney らの方法の評価に取り組み，その問題点を明らかにし，その改善・拡張に取り組むことにした．以下本論文の 2 章では心理音響モデルを利用したオーディオ信号への電子透かし法の実装と評価について論じる．

1.5.2 音によるメディア複合の研究とその課題

表 1.1 に示したように，機械向けの音と人向けの音を複合したメディアの産業的な応用は幅広く，多くの波及効果が期待できる．メディアの複合化が目的の場合には，秘匿性・安全性・耐改ざん性よりも通信特性の重視が重視され，人間向けの音信号も機械向けの音信号も同等に重視される．つまり，ビットレート，低い通信遅延などを確保しつつ，人間への聞こえも考慮したデータの埋め込み手法を採用する必要がある．

データを音に変調して伝送する考え方自体は通信用 MODEM，プッシュ式電話の DTMF (Dual Tone Multi Frequency) など古くから実用化されている。またデータ用と音声用に周波数帯域を分割して伝送する技術についても，検討が行われてきているが，さらに，人間への聞こえも考慮してデータと音を変調して伝送する検討として，Gerasimov，Lopes ら，さらに前節で述べたような電子透かしの研究がある。Gerasimov らは，機器間の通信簡易な手段という観点から簡便な通信手段として音伝送に注目し，計算量，伝送効率，そして人間への聞こえの観点からさまざまな変調方式の網羅的な検討を行っている [18]。Lopes らも同様の検討を行っており，人間への聞こえが音楽や楽音に近づくような変調のパラメータを見出している [19]。前節 1.5.1 で述べたように，音声の電子透かしの検討もある。また本研究以後，最近になって松岡らも，音声・音楽に OFDM (Orthogonal Frequency Division Multiplexing) 変調した伝送信号を重畳して音波で伝送する音波情報伝送方式を提案している [20]。

Gerasimov，Lopes らの一連の研究では，「機械向け」の音信号の聞こえをどう人にとって聞こえやすくするかという変調方式に関する検討がなされており，本研究のテーマのように「人間向けの音」にどのように「機械向けの信号」を重畳し流通させるのかといった視点は含まれていない。「電子透かし」では送信できる情報量は，現状では 10～40 bps 程度であり，また，デコードに大きな遅延量が必要であったり，また先験的知識（元信号など）が必要である。このような理由のため，これらの研究成果をメディア複合化の分野に適用させることはかなりの困難がある。また松岡らの研究では，人向けの音の 6 kHz 以上の高周波帯域を OFDM 変調した機械向けの音信号で置き換えるため，人向けの音の品質は劣化することは避けられない。さらに，これらの一連の研究では，音信号の変調方式に関する検討が中心で，複合化されたメディアの流通およびその促進の視点は含まれておらず，その意義と応用例に関する検討も不十分であるといえる。

そこで，本論文では，昨今のロボットブームに注目し，ロボット動作データ（モーションメディアコンテンツ）で変調された音信号を重畳し流通させるというプラットフォーム構想（音によるモーションメディアコンテンツ流通）の提案・設計・実装を行い，「音」メディアによるメディア複合の意義，応用例を示したい。

以下、3章では、まず筆者がこれまで提案してきたモーションメディアコンテンツ流通プラットフォームの基本的な考え方を述べ、そのプラットフォームに対する要求条件を提示する。そして「機械向けの情報」と「人向けの情報」を複合した「音」メディアを使うことでそれらの要求条件をシンプルに満たすことが可能であることと、その根本的なメリットについて記述する。このことにより、本研究の最初のターゲットとして「ロボット制御情報の伝送」を取り上げること理由を明らかにする。次に、音信号への重畳方式として、有線・無線の2つの実現手法を提案し、それぞれの方式に基づいたコンテンツ生成、配信、再生方式や実装例について詳述する。その後、上記メリットを生かしたネットワークサービス応用例とビジネス導入事例を紹介することで本方式の実用性を示す。また4章では、3章の発展として、情報の変復調アルゴリズムの改善、およびUSB Audio Interfaceに基づく、多自由度のロボットの制御例を示す。

なお、これらの「音によるロボット動作データ伝送」の試みは、ロボットの学問分野における「ロボットの動作データの伝送手段」という観点においても新規の試みである。以下では、ロボット研究・ビジネスの現状とロボット動作データの伝送・流通上の課題を明らかにする。また、ロボット制御情報の伝送の関連研究について述べ、ロボット動作データで変調された音信号を人向けの音コンテンツに重畳し流通させることの新規性を示す。

昨今、家電・コミュニケーション・エンターテインメント分野等を中心に、様々なロボットやコンセプトが提案されてきている [21, 22, 23, 24, 25, 26, 27, 28, 29]。これらは単なる一過性のブームに留まらず、ユビキタス社会への発展の一環として我国の重点化技術項目として位置づけられており [30]、数多くの企業・大学により精力的に研究開発が進められている。しかしながら、冷静にこれら一連の活動を眺めてみると実用的・ビジネス的には未だ黎明期であり、企業利潤を生む実産業としての市場が立ち上がっているとは決して言えない。コミュニケーションロボットのビジネスが膠着している最大の理由は、それらロボットを活用したキラーアプリケーション・キラーサービスが提案されていないことに尽きる。これは見方を変えると、機能と価格のバランスに購入者が納得感を持てるような商用サービスが未だ市場に投入されていないからだと筆者は考えている。キラーサー

ビスの創出確率を高める方策としてひとつは、まずコミュニケーションロボット利用の議論をロボット研究者に閉じずに一般に開放することだと考えている。

つまり、コミュニケーションロボットの利用法に関する現実的・多面的な議論を、ロボット研究者や開発メーカだけに留めず、マルチメディアコンテンツクリエイター、デザイナー、ネットワークキャリア、そして何より一般のユーザを含めた様々な立場の人に広げていくことが重要であると考えている。そのためには、ハードウェアとしてのロボットに加え、それを動かすためのプログラムや動作データ、すなわちコミュニケーションロボットに用いられるコンテンツが、誰にでも使える形でネットワーク上を流通する仕組みが必須になってくると思われる。

このような考え方は、近年のインターネットにおけるデジタルメディア（テキスト、音、画像、映像）のコンテンツの急速な広がりを見たときに想起されたものである。これら既存のメディアにおいては、興味のあるユーザがだれでもそのコンテンツを作成し、特定の人に送信したり、あるいは一般に公開することで第三者がダウンロードし再生するという、ユーザドリブンなコンテンツ流通の枠組みと文化が既に浸透しつつある。それゆえ、良質なコンテンツが創出・発見される確率が高まり、さらに関連技術の共通化・標準化が促進され、ユーザのみならず、コンテンツの生成、配信、蓄積、再生に関わる様々なメーカやベンダも潤うという好循環が生まれている。コミュニケーションロボットにおいても、そのような流通の枠組みがもし存在するならば、キラーサービス発生の確率が高まって行くのではないだろうか。そして一旦キラーサービスが発生すればそれに類似したサービスが次々に誕生し、ひいてはロボット市場全体が活性化していくのではないかと筆者は予想している。

筆者はこれまで、コミュニケーションロボットの動きを、従来のテキスト、音声、画像、映像の次に来る第5のメディア：モーションメディアと位置づけ、そのモーションメディアを含むコンテンツをモーションメディアコンテンツと提唱してきた [31]。さらに、そのモーションメディアコンテンツをネットワーク上で流通する枠組みをモーションメディアコンテンツ流通プラットフォームと呼んで、その意義と概念的な要件について検討してきた [31][32]。

上記と関連する研究として、ロボットを一種のメディアと捉える考え方が古く

から提唱されている [33, 34, 35, 36, 37] . また , コミュニケーションロボットのプログラムやデータをネットワーク上で流通させようという試みも数多く報告されている [32, 38, 39, 40, 41, 42, 43, 44, 45, 46] .

しかしながらこれらの研究では , 既存のネットワーク資源すなわち , 既存コンテンツや既存インフラストラクチャを積極的に再利用しようという発想に乏しい . 加えて , コンテンツの生成から再生までネットワーク上での流通全体を視野に入れたフィージビリティに関する検討が十分ではないので , 早期に世の中全体に広まることは困難であると考えられる . モーションメディアコンテンツを早期に流通させるためには , それ専用のネットワーク環境やコンテンツを全く新規に作り上げていくのは現実的ではない . むしろ , 現存するデジタルコンテンツ流通のためのネットワークインフラストラクチャ , 端末 , コンテンツ等などの既存資源を有効活用・再利用することが加速のポイントではないかと筆者は考えている .

そこで本研究のテーマ「音」メディアの考えを活用し , モーションメディアコンテンツの流通を加速する有用性の高い方法を新たに提案したいと考える [47] .

1.5.3 音による秘匿通信 (ステガノグラフィー) の課題と現状 , 本研究テーマとの関連

河口らは「人類は , 他人に気づかれることなく特定の相手とこっそり情報交換する技を太古の昔から探し求めてきた」と述べ「ステガノグラフィーは真に伝えたい情報を別の媒体の中に埋め込んで伝送することによって , 通信当事者以外に対して通信の存在そのものを秘匿することを目的とした研究分野である」と定義している [3] .

つまり , その目的から , 電子透かし・メディアの複合化分野とは , 下記の観点で異なる要求があると考えられる .

- ステガノグラフィーにおいて主役は埋め込みデータであり , 埋め込みデータを隠すメディアは攻撃者にとって不自然に見えない媒体であればなんでもよい . また比較的大規模な埋め込み容量があることが望ましい . 電子透

かし，メディア複合化の場合には埋め込みデータを隠すメディアについてもその種類・品質を考慮する必要がある．

- ステガノグラフィーの場合は秘密の情報が埋め込まれた媒体は，秘密の情報を含んでいないように見せかけなければならない．一方，電子透かし・メディアの複合化の場合は，秘密の情報が含まれているように見えても支障はない¹．
- ステガノグラフィーの場合は，秘密の情報が埋め込まれた媒体はそのまま伝送されることを仮定している研究例が多く，データを消去しようとする積極的な攻撃，特性が未知の伝送路の伝播を想定していない（攻撃を受けた場合や別の伝送路を経由した場合は情報が伝わらないほうがデータの保全性（化けたデータが伝わらない性質）が保てるという考えである）．一方，電子透かし，メディアの複合化の場合は，攻撃や，伝送路の特性を考慮し，比較的多様な伝送路・機器への対応が求められる

なお，埋め込むデータを隠すメディアのことをカバーメディア（cover media）あるいは、ダミーデータ，埋め込みデータをカバーメディアに隠したときにできる，埋め込みデータとカバーメディアが一体化したカバーメディアのことをステゴメディア（stego media）とステガノグラフィー分野では呼ばれている [3]．

音メディアをカバーメディアとした，ステガノグラフィーの研究は古く，1980年代のはじめころには，音声情報にデータ情報を埋め込んで伝送する研究がなされている [48, 49, 50]．Steeleらは，電話のスクランブラで使われる原理を応用し，伝送されるデータのビットの系列で音声の周波数成分を反転させることで，データの伝送を行っている [48]．Wongらは，送信側音声信号を直交変換し，その位相成分の一部を埋め込みデータに応じて変更しデータの伝送を行う方式を提案している [49]．いずれの方法でも，元の電話音声の品質は劣化すると考えられるが，「電話には雑音があたりまえ」という人の心理を付いたステガノグラフィーの先駆的な試みである [50]．

¹ 電子透かしが挿入されていると宣言した画像データが多く流通していることから明らかである

1990年代の中ごろには、デジタル伝送系を主眼に置いた検討が行われてきている。岩切らは、適応差分PCM方式および、CS-ACELP (Conjugate Structure Algebraic Code Excited Linear prediction Audio Codes) において音声符号の一部を利用してテキスト情報を伝送する方法を提案し、音質を損なわず、100 bits/s程度のデータ伝送できたと報告している [51, 52]。また、井上らはMIDIファイルデータを埋め込む方法を提案し、MIDIファイルのサイズに対して、平均1%程度の情報を埋め込む能力があることを実証している [53]。Cvejicらは、高品質なデジタルオーディオファイルのLSB (least significant bit) に音量に応じて情報を埋め込むことにより、44.1 kbits/sもの大量のデータを音質を損なわずに達成したと報告している [54]。

これらの研究は、電子透かし、メディアの複合化分野の観点から見ると、音質が損なわれる（雑音が加わる）[49, 50]、デジタルオーディオのフォーマットに依存した形式であるため、フォーマットの変換によりデータが取り出せなくなる [51, 52, 53]、ボリュームの変更に弱い [54] などの欠点があり、そのまま電子透かしおよびメディア複合化の手法として適用することができない。

また近年では、情報が埋め込まれた証拠を検出しようとする研究（ステガナリシス (Steganalysis)）も注目を集めている。Johnsonらは、LSBに情報を埋め込んだ音声ファイルの統計的解析を試み、LSBに情報が埋め込まれているかどうかの認識器を構築している [55]。本分野は、音信号の高サンプリングレート化、高ビット化、多チャンネル化²の進展に伴い今後の発展が期待されると考えられる。

1.6 本論文の構成

以降の本論文の構成は以下の通りである。第2章では、MPEG心理音響モデルを利用した音声電子透かしについて述べ、高品質で頑健な音声電子透かしシステムが構築できることを実験により示す。第3章では、音信号のロボット制御伝送への応用について述べ、人間向けの音信号にロボット制御信号に複合させることのメリット、優位性について豊富な事例を元に実用性を示す。第4章では、第3

² DVD などでは、192 kHz サンプリング、24 ビット、5.1 チャンネルである。

章の発展系としての多自由度ロボットへの応用について述べる．第5章では，本論文のまとめと今後の検討課題を述べる．

第2章 心理音響モデルに基づく音声 電子透かし

2.1 はじめに

聴覚的に聴こえないという観点からみると，心理音響モデルを用いてマスキングレベルを計算し，そのマスキングレベル以下になるように透かしを埋め込もうとする Boney らの方法 [16] [17] は，有効な手段だと考えられる．まず Boney らの提案する手法を試みて，その有効性・欠点を明らかにする．その後 Boney らの手法の改良・評価を行う．

2.2 心理音響モデルを用いた方法 (Boney らの方法) の追試

2.2.1 心理音響モデルを用いた埋め込みアルゴリズム (Boney らの方法)

電子透かし埋め込みアルゴリズム [16][17] を図 2.1 に示す．このアルゴリズムは，MPEG の心理音響モデルを用いてマスキングレベルを計算し，そのマスキングレベル以下になるように透かし信号を埋め込むことを基本的なアイデアとしている．以下に処理の概要を説明する．

1. オーディオ信号に対して，512 点の窓をかけて周波数分析を行ない，パワースペクトルを計算する．パワースペクトルの最大値 (32768) は 96 dB SPL として正規化するものとする．

2. パワースペクトルから心理音響モデルを用いてマスキングレベルを計算する .
3. 計算されたマスキングレベル 512 点を 32 の帯域に分割しその 32 の帯域ごとに , その帯域の中でもっとも低いマスキングレベルを求める .
4. 32 点のマスキングレベルを 25 次の極 , 1 次の零を持つフィルタで近似する . 二乗誤差が最小になるようにフィルタを設計する¹ .
5. フィルタと PN 系列 (ランダムなビット列)[56][57] を畳み込み , マスキングレベル以下の PN 系列を作成する .
6. 音信号の各サンプル値を二乗し 1024 点のハミング窓で平滑化したものの平方根を包絡とし , PN 系列の時間軸の重み付けを行なう . 得られた系列を透かし信号として 16 bit で量子化し , 元の音楽に加え合わせて透かしの埋め込んだ音楽を得る .

以上の 1 から 6 を音楽信号の最後に到達するまで行なう .

図 2.2 に音楽のパワースペクトル , 心理音響モデルより計算されたマスキングレベル , 設計されたフィルタの周波数特性の一例を示す .

2.2.2 Boney らの検出アルゴリズム

透かしが埋め込まれた受信信号を $y(i)$, 原信号を $s(i)$, 原信号より計算した透かし信号を $w(i)$, 受信信号から計算した透かし信号を $x(i)$ とすると , 検出アルゴリズムは以下のような手順で表せる [16][17] .

1. $x(i) = y(i) - s(i)$ をフレーム単位で求める . この $x(i)$ は , 量子化歪みや攻撃 (透かしの消そうと $y(i)$ を操作すること) の影響を受けた透かしが含まれていると考えられる .

¹ [16][17] で示されている方法では , 10 次の極フィルタが用いられている . しかし著者の追試によれば , 近似精度が低く , 原音の品質が損なわれていた . そのため高次のフィルタを用いることにした .

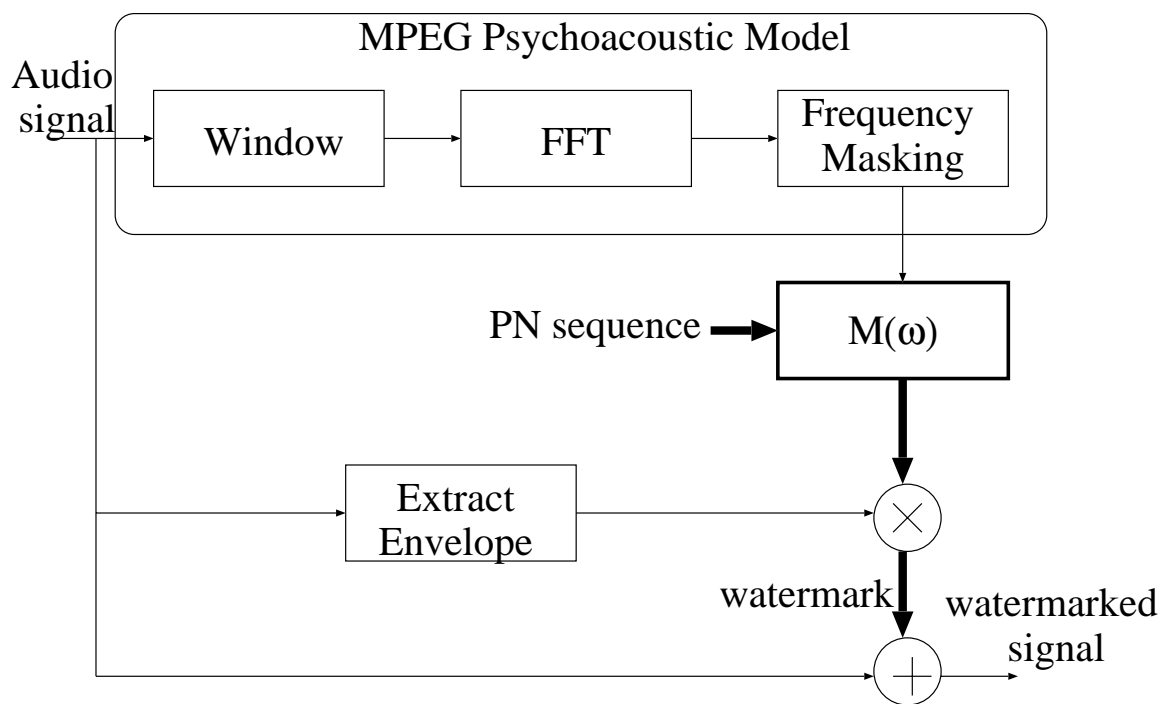


図 2.1 Boney らの提案する電子透かし埋め込みアルゴリズム

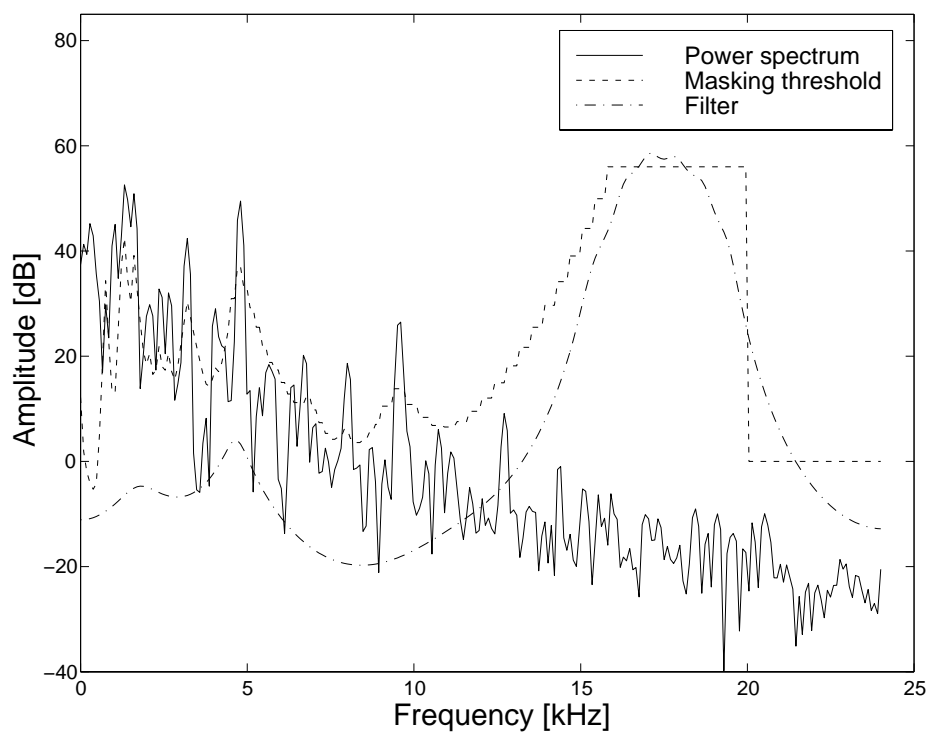


図 2.2 マスキングレベルと透かし

2. $x(i)$ と $w(i)$ の相互相関関数を求め，その相関関数の 0 次の値が，あるしきい値より大きければ，透かし $w(i)$ が入っていたと判断する．

2.2.3 透かし検出実験

実験条件・実験方法

透かしアルゴリズムの有効性を確認するために，以下のような透かしの検出実験を行なった．実験に用いた音楽サンプルは，標本化周波数 48 kHz，長さ 10 秒程度のモノラルのビオラ，カスタネット，ピアノ，女声の歌である．埋め込みレベルはおよそマスキングレベルより -25 dB 以下になっている．

実験 1 透かしの入った信号に MPEG 心理音響モデル用いて計算したマスキングレベル以下になるガウス雑音を加える．その後，透かしの検出を行なう．これは理論的には聴こえない雑音を攻撃として加えたことに相当する．また埋め込みアルゴリズムで複数の PN 系列を埋め込まれた場合における透かし検出のシミュレーションとも考えることができる．

実験 2 透かしの入った信号を MPEG-1 Audio Layer I (MUSICAM)，Psychoacoustic model 1，384 kbits/s) で符号化し，符号化した信号に対し透かしの検出を行なう．384 kbits/s での圧縮は，多くの音楽サンプルで CD と同等の品質をもつといわれており [58]，伝送路途上で透かしの入った音楽データがこの種の符号化処理を受ける可能性は高い

実験結果

検出においては，しきい値の設定が課題となる．そこで，透かし以外の要因（本実験の場合，雑音と量子化誤差）と透かしとの相関と，音楽から取り出した透かし $x(t)$ と原信号から求めた透かし $w(t)$ との相関値から，二つの判断の誤り（「透かしが埋め込まれているのに，埋め込まれていない」「透かしが埋め込まれているのに，埋め込まれていない」と判断する）の割合が等しくなるように設定した [59]．

表 2.1 雑音の攻撃を受けた場合の透かし検出

Music Source	Threshold	Detection rate [%]	Error rate [%]
Viola	0.48	99.8	0.15
Castanet	0.27	96.9	3.18
Piano	0.47	99.9	0.08
Female singing voice	0.26	96.9	3.06

表 2.2 MPEG 符号化後の透かし検出

Music source	Threshold	Detection rete [%]	Error rate [%]
Viola	0.000	50.3	49.7
Castanet	0.001	54.1	45.9
Piano	0.002	53.3	46.7
Female singing voice	0.001	51.1	48.9

実験 1 結果を、表 2.1 に示す。ノイズを加えた場合でもしきい値を適切に設定することで、ほぼ確実に検出できる。本手法は、雑音攻撃に対して、ロバストであることが分かる。

実験 2 結果を表 2.2 に示す。MPEG 符号化のため、ほとんどの透かしは消去されてしまう。検出率に統計的な有意性はない。

表 2.3 原音と透かし入り音楽とのセグメンタル SNR

Music source	Segmental SNR [dB]
Viola	22.1
Castanet	17.1
Piano	27.7
Female singing voice	21.3

2.2.4 音質評価実験

評価方法

透かしが埋め込まれることによる影響を，物理的な距離尺度（セグメンタル SNR）と，主観評価（対比較法）により評価する．音楽のサンプルは，透かし検出実験時に使用したサンプルを使用した．

セグメンタル SNR による客観評価実験

透かしアルゴリズムが波形に及ぼす影響を調べるため，原音と透かしが埋め込まれた音とのセグメンタル SNR を計算した．計算結果を表 2.3 に示す．カスタネットを除くと，セグメンタル SNR は 20 dB を越えており，ある程度の品質を保っていると推測できる．

対比較法による主観評価実験

聴感上，電子透かしがどのような影響をもっているか調査するために対比較による主観評価実験を行なった．被験者は成人男女計 13 名であり，防音室内にて刺激をヘッドホン (Sennheiser HDA 200)[60] により両耳に提示した．実験に用いた音楽サンプルは，透かし検出実験に用いたものと同じである．透かしの入っているサンプル (4 種類) と透かしの入っていないオリジナルサンプルのペアを，各ペアにつき 6 回，全 24 ペアの刺激を作成し，ランダムに提示した．音量の調節は，

表 2.4 対比較によるプリファレンススコア

Music source	Original [%]	Watermark embedded [%]
Viola	56.1	43.9
Castanet	56.1	43.9
Piano*	72.7*	27.3*
Female singing voice	50.0	50.0
Total	58.7	41.2

(Asterisks indicate the significant differences in perceptual quality of the sounds ($p < 0.05$).)

あらかじめ被験者にいくつかの音楽サンプルを提示し、各自にいつも音楽を聞いている時の音量に設定してもらった。

実験結果を表 2.4 に示す。表中の比率は、その音の方が品質が良いと被験者が判断した割合である。原音と透かしを埋め込んだ音に品質の差がなければ 50% 付近の値を示すと考えられる。

適合度検定の結果、ピアノを除くと音質に統計的に有意な差はなく ($p > 0.05$)、Boney らの埋め込み手法が有効な手段であることを示している。

表 2.3、表 2.4 との比較より、本手法のように聴覚的に重要でない部分に埋め込むという手法の場合、セグメンタル SNR が客観評価値としては有効な手段ではないことが分かる。

2.2.5 本節のまとめ

本節では、Boney らの提案する MPEG 心理音響モデルを用いた手法の追試および、評価結果について述べた。その結果、本手法は雑音の攻撃に対しては良好な特性を示し、透かしの検出割合は 96% 以上であった。しかしながら、MPEG 符号化に対しては、脆弱であることが明らかになった。透かしが音質に与える影響は少ないことが分かった。

2.3 埋め込みアルゴリズムの改良

2.3.1 はじめに

前章で述べた Boney らの方法が MPEG 符号化に脆弱であった理由を探り、MPEG 符号化に対してロバストにする方法について述べる。さらに継時マスキングの導入を試みた結果について述べる。また Boney らのアルゴリズムが想定していない、DA/AD 変換およびクロッピング攻撃に対する耐性を高めるために元信号と透かし入り信号の同期検出を白色化相互相関法 (Cross-power Spectrum Phase)[61] により実装した。

2.3.2 透かしを埋め込む周波数帯域の制限

透かしが消える原因は、MPEG 符号化が伝送レートを下げるために高周波成分 (MPEG-1 Audio Layer I 384 kbits/s 場合 15 kHz 以上の周波数成分、MPEG-1 Audio Layer III 64 kbits/s の場合 12 kHz 以上の成分) を伝送しないためであった。これに対応するために、透かしを埋め込む帯域を 12 kHz 以下になるように制限する。

また、心理知覚実験の結果によれば、多くの人が 16 kHz で帯域制限された音声および音楽信号とそれぞれの原信号と区別がつかないことが知られている [62][63]。このため全帯域に透かし信号を埋め込んでも、低域通過フィルタリングを行なうことで、音質を劣化させずに高域に埋め込まれた透かしを除去することができる。帯域を制限することは、このような低域通過フィルタによる攻撃に有効である。

2.3.3 マスキングを近似するフィルタの実現方法の改善

Boney らの方法は、パラメトリックな極・零フィルタを用いマスキングフィルタを近似していた。このため複雑な形状をしているマスキング特性をうまく近似できないこともあった。また実際の心理音響モデルから計算されているマスキング特性をフィルタ設計の難しさから単純な形にする必要があった。そこで、フィ

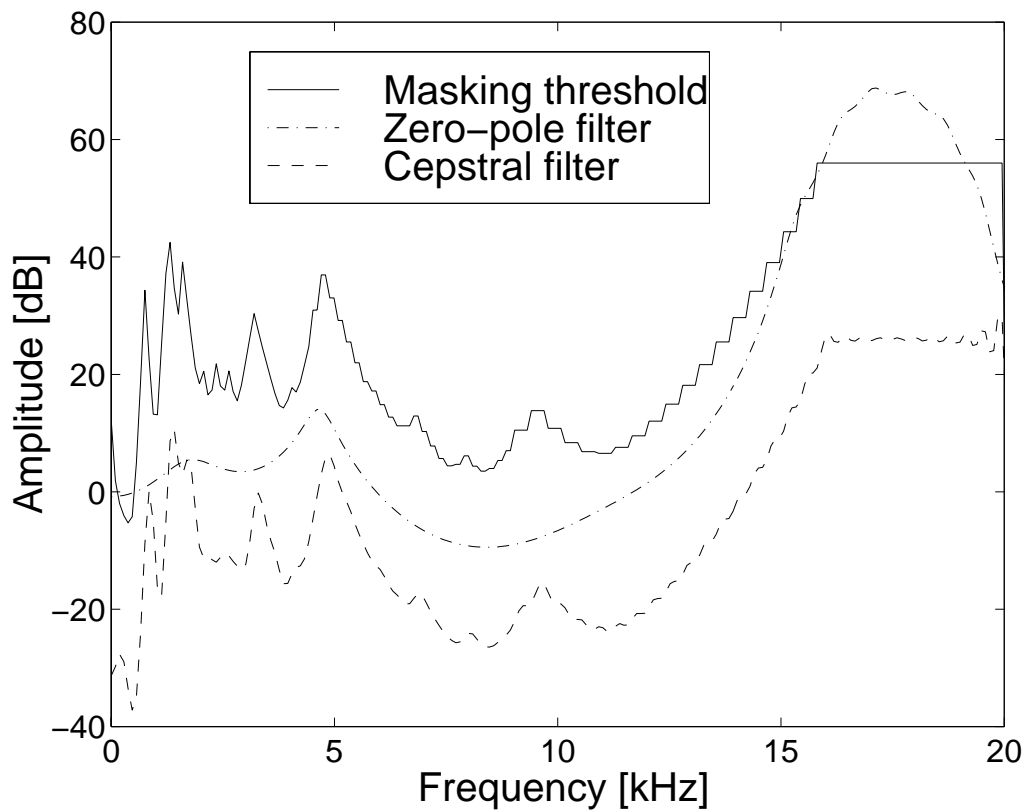


図 2.3 ケプストラムフィルタによるマスキングレベルの近似精度改善

ルタ形状をケプストラム分析 / 合成の考えを用いて近似する [64] . 従来法のように 32 帯域に分割を行い最小値を求めて近似するのではなく , 計算されたマスキングレベルから , 最小位相インパルスを求めて , マスキングレベルを近似するようなフィルタを作成する . そのフィルタ形状を 15 dB ずつシフトして図 2.3 に示す . 極・零フィルタによるものより , 近似精度が高まっていることが分かる . フィルタの安定性 , 計算量の削減の点からも本手法は有利である .

2.3.4 継時マスキングの導入

聴覚系での順向マスキングにおいては，マスキングと信号音の時間間隔の増大にともなって，マスキング量が減少すると共に，マスキング特性が周波数方向に広がることが明らかになっている [65]．本研究では，音楽を聞いている場合でも過去に聞いている音楽によるマスキングの影響があるとして [66]，継時マスキングモデルの定式化を行なう．なお，このような定式化は音声認識の入力部として用いられることがある [67]．

フレーム番号 t のパワースペクトルより，MPEG 心理音響モデルを用いて計算した同時マスキング量を $M(\omega, t)$ とする．ここで ω は周波数である．すると継時マスキングを考慮したマスキング量 $M_t(\omega, t)$ は次式のように表される．

$$\begin{aligned} M_t(\omega, t) = & M(\omega, t) + \exp(-1/\tau)M(\omega, t-1) \otimes S(1, \omega) \\ & + \exp(-2/\tau)M(\omega, t-2) \otimes S(2, \omega) + \dots \\ & + \exp(-k/\tau)M(\omega, t-k) \otimes S(k, \omega) + \dots \end{aligned} \quad (2.1)$$

本式は「ある時点 t の現在の実際のマスキング量」は「過去のフレームのマスキング量を指数関数的に減衰させ・周波数領域になだらかにしたもの」と「現在の同時マスキング量」との「和」であることを意味している．ここで τ は減衰の量を決定する定数である． k はどの程度過去フレームが現在のマスキング量に影響を及ぼすかを決定する定数である． \otimes は ω に関する畳み込みを表す．

$S(k, \omega)$ は時間の経過とともにマスキングカーブが周波数方向に平滑化されることを表す関数で，次式のものを用いた．

$$S(k, \omega) = \frac{0.54 + 0.46 \cos\left(\frac{\pi\omega}{\nu \cdot k + N}\right)}{\sum_{\omega=-(N+\nu \cdot k)}^{\omega=(N+\nu \cdot k)} 0.54 + 0.46 \cos\left(\frac{\pi\omega}{\nu \cdot k + N}\right)} \quad (2.2)$$

ただし ω の変域は $-(\nu k + N) \leq \omega \leq \nu k + N$ である．ここで ν はマスキングと信号の時間差が広がった時に，どの程度窓の幅を広げるかを決定する定数である．

また N はマスキングパターンの広がり初期値を与える．分母は，平滑化することによってマスキング量が変化しないようにするための係数である．

求められた，継時マスキングを考慮したマスキング量 $M_t(\omega, t)$ に MPEG 心理音響モデルで定義されている静寂時最小可聴値を加えることにより継時マスキングを考慮したマスキングレベルが求められる．

2.3.5 白色化相互相関法による同期検出

透かし検出の際，同期を自動的にとることは，クロッピング攻撃（信号の一部を切りとること）や DA/AD 変換への耐性を高めるために必要である．雑音による攻撃や，符号化の存在，同期を取るべき対象が準周期的な音声 / 音楽信号であることを考えると，従来から信号処理の分野で同期に用いられてきた matched-filter による方法は得策ではない．

ここでは，マイクロホンアレーの分野で信号の同期手段として用いられている白色化相互相関法 (CSP)[61] を使用することにする．この方法は，雑音やその他の信号の歪みに対しても頑健に同期の検出が可能であることが知られている（計算方法については，付録 A.3 参照）．

2.4 従来法からの検出率改善の評価

改善効果の評価を行なうため，もともとの Boney の方法，帯域制限を導入した時，ケプストラムによるフィルタ近似，そしてそれに継時マスキングを導入した時の検出率を比較評価する．継時マスキングを考慮した時の透かしのパラメータとしては， $N = 11$ ， $\nu = 1$ ， $\tau = 1$ ， $k = 3$ を用いた．これらのパラメータは，参考文献 [65]，[67]などを参考にし，実験的に定めた．また，埋め込みのレベルはマスキングレベルより -30 dB とした．それぞれの方法で透かしを埋め込んだ音楽を，MPEG-1 Audio Layer I 192 kbits/s で符号化したものに対して検出実験を行なった．その結果を表 2.5 に示す．検出時のしきい値の設定は，2 節の場合と同じく 2 種類の誤りが等しくなるところに設定した．この結果より

1. 帯域を制限することで透かしが検出できるようになる。
2. ケプストラム分析で実現したフィルタにより、より安定した頑健な透かしが埋め込み可能である。
3. 継時マスキングの導入は、カスタネットなどの突発的な信号に対して有効である。

であることが分かった。後述の表 2.10 および表 2.11 から埋め込みレベル -30 dB の場合、原音と比べて音質の劣化が少ないにもかかわらず、表 2.5 に示すように、従来法に比べ検出率が向上しており、提案法が有効であることがわかる。

2.5 攻撃への耐性評価

2.5.1 雑音攻撃への耐性

透かしの入った信号に MPEG 心理音響モデルおよび継時マスキングを考慮したモデルを用いて計算したマスキングレベル以下になるようにガウス雑音を加える。その後、透かしの検出を行う。これは理論的には聞こえない雑音を攻撃として加えたことに相当する。

埋め込む雑音の量は計算したマスキングレベルに対して 0 dB から -50 dB まで -10 dB 刻みでの 6 通りである。その結果を図 2.4 に示す。この図より、攻撃の雑音レベルがマスキングレベルに対して、 -20 dB より大きくなると、透かしの検出率が大きく落ち始めることがわかる。 -30 dB より小さな雑音は検出率に与える影響は小さい。

2.5.2 MPEG-1 Audio Layer III に対する耐性評価 (2つのPN系列を埋め込んだ時)

MPEG Audio Layer III[68] に対する耐性を、より実運用に近い環境での評価を目的として、2つのPN系列を埋め込んだ場合で評価する。

表 2.5 埋め込みアルゴリズムの改善の効果

Conventional method	Boney method		
	Music source	Threshold	Detection rate [%]
	Viola	-0.000	46.2
	Castanet	-0.001	51.7
	Piano	-0.002	43.2
	Female singing voice	+0.001	50.0
Proposed method	Band width restriction		
	Viola	+0.074	60.1
	Castanet	—	—*
	Piano	+0.082	65.9
	Female singing voice	+0.137	60.0
	Band width restriction+cepstrum filtering		
	Viola	+0.117	83.1
	Castanet	+0.058	57.4
	Piano	+0.132	84.5
	Female singing voice	+0.038	68.5
	Band width restriction+cepstrum filtering+temporal masking		
	Viola	+0.138	84.7
	Castanet	+0.073	72.7
	Piano	+0.142	84.7
	Female singing voice	+0.063	73.7

* 極・零フィルタの近似精度低下のため安定して埋め込みを実現できず。

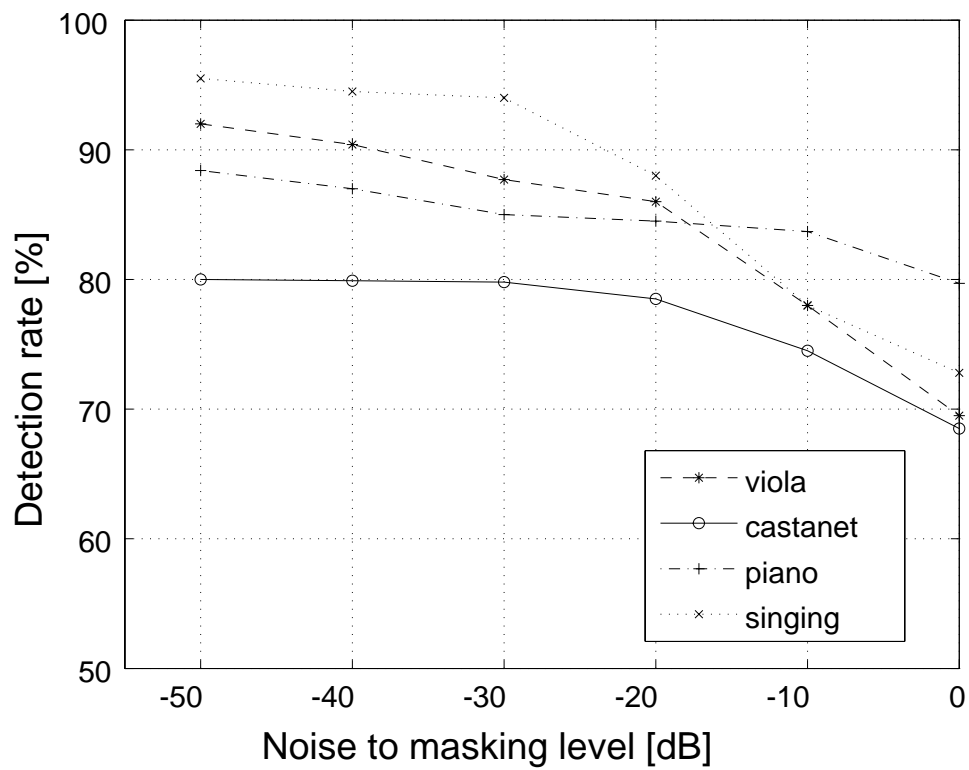


図 2.4 雑音攻撃に対する耐性

透かしを埋め込んだ音楽を，MPEG-1 Audio Layer III でさまざまに伝送レートで符号化しその後で検出を行なった．しきい値の設定は，2つのPN系列の復号誤りが最小になるように設定した．ケプストラムを用いたフィルタを使用し，同時マスクングのみ考慮した場合と，継時マスクングも考慮した場合の埋め込みアルゴリズムを使用した．埋め込みのレベルはマスクングレベルより -30 dB，継時マスクングを考慮した時の透かしのパラメータとしては， $N = 11$ ， $\nu = 1$ ， $\tau = 1$ ， $k = 3$ を用いた．

実験結果を図 2.5 に示す．これは一フレームに対して，一つのPN系列を埋め込んだ場合の結果である．”Baseline” は符号化を通さず（攻撃を行っていない状態と等価）検出を行なった時の結果を表し，検出率の上限を表す．符号化のビットレートの減少にともない，透かし検出は難しくなるが，128 kbits/s までの符号化に対しては十分ロバストである．また継時マスクングを導入することにより，数パーセント程度検出率が改善することが分かった．

2.5.3 結託した攻撃への耐性評価（多くの数のPN系列を埋め込んだ場合の評価）

8つのPN系列を同時に埋め込んだ時の，1つのPN系列の検出特性を評価する．同時マスクングのみ考慮し，マスクングレベルより -30 dB に埋め込んだ．しきい値の設定は，PN系列の復号誤りが最小になるように設定した．結果を表 2.6 に示す．

2つのPN系列を埋め込んだ場合と変わらない検出率を保っている．攻撃のために，さらに多くのPN系列を加算することも考えられるが，音質が損なわれるため，PN系列を加算することによる攻撃への耐性は備えていると考えることができる．

2.5.4 DA/AD変換での攻撃に対する耐性評価

同時マスクングのみを考慮してマスクングレベルより -30 dB に透かしを埋め込んだ信号を DA 変換し，同軸ケーブルで 3 m 伝送したあと，再び AD 変換し，

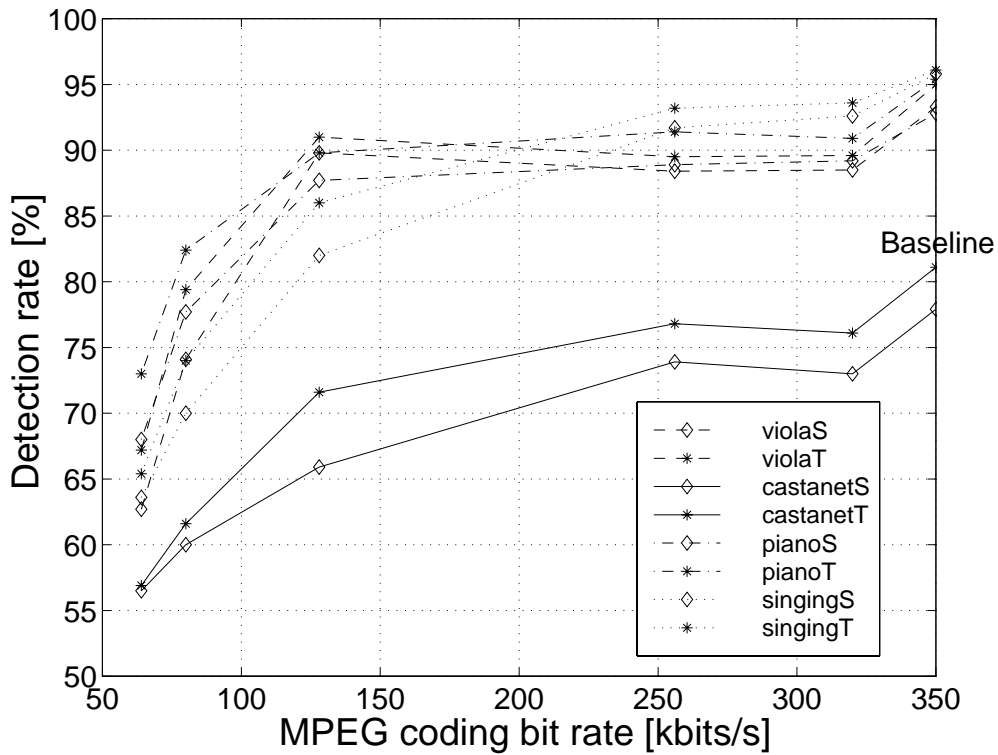


図 2.5 MPEG Audio Layer III に対する耐性 (S : 同時マスキングのみ考慮, T : 継時マスキングも考慮したことをあらわす)

表 2.6 8つのPN系列を埋め込んだ場合の評価 (マスキングレベルより -30 dB に埋め込み)

Music source	Threshold	Detection rate [%]
Viola	0.172	85.2
Castanet	0.186	80.2
Piano	0.146	85.5
Female singing voice	0.180	92.3

表 2.7 DA/AD 変換に対する耐性 (マスキングレベルより -30 dB に埋め込んだ場合)

Music source	Threshold	Detection rate [%]
Viola	0.070	77.4
Castanet	0.057	63.8
Piano	0.075	77.6
Female singing voice	0.031	77.7

その信号に対して透かしの検出を行なった。2つのPN系列を埋め込んだ時の、検出特性を評価する。しきい値の設定は、2つのPN系列の復号誤りが最小になるようにした。

DA/AD変換に使用した機器はSONY DATデッキDTC-2000ESである。実験に先立ちDATのDA/ADの変換値の差ができるだけなくなるように録音レベルを調節した。

検出結果を表2.7に示す。やや正解率が減少しているがこれは白色化相互相関法では同期検出できない、サンプリングポイント以下のずれが存在するためである。今回使った機器の組合せでは、10秒につき20サンプリングポイント程度のずれが生じた。オーバーサンプリング等による処理を用いることで、さらに検出率は向上させることができると考えられる。

2.6 冗長性導入による効果

一つのPN系列を埋め込む時に冗長に複数のフレームを使用することも透かしの強度を高める上で有効な手法であると考えられる。そこで、一つのPN系列を埋め込むのに使用するフレーム数を増加させた時の検出率の変化を調査した。音源はビオラ、マスキングレベルより -30 dBに透かしを埋め込んだ場合において、MPEG符号化、DA変換を行った音源に対して評価を行った。結果を表2.8、表2.9に示す。MPEG符号化および、DA/AD変換に対しても冗長性を導入するこ

表 2.8 MPEG 符号化時 (64 kbits/s) の冗長性導入の効果 (音源はビオラ, マスキングレベルより -30 dB に透かしを埋め込んだ場合)

Number of frames to use embedding one watermark (time to correspond)	Number of embedded watermarks	Number of detected watermarks	Detection rate [%]
1 (10 msec)	1299	801	62.7
5 (53 msec)	256	188	73.4
10 (107 msec)	128	106	82.8
20 (213 msec)	64	58	90.6
40 (427 msec)	32	30	93.8

とでより頑健な透かしを実現できることがわかる。

2.7 音質の主観評価実験

2.7.1 主観評価方法

より精密に音質の主観評価を行うため, 隠れ基準付三刺激二重盲検法 (triple stimulus/hidden reference/double blind method) [69][70] による, 成人男女計 15 名を用いた主観評価実験を行なった。音源は 48 kHz サンプリングの 10 秒程度のビオラ, カスタネット, ピアノ, 女声の歌である。評価対象はマスキングレベルより -20 dB および -30 dB のレベルで透かしを埋め込んだ音楽である。実験は防音室内にてヘッドホン (STAX SRM-Xh)[60] で両耳受聴により行った。評価方法は, 原音-処理音-原音, 原音-原音-処理音のどちらかの順序で, ランダムに三つ組で提示し, 被験者は 2 番目, もしくは 3 番目の音に対して, 原音と感じた音に 5 点, その対の音に原音と感じた音との差を「5: 差が分からない, 4: 分かるが, 気にならない, 3: やや気になる, 2: 気になる, 1: 非常に気になる」に従って点数をつけた。音量の調節は, あらかじめ被験者にいくつかの音楽サンプルを提示

表 2.9 DA/AD 変換への冗長性導入の効果（音源はビオラ，マスキングレベルより -30 dB に透かしを埋め込んだ場合）

Number of frames to use embedding one watermark (time to correspond)	Number of embedded watermarks	Number of detected watermarks	Detection rate [%]
1 (10 msec)	1301	1007	77.4
5 (53 msec)	200	192	96.0
10 (107 msec)	100	100	100.0

し，各自にいつも音楽を聞いている時の音量に設定してもらった．

主観評価結果

実験結果を表 2.10 に示す．表中の値は，透かしの埋め込まれた音に対する平均点から，原音の平均点を減じた値を示している．原音と透かしが埋め込まれた音との間に品質の差がなければ，零付近の値を示す．

原音に対する平均評価値と，透かしが埋め込まれた音に対する平均評価値の間に有意な差があるかどうかの t 検定 [71] を行なった．表中*印が，有意な差 ($p < 0.05$) があったことを示している．

マスキングレベルより -30 dB に埋め込んだ場合は，ピアノを除くと統計的に有意な差はなく ($p > 0.05$) 透かしが埋め込まれた音と原音との差がないことを示している．

マスキングレベルより -20 dB 以下に埋め込んだ場合は，女声の歌を除いて，原音との間に統計的に有意な差 ($p < 0.05$) が生じた．この場合でも主観評価値は「差が分からない」と「差が分かるが，気にならない」との間のスコアを示した．またマスキングレベルより -30 dB 以下に埋め込んだ場合は，ピアノを除くと統計的に有意な差はなかった．これはピアノが歪みが検知され易い音源であるためと考えられる [72]．どの程度の透かし埋め込み量に対して，音質の劣化が検

表 2.10 透かしが埋め込まれた音に対する主観評価値

Simultaneous masking only		
Music source	Below -20 dB	Below -30 dB
Viola	-0.46*	-0.008
Castanet	-0.38*	-0.016
Piano	-0.76*	-0.36*
Female singing voice	+0.19	+0.16
Simultaneous and Temporal masking		
Music source	Below -20 dB	Below -30 dB
Viola	-0.59*	-0.025
Castanet	-0.75*	-0.18
Piano	-0.69*	-0.36*
Female singing voice	+0.09	+0.16

(Asterisks indicate the significant differences in perceptual quality of the sounds ($p < 0.05$).)

出されないかをピアノに対して再実験を行った。マスキングレベルの -20, -30, -35, -40 dB 以下に埋め込んだ場合に対して、別の 8 人の被験者による主観評価実験を行った。その結果を表 2.11 に示す。再実験においては、-30 dB 以下では、有意な差が検出されなかった。これらより、本埋め込み手法が有効な手段であることが分かる。

2.8 耐性評価および音質評価のまとめ

透かしを埋め込む帯域を制限することで MPEG 符号化に頑健にすることができた。またフィルタの設計方法を変更することにより、より安定して透かしを埋め込むことができることを示した。また継時マスキングの簡単なモデリングを行ない実装した。同期検出を取り入れることで、DA/AD 変換にロバストな透かし

表 2.11 ピアノ音源に対する主観評価値（再実験）

Embedding intensity [dB]	-20	-30	-35	-40
Simultaneous masking only	-0.45*	-0.05	-0.10	-0.13
Simultaneous and temporal masking	-0.63*	0.08	0.03	-0.05

(Asterisks indicate the significant differences in perceptual quality of the sounds ($p < 0.05$).)

を実現できた．さらに冗長性を取り入れることで透かしがより頑健になることが分かった．また主観評価実験の結果，透かしを埋め込むことによる音質の低下は少ないことが分かった．

2.9 原信号を必要としない検出方法

2.9.1 はじめに

透かし検出に原音が必要であることは電子透かしの応用を考えた時に不利な条件となる．そこで本節では，原音を必要としない検出方法について検討する．検討したアルゴリズムは，長い系列に対して相関を取る方法と，信号の統計的性質を利用する方法である．

2.9.2 原音を必要としない透かし検出アルゴリズム

透かしが埋め込まれた受信信号を $y(i)$ ，原信号を $s(i)$ ，原信号より計算した透かし信号を $w(i)$ ，受信信号から計算した透かし信号を $w'(i)$ とすると，これから提案する検出アルゴリズムは以下のような手順で表せる．

[長い系列に対して相関を取る方法]

透かしを埋め込むことは，次のように表せる．

$$y(i) = s(i) + w(i)$$

ここで, $y(i)$ と, $w'(i)$ の相関を求める .

$$\begin{aligned}\sum_{i=1}^N y(i)w'(i) &= \sum_{i=1}^N \{s(i) + w(i)\}w'(i) \\ &= \sum_{i=1}^N \{s(i)w'(i) + w(i)w'(i)\}\end{aligned}$$

$s(i)$ と $w'(i)$ が無相関であると仮定すると

$$\sum_{i=1}^N y(i)w'(i) \approx \sum_{i=1}^N w(i)w'(i)$$

これを利用して, 透かしを検出する . $w(i) \approx w'(i)$ として, 上記の値を $\sum_{i=1}^N w'^2(i)$ で正規化し, C と定義する .

$$C \triangleq \frac{\sum_{i=1}^N y(i)w'(i)}{\sum_{i=1}^N w'^2(i)}$$

N が十分大きいとした場合

- $C \approx 1$: 透かしが入っている場合
- $C \approx 0$: 透かしが入っていない場合, もしくは $w(i)$ 以外の透かしが入っている場合

となる .

[信号の統計的性質を利用する方法]

二つの透かし信号が埋め込まれると仮定し, $s(i)$ に対して, $w_1(i)$ という透かしの埋め込んだとする .

$$y(i) = s(i) + w_1(i)$$

$y(i)$ から, $w'_1(i)$, $w'_2(i)$ を減算する. その標準偏差を求め, $y(i)$ の標準偏差で正規化する. $\sigma_N(\cdot)$ は N 個のサンプル値から標準偏差を求める操作を表す.

$$\begin{aligned} S_1 &= \frac{\sigma_N\{y(i) - w'_1(i)\}}{\sigma_N\{y(i)\}} \\ &= \frac{\sigma_N\{s(i) + w_1(i) - w'_1(i)\}}{\sigma_N\{y(i)\}} \\ S_2 &= \frac{\sigma_N\{y(i) - w'_2(i)\}}{\sigma_N\{y(i)\}} \\ &= \frac{\sigma_N\{s(i) + w_1(i) - w'_2(i)\}}{\sigma_N\{y(i)\}} \end{aligned}$$

$w_1(i) \approx w'_1(i)$ であるとする, 上式より $S_1 < S_2$ である. これを利用して, 透かしの検出を行なう.

2.9.3 検出アルゴリズムの有効性の検討

CD のクラシック音楽を 2 分間サンプリングしたものを, 検出実験に用いた. 2 つの PN 系列を埋め込んだ時, 正しく復号された割合を評価する. 埋め込みアルゴリズムとしては, 同時マスキングのみ考慮する方法で, マスキングレベルの -20 dB 以下になるように透かしを埋め込んだ. 結果を表 2.12 に示す. 両手法とも有効な検出方法だといえる. フレーム内のサンプル数 $N = 163840$ (3.7 秒程度) であれば 85% 以上の高い検出率が得られている. またその時の C の値, および S_1 , S_2 値の変化をそれぞれ, 図 2.6, 図 2.7 に示す. 透かし信号が原信号より生成されているため, 完全に透かし信号と音楽信号の相関が零であるわけではない. そのため, 理想的な値よりずれが生じていると考えられる.

次に, MPEG-1 Audio Layer III 64 kbits/s において透かしを埋め込んだ音楽を符号化したものに対して, 透かし検出を行なった. 1 チャンネルあたり, 64 kbits/s での符号化は多くの音楽サンプルで, CD に近い音質を保つことが知られている [58]. 実験条件は, 先ほどの実験と同じである.

結果を表 2.13 に示す. 透かしを埋め込んだ信号をそのまま検出するよりも, 若干の検出率の低下が見られるものの, この検出方法は MPEG-1 Audio Layer

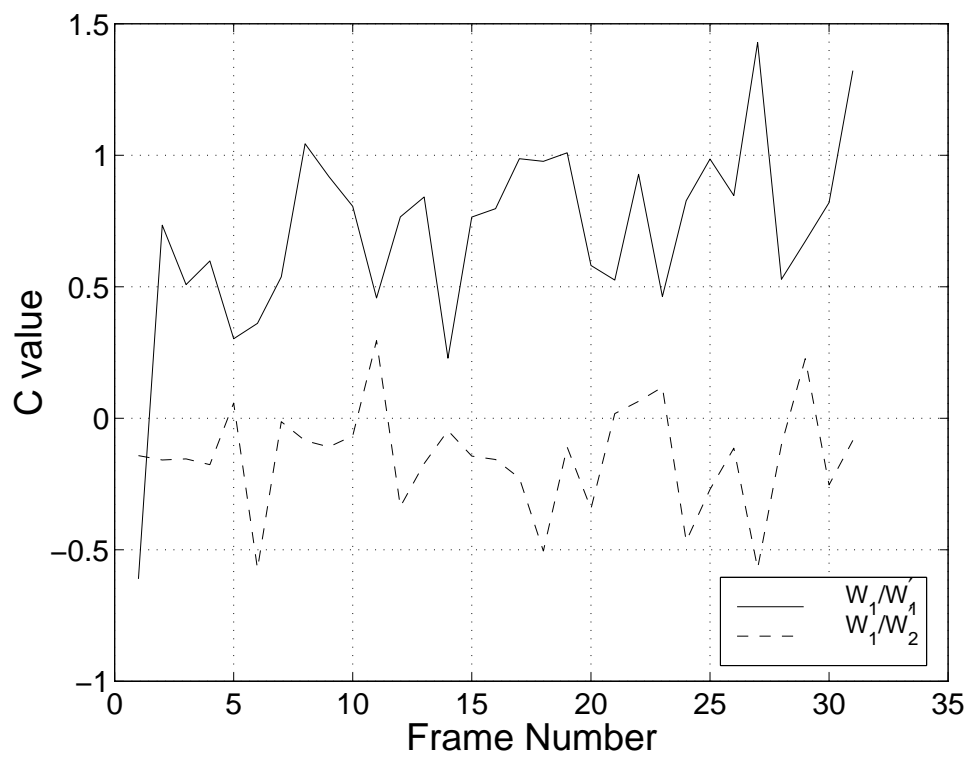


図 2.6 長い系列の相関を取る方法の場合の検出

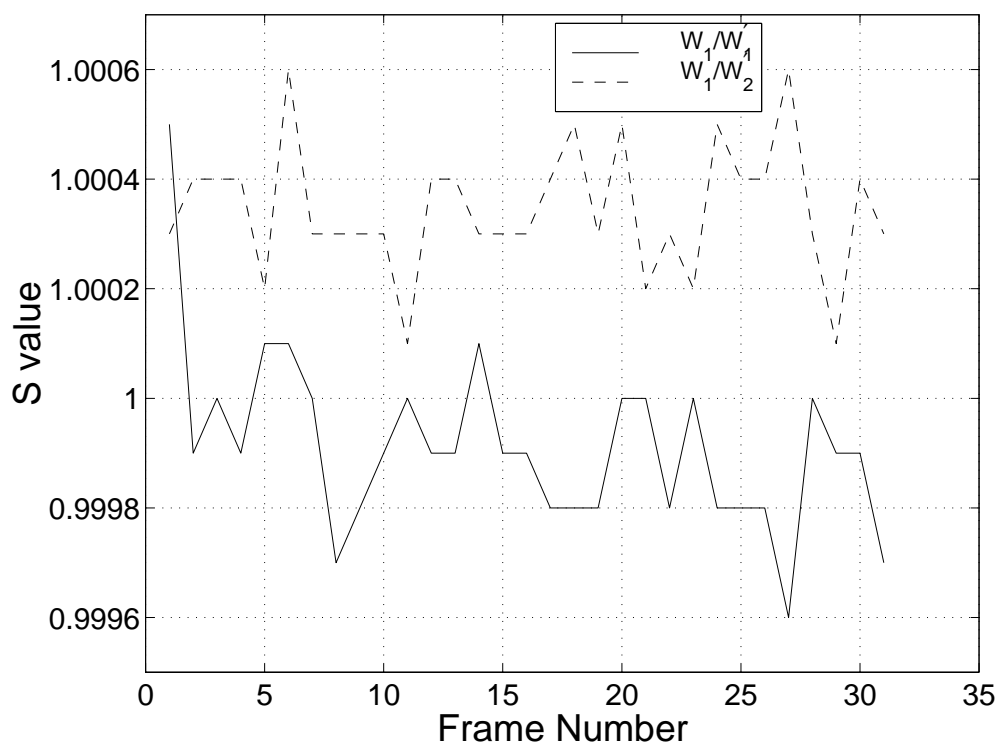


図 2.7 信号の統計的性質を用いる場合の検出

表 2.12 原音なしで透かしを取り出す方法の評価 (−20 dB 以下に透かしを埋め込んだ場合)

Number of frames to use embedding one watermark (time to correspond)	Number of embedded watermarks	Number of detected watermarks	Detection rate [%]
Long term temporal sequence correlation method			
81920 (1.9 sec)	63	54	85.7
163840 (3.7 sec)	31	30	96.8
Long term temporal sequence deviation method			
81920 (1.9 sec)	63	55	87.3
163840 (3.7 sec)	31	30	96.8

III の符号化にも十分有効な手法であることが確かめられた。

2.9.4 本節のまとめ

信号の長い系列の相関または、信号の統計的性質を利用することにより原音なしで透かしを検出することができることを示した。本埋め込みアルゴリズムの適用範囲を広げることができた。またこの検出アルゴリズムは MPEG 符号化にも頑健な手法であることが分かった。

2.10 本章のまとめ

高品質かつ頑健なオーディオ信号への電子透かしの実現を目指して、研究を進めた。そこで、聴こえないという条件に着目しマスキング現象を利用することを考えた。第 2.2 節ではマスキング現象を利用している Boney らの方法の追試を行ないその性能評価を行なった。その結果雑音に対しては良好な特性を示すものの、MPEG 符号化に対して非常に弱いという欠点を持つことが分かった。また音質

表 2.13 MPEG 符号化時の原音なしで透かしを取り出す方法の評価 (−20 dB 以下に透かしを埋め込み MPEG-1 Audio Layer III 64 kbits/s で符号化)

Number of frames to use embedding one watermark (time to correspond)	Number of embedded watermarks	Number of detected watermarks	Detection rate [%]
Long term temporal sequence correlation method			
81920 (1.9 sec)	63	50	79.4
163840 (3.7 sec)	31	28	90.3
Long term temporal sequence deviation method			
81920 (1.9 sec)	63	51	81.0
163840 (3.7 sec)	31	29	93.5

の主観評価実験の結果，良好な特性を示すことが分かり Boney らの埋め込み手法は，有望であることが確かめられた。

第 2.3 節から 2.8 節では，埋め込みアルゴリズムの改善に取り組んだ。まず透かしを埋め込む周波数帯域の制限を行なった。そしてマスキングを近似するフィルタの実現方法の改善に取り組んだ。さらに継時マスキングをモデル化し埋め込みアルゴリズムに導入した。また同期検出機構を実装し，DA/AD 変換およびクロッピング攻撃に耐える透かしの実現を目指した。その結果，MPEG および，DA/AD 変換に対しても頑健な電子透かしが実現できた。また主観評価実験の結果，音質の劣化は認められなかった。

第 2.9 節では，透かし検出の際に原音を必要としない検出方法を検討した。長い系列の相関を取ること，また信号の統計的性質を利用することにより原音を必要としない検出方法を実現することができた。これにより本埋め込み手法の適用範囲が広がられた。またこの検出方法は MPEG 符号化に対しても頑健であることが分かった。

今後の課題としては，透かしの安全性の検討，また事後的に定めている検出しきい値の自動設定などがある。

第3章 「音」メディアによるモー ションメディアコンテンツ流 通プラットフォームへの適用

3.1 はじめに

本章では、まず「音」メディアによるメディア複合化研究の前提として、筆者がこれまで提案してきたモーションメディアコンテンツ流通プラットフォーム（ロボット制御情報伝送のためのプラットフォーム）の基本的な考え方を述べ、そのプラットフォームに対する要求条件を提示する。「人向けの情報」と「機械向けの情報」とを複合した「音」メディアを使うことでそれらの要求条件を簡潔なシステムで満たすことが可能であることを示す。そのメリットについて記述する。これらの記述を通して、本研究の最初のターゲットとして「ロボット制御情報の伝送」を取り上げることの理由を明らかにする。

次に、音信号へのロボット制御情報の重畳方式として、有線・無線の2つの実現手法を提案し、それぞれの方式に基づいたコンテンツ生成、配信、再生方式や実装例について詳述する。その後、上記メリットを生かしたネットワークサービス応用例とビジネス導入事例を紹介することで本方式の実用性を示す。

3.2 モーションメディアコンテンツ流通プラットフォームとそのプラットフォームに望まれる要件

本節では、筆者がこれまで提案してきたモーションメディアコンテンツ流通プラットフォームの基本的な考え方を述べ、そのプラットフォームに対する要求条件を提示し、メディア複合化の研究の導入としたい。

筆者は、モーションメディアコンテンツ流通のためのモジュール群を「モーションメディアコンテンツ流通プラットフォーム」と呼び、検討してきた [31, 32]。このイメージ図を図 3.1 に示す。このモーションメディアコンテンツ流通プラットフォームは、ネットワーク基盤、課金・決済などの情報流通のための基礎的な情報流通プラットフォーム上に置かれることを想定し、生成・配信・再生・記述・管理・蓄積のコンポーネントから構成されている。「管理」「蓄積」「記述」については、すでに議論されているので [31, 32]、本章では、プラットフォームのキーコンポーネント「生成」「配信」「再生」に焦点を当て論じる。

本節の理解を助けるために、キーコンポーネント「生成」「配信」「再生」「記述」を音楽の例で説明する。生成は、音楽を作曲するためのコンポーザソフトなどの道具に対応する。また配信は音楽を配信する仕組み、例えば、ストリーミングのためのサーバや、ネットワーククライアントソフト、CD などの媒体に対応する。再生はデコーダソフトやプレイヤーなどからなる。また記述は、MIDI や MP3、楽譜などの音楽を記述するためのフォーマットや規約にあたる。

このように音コンテンツ流通のためのプラットフォームとして、さまざまな標準的なシステム・ソフトウェア・規約が存在しているために、音コンテンツは積極的に流通されているといえる。モーションメディアコンテンツにおいても同様なコンテンツ流通のためのプラットフォームが重要であると筆者は考える。同時に、音楽におけるスピーカに相当する出力機器としてのロボットの多様性（形状、関節数、自由度、入出力インタフェースのバリエーションなど）、またそれら多様なロボット間の相互接続性・ロボット間のコンテンツの互換性、モーションのコンテンツとしての使用方法・位置づけ（モーション単独ではなく、他メディアと連携・同期して使用される、専門家ではなく一般ユーザが使用）、プラットフォーム

構築にかかるコスト，にかんがみて，モーションメディアコンテンツ流通プラットフォームが満たすべき要件を下記のように考察してきた [31, 32] .

1. モーションメディアコンテンツの生成の観点
 - (a) 簡易なアプリケーションの作成環境の提供
 - (b) 他メディアとの同期を含むモーションメディアコンテンツの簡易な作成・編集環境の提供
2. モーションメディアコンテンツの配信・再生の観点
 - (a) 既存ネットワーク資源との親和性
 - (b) 従来メディアとの同期
 - (c) 各ロボットに応じた動作の変換
3. モーションメディアコンテンツの記述の観点
 - (a) 他メディアとの同期も考慮したロボット動作記述の統一化・標準化
 - (b) インタフェースの統一化・標準化

モーションメディアコンテンツ流通はいまだ黎明期であるとの認識のもとで，コンテンツ流通を加速する視点にたつと，簡便でかつ魅力的なコンテンツを生成・配信・再生できるプラットフォームの構築が標準化・統一化に比して，特に重要であると考えられる。すなわち (1) - a, b (2) - a, b の要件を満たすプラットフォームの構築が黎明期のモーションメディアコンテンツ流通を加速する意味で重要であると考えられる。次節以降で (2) - a, b をオーディオ技術をベースとして実現する方式について詳細に説明し (1) - a, b についても，具体例を述べる。なお (2) - c (3) - a, b については別論文で議論しているので参照されたい [31, 32] .

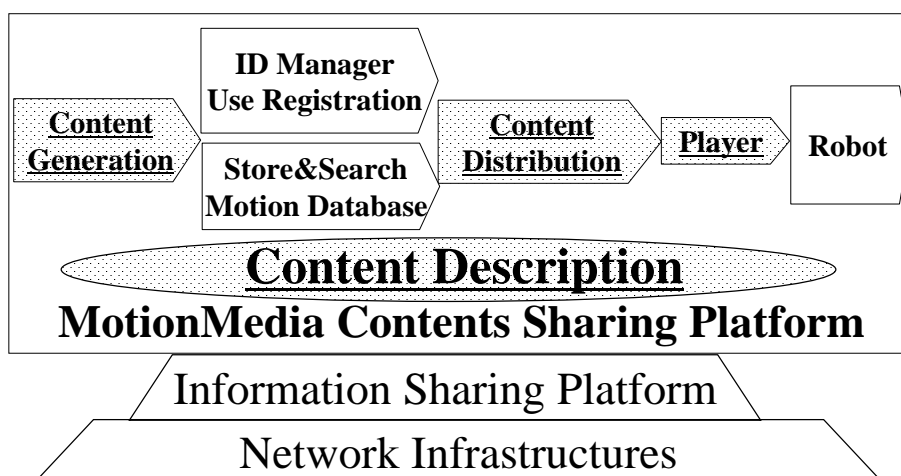


図 3.1 モーションメディアコンテンツ流通プラットフォーム

3.3 「音」メディアを用いたモーションメディアコンテンツ流通の提案とその特長

前節の(1)-a, b(2)-a, bの要件を満たすひとつの方法として、筆者は、「音コンテンツ」の中に「モーションメディアコンテンツで変調された音信号」を重畳し流通させることを提案する。その概念図を図3.2に示す。送信側では、動作信号(モーションメディアコンテンツ)で音信号を変調し、既存の音インフラストラクチャに送出される。音インフラストラクチャにおいては、モーションメディアコンテンツは音信号として扱われ、流通される。一方、受信側では、送出された信号を受信し、復調し再び元のモーションメディアコンテンツを再生する。

人間や動物が何らかの感情や意思を表現する場合、身振り・行動と言語または鳴声が組み合わされた形で発現される。また、舞踊は動作と音楽が結びついた表現である。このように「モーションメディア」と「音メディア」との結びつきは自然なものであり、音コンテンツの中に音コンテンツと関連の深い「モーションメディアコンテンツ」を重畳し一緒に流通させることは自然な発想であろう。音を用いたモーションメディアコンテンツ流通方式の特長を以下にまとめる。

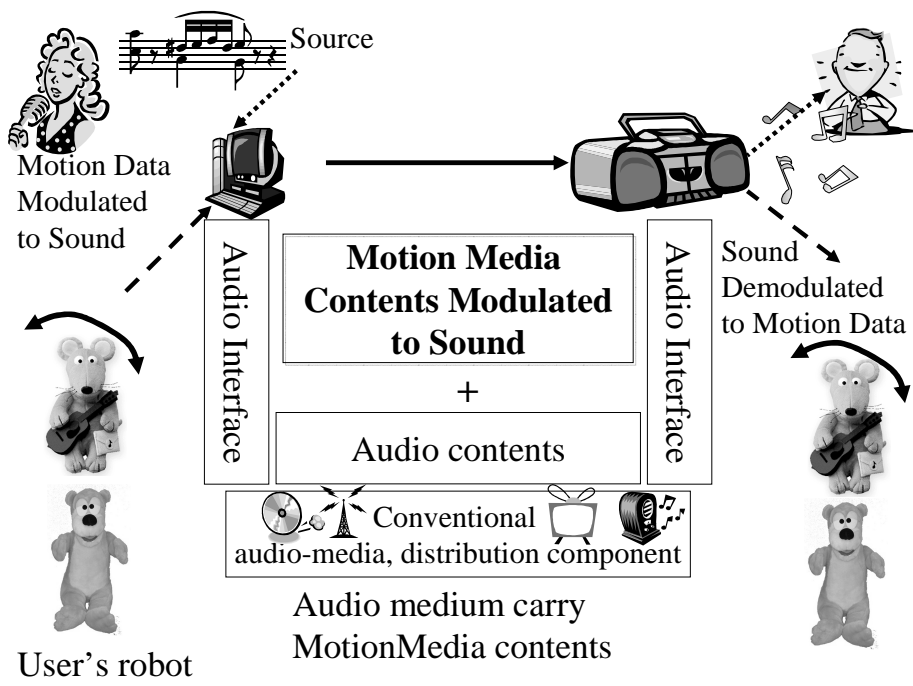


図 3.2 音によるモーションメディアコンテンツ流通プラットフォームの構想とそのコミュニケーションサービスへの応用

- 既存の音資源との親和性

豊富に存在している，音のインフラストラクチャ，機器あるいはソフトウェアなどが利用可能である．すなわちインターネット・放送・電話などの音インフラストラクチャ，CD・MD・音声関連 API・音声関連プラグインなどの機器・端末・端末ソフトウェアが利用できる．これらを利用することにより，モーションメディアコンテンツが世の中に流通・配信・蓄積していく機会が広がり，多様なコミュニケーションスタイルを構築できる．

- 音メディアとの同期性

音信号そのものにモーションメディアコンテンツが内在しているので，本質的に音コンテンツとモーションメディアコンテンツの同期をとることができる．これらは，踊りなどのように音と動きがぴったり合うことが要請されるコンテンツには重要な性質である．また面倒なメディア間同期をコンテンツ再生時に考慮しなくてよい．

- 物理インタフェースの簡便さ

アナログ音声端子・スピーカはもっともありふれたインタフェースである．この音声端子だけのシンプルな接続だけでロボットを動作させることができる．また従来 of ロボットのように，ロボット利用者に接続のわずらわしさを感じさせることはない．

- モーションメディアアプリケーション作成の容易さ

音が出る仕組みを有する環境があれば，簡単にモーションメディアを利用する音と動きの連動アプリケーションが作成できる．また音が出る既存アプリケーションの，音が生成される処理部分あるいは，アプリケーションの音ファイル（メール着信音，アプリケーション起動音など）を交換するだけで，簡単に PC 画面のテキストや画像，映像とロボットが連動するアプリケーションソフトへ変更できる．また音ファイルを再利用する要領で，モーションコンテンツの再利用なども可能となる．

このように音を用いたモーションメディアコンテンツ流通は，前節で述べたモーションメディアコンテンツ流通プラットフォーム構築の重要な望ましい要件，

すなわち(1)簡易なアプリケーション作成環境の提供(2)従来メディアとの同期を含む簡易なコンテンツ作成環境の提供(3)既存ネットワーク技術との親和性(4)配信時での従来メディアとの同期,を満足する解であるといえる。さらに,生成・配信・再生・記述をすべて音メディアで境目なく行う非常に簡便な構築方法ともいえる。そのため音に関連する技術,例えば,音声認識・合成処理,自然言語処理等と組み合わせてネットワークを介した人と人,人と機械のロボティックコミュニケーションインタフェースの実現技術としての高い可能性をもっている。本節で述べたこれらの利点と将来性が,本研究の最初のターゲットとしてロボット制御情報の伝送に取り組む理由である。

3.4 「音」メディアを用いたモーションメディアコンテンツ流通の要件と実現方式

「音」メディアを用いたモーションメディア流通の実現方式は,応用サービス,使用するロボット(自由度,制御精度など),そのサービスを実現するために使用可能な音の生成・配信・再生・記述プラットフォーム,モーションメディアコンテンツの音への変調・重畳システム,音からのモーションメディアコンテンツへの分離・復調システムなどのすべての面での特質を考慮し,これら全体の開発,すなわち「音プラットフォームの適切な選択」「変復調アルゴリズムの実装」が行わなければならない。

特にそのデータの変復調アルゴリズムについては,音楽・音声に同期させて滑らかにロボットを駆動する,幅広い流通形態(ストリーミング,ラジオ,テレビ等のさまざまな音流通メディア,さまざまなオーディオ機器の特性)に耐えうるという想定を行うと,以下のような要求条件を同時に満たすことが求められる。

- 信号を埋め込むことによって,音質が損なわれないこと
- DA変換に耐えられること
- 音量の変更にも耐えられること

- 伝送路上での符号化によるひずみ，あるいは，加法的な雑音にも十分強いこと
- 受け取った信号のみからデータが解読可能なこと，音が途切れても，データの伝送が速やかに復旧すること．
- 復号までの時間が短いこと，解読までの遅延量が一定量であること（例えば大半のポピュラー音楽をカバーするために，テンポ 180 で，8 ビートの音楽 [74] に同期して制御させるとすれば， $60 \text{ 秒} / 180 / 2 = 0.17 \text{ 秒}$ の遅延のみ許容．音声に同期させて駆動することを考慮すると，日本語母音の平均継続時間長（100 ミリ秒 [73]）以下の遅延であること）
- 音楽・音声に同期させかつロボットに対して十分滑らかな動きを持つようにビットレートを持つこと（サーボモータの代表的な仕様によれば，20 ミリ秒ごとに角度を制御する必要があり，また角度は，連続的に $-90 \sim +90$ 度まで制御可能である）

これまで指摘してきたように，電子透かし技術，音声秘匿通信技術とも未だこの上記の要件を同時に満たすアルゴリズムは提案されていない．本節では，上記の条件を考慮した，二つの音を用いたモーションメディアコンテンツ流通プラットフォーム実現方式をそのデザインコンセプトとともに示す．一つ目は，ワイヤレスでのロボット制御を志向して，原音そのものに制御信号を重ねて送出する変復調方式で構成された実現方式，そして二つ目は，よりロバストで，ひずみに強い方式を志向し，音チャンネルを音情報と制御信号に分離して送出する変復調方式で構成された実現方式である．筆者は，ロボットが Cyber 空間と実空間をつないで音信号の基盤の上に動作するものであるところから以下ではこれらのプラットフォーム実現方式をそれぞれ，CyberPerformer/Hyper，CyberPerformer/Audio と呼ぶ．なお，本章では，単一自由度のロボットのための実現方式の実装を行った．

3.4.1 CyberPerformer/Hyper

CyberPerformer/Hyper のデザインコンセプト

CyberPerformer/Hyper は、「聴衆とロボットに同時に影響を及ぼす単一のオーディオストリーム」によるモーションメディアコンテンツ流通の実現を狙ったものである。すなわち、聴衆には音楽・ガイダンス等の普通の音として聞こえながら、ロボットには制御信号が受信されるという同時性とオーディオケーブルを必要としないワイヤレス化を狙いとしている。

CyberPerformer/Hyper の向け変復調システム

人間に聞こえないように音に情報を埋め込む方式としては、従来、電波などで用いられている変復調方式を適用する方法と、著作権保護などで使用されている電子透かしを利用する方法がある。前者では、AM 変調方式・FM 変調方式・位相変調方式 [18] また、後者では、位相を変化させる方法 [11, 13]、エコー成分を付加する方法 [12]、周波数拡散による方法 [13]、マスキング現象を利用する方法 [16] などの方法が提案されている。筆者は CyberPerformer/Hyper のコンセプトの「ワイヤレスでの伝播でのロバスト性・人間に対してのロボット制御信号音の不可聴性」を満足する方法として、雑音に強く、かつ、音量の変更にも頑健な、FM 変復調方式を選択し、人間の耳の周波数特性を考慮した方法を開発した。また本方式には、ソフトウェア的に復調器が製作しやすいという利点もある。

FM 変調信号の中心周波数および、最大周波数偏移量、振幅については以下の技術的観点からバランスをとって定めた。

中心周波数：人間の耳の最小可聴域は周波数に対して一様ではなく、高周波領域において閾値は上昇すること [76] が知られている。そのためできるだけ高い周波数を使用することが望ましい。一方、比較的安価なオーディオ機器（Hi-fi 対応でないスピーカの場合）のカットオフ周波数特性¹ を考えると、できうる限り低

¹ 筆者の調査によれば、Hi-fi 対応でないスピーカは 14~16 kHz 付近にカットオフ周波数を持つものが大半であった。マイクロホンに関しては、調査したすべてのコンデンサマイクロホンが 20 kHz 付近まで感度があることが分かった

い周波数を用いることが望ましい。PC用のサウンドボードによっては、18 kHz 付近より上の帯域に対して時間的ジッターを持っている機種も複数確認された。今回は、以上の制約条件を考慮し、18 kHz 中心周波数を持つ動作情報を変調した FM 信号を使用することとした。

最大周波数偏移量：最大周波数偏移量は、ノイズに対する耐性、ソフトウェアでの復調器の作りやすさなどを考えるとできる限り大きいほうが望ましい。しかしながら、大きくすることにより、FM 信号の帯域幅が、可聴域に近づくことも考慮にいれなければならない [75]。以上の制約条件を考慮し、最大周波数偏移 100 Hz と決定した。

FM 信号の振幅：FM 信号の振幅は、人間の 18 kHz 近傍における最小可聴域、想定される使用環境下での雑音のレベル、原音の音量レベル、復調側での信号入力電圧を考慮し、実験的に定めた。16 ビット DA 変換器（最大振幅は 32767）を持つオーディオ機器の場合は、振幅 12000 とした。なおこの項目に関しては、復調器側のハードウェア、およびファームウェアとの関連は低いので、事後（信号埋め込み時）に設定可能である。

以上の条件を考慮した伝送方式のブロック図を図 3.3 に示す。変調側においては、モーションメディアコンテンツ（ -90 度から $+90$ 度の一関節の角度時系列データ）は 100 Hz の変調幅で FM 変調され、カットオフ周波数 11 kHz でローパスフィルタリング処理された音楽信号に埋め込まれる。

FM 信号の発生は次式に従って行った。

$$s_{FM}(t) = A \cos \left(\omega_c t + k_f \int f(t) dt \right) \quad (3.1)$$

ただし A ：FM 信号の振幅， ω_c ：中心周波数（ $2\pi * 18000$ ）， k_f ：定数（ $100/44100 * 2\pi * 1/90$ ，最大周波数偏移 100 Hz）， $f(t)$ （値域： $-90 \leq f(t) \leq +90$ ）：角度情報， t ：時間を表している（3.1）式によって指令値 -90 度に対応する瞬時周波数は 17900 Hz，同 $+90$ 度に対応する瞬時周波数は 18100 Hz となる。

復調側では 18 kHz 近傍の帯域のみバンドパスフィルタで抽出した後、抽出された信号に同調した矩形波を PLL（Phase Locked Loop）回路により生成する。次に、FPGA（Field Programmable Gate Array）に実装されたカウンタ回路を用

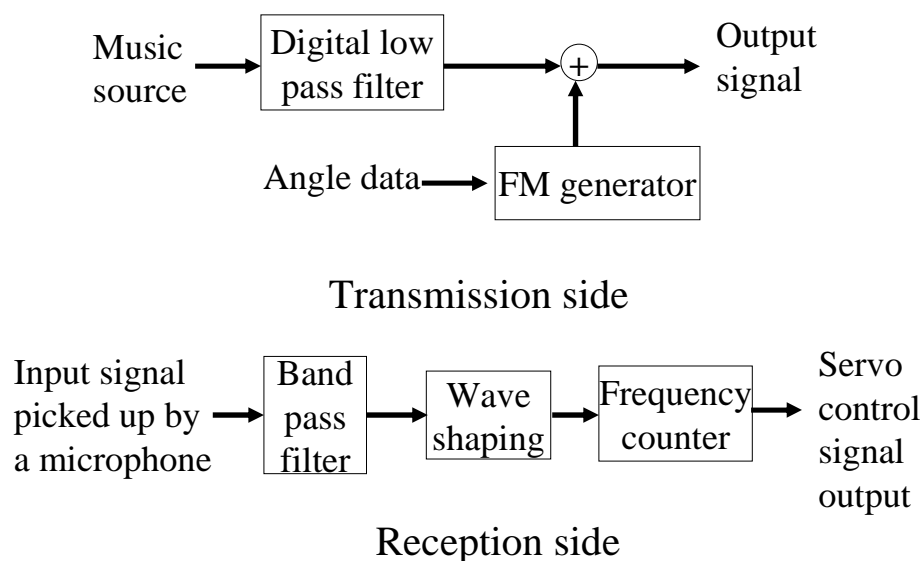


図 3.3 CyberPerformer/Hyper の処理の概要

いて、20 ミリ秒毎に、同調した矩形波の 50 個の立ち上がりを計数し、その計数に要した時間で、市販の RC (Radio Control) サーボの制御信号である繰り返し矩形波 (繰り返し周期 20 ミリ秒) のデューティ比を制御する。実現したシステムでは、20 ミリ秒間隔で 1.5 度の精度で一個のモータ制御を実現したが、FM 信号の多重化などにより、複数のモータの制御への拡張も可能である。

試作した変復調システムの評価実験

実用性の評価のために、一般的環境下での動作確認、ロバスト性、汎用性確認、音質確認実験を行った。

原理確認実験：暗騒音レベル 65 dB(A) のオフィス環境下で動作確認実験を行った。あらかじめ動作情報が重畳された音楽信号は通常の CD-R レコーダを用いて書き込まれた。その信号はスピーカ (Panasonic EAB-MPC57S) で放射され、マイクロホン (SONY ECM-MS907) でピックアップされ復調システムに入力された。スピーカはハイファイ仕様 (カットオフ周波数 30 kHz) で、マイクは高音質收音マイクである。音楽信号の平均提示レベル 75 dB(A) の場合で、マイクロホ

ン、スピーカの距離5メートルにおいても安定して動作することを確認された。

ロバスト性確認：本手法のノイズへのロバスト性を調査するために、オーディオケーブルで実験システムを音源（DAT デッキ）に接続し、あらかじめ白色雑音を重畳して作成したさまざまな SNR (Signal-to-Noise Ratio) の 20 秒間の振幅 12000 の FM 信号（指令値 0 度一定）を入力し、回転角度の制御値との最大誤差と SNR との関係を求めた。図 3.4 に結果を示す。最大の誤差は、SNR が 6[dB] であっても、15 度に収まっている。通常の子屋の背景雑音のエネルギー密度は周波数の増加とともに減少すること、またロボットの関節の「遊び」、ロボットの機械的構造にかぶせる「ぬいぐるみ」などで、微小な（10 度程度）の誤差は隠されることで実用上の問題は生じないと考えることができる。またオーディオ用コーデックへのロバスト性を調査するために動作情報が重畳された音源を CD と遜色ない品質で圧縮できるといわれている、MPEG-1 Audio Layer III (128 kbits/s) で符号化した安定した動作を行うことができなかつた。

汎用性確認：本手法の汎用性を確かめるために、6 種類のスピーカ（ハイファイスピーカ、デスクトップ PC スピーカ、安価なオーディオ用スピーカ）と 7 種類のコンデンサ型マイクロホン（ヘッドセットマイク、スタンドタイプ、ボカールピックアップ用マイク）の特性測定を行った。その結果、7 種類のマイクロホンは動作に必要な十分な 18 kHz 付近での十分な感度を得られた。しかしながら、スピーカの周波数特性は様々であり、カットオフ周波数 16 kHz 以上が、適用できるかどうかのひとつの目安となることがわかつた。

音質確認：音質に与える影響をチェックするために 20 から 40 歳代の男女、92 名の人にオフィスにて、動作信号の埋め込まれた音を視聴してもらつた。90 名の方は、高い周波数の FM 信号の存在に気がつかなかつた。また存在に気づいた人も、「かすかに聞こえる程度」との回答が寄せられ、大きく原音の品質を傷つけていないことが確認できた。

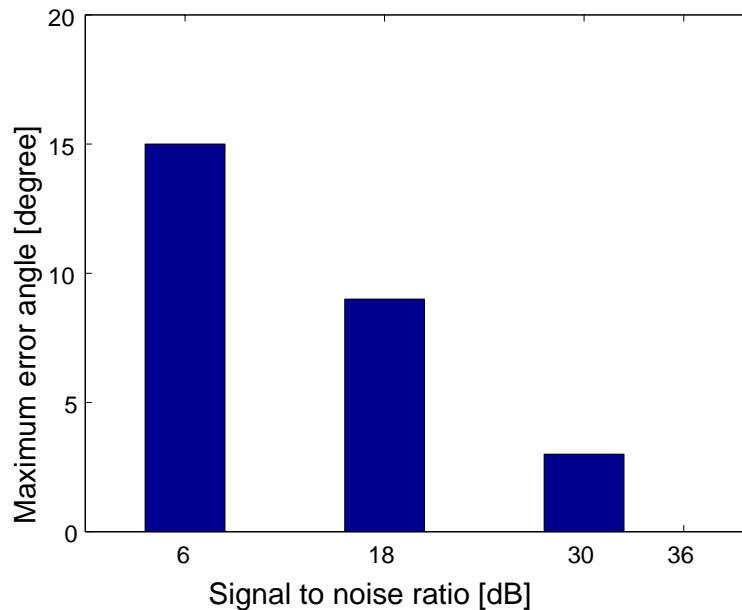


図 3.4 CyberPerformer/Hyper のノイズへのロバスト性評価

CyberPerformer/Hyper のための音の生成・配信・再生・記述プラットフォーム選択の指針

前期実験結果などからプラットフォーム設計において次のような知見・指針を得た。

1. 生成：18 kHz 近傍の FM 変調波を精密に制御し生成できる能力が必要であり，その要件をみたす手段としては，例えば 44.1 kHz サンプルング以上対応のデジタルオーディオ機器などがあげられる。
2. 配信：18 kHz 近傍の高周波成分をひずみなく伝える能力が必要であり，適切な手段としては CD 媒体などがあげられる。オーディオ圧縮などは高周波成分の成分をひずませる場合が多く適さない。また AM ラジオ放送（10 kHz 程度まで伝送），FM ラジオ放送（16 kHz 程度まで伝送）なども周波数帯域の面から適さない。

3. 再生：18 kHz 近傍の高周波成分をひずみなく再生できる能力が必要であり，ハイファイ対応オーディオ機器，サウンドボード・高い忠実度を持つ音声プレイヤー，プラグインや API などが適切である．
4. 記述：高周波成分を記述できるフォーマットとしては，Linear PCM (Pulse Code Modulation) 表現 (Microsoft Windows の .wav などでの標準形式) などが，適切である．

3.4.2 CyberPerformer/Audio

CyberPerformer/Audio のデザインコンセプト

CyberPerformer/Hyper より簡便で機構でそして，雑音や信号のひずみに対しても確実な動作，狭帯域オーディオ信号でも動作するモーションメディア流通の実現を狙ったものである．さらにコスト削減のために，動作信号のためだけに一つの音声チャンネルを占有させて使用することとした．

CyberPerformer/Audio 向け変復調システム

我々は，ステレオ音声ケーブルを使用するモーションメディア変復調システムを構築した．図 3.5 に，そのシステム詳細図を示す．本方式では，ステレオ音声の内，左チャンネルは通常の音声信号，そして右チャンネルにロボット制御信号を載せるように構成されている．左チャンネルの音声信号はそのまま増幅されスピーカから放射される．右チャンネルのロボット制御信号としては，デコーダが安価でかつ雑音に強い DTMF (Dual Tone Multi Frequency) 信号を用いた．DTMF 信号はプッシュホン電話に用いられているため高性能・安価なデコーダが入手が容易である．DTMF (16 種類) 信号をそれぞれロボットのモータの角度指令値 (図 3.6 参照) と対応付けている．

なお，受信側が比較的簡単に構築可能なほかの変調方法，例えば PSK (Phase Shift Keying) を使うことも考えられたが，サウンドボードによっては，正しく PSK 信号が出力できない場合が見られ，またステレオ分離度が著しく低い (左右

の音信号が分離されず混ざり合って出力される)機種があったこと、また、仮に利用者が機械向けの音信号を聴取した場合に異常な音だと勘違いされる恐れがあることなどから、DTMF 変調方式を採用した。

変調側では、入力された角度データを 80 ミリ秒毎にサンプリングし、その値に応じた DTMF 信号を生成する。DTMF 信号は、図 3.6 のように DTMF デコーダの規格 [77] を満足しもっとも伝送レートが高くなるように、40 ミリ秒持続し、80 ミリ秒間隔で配置する。DTMF 信号の振幅は、左チャンネルとの音声信号とのバランス、DTMF デコーダへの適正入力電圧の確保、出力機器の大振幅出力時のひずみを考慮し、実験的に 12000 とした。なお、この値は復調側のハードウェアおよびファームウェアと関連が低いため事後に設定可能である。

復調側では、受け取った DTMF 信号を DTMF デコーダ IC でデコードする。その結果は Peripheral Interface Controller (PIC) 内部で、RC サーボへの制御信号に変換される。PIC 内部ではその時点のデコード結果とその直前の DTMF 信号のデコード結果を用い、80 ミリ秒毎の角度指令値を線形補間してより細かい 20 ミリ秒毎の角度指令値を生成し、その角度指令値に基づいて RC サーボの制御信号を生成する。これによりサーボモータのなめらかな動きを実現できた(図 3.7 参照)。

CyberPerformer/Hyper の変復調システムとは対照的に、オーディオモーションメディアの実現のためにわずか 1.7 kHz 程度しかオーディオの帯域を使用しない。また制御情報は別のチャンネルに埋め込まれているため、音声信号へのひずみは本質的に存在しない。

試作した変復調システムの評価実験

実用性の評価のために、動作・汎用性確認、ロバスト性確認実験を行った。

動作・汎用性確認：開発したシステム(図 3.8)を、20 種類の PC やオーディオ機器に接続し、雑音や音量に対するシステムの頑健性を評価した。5 人のユーザがそれぞれシステムの操作をおこなった。PC やオーディオ機器の音量は、各ユーザが好ましいと考える音量に設定してもらった。テストの結果、ユーザと PC、オーディオ機器いずれの組み合わせにおいても、システムは正常に動作することが確

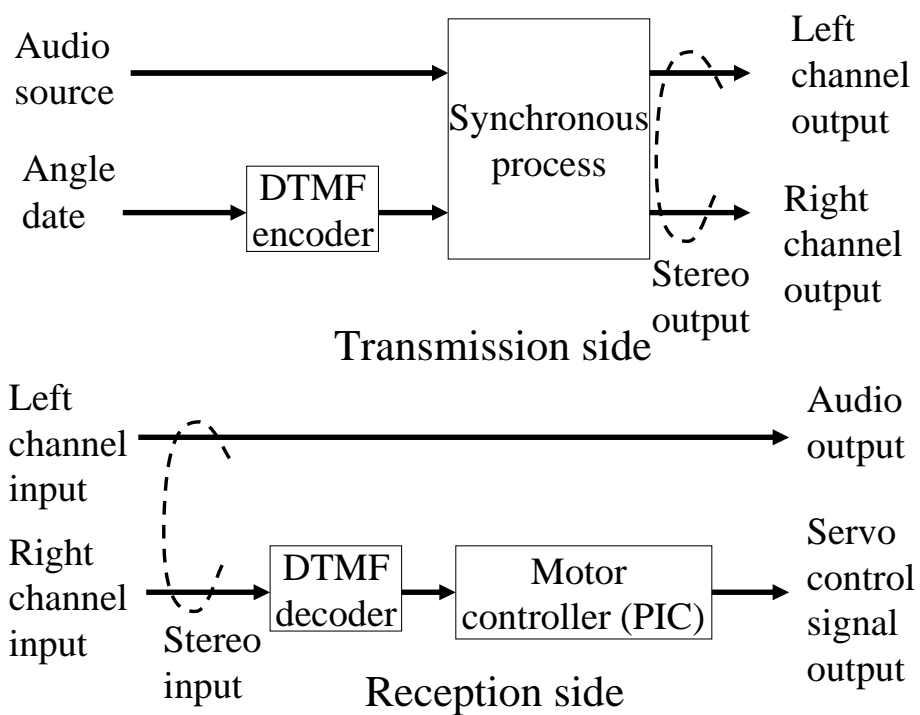


図 3.5 CyberPerformer/Audio の処理の概要

Angle [degree]	-30	-26	-22	-18	-14	-10	-6	-2
DTMF code	D	1	2	3	4	5	6	7
	+2	+6	+10	+14	+18	+22	+26	+30
	8	9	0	*	#	A	B	C

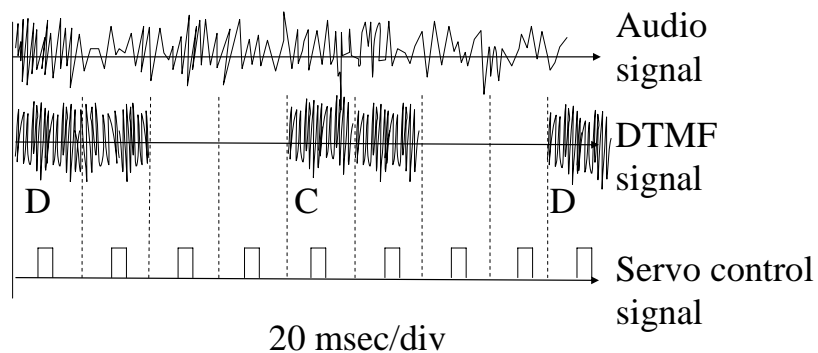


図 3.6 DTMF コードと角度情報への対応/サーボ制御信号と DTMF 信号のタイミングチャート

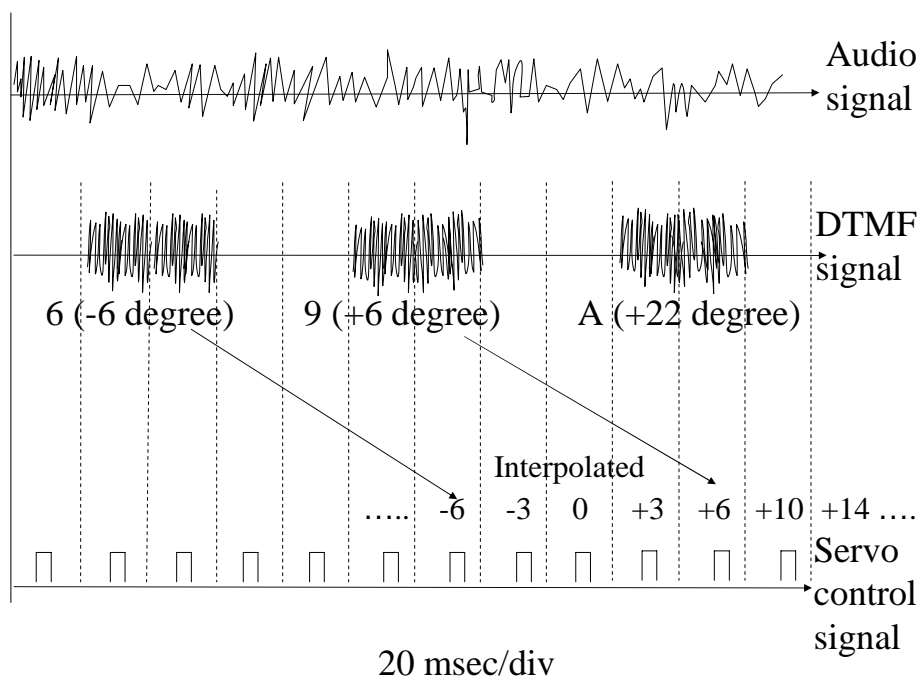


図 3.7 モータ角度情報の補間

かめられた。

ロバスト性確認：本実現例のノイズへのロバスト性を調査するために、白色雑音を重畳したさまざまな SNR の DTMF 信号（雑音重畳前の最大振幅 12000）を用いて、エラー（デコードエラー、脱落エラー）と SNR との関係を求めた。実験では音源として DAT デッキを使用し、おのおのの SNR の DTMF を 100 個復調システムに入力し、復調システムの出力を記録することで行った。図 3.9 に結果を示す。実使用環境ではあり得ない、SNR が 6 dB という悪条件であっても正常にシステムは動作することが確認できた。オーディオ出力が多くのノイズを含んでいたり、ひずみを含んでいても、正常に動作することが確かめられた。音声信号に DTMF 信号が直接重畳されたような環境でも動作可能であった。またオーディオ用コーデックへのロバスト性を確認するために、REALNETWORKS 社のコーデックを用いて動作実験を行った。左チャンネルは通常の音声、右チャンネルにはロボット制御 DTMF 信号が付加されているステレオ信号を、さまざまなビットレートで圧縮したストリームを用意し、それを再生することで動作試験を行った。その結果、たかだか 20 kbits/s のストリームの場合でも正常に動作することが確認できた。

CyberPerformer/Audio のための音の生成・配信・再生・記述プラットフォーム選択の指針

前期実験結果などからプラットフォーム設計において次のような知見・指針を得た。

1. 生成：DTMF 信号の音の高さを $1.5\%+2$ Hz ほどの誤差で正確に発信する能力が必要であり [77]、その要件をみたす手段としては、サンプリングレート 4 kHz 以上のデジタルオーディオ機器、DTMF トーン発生器、MIDI 楽器（具体的な方法は後述）などがあげられる。
2. 配信：2 kHz 近傍の周波数成分を伝える能力があれば十分であり、適切な手段としては電話・AM 放送・CD 媒体・FM 放送・ネットワーク上でのストリーミングなどがあげられる。オーディオ圧縮などで、信号が多少ひず

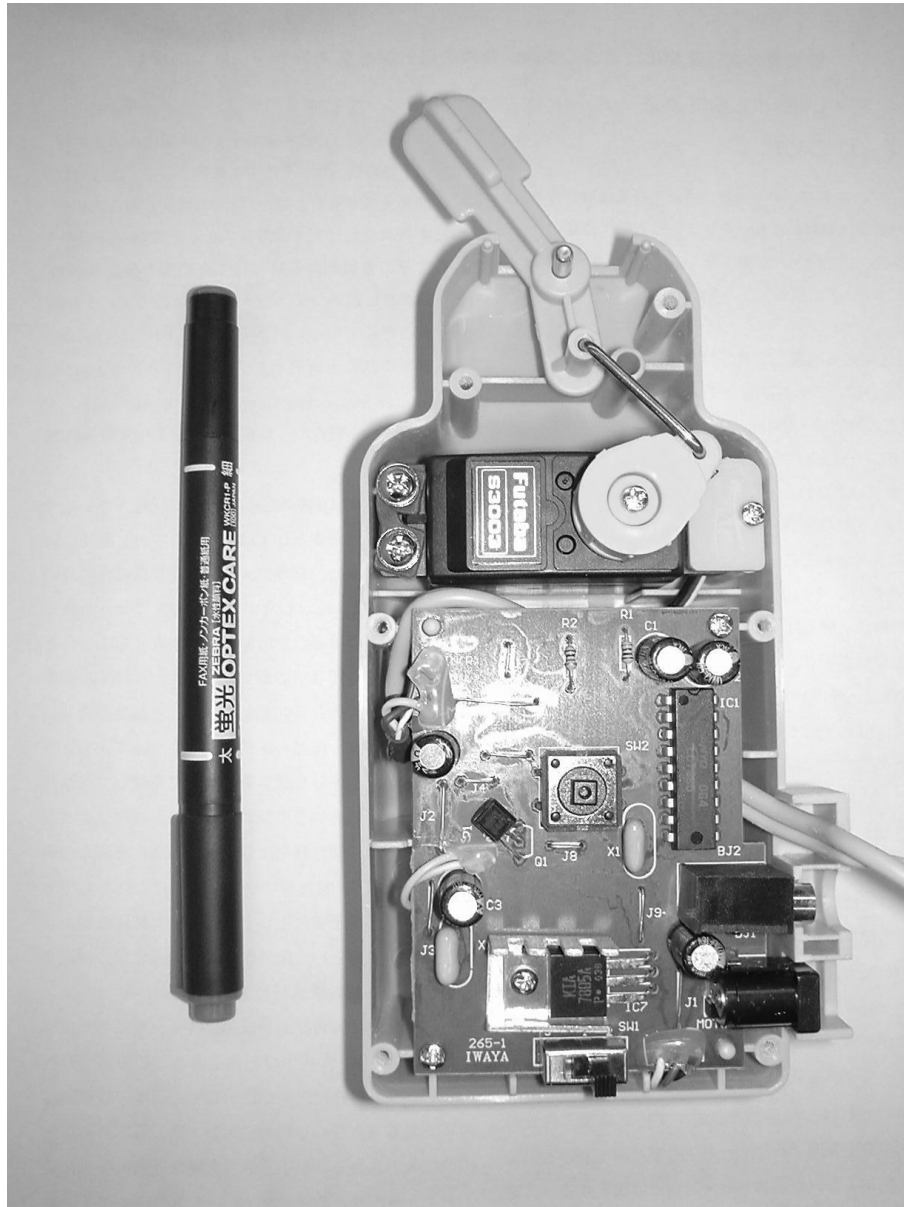


図 3.8 CyberPerformer/Audio を用いたプロトタイプコミュニケーションロボット

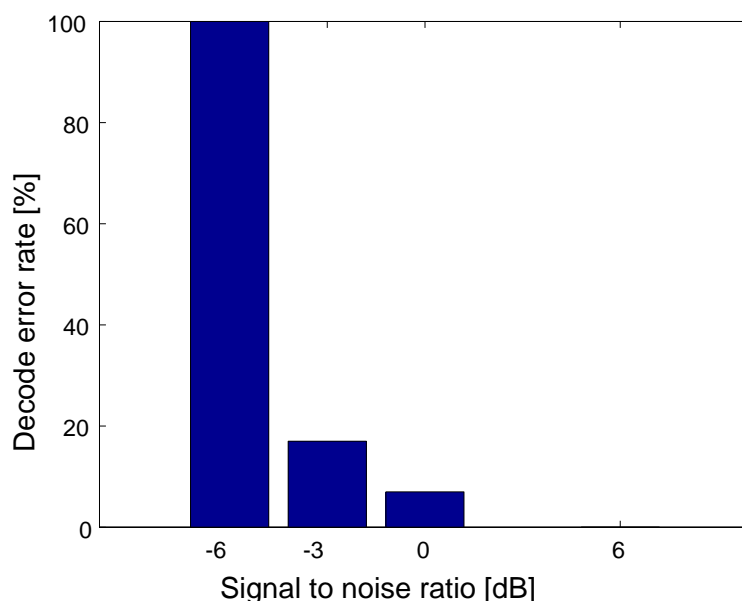


図 3.9 CyberPerformer/Audio のノイズへのロバスト性評価

むほどの場合でも十分動作を保證できるため，音声の歪が感じられるような配信媒体であっても適用可能である．

3. 再生：2 kHz 近傍の音を出力できる手段であれば，ほぼ適用可能であると考えられる．ハイファイ対応ではないオーディオ機器，携帯電話，サウンドプラグインなどでも適用可能である．
4. 記述：ADPCM などのひずみが伴う音声サンプル値の記述，MP3 などの圧縮フォーマットでも十分な性能を發揮できる．

3.4.3 各プラットフォームの比較

本節でのデザインコンセプト・提案・検討・評価実験結果を，表 3.1 に示す．「制御精度が高く，ワイヤレス，多自由度への適用が可能なプラットフォーム実現方式」CyberPerformer/Hyper と「頑健で，安価なプラットフォーム実現方式」CyberPerformer/Audio とまとめることができる．

表 3.1 二つの実現手法 CyberPerformer/Hyper・Audio の比較

	CyberPerformer /Hyper	CyberPerformer /Audio
Wireless		×
Stereo output		×
Cost for decoding	×	
Cost for sharing	×	
Noise robustness		
Multiple degree of freedom		
Control precision		×

3.4.4 簡易なモーションメディアコンテンツの生成と音への変調・重畳のための GUI 編集システム

簡易なモーションメディアコンテンツ生成・変調・重畳ツールとして，図 3.10 に示されるような，GUI タイプの動作エディタ（モーションエディタ）を開発した．この図において，上部の波形は音声信号，下部の波形はロボットのサーボモータへの角度指令値を表している．各タイミングでの指令値をマウスで指定すると対応した角度指令値の波形が表示される．角度指令値の調整は時間軸方向の移動も含め，波形の端点の Drag & Drop 操作により簡単に行える．「Make」ボタンの押下により，CyberPerformer/Audio もしくは CyberPerformer/Hyper 向けの変調処理が瞬時に行われる．その後「Play」ボタンの押下により，モーションメディアコンテンツを含んだ音声信号が再生されロボットが動作する．角度指令値の調整，信号の生成，再生の一連の過程の反復を通じて，ユーザは簡単に音に合わせたロボット動作の作りこみ（振り付け）を容易に行える．

3.4.5 本節のまとめ

モーションメディアコンテンツ流通プラットフォームの実現方式として，高周波の FM 信号を重畳する方法で構成したプラットフォーム実現方式「CyberPerformer/Hyper」および，DTMF 信号を付加する方法で構成したプラットフォーム

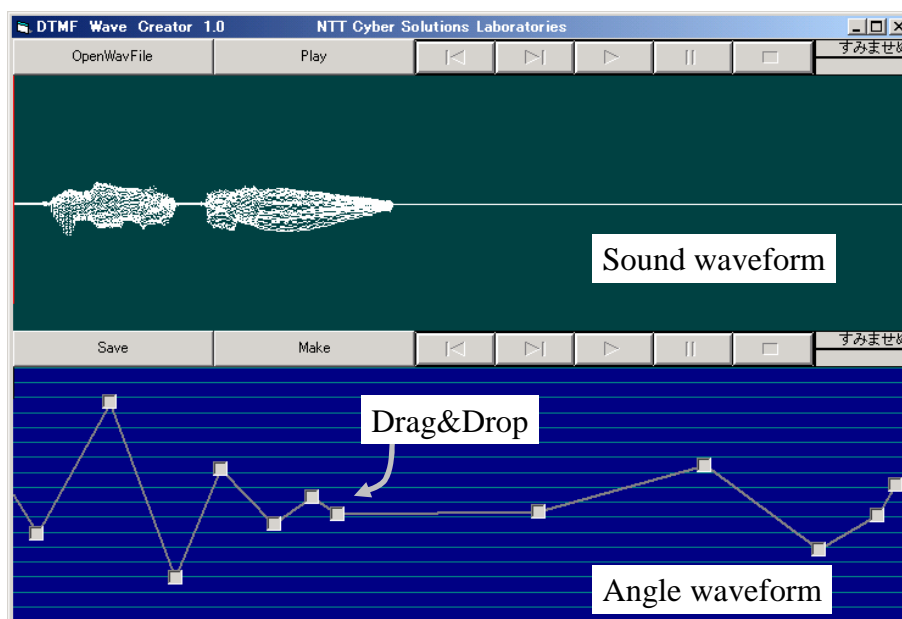


図 3.10 試作したモーションエディタ

実現方式「CyberPerformer/Audio」について提案・実装を行った．またそれらのプラットフォームの汎用性の確認，雑音へのロバスト性の評価を行い，その実用性を確認した．そして，プラットフォームの特質について，比較し整理した．各プラットフォームに合わせた，モーションメディアコンテンツ生成・変調・重畳のための GUI ツールとして，簡易な動作エディタ「MotionEditor」についても紹介した．

3.5 応用例

本節では「音」メディアを用いたモーションメディアコンテンツ流通プラットフォームのサービス例を例示することで，その利便性・実用性・汎用性を明らかにする．

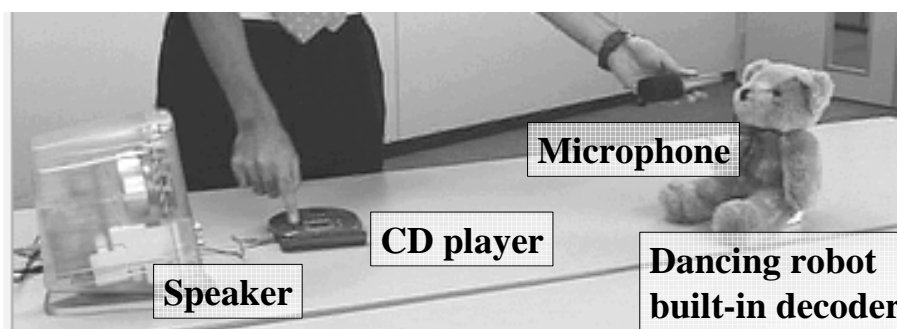


図 3.11 CyberPerformer/Hyper ダンシングロボットのパフォーマンス

3.5.1 CyberPerformer/Hyper の応用例

CyberPerformer/Hyper ダンシングロボット

CyberPerformer/Hyper を図 3.11 に示すような、玩具ロボットに応用した。この人形は頸部にサーボモータが内蔵されており、内蔵されたマイクロホンをつかって、あらかじめ CD に音楽信号とともに重畳されている FM 信号を検出することで、振り付けに従って首の位置を振ることができる。こうして聴衆は音と同期したロボットのダンスを楽しむことができる。単なる音圧に反応するだけではなく、送信者の意図に合わせた動作をする「踊りロボット」を実現している。さらにユーザはスピーカの近傍の場所にロボットを自由に置くことができる。音楽データとともにモーションメディアコンテンツを送りあう新しいコミュニケーションサービスの一例である。

3.5.2 CyberPerformer/Audio の応用例

メール読みマウスへの組み込み

CyberPerformer/Audio は、音声合成エンジン Flute[78, 79] と組み合わせられメール読みマウス 4 (NTT 東日本の ISDN 販売促進キャンペーンマスコットロボット) に搭載された。一万人のユーザに配布され、ロボットを利用した新しいコミュニ

ケーション方式として実利用された。このメール読みマウスの特徴的な機能は、ただ単にメールの文章を読み上げるだけではなく、メールのテキストに存在する喜怒哀楽に関するスマイリー（(^ o ^) (T__T) など）を抽出し、その感情にふさわしいロボットの首の動き・合成音声を表現することである。この機能により、メールの送信者は、メール読みマウスを通じてよりテキスト情報のみならず、モーションも伝えることができる。スマイリーに対応した抑揚付き音声合成音（「にっこり」(笑)「ばいばいー」(お別れ)「やったー」(歓喜)「とほほ」(落胆)「あー」(落胆)「あうち」(痛み)「ねえねえ」(呼びかけ)「すみませぬー」(謝罪)「はいー?」(疑問)「目が点」(驚き)「ぷんぷん」(怒り)「えーんえーん」(悲しみ))とそれぞれの音声に応じた首振り位置のデザイン(振り付け)により、喜怒哀楽の表現を試みた。音声合成音とその対応した動作をあらかじめ作りこんでおき、その制御信号を含む音声合成音(ステレオ信号)を読み上げテキスト中の適切なタイミングで再生することで実現したため、メール読みマウスのソフト用に新たにデバイスドライバ作成などは必要ではなかった。

モーションメディアコミュニケーションポータル

このメール読みマウス4を利用して、開発したモーションメディアコミュニケーションポータル(モーションポータルサービス)の一部を図3.12に示す。このサイトではゲーム、動画配信、占いなどのコンテンツ提供、会員間コミュニケーションサービスを提供している。社内で試行サービス実験を行った。それぞれのコンテンツ(QuickTimeムービー、Shockwave、RealVideo等)の音信号としてCyberPerformer/Audioの音信号が埋めこまれており、それぞれのコンテンツと連動してマウスが動く仕組みとなっている。ユーザは従来のウェブページでのテキスト・イメージ・動画・音コンテンツに加え、モーションメディアコンテンツを楽しむことができる。

従来のコンテンツと新しいモーションメディアコンテンツの同期再生が、既存技術とインフラストラクチャ(HTTPサーバ(Apache)、Realサーバ、FTPサーバ、ネットワーク)を用いることで実現できている。

我々はこのようなコンテンツを作成するためのいくつかの初歩的ツールをポー

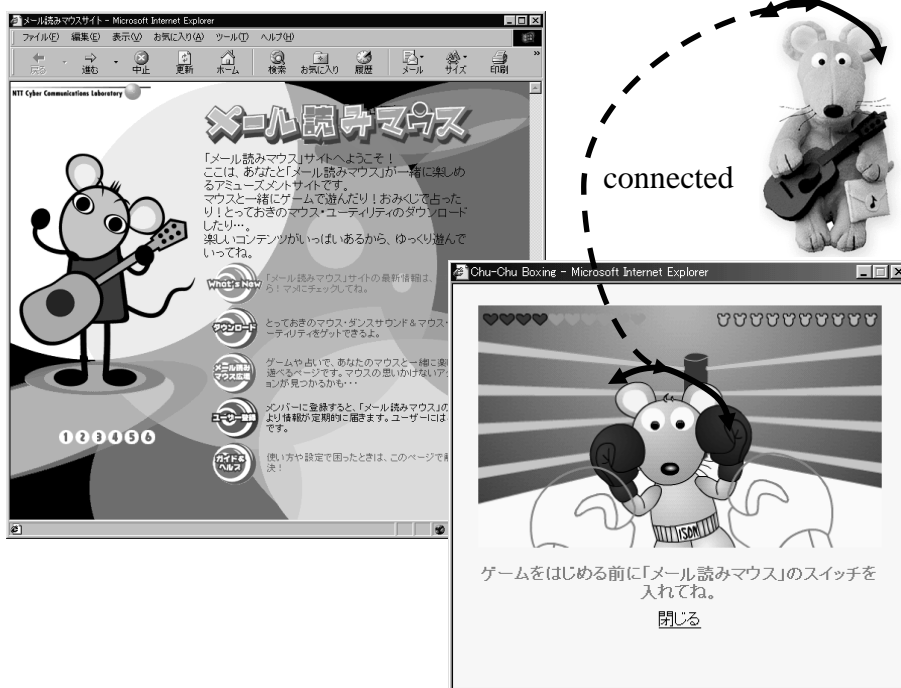


図 3.12 モーションポータルサービスの Web コンテンツの一例

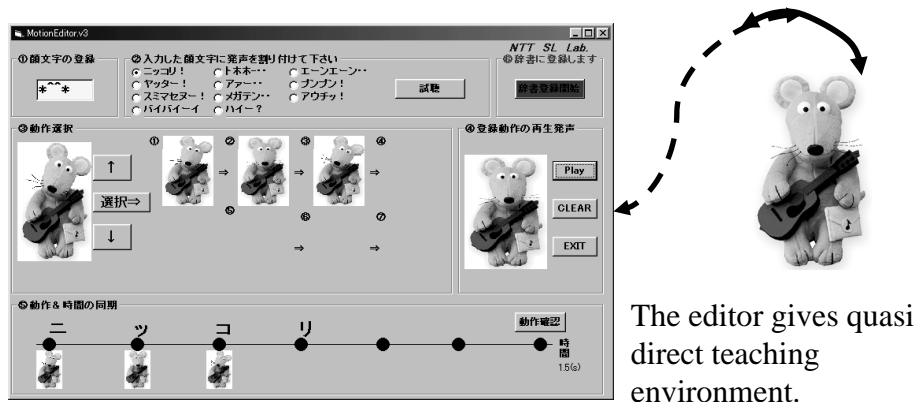


図 3.13 スマイリーに特化した，一般ユーザを対象とした擬似的にダイレクトな動作教示が可能な動作エディタの一例

タルサイトで提供した．ユーザ同士のモーションメディアコミュニケーション促進を意図して，図 3.13 は「スマイリー」に特化したモーションエディタの例である．ユーザは新しくスマイリーを登録でき，またそれに応じた振り付けをすることができる．図 3.10 に示したエディタにおいては首の角度は波形の表示によって行っているが，本エディタでは，首が傾いたメール読みマウスの画像を並べることで，より直感的に教示が行える仕組みを提供している．

図 3.14 はビデオメッセージの例である．このビデオメッセージには音声と動画のほか，音信号として CyberPerformer/Audio の音信号が埋めこまれており，送信者とともに連動して首を振る演出をしたり，あたかも掛け合い漫才をしているかのような演出を行うことができる．このようなコンテンツは図 3.10 で示したようなエディタや DTMF 発生機能のついた音声エディタやいくつかの動画編集フリーウェアを使用して容易に作成することが可能である．

Motion 着メロ

DTMF デコーダの規格に近い周波数およびタイミングを出すことで，MIDI 楽器（携帯電話）などからでもマウスを駆動すること，すなわち，CyberPerformer/Audio

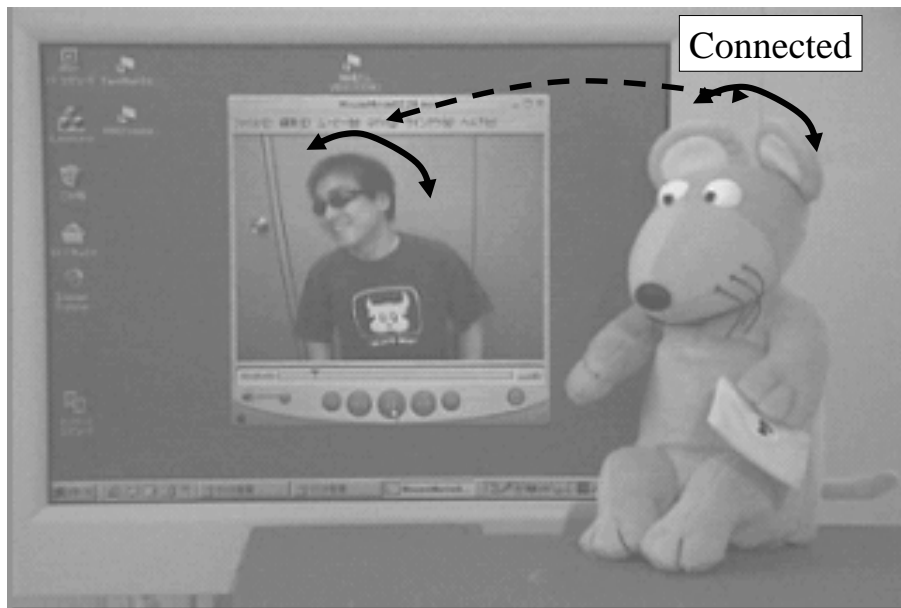


図 3.14 モーションメディアによって強調されたビデオメッセージの一例

プラットフォームの「生成・配信・再生・記述」の音のプラットフォームとして MIDI が適用可能である。表 3.2 に DTMF の 8 個の基本周波数と、音名およびその周波数との対応を示す。DTMF デコーダ動作のための許容受信周波数「基本周波数から $1.5\%+2 \text{ Hz}$ 」[77] 程度の誤差で音を生成することが可能であることがわかる。また標準的なテンポ 120 では最低、32 分音符程度の持続時間 (62.5 ミリ秒) を持つ DTMF コードに対応する 2 音を演奏すればマウスを駆動することが可能となる。筆者の実験によれば、携帯電話の着信メロディ、メール添付メロディによるロボットの駆動も可能であった。

3.6 本章のまとめ

本章では、まず筆者がこれまで提案してきたモーションメディアコンテンツ流通プラットフォームの基本的な考え方を述べ、そのプラットフォームに対する要求条件を提示した。そして「音」メディアを「モーションメディアコンテンツ流

表 3.2 DTMF の 8 個の基準周波数と音名との対応

[Hz] (Pitch Name, Actual Frequency)	1209 (D6♯, 1245)	1336 (E6, 1319)	1477 (F6♯, 1480)	1633 (G6♯, 1661)
697(F5, 698)	1	2	3	A
770(G5, 784)	4	5	6	B
852(A5, 880)	7	8	9	C
941(A5♯, 932)	*	0	♯	D

通の手段」として使うことでそれらの要求条件をシンプルに満たすことが可能であることと、その根本的なメリットについて記述した。次に、音へのモーションメディアコンテンツ重畳方式として、有線・無線の2つの実現手法を提案し、それぞれの方式に基づいたコンテンツ生成、配信、再生方式や実装例について詳述した。その後、上記メリットを生かした E-mail, Web, 携帯着信メロディなどの既存コミュニケーションサービスに適用し、本方式によるモーションメディアコンテンツ流通プラットフォームの有用性を示した。

今回は「動き」情報の一方向の送信・単一自由度についてのみ取り組んできたが、今後の課題として、双方向での動き情報のやりとりを含めた双方向多自由度モーションメディアコンテンツ流通方式の実現、またモーションメディア放送の実現、モーションメディアコンテンツの評価などが挙げられる。

第4章 「音」メディアを用いた多自由度モータコントローラへの展開

4.1 はじめに

前章の手法では，ロボットへの動き情報の出力には対応できていたものの，ロボットに入力されたセンサー情報の流通に対応できないという問題点があった．また多チャンネル・高精度のロボットのコントロールに対して，十分なビットレートが確保できないという問題点もあった．そこで本章では「音」メディアのよさを生かしつつ，上記の問題点を解決しうる USB オーディオインタフェースを使った多自由度ロボットコントローラの開発と，CyberPerformer/Hyper の発展形としての多自由度ロボットコントローラの開発について述べる．USB オーディオインタフェースは，計算機上で音声のデジタル入出力を可能にするインタフェースであり，デジタル信号，アナログ信号の相互の変換機能などを持つ．さらに，多くのオペレーティングシステム（OS）上でサポートされているため，簡便に使用することが可能である．今回の提案では USB オーディオインタフェースの音声のデジタル信号出力機能を用いてモータの制御コマンドの伝送を試みる．また CyberPerformer/Hyper の発展系では，機械情報伝送用の音信号の変復調方法の工夫により，多自由度のロボット制御を可能にする．

なお，USB を利用してロボットのモータを制御する方法は [80, 81, 82] などがあり，市販品もある [83]．これらの研究では，USB によるモータコントロールの実現に関する検討が中心であり，専用のドライバを使用しており，USB オーディオインタフェースを使用するものではない．

以下、本章では、まず、USB オーディオインタフェースを使うことのメリットについて記述する。次に、音への重畳方式やその実装例について詳述する。その後、実装例の評価を行うことで本方式の実用性を示す。また、CyberPerformer/Hyperの発展形としての音への重畳方式、その実装例について詳述する。

4.2 USB オーディオインタフェースを利用した多自由度ロボットコントローラの開発

4.2.1 USB オーディオインタフェース使用のメリット

USB オーディオインタフェースを用いたモーションメディアコンテンツ流通方式の特長を「音メディア使用」と「USB 使用」のそれぞれの観点からまとめる。

既存の音資源との親和性： 音のインフラストラクチャ、機器あるいはソフトウェアなどが利用可能である。これらを利用することにより、モーションメディアコンテンツが世の中に流通・配信・蓄積していく機会が広がる。

音メディアとの同期性： 音信号そのものにモーションメディアコンテンツが内在しているので、本質的に音コンテンツとモーションメディアコンテンツの同期をとることができる。これらは、踊りなどのように音と動きがぴったり合うことが要請されるコンテンツには重要な性質である。また面倒なメディア間同期をコンテンツ再生時に考慮しなくてよい。

モーションメディアアプリケーション作成の容易さ： 音が出る仕組みを有する環境があれば、簡単にモーションメディアを利用する音と動きの連動アプリケーションが作成できる。アプリケーションの音ファイルを交換するだけで、簡単に PC 画面のテキストや画像、映像とロボットが連動するアプリケーションソフトへ変更できる。

物理インタフェースの簡便さ： USB 端子はありふれた PC のインタフェースである。この端子だけのシンプルな接続だけでロボットを動作させることができる。ドライバも特殊なドライバではなくオペレーティングシステムに標準で搭載された音声ドライバを使用することができる。

多チャンネルへの拡張性：USBの特徴である最大デバイス数 127 台などの特徴は継承される。

このように USB オーディオインタフェースによるモーションメディアコンテンツ流通は、「音メディアのよさ」と「USB の良さ」を併せ持った方式といえる。

4.2.2 実現方式

本節では、コンテンツ流通のためのキーコンポーネントであるモータコントローラについて詳述する。

概要

本モータコントローラは、USB インタフェースを介して PC との通信を行い、音声を出力する USB digital to analog converter (USB DAC, Philips 製 UDA1321PS)、また USB DAC IC からの Inter IC Sound Bus (I2S バス) 信号を受信し、市販の RC (Radio Control) サーボの制御信号である繰り返し矩形波 (周期 20 ミリ秒) のデューティ比を制御する FPGA (Field Programmable Gate Array, ALTERA 製 EPM7160STC100-10) よりなる。システム構成を図 4.1、システム写真を図 4.2 に示す。

USB DAC IC では、音情報の含まれた USB 信号を受け取り、その信号から元の音信号を D/A 変換し、ステレオアナログオーディオ信号として出力するとともに、I2S のフォーマットで左右チャンネルの信号をデジタル信号としても出力する。FPGA では、右チャンネルの信号を抽出し後述のフォーマットに従って、カウンタ回路を利用し、RC サーボの制御信号を生成する。本試作ボードでは 8 個のサーボモータを制御可能である。-90 ~ +90 度 (パルス幅に換算して、2112 ~ 316 マイクロ秒まで) まで可動させることができる。

信号フォーマット

モータ制御コマンドは 16 ビットのオーディオ信号の振幅値に対応させた。モータ制御コマンドは、同期のためのコード (任意だがここでは AA55H とした) とそ

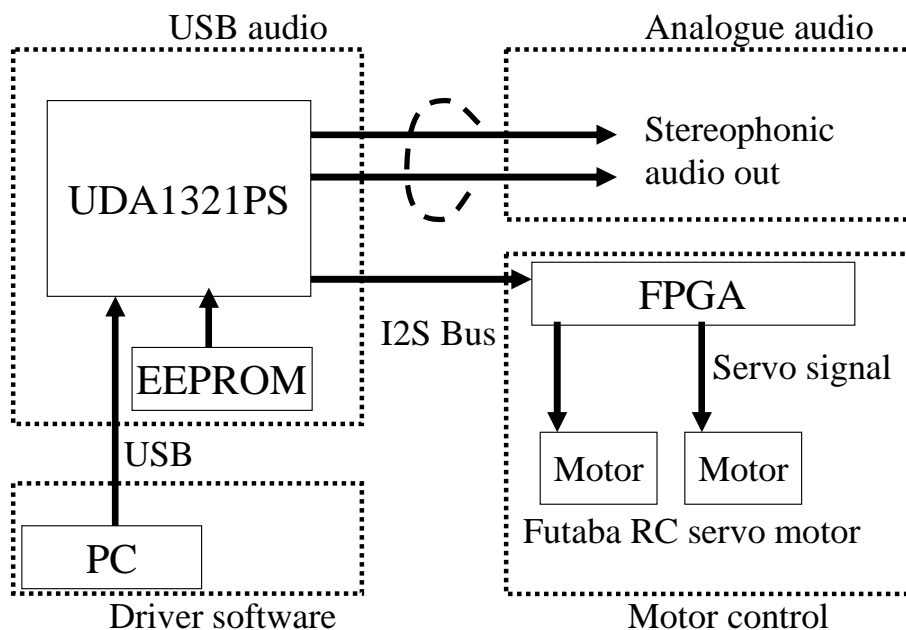


図 4.1 USB audio によるモータコントローラの概要

の直後に続くコントロール信号からなる．下記にコントロール信号のフォーマットを示す．

$$0^{15}0^{14}P^{13}1^{12}C_2^{11}C_1^{10}C_0^9D_8^8D_7^7D_6^6D_5^5D_4^4D_3^3D_2^2D_1^1D_0^0$$

P : 奇数パリティビット (13bit)

$C_2 \sim C_0$: チャンネル番号指定 0 ~ 7

$D_8 \sim D_0$: 角度指定データ 000H ~ 1C1H

4.2.3 評価

サンプリング周波数 5 ~ 48 KHz まで変化させて，8 個のモータの正常な動作を確認した．システムの頑健性を調査するために，Windows のオーディオプロパティの操作を行い，動作の確認を行った．その結果として，ボリュームコントロール（音量，バランス，トーン）の操作はモータの制御信号に影響を及ぼさないが，WAVE コントロール（音量）は常に最高値でないと制御信号を正しく伝送できな

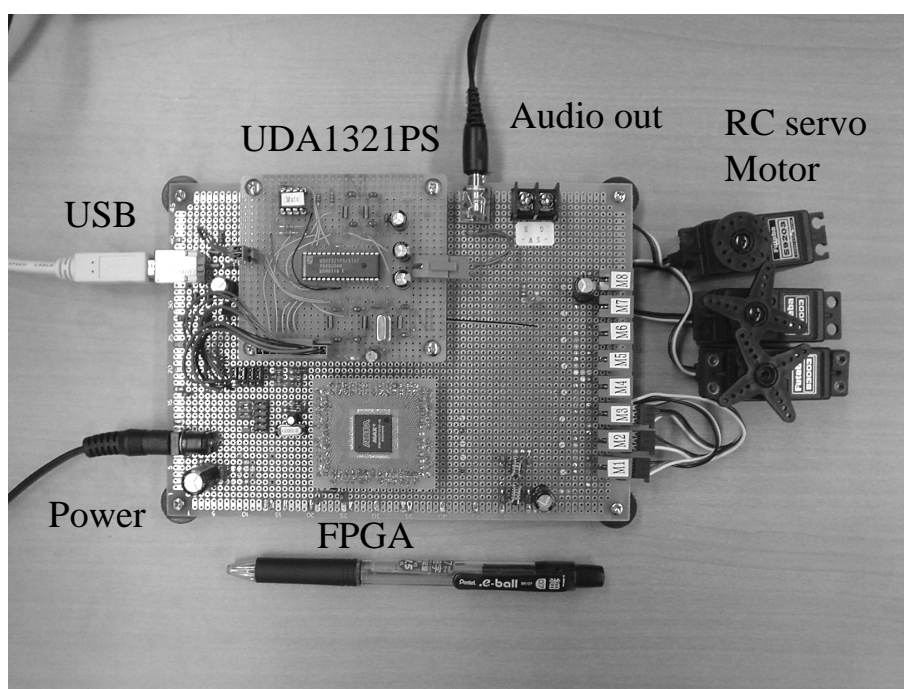


図 4.2 モータコントローラのプロトタイプボード

いことが分かった。また，WAVE コントロール（トーン）の操作はモータの制御信号に影響を及ぼさないが，WAVE コントロール（バランス）では，右に傾ける操作に関しては影響がないが，左に傾ける操作に関しては，影響があることがわかった [84]。

4.3 CyberPerformer/Hyper の発展形としての多自由度ロボットコントローラ

4.3.1 概要

本モータコントローラでは，音信号に変調されたロボット動作信号を入力とし，4 個のサーボモータの制御が可能である。実現したシステムでは，20 ミリ秒間隔で 0.8 度の精度で 4 個のモータ制御を実現している。

変調方法として，FM (Frequency Modulation) を用い，多チャンネルの信号を扱うため，各チャンネル毎の信号は時分割され多重化されている。FM 変調波の中心周波数は 18 kHz とした。

4.3.2 回路構成とその動作

本モータコントローラの回路構成を図 4.3，システム写真を図 4.4 に示す。入力した音信号から所望の FM 高周波信号を取り出すバンドパスおよび増幅を行うオペアンプ群，その信号の所定周期にかかった時間を計測するカウント用の PIC (Peripheral Interface Controller, Microchip 製 PIC16F628A I/P)，およびそのカウント結果によりサーボの制御信号である繰り返し矩形波を生成する PIC より構成されている。

本コントローラでは 18 kHz 近傍の帯域のみバンドパスフィルタで抽出した後，抽出された信号に同調した矩形波をオペアンプを飽和領域まで使用することにより生成する。次に，PIC を用いて，矩形波の 57 個の立ち上がり（約 3.2 ミリ秒に相当）を計数し，その計数に要した時間を，シリアル通信によって，サーボの制

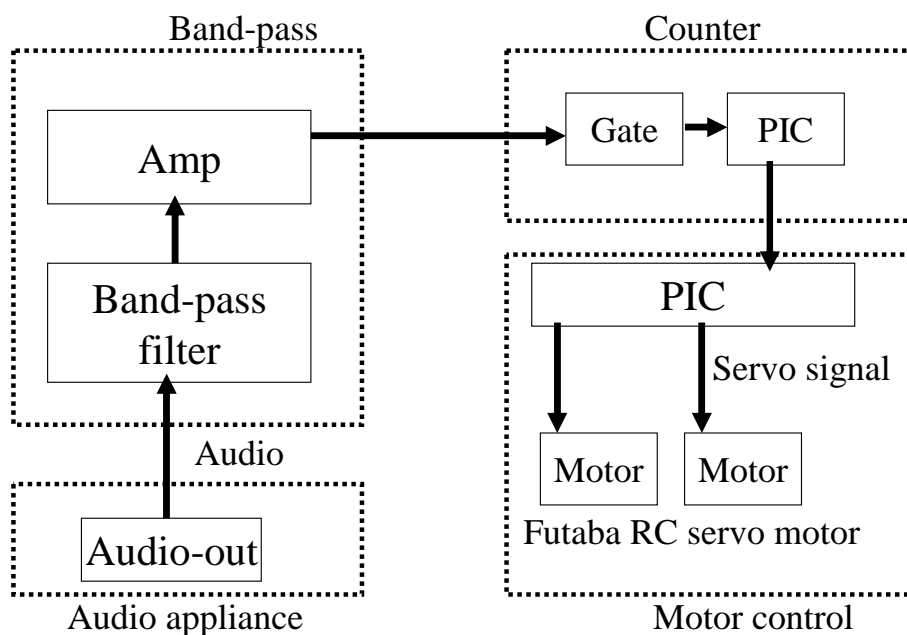


図 4.3 音による 4ch モータコントローラボードの概要

御信号生成用の PIC に送信する．生成用の PIC では受け取った時間幅に基づいて，市販の RC (Radio Control) サーボの制御信号である繰り返し矩形波（繰り返し周期 20 ミリ秒）のデューティ比を制御する．パルス幅はおおよそ 320 マイクロ秒から 2120 マイクロ秒まで，225 段階の分解能で変化させることが可能である．（180 度の稼動範囲に相当，分解能 0.8 度相当）．

モータ制御用の音信号は，200 マイクロ秒以上の無音部分と，4 ミリ秒の信号部分の繰り返しを行うように生成され，200 マイクロ秒の無音部分の検出を行うと，周波数カウント用のタイマがクリアされ，カウンター回路が信号待ち受け状態に遷移する．4 ミリ秒の信号部分の立ち上がりエッジでカウントを開始する．

4.3.3 時分割多重された FM 変調波の信号フォーマット

モータ制御用の音信号は，200 マイクロ秒以上の無音部分と，4 ミリ秒の正弦波部分の繰り返しで構成される．正弦波部分の「周波数，遷移幅」と各モータへ

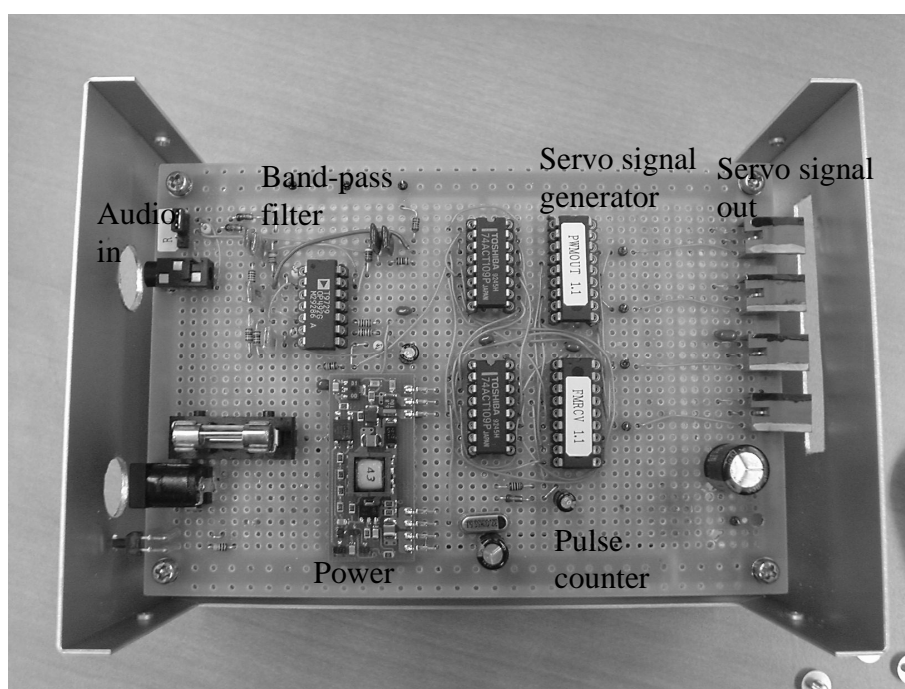


図 4.4 音による 4ch モータコントローラプロトタイプボード

表 4.1 各モータチャンネル毎の FM 変調波の周波数，遷移幅および対応するパルス幅

Frequency [Hz]	Control channel	Angle output
~ 17895	none	-
17895 ~ 17900	Motor A	minimum
17900 ~ 17940	Motor A	0.8 degrees step
17940 ~ 17945	Motor A	maximum
17945 ~ 17950	Motor B	minimum
17950 ~ 17990	Motor B	0.8 degrees step
17990 ~ 17995	Motor B	maximum
17995 ~ 18000	Motor C	minimum
18000 ~ 18040	Motor C	0.8 degrees step
18040 ~ 18045	Motor C	maximum
18045 ~ 18050	Motor D	minimum
18050 ~ 18090	Motor D	0.8 degrees step
18090 ~ 18095	Motor D	maximum
18095 ~	none	-

のサーボ信号のパルス幅の対応関係を以下の表 4.1 にまとめる。

4.3.4 評価

4 個のモータの正常な動作を確認した。システムの頑健性を調査するために、18kHz 付近の高域をフィルタによりカットした楽音と制御信号をミキシングした信号についてコントローラに入力しコントローラボードが正常に動作することを確認した。

4.4 本章のまとめ

USB オーディオインターフェースを用いて、モータコントローラの試作を行った。音信号により、8 個の RC サーボモータの制御が可能である。これにより、USB インターフェースと音信号の利点を継承したモーションメディアコンテンツ流通プ

プラットフォームの実現を可能とした。またセンサ情報を音に変調することで、同様に両方の利点を継承したシステムの実現が期待できるが今後の課題としたい。

また、CyberPerformer/Hyper の発展系として FM 高周波信号により、4 個の RC サーボモータの制御が可能なモータコントローラの試作を行った。これにより、より自由度の高いモーションメディアコンテンツ流通プラットフォームの実現を可能とした。このコントローラを使ったロボット作成およびそのコンテンツ作成を行い評価を行っていきたい。

第5章 結論

5.1 本論文のまとめ

本論文では、音楽や音声など可聴音に対し、伝達すべき情報をさりげなく重畳した「音」メディアについて論じた。人間にとっては普通の音に聞こえると同時に、機械にとっては内部に埋め込まれた情報（信号）が受信することが可能となるような、音の設計法、およびその応用法について提案をした。特に、そのような「音」メディアの著作権管理への分野での応用を目指し、MPEG 心理音響モデルに基づいたオーディオ信号への音声電子透かし、メディア複合化分野へのそれとして、ロボット制御情報の「音」メディアによる生成・伝送・再生の応用について、検討を行った。

第1章では、「音」メディア研究の重要性、応用例、また関連研究について述べ、本研究の位置づけ、背景と目的について述べた。

第2章では、Boney らの提案する MPEG 心理音響モデルを用いた電子透かし方法に注目し、評価実験を行い、いくつかの課題を明らかにし、ロバスト性を高めつつ、音質も高く保つ手法を提案した。MPEG 符号化に対して電子透かしの耐性を高めるためにいくつかの改良を行ない、また心理音響実験の知見を用いて継時マスキングを導入した。これらの改良により、MPEG 符号化に対してロバストな透かしが実現できた。次に DA/AD 変換・クロッピング攻撃に対する耐性のため、白色化相互相関法を用い、同期検出を実現した。これにより DA/AD 変換に耐性をもつ透かしが実現できた。また受けとった信号のみから透かしを検出する方法についても考察し、有効な検出方法を提案した。さらに、主観評価実験の結果を示し、透かしの埋め込んだことによる品質の劣化は少ないことを確認した。

第3章では、実現性・有用性の観点で近い将来最も有望なコミュニケーション

ロボットに焦点をあて、その利用場面を早期に広げることを狙いに、そのための方策として「音」メディアを用いる方法について検討した。筆者がこれまで提案してきたモーションメディアコンテンツ流通プラットフォームの基本的な考え方を述べ、そのプラットフォームに対する要求条件を提示した。そして「機械向けの情報」と「人向けの情報」を複合した音を使うことでそれらの要求条件をシンプルに満たすことが可能であることと、その根本的なメリットについて記述した。次に、音への重畳方式として、有線・無線の2つの実現手法を提案し、それぞれの方式に基づいたコンテンツ生成、配信、再生方式や実装例について詳述する。その後、上記メリットを生かしたネットワークサービス応用例とビジネス導入事例を紹介することで本方式の実用性を示した。

第4章では、第3章の発展系としての多自由度ロボットへの応用について述べた。本章では「音」メディアのよさを生かしつつ、十分なビットレートを確保し、多自由度のロボットの制御を可能とする、USB オーディオインタフェースを使った多自由度ロボットコントローラの開発と、CyberPerformer/Hyperの発展形としてのFM高周波信号の変調方式を提案し、その開発について述べた。本手法により、3章の音を使う利点を継承し、多自由度のロボットの制御が可能とした。

以上、本論文では、情報をさりげなく重畳した「機械」「人間」の両者に有益に作用する「音」メディアの設計とその応用について論じた。本研究により、電子透かしおよびロボット端末の応用先の拡大に寄与すると考えられる。

5.2 今後の課題

本研究により、人と機械の両者に作用を及ぼす音の設計とその応用に関して、いくつかの技術・応用例の開拓を行うことができたが、本格的な実用化および、さらなる発展を目指した場合にはいくつか課題が残されている。本技術をより発展させるための検討課題をいくつか以下に列挙する。

5.2.1 電子透かし分野に関して

構築したシステムの今後の課題としては、透かしの安全性、より多くの攻撃手法に対する耐性の検討、また事後的に定めている検出しきい値の自動設定などがある。また他手法と比べての安全性・音質の比較・評価なども残された課題である。

さらに今後の発展として、これまでに実用化された電子透かし技術の性能評価およびアルゴリズムについては明らかになっていないことも多く、これらの分野の研究をさらに飛躍させるためにも、必要な要件の詳細な定義 [85]、標準的な安全性・秘匿性・音質評価技術の構築、それらの評価用ツール（攻撃用のツールも含む）の整備についても重要な課題である [10, 86]。

5.2.2 メディアの複合化分野に関して

専用/ハイクオリティ音源チャンネルを必要としない方式の考案

3章、4章で提案した手法では、機械向け情報の伝送のために、専用の一チャンネルを割り当てる、あるいは、ハイクオリティな音源チャンネルが必要となった。これらの欠点を解消するために新たな埋め込み技術の開発を行う必要がある。具体的には、変更しても人間が感知しにくい音源による雑音成分等が多い、音声の摩擦音などへPN系列信号を埋め込む方法が考えられる。

また、ステレオチャンネルを利用した音源への埋め込み技術については、例えば、ステレオ音源の定位が低周波領域では音圧ではなく、位相差で行われていることを利用して、低周波側の音量を埋め込むデータに合わせて変化させるなどの手法が考えられる [87]。ステレオの左右のチャンネルの信号の相関が高い場合は、オールパスフィルタなどを使用し、左右のチャンネルの信号を聴覚上の検地限以下で、相対的に時間軸方向で波形をシフトさせることによって、データの埋め込みなどを行うこともできるであろう。

Lopes らの手法の拡張を行い、より人間に心地よい「音」となるように発展させることも興味深い試みである。

ロボットコントロールに対して十分なビットレートが確保できる手法の確立

開発した USB コントローラボードでは，スピーカ音量の調整に頑健でないという欠点もあり，今後は，符号化手法の改良を行い，より実用的な方法としなければならない．

センサ情報のアップロードを含めた音によるモーションメディアコンテンツ流通プラットフォームの構築

センサからの信号を音に変調し取り扱いをした例は少ないが，例えば，片寄らが，オーディオポートを介して曲げセンサ情報をコンピュータに入力する機能を実現している．この研究では，低周波信号を出力するセンシングデバイスにも対応出来るように，簡単に製作できるモジュレータ（変調）設計法を示したうえで，ソフトウェアによる復調機構を製作し，マルチメディアコンテンツ作成用ビジュアル言語 MAX で使用できるオブジェクトを用意している [88]．今後片寄らの手法を検討するとともに，音声のネットワーク伝送ソフトウェア，サーバプログラムを組み合わせ，多地点間の「音」と「動き」の通信システムの実装を行う．これにより，ダウンロードのみでなく，「音」と「動き」を「実時間」で「相互」にやりとりが可能な，新しいコミュニケーションサービスの実現を行うことも課題である．

謝辞

本論文は、筆者が奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士前期課程在学中に行った「心理音響モデルに基づくオーディオ信号への電子透かし」の研究と、日本電信電話株式会社 NTT サイバーソリューション研究所 ビジネスイノベーション SE プロジェクトにて行った「音を用いたモーションメディアコンテンツ流通方式の提案とそのネットワークコミュニケーションサービスへの応用」の研究をまとめたものです。本研究を遂行するにあたり、多くの方々に援助を頂きました。ここに感謝の意を表したいと思います。

はじめに、主指導教官であり、筆者が博士前期課程在学および博士後期課程にいたるまでご指導いただいた奈良先端科学技術大学院大学 情報科学研究科 音情報処理学講座の鹿野 清宏 教授に深く感謝いたします。鹿野先生の懇切なる継続的なご指導・ご助言により、本研究をまとめあげることができました。博士後期課程への入学後も社会人としての私のことを常に気にされ、ご配慮を頂いたこと、くじけそうな時に適切なお助言をいただいたことに対する気持ちは言葉では言い尽くすことができません。心のうち、行動のすべてを見透かしていらっしゃるような、鋭いコメントは、励みとなるとともに、本研究を遂行する上で大きな助けとなりました。そして本論文に関する有益なお助言とご指導を賜った本論文審査委員である、ロボティクス講座 小笠原 司 教授、音情報処理学講座 猿渡 洋 助教授に厚く御礼申し上げます。本研究について、忌憚のない意見を賜り、本研究をより良い方向へ進歩させることができました。

そして筆者が博士前期課程在学時に、本研究に関して有益なお助言を頂いた音情報処理学講座の中村 哲 助教授（現：ATR 音声言語研究所 所長）に感謝いたします。研究上の迷いを解消するような助言、また研究の立場・方向性について有益なお助言を賜ることができました。また研究者としての振る舞い、研究アイ

ディアの出し方についていただいたご助言は今も私の財産になっています。また研究遂行上の議論・実際のプログラミングに至るまで非常に多くの時間をさいて頂きました陸 金林 助手（現：愛知県立大学 助教授）には、ここに心より感謝いたします。研究室セミナーなどで、研究の内容についてご助言頂いた、伊勢 史郎 助手（現：京都大学 助教授）に感謝いたします。お忙しい中、博士前期課程在学時に副指導教官になっていただき、多くの貴重な御助言を賜った計算機言語学講座の関 浩之 教授に深く感謝します。

奈良先端科学技術大学院大学 音情報処理学講座では、多くの諸先輩方、学生の皆様に助けて頂きました。感謝いたします。音声合成・符号化班の先輩として、研究の議論のみならず、計算機の使い方、聴取実験の方法、発表の仕方などに至るまで丁寧にご教授頂いた、坂野 秀樹 氏（現：名城大学 講師）、東山 恵祐 氏（現：ソニー株式会社）に感謝いたします。

音によるモーションメディアコンテンツ流通方式の研究の機会を賜りました、柿崎 隆夫 プロジェクトマネジャー（現：日本電信電話株式会社 第三部門 R&D チーフプロデューサー）に感謝いたします。NTT サイバーソリューション研究所 ビジネスイノベーション SE プロジェクトでは多くの上司・先輩方に助けて頂きました。研究の当初より、十分な予算と場所・時間を確保して下さい、自由に研究できる環境を提供し、そして議論頂いた、奥平 雅士 主幹研究員（現：武蔵工業大学 教授）に深く感謝いたします。筆者が迷った時に高所から悩みを吹き飛ばすアドバイスいただいたことにより、本研究の質を高めてくださいました。メール読みマウス4の事業化・プロモーション活動にご尽力いただき、さらには、学位取得を勧め、励まして下さった、武田 英昭 主幹研究員、下倉 健一郎 主幹研究員に感謝いたします。お二方のご尽力がなければ、研究成果が世の中にできることはありませんでした。さらに、メール読みマウス4開発を共に行った、NTT 東日本研究開発センターの開発メンバー、およびNTT サイバースペース研究所メディア処理プロジェクトの開発メンバーの皆様にも深く感謝いたします。新入社員時代より丁寧に会社での生活・研究生活の指導を賜り、研究でも熱く議論させて頂いた、岩城 敏 主任研究員に深く感謝いたします。本論文の3章で主に論じた着想および研究は、岩城 主任研究員のご尽力なければ、完成できなかったと思いま

す。会社における研究での、アイデア出し、研究企画、実行、刈り取りの一連のプロセス、進め方を学ばせていただきました。入社時より同じ研究グループに属し、常日頃から指導、研究議論、サポートをしていただいた町野 保氏（現：研究主任）、北岸 郁雄氏（現：ヤフー株式会社サービス統括部）に感謝いたします。初めての社会人生活で、上司・先輩としてお三方から学んだこと、そしてメール読みマウス⁴を共に作りあげる経験をしたことが、今の会社・研究生生活の糧となっています。

筆者は、日本電信電話株式会社、東日本電信電話株式会社、NTT コミュニケーションズ株式会社に在籍しながら、奈良先端科学技術大学院大学 博士後期課程への入学・通学という機会を頂き、本研究を完成させることができました。業務の忙しい時期に入学を薦めてくださり、筆者の入学をご快諾いただいた、日本電信電話株式会社 NTT サイバーソリューション研究所 ヒューマンインタラクションプロジェクト 小林 稔 主幹研究員に感謝いたします。異動した後、通学に配慮そして支援・応援いただきました東日本電信電話株式会社、NTT コミュニケーションズ株式会社の上司・諸先輩・同僚の皆様方に感謝いたします。

最後に、ここに紹介できなかった、私の研究上の議論につきあって頂き、情報収集/提供の援助をし、精神的に私を支え、また研究に最適な心の状態を保たせ、事務手続きに至るまで、いろいろな援助して下さった学生・友人・先輩・同僚・秘書・アシスタントの皆様方に厚く感謝します。

参考文献

- [1] 中山 彰, 岩城 敏: “HyperAudio: 音を媒介にしたマンマシンインタフェースの一手法,” 情報処理学会研究報告ヒューマンインタフェース, HI-88-01, 2000.
- [2] 鹿野 清宏, 中村 哲, 伊勢 史郎: “音声・音情報のデジタル信号処理,” 昭晃堂, 1997.
- [3] 河口 英二, 野田 秀樹, 新見 道治: “画像を用いたステガノグラフィ,” 情報処理, Vol. 44, No. 3, pp. 236-241, 2003.
- [4] Liquid Audio: “Music on the Net, A Topographic Tour of the Online Music World,” 1997.
- [5] 名和 小太郎: “サイバースペースの著作権,” 中公新書 1320, 中央公論社, 1996.
- [6] 高橋 史忠: “電子透かしがマルチメディア時代を守る,” 日経エレクトロニクス, No. 683, pp. 99-162, 1997.
- [7] Ingemar Cox, Joe Kilian, Tom Leighton and Talal Shamoan: “Secure Spread Spectrum Watermarking for Multimedia,” Tech. Rep. 95-10, NEC Research Institute, 1995.
- [8] 越前 功: “電子透かし技術とその応用,” 情報処理, Vol. 47, No. 11, pp. 1243-1249, 2006.
- [9] 井上 彰, “マルチメディア時代の暗号システム - 電子透かし,” pp. 78-81, 丸山学芸図書, 1997.

- [10] 岩切宗利, 松井甲子雄: “音声・音楽を用いたインフォメーションハイディング,” 情報処理, Vol. 44, No. 3, pp. 242–247, 2003.
- [11] Yasemin Yardimci, A. Enis Cetin and Rashid Ansari: “Data Hiding in Speech Using Phase Coding,” ESCA. Eurospeech97, pp. 1679–1682, 1997.
- [12] Daniel Gruhl, Anthony Lu and Walter Bender: ”Echo Hiding,” Lecture Notes in Computer Science, Vol. 1174, Information Hiding, pp. 295–315, Springer, 1996.
- [13] Walter Bender, Danien Gruhl, Norishige Morimoto and Anthony Lu: “Techniques for data hiding,” IBM SYSTEMS JOURNAL, Vol. 35, NOS 3&4, pp. 313–336, 1996.
- [14] 岩切 宗利, 松井 甲子雄: “音楽ソフトへの電子透かしの一方式,” SCIS98-8.2.C, 1998.
- [15] 松井 甲子雄, 中村 康弘, ナタウット サムパイブーン: “音声通信への文字情報の埋め込み,” 第18回情報理論とその応用シンポジウム, pp. 389–392, 1995.
- [16] Laurence Boney, Ahmed H. Tewfik and Khaled N. Hamdy: “Digital Watermarks for Audio Signals,” IEEE Intl. Conf. on Multimedia Computing and Systems, Hiroshima, pp. 473–480, 1996.
- [17] Laurence Boney, Ahmed H. Tewfik and Khaled N. Hamdy: “Digital Watermarks for Audio Signals,” European Signal Proc. Conf. , Trieste, Italy, September, 1996.
- [18] Vadim Gerasimov and Walter Bender: “Things that talk: Using sound for device-to-device and device-to-human communication,” IBM SYSTEMS JOURNAL, Vol. 39, NOS 3&4, pp. 530–546, 2000.
- [19] Cristina Videira Lopes and Pedro M. Q. Aguiar: “Acoustic modems for ubiquitous computing,” IEEE Pervasive Computing, Vol. 2 , No. 3 , pp. 62–71, 2003.

- [20] 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—,” 電子情報通信学会応用音響研究会報告, EA2006-24, pp. 25-29, 2006.
- [21] Hideaki Kuzuoka and Saul Greenberg: “Mediating Awareness and Communication through Digital but Physical Surrogates,” Proc. CHI'99 Extended Abstracts, pp. 11-12, 1999.
- [22] Hideaki Kuzuoka, Shinya Oyama, Keiichi Yamazaki, Kenji Suzuki and Mamoru Mitsuishi: “GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions,” Proc. of CSCW2000, pp. 155-162, 2000.
- [23] Eric Paulos and John Canny: “PRoP: Personal Roving Presence,” Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 296-303, 1999.
- [24] Dairoku Sekiguchi, Masahiko Inami and Susumu Tachi: “RobotPHONE: RUI for Interpersonal Communication,” CHI2001 Extended Abstracts, pp. 277-278, 2001.
- [25] Scott Brave and Andrew Dahley: “inTouch: A Medium for Haptic Interpersonal Communication,” Extended Abstracts of CHI97, pp. 115-116, 1997.
- [26] 米澤 朋子, ブライアン クラークソン, 間瀬 健二: “文脈適応型音楽生成をともなうぬいぐるみインタラクション,” 情報処理学会論文誌, Vol. 43 No. 8, pp. 2810-2820, 2002.
- [27] Tomio Watanabe, Masashi Okubo and Hiroki Ogawa: “A Speech Driven Embodied Interaction Robots System for Human Communication Support,” Proc. of 2000 IEEE International Conference on Systems, Man, and Cybernetics (SMC.2000), pp. 852-857, 2000.
- [28] Michael Kaminsky, Paul Dourish, W. Keith Edwards, Anthony LaMarca,

Michael Salisbury and Ian Smith: “SWEETPEA: Software Tools for Programmable Embodied Agents,” CHI99, pp. 144–151, 1999.

- [29] 富士通: “CHOCOA Net 報道発表,” <http://www.labs.fujitsu.com/News/1999/Nov/15.html>, 1999.
- [30] 総務省 ネットワーク・ロボット技術に関する調査研究会: “ネットワーク・ロボット実現に向けた取り組み ~ユビキタスネットワークとロボットのフラッグシップテクノロジーの集結による日本発新ITの創出~,” http://www.soumu.go.jp/joho_tsusin/policyreports/chousa/netrobot/pdf/030711_2b.pdf, 2003.
- [31] 岩城 敏, 町野 保, 北岸 郁雄, 中山 彰, 奥平 雅士: “ネットワークコンテンツとしてのロボットモーション-モーションメディアコンテンツ流通プラットフォーム構想-,” 第7回ロボティクスシンポジウム予稿集, pp. 297–304, 2002.
- [32] Ikuo Kitagishi, Tamotsu Machino, Akira Nakayama, Satoshi Iwaki and Masashi Okudaira: “Development of Motion Data Description Language for Robots Based on eXtensible Markup Language – Realization of Better Understanding and Communication via Networks–,” *Journal of Robotics and Mechatronics*, Vol. 14, No. 5, pp. 471–478, 2002.
- [33] 平岩 明, 篠沢 一彦, 内田 典佳, 曾根原 登: “モーションメディアのコンセプト,” *情報処理学会研究会報告, 情報メディア*, Vol. 14, No. 7, pp. 39–46, 1994.
- [34] 藤井 輝夫: “情報ネットワークにおける物理メディアとしてのロボットシステム,” *日本ロボット学会誌*, Vol. 17, No. 4, pp. 462–466, 1999.
- [35] 安西 祐一郎: “実世界指向メディアとしての次世代ロボット,” *日本ロボット学会誌*, Vol. 16, No. 1, pp. 15–18, 1998.
- [36] 岩田 洋夫: “メディアとしてのロボット,” 第19回日本ロボット学会学術講演会講演論文集, pp. 13–14, 2001.

- [37] 妻木 勇一: “メディアとしてのウェアラブルロボット,” 日本ロボット学会誌, Vol. 20, No. 8, pp. 819–821, 2002.
- [38] 十時 伸, 安川 祐介: “ロボット&パソコン連携システムと応用例,” 情報処理学会研究報告, 2001–HI–95, pp. 139–146, 2001.
- [39] 社団法人 日本ロボット工業会: “RoboLink プロトコル仕様書 パブリックドラフト 1.1,” <http://www.osl.fujitsu.com/osl/contents/RoboLink/>, 2003.
- [40] 広瀬 健志郎, 佐竹 聡, 川島 英之, 今井 倫太: “コミュニケーションロボットによる WWW 上のコンテンツ閲覧へ向けたモーション記述言語の設計と実装,” 情報処理学会第 66 回全国大会予稿集, vol. 3, pp. 151–152, 2004.
- [41] 岡田 慧, 稲葉 雅幸, 井上 博允: “ビヘイビアコンテンツのためのエンドユーザ指向ヒューマノイドシミュレーション環境 –Web ブラウザプラグインによるプロトタイプシステムの開発–,” 日本機械学会ロボティクス・メカトロニクス講演会’03 講演論文集, pp. 2P1–1F–F8, 2003.
- [42] SONY Corp.: “AIBO Master Studio,” <http://www.jp.aibo.com/ams/>, 2002.
- [43] The LEGO Group: “LEGO MINDSTORMS,” <http://mindstorms.lego.com/japan/default.asp>, 2000.
- [44] BANDAI Co.,LTD: “ロボットワークス,” <http://www.roboken.channel.or.jp/borg/program/index.html>, 1999.
- [45] BANDAI Co.,LTD: “プログラミングロボット BN-1,” <http://www.bn-1.channel.or.jp/index02.html>, 2000.
- [46] 国際電気通信基礎技術研究所: “ロボット動作編集ソフトウェア「シナリオエディタ」,” http://www.atr.co.jp/html/topics/press_031028_j.html, 2003.

- [47] Akira Nakayama, Tamotsu Machino, Ikuo Kitagishi, Satoshi Iwaki and Masashi Okudaira: “Rich Communication with Audio-Controlled Network Robot,” Proceedings of 11th IEEE Intl. Workshop on Robot and Human Interactive Communication, pp. 548–553, 2002.
- [48] Raymond Steele and D. Vitello: “Simultaneous transmission of speech and data using code-breaking techniques,” The Bell system Technical Journal, Vol. 60, No. 9, pp. 2081–2105, 1981.
- [49] Wai C. Wong, Raymond Steele and Costas S. Xydeas: “Transmitting data on the phase of speech signals,” The Bell system Technical Journal, Vol. 61, No. 10, pp. 2947–2970, 1982.
- [50] 松井 甲子雄: “画像深層暗号 –手法と応用–,” pp. 1–17, 森北出版, 1993.
- [51] 岩切 宗利, 松井 甲子雄: “適応差分 PCM 符号化における音声符号へのテキスト情報の埋込み,” 情報処理学会論文誌, Vol. 38, No. 10, pp. 2053–2061, 1997.
- [52] 岩切 宗利, 松井 甲子雄: “共役構造代数符号励振線形予測による音声符号へのテキスト情報の埋込み,” 情報処理学会論文誌, Vol. 39, No. 9, pp. 2623–2630, 1998.
- [53] 井上 大介, 松本 勉: “スタンダード MIDI ファイルステガノグラフィとその能力,” 情報処理学会論文誌, Vol. 43, No. 8, pp. 2489–2501, 2002.
- [54] Nedeljko Cvejic and Tapio Seppänen: “Reduced distortion bit-modification for LSB audio steganography,” Proceedings of ICSP 2004, Vol. 3, pp. 2318–2321, 2004.
- [55] Michah K. Johnson, Siwei Lyu and Hany Farid: “Steganalysis of Recorded Speech,” Proceedings of the SPIE, vol. 5681, pp. 664–672, 2005.
- [56] 柏木 潤: “M 系列とその応用,” 電子情報通信学会誌, Vol. 76, No. 12, pp. 1337–1340, 1993.

- [57] William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery: “NUMERICAL RECIPES in C,” Cambridge University Press, 1988. (邦訳) 丹慶 勝市, 奥村 晴彦, 佐藤 俊郎, 小林 誠 訳: “ニューメリカル レシピ・イン・シー,” 技術評論社, 1993.
- [58] Peter Noll: ”MPEG Digital Audio Coding,” IEEE Signal Processing Magazine, September, pp. 59–81, 1997.
- [59] 古井 貞熙: “音声信号処理,” pp. 133–135, 森北出版, 1998.
- [60] 平原 達也: “聴覚実験に用いられるヘッドホンの物理特性,” 日本音響学会誌, 53 巻, 10 号, pp. 798–806, 1997.
- [61] Maurizio Omologo and Piergiorgio Svaizer: “Acoustic Source Location in Noisy and Reverberant Environment Using CSP Analysis,” Proc. ICASSP96, pp. 921–924, 1996.
- [62] Brian C. J. Moore: “An Introduction to the Psychology of Hearing,” Academic Press, pp. 312–314, 1997.
- [63] Ohgushi Kengo: “Recent Research on Hearing in Japan,” J. Acoust Soc. Jpn. (E), vol. 5, pp. 127–133, 1984.
- [64] Alan V. Oppenheim and Ronald W. Schaffer: “Digital Signal Processing,” New Jersey, 1975. (邦訳) 伊達 玄 訳: “デジタル信号処理 (上, 下),” コロナ社, 1978.
- [65] 宮坂 栄一: “急激な立ち上がり, 立ち下がりを有する正弦波信号に対する聴覚マスキングの時空間特性,” 日本音響学会誌, 39 巻, 9 号, pp. 614–623, 1983.
- [66] Tetsuya Hirahara: “Internal Speech Spectrum Representation by Spatio-temporal Masking Pattern,” J. Acoust Soc. Jpn. (E), vol. 12, pp. 57–68, 1991.

- [67] 小原 和昭, 相川 清明, 河原 英紀: “時間周波数マスキング特性を模擬した聴覚フィルタモデルによる音声認識,” 日本音響学会誌, 50 巻, 5 号, pp. 345-351, 1994.
- [68] ISO/IEC JTC1/SC29: “Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5Mbit/s - IS 11172 (Part3, Audio),” 1992.
- [69] 岩上 直樹, 守谷 健弘, 三樹 聡: “周波数領域重み付けインタリーブベクトル量子化 (Twin VQ) によるオーディオ符号化,” 日本音響学会講演論文集, 1-P-1, 1994.
- [70] David Kirby and Kaoru Watanabe: “Overview of the Report on the Formal Subjective Listening Tests of MPEG-2 NBC Multichannel Audio Coding,” ISO/IEC JTC1/SC29/WG11, N1420, 1996.
- [71] 森 敏哲, 吉田 寿夫 編著: “心理学のためのデータ解析テクニカルブック,” 北大路書房, 1990.
- [72] 辻本 廉: “音響機器の品質評価,” 日本音響学会誌, 42 巻, 10 号, pp. 816-821, 1986.
- [73] 中川 聖一, 東倉 洋一, 鹿野 清宏: “音声・聴覚と神経回路網モデル,” p.18, オーム社, 1990.
- [74] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: “RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース,” 情報処理学会論文誌, Vol. 45, No. 3, pp. 728-738, 2004.
- [75] B.P. Lathi (原著), 山中 惣之助 (翻訳), 宇佐美 興一 (翻訳): “通信方式 情報伝送の基礎,” 朝倉書店, pp. 149-160, 1995.
- [76] Brian C. J. Moore: “An Introduction to the Psychology of Hearing,” pp. 50-52, Academic Press, 1997.

- [77] 三洋半導体ニュース: “LC7385, 7385M CMOS LSI DTMF レシーバデータシート,” No. 3003C.
- [78] Osamu Mizuno and Shin'ya Nakajima: “Synthetic Speech/Sound Control Language: MSCL,” Third ESCA/COCOSDA Workshop on Speech Synthesis, pp. 21–26, 1998.
- [79] Kazuo Hakoda, Tomohisa Hirokawa, Hajime Tsukada, Yuki Yoshida and Hideyuki Mizuno: “Japanese text-to-speech software based on waveform concatenation method,” in AVIOS '95, pp. 65–72, 1995.
- [80] 徳山 陽人, 精廬 幹人, 橋本 周司: “USB インタフェース搭載の小型モータコントローラの開発,” 第 18 回日本ロボット学会学術講演会予稿集, pp. 119–120, 2000.
- [81] 徳山 陽人, 精廬 幹人, 橋本 周司: “USB インタフェースによるモータコントローラのカスタマイズとその応用,” インタラクシオン 2001 論文集, pp. 45–46, 2001.
- [82] 光永 法明, 細田 耕, 浅田 稔: “自立ロボットシステムのための USB 接続による小型モータコントローラの開発,” 第 19 回日本ロボット学会学術講演会予稿集, pp. 1189–1190, 2001.
- [83] Phidgets Inc.: “4-Motor PhidgetServo,” <http://www.phidgets.com/>
- [84] 池田 亜希, 金田 豊, 阪内 澄宇: “音響計測を目的とした PC オーディオ系の基本検討,” 日本音響学会秋季大会講演論文集, pp. 609–610, 2003.
- [85] International Federation of the Phonographic Industry: “Request for Proposals – Embedded Signalling Systems Issue 1.0.” 54 Regent Street, London W1R 5PJ, June 1997.
- [86] Fabien A. P. Petitcolas, Ross J. Anderson and Markus G. Kuhn: “Attacks on Copyright Marking Systems,” Second workshop on information hiding, Vol. 1525, Lecture Notes in Computer Science, pp. 218–238, 1998.

- [87] 中山 彰, 岩城 敏: “情報伝達方法および情報伝達装置,” 特開 2001-186093.
- [88] 片寄 晴弘: “パフォーマンスのためのマルチメディア情報利用,” マルチメディア情報学 10 巻, 自己の表現, pp. 67-113, 岩波書店, 2000.

付録

A.1 セグメンタルSNR

セグメンタルSNR (SNR_{seg}) は、短区間ごとに測定したSNRのdB値を長時間の音声区間で平均したもので表す、時間領域のひずみの評価法である。

$$\text{SNR} = 10 \log_{10} \frac{\sum_n S^2(n)}{\sum_n (S(n) - \hat{S}(n))^2} \text{ [dB]} \quad (\text{A.1})$$

$$\text{SNR}_{\text{seg}} = \frac{\sum_{j=1}^N \text{SNR}_j}{N} \text{ [dB]} \quad (\text{A.2})$$

ただし、 $S(n)$ は原音を、 $\hat{S}(n)$ は処理音を、 N は全フレーム数を表す。

A.2 ケプストラム合成によるマスキングカーブの近似法

以下に示す方法で、マスキングカーブを近似したインパルスレスポンス（時間軸波形）を求めた。計算方法の概念を図A.1に示す。

1. マスキングカーブを求めて、左右対称の形にする（周波数平面の信号表現）
2. 逆フーリエ変換を求める（時間軸平面の波形が得られるいわゆるケプストラムに相当）
3. 最小位相系を構成するようにケプストラムの係数を加工する。

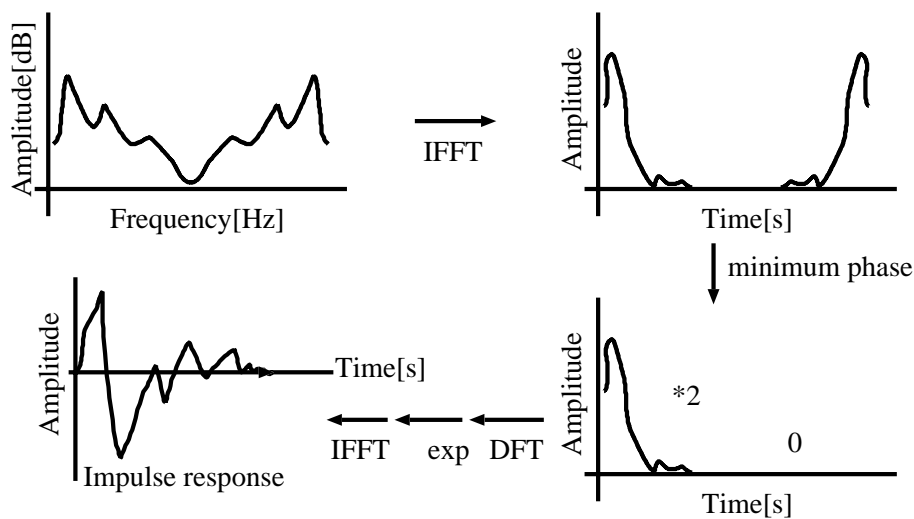


図 A.1 ケプストラム合成による最小位相フィルタの構成法

4. フーリエ変換を行う（振幅がログ表記の周波数平面の信号表現）
5. 指数を計算し，振幅を実数表記にする．
6. 逆フーリエ変換を求める．
7. 時間軸のマスキングカーブを近似したインパルスレスポンスが求められる．

A.3 白色化相互相関の計算方法

オリジナルの信号を $s(n)$ ，受け取った信号を $r(n)$ とすると，白色化相互相関関数 $CSP(k)$ は，以下の数式で与えられる．もっとも大きな白色化相互相関係数をもつ k が，2 信号間の遅延量 τ を与える．準定常波形の繰り返しのような信号を与えた場合でも，理想的には CSP としてインパルス関数が得られる．電子透かしの検出の際には，遅延量 τ を計算し，受信信号 $r(n)$ から，遅延を補正した $s(n)$ 信号の差を求めることで，透かし信号 $w(n)$ を得た．

$$\text{CSP}(k) = \text{DFT}^{-1} \left(\frac{\text{DFT}(s(n))\text{DFT}(r(n))^*}{|\text{DFT}(s(n))| |\text{DFT}(r(n))^*|} \right) \quad (\text{A.3})$$

$$\tau = \underset{k}{\text{argmax}}(\text{CSP}(k)) \quad (\text{A.4})$$

A.4 電子透かし/ネットワーク・デジタル放送と著作権に関連した新聞報道

1. 朝日新聞, 1998年4月22日朝刊, “著作権料設定でルール作り難航 インターネットの音楽, 一曲いくら”
2. 朝日新聞, 1998年6月23日朝刊, “デジタル音楽情報に「透かし」”
3. 読売新聞, 1998年7月29日朝刊, “CS放送会社を提訴 著作権紛争 「音楽流すな」 デジタル録音で即CD化”
4. 朝日新聞, 1998年7月29日朝刊, “コピーが危ない(1) 米国著作権論争”
5. 朝日新聞, 1998年7月30日朝刊, “コピーが危ない(2) 米国著作権論争”
6. 朝日新聞, 1998年7月31日朝刊, “コピーが危ない(3) 米国著作権論争”
7. 読売新聞, 1998年9月30日朝刊, “MP3に音楽業界ピリピリ 違法サイト続々”
8. 朝日新聞, 1998年10月21日朝刊, “音質落とさず ネットで音楽配布可能 MP3ブーム 頭痛い!”

研究業績

論文

1. 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, Vol. J83-D-II, No. 11, pp. 2255–2263, 2000. (1章, 2章の内容に関連)
2. Ikuo Kitagishi, Tamotsu Machino, Akira Nakayama, Satoshi Iwaki and Masashi Okudaira: “Development of Motion Data Description Language for Robots Based on eXtensible Markup Language –Realization of Better Understanding and Communication via Networks–,” Journal of Robotics and Mechatronics, Vol. 14 No. 5, pp. 471–478, 2002.
3. 細田 真道, 中山 彰, 小林 稔, 岩城 敏: “生体情報に基づく会議状態推定 共鳴コンセプトの提案,” 日本 VR 学会論文誌, Vol. 9, No. 2, pp. 151–159, 2004.
4. 中山 彰, 町野 保, 北岸 郁雄, 岩城 敏, 奥平 雅士: “音を用いたモーションメディアコンテンツ流通方式の提案とそのネットワークコミュニケーションサービスへの応用,” 日本ロボット学会誌, Vol. 23, No. 5, pp. 602–611, 2005. (1章, 3章の内容に関連)

国際会議

1. Akira Nakayama, Jinlin Lu, Satoshi Nakamura and Kiyohiro Shikano: “Robust digital watermarks for audio signal,” Second International Conf. on Information, Communications and Signal Processing (ICICS'99), 1999.

2. Ikuo Kitagishi, Tamotsu Machino, Akira Nakayama, Satoshi Iwaki and Masashi Okudaira: “Motion data description language for robots based on eXtensible Markup Language,” Proceedings of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems EPFL, Lausanne, Switzerland, Vol. 1, pp. 1145–1151, 2002.
3. Akira Nakayama, Tamotsu Machino, Ikuo Kitagishi, Satoshi Iwaki and Masashi Okudaira: “Rich Communication with Audio-Controlled Network Robot,” Proceedings of the 2002 IEEE Intl. Workshop on Robot and Human Interactive Communication, pp. 548–553, 2002.
4. Masamichi Hosoda, Akira Nakayama, Minoru Kobayashi, Satoshi Iwaki: “Conference State Estimation by Biosignal Processing - Observation of Heart Rate Resonance -,” CHI2004 Extended Abstracts, pp. 1187–1190, 2004.
5. Satoshi Iwaki, Akira Nakayama and Masashi Okudaira: “MotionMedia to Augment Content’s Expressivity and its Evaluation Tests Using a Biological Sensor,” Proc. of SCI 2004, 2004.
6. Yuriko Suzuki, Satoshi Iwaki, Minoru Kobayashi, Akira Nakayama and Yoshihiro Shimada: “Untethered Force Feedback Interface That Uses Air Jets,” SIGGRAPH 2004 Emerging technologies, 2004.
7. Masamichi Hosoda, Akira Nakayama, Minoru Kobayashi and Satoshi Iwaki: “Meeting State Estimation using Head Movement,” Proc. AMT2005, pp. 40–43, 2005.
8. Akira Nakayama, Masamichi Hosoda, Minoru Kobayashi and Satoshi Iwaki: “Constructing a Multi-Point Video Conference Corpus Labeled by Participants with Reference to Their Feelings Towards “Video Conference-Grasper,”” Proc. CollabTech2005, pp. 17–21, 2005.

研究会

1. 中山 彰, 陸 金林, 中村 哲, 鹿野 清宏: “時間・周波数マスキングに基づくオーディオ信号への電子透かし,” 電子情報通信学会 / 音声研究会 (デジタル信号処理), SP98-66 (DSP98-87), pp. 57-62, 1998.
2. 陸 金林, 中山 彰, 中村 哲, 鹿野 清宏: “マスキングモデルを利用した電子透かし法とその性能評価,” 1999年暗号とセキュリティシンポジウム (SCIS'99), T2-2.3, pp. 515-520, 1999.
3. 中山 彰, 岩城 敏: “HyperAudio: 音を媒介にしたマンマシンインタフェースの一手法,” 情報処理学会研究報告ヒューマンインタフェース, HI-88-01, 2000.
4. 岩城 敏, 町野 保, 北岸 郁雄, 中山 彰, 奥平 雅士: “ネットワークコンテンツとしてのロボットモーション-モーションメディアコンテンツ流通プラットフォーム構想-,” 第7回ロボティクスシンポジウム予稿集, pp. 297-304, 2002.
5. 中山 彰, 細田 真道, 犬童 拓也, 小林 稔, 岩城 敏: “多地点ビデオ会議コーパスに基づく会議途中参加支援機能の研究,” 情報処理学会研究会報告, 2003-GN-49 (20), pp. 115-120, 2003.
6. 犬童 拓也, 細田 真道, 中山 彰, 小林 稔, 岩城 敏: “会議途中参加支援を目的とした会議情報検索インタフェース - GijiLook の提案 -,” 情報処理学会研究会報告, 2003-GN-49 (21), pp. 121-126, 2003.
7. 岩城 敏, 北岸 郁雄, 中山 彰, 小林 稔: “画像メディアとモーションメディアの融合 ScreenServo 方式の提案,” 電子情報通信学会研究会報告, ヒューマンコミュニケーション基礎, HCS-2003-1, pp. 19-20, 2003.
8. 中山 彰, 細田 真道, 小林 稔, 岩城 敏: “会議参加者の発話継続時間長による遠隔会議対話セグメンテーション,” 信学技報, MVE2004-50, pp. 1-6, 2004.

9. 細田 真道, 中山 彰, 小林 稔, 岩城 敏: “生体情報に基づく会議状態推定
心拍変動共鳴現象の観測,” 情処研報, 2004-GN-50 (7), 2004-EIP-22 (7),
VR 学研報, CSVC2004-7, pp. 43-48, 2004.
10. 細田 真道, 中山 彰, 小林 稔, 岩城 敏: “頭部動作に基づく会議状態推定
推定アルゴリズムの検討,” VR 学会サイバースペースと仮想都市研究会,
CSVC2005-4, pp. 21-26, 2005.
11. 中山 彰, 細田 真道, 小林 稔: “飲みコミュニケーションとパーティコーパスの収
集と分析,” 信学技報, MVE2005, MVE-2005-41, pp. 83-88, 2005.
12. 石井 陽子, 小林 稔, 中山 彰, 細田 真道: “情報との出会いを演出する手の
平表示インタフェース,” 信学技報, MVE2005, MVE-2005-42, pp. 89-93,
2005.
13. 中茂 睦裕, 井原 雅行, 中山 彰, 木村 篤信, 小林 稔: “状態量を熱表現する
「ひやあつメディア」の検討,” 第 27 回 サイバースペースと仮想都市研究会,
2006.

大会発表

1. 中山 彰, 林 昭博: “音の $1/f$ ゆらぎに関する研究,” 電気学会関西支部第 2 回
高専卒業研究発表会論文集, pp. 40-41, 1995.
2. 中山 彰, 陸 金林, 中村 哲, 鹿野 清宏: “MPEG の心理音響モデルを利用した
オーディオ信号への電子透かしの検討,” 日本音響学会講演論文集, 2-P-30,
pp. 345-346, 1998.
3. 中山 彰, 陸 金林, 中村 哲, 鹿野 清宏: “MPEG 心理音響モデルを利用した電
子透かし法の改良,” 日本音響学会講演論文集, 1-3-22, pp. 223-224, 1999.
4. 中山 彰: “オーディオ信号への電子透かし,” 日本音響学会関西支部若手研
究者交流研究発表会, 1998.

5. 北岸 郁雄, 中山 彰, 岩城 敏, 小林 稔: “XML をベースとした異構造アバタ動作記述言語の開発 - 高度意思疎通メディアとしてのモーション - ,” 計測自動制御学会 SI 部門学術講演会講演論文集 I, 1P71-02, pp. 357-358, 計測自動制御学会, 2002.
6. 岩城 敏, 北岸 郁雄, 中山 彰, 小林 稔: “未来型コミュニケーションにおける第 5 のメディア ; モーションメディアの役割,” 計測自動制御学会 SI 部門学術講演会講演論文集 I, 1P71-01, pp. 355-356, 計測自動制御学会, 2002
7. 岩城 敏, 中山 彰, 奥平 雅士: “生体センサによるモーションメディアコンテンツ表現力の測定,” 計測自動制御学会 SI 部門学術講演会講演論文集 SI2003, 1F4-6, 2003.
8. 町野 保, 中山 彰, 岩城 敏, 下倉 健一郎: “携帯用簡易 MIDI ファイルによるモーションコンテンツ配信,” 計測自動制御学会 SI 部門学術講演会講演論文集 SI2004, 1F4-3, 2004.
9. 中山 彰, 町野 保, 岩城 敏: “USB オーディオインタフェースによるモーションメディアコンテンツ流通,” 計測自動制御学会 SI 部門学術講演会講演論文集 SI2004, 1F4-2, 2004.
10. 細田 真道, 中山 彰, 小林 稔, 岩城 敏: “生体情報による会議状態推定 切断・接続による生体情報共鳴現象の変化 ,” 信学総大, A-16-12, p. 321, 2004.
11. 中山 彰, 細田 真道, 小林 稔, 岩城 敏: “参加者の発話量変化に基づく多地点遠隔会議対話セグメンテーションの検討,” 音響学会春季大会予稿, pp.75-76, 2004.
12. 細田 真道, 中山 彰, 小林 稔, 岩城 敏: “生体情報に基づく会議状態推定 心拍変動共鳴現象と会議参加者による主観評価値の比較 ,” FIT2004, M-007, pp. 105-106, 2004.

13. 細田 真道, 中山 彰, 小林 稔, 岩城 敏: “顔画像に基づく会議状態推定 顔位置変動と会議状態の比較,” VR学会大会, 2D1-3, pp. 323-326, 2004.
14. 中山 彰, 細田 真道, 小林 稔, 岩城 敏: “多地点遠隔会議対話セグメンテーションのための発話長分布の分析,” 日本音響学会講演論文集, 2-P-29, pp. 419-420, 2004.
15. 細田 真道, 中山 彰, 石井 陽子, 小林 稔: “頭部動作に基づく会議状態推定 頭上ステレオカメラによる頭部動作取得,” 画像電子学会年次大会予稿集, 05-22, p. 43-44, 2005.
16. 岩城 敏, 町野 保, 中山 彰, 下倉 健一郎: “モーションメディアに関する研究活動概要報告,” 第37回計測自動制御学会北海道支部学術講演会予稿集, B6, 2005.
17. 細田 真道, 中山 彰, 小林 稔: “頭部動作に基づく会議状態推定 実空間会議のための頭部動作取得,” 信学会総合大会, A-15-19, p. 262, 2005.
18. 細田 真道, 中山 彰, 石井 陽子, 小林 稔: “会議室における頭部動作取得,” FIT2005 予稿集, K-047, 2005.
19. 岩城 敏, 町野 保, 中山 彰: “音によるモーションメディアコンテンツ流通のための4チャンネルサーボモータコントローラ,” 計測自動制御学会 SI 部門学術講演会講演論文集 SI2005, 1D3-2, 2005.
20. 岩城 敏, 町野 保, 中山 彰: “モーションメディアコンテンツのデザインに関する考察 -第1回モーションメディア作品コンテストの概要紹介-,” 第38回計測自動制御学会北海道支部学術講演会予稿集, C12, 2006.
21. 中茂 睦裕, 中山 彰, 小林 稔: “テーブルトップへの画像刺激による遠隔コミュニケーションの分析,” ヒューマンインタフェースシンポジウム 2006 論文集 2, pp. 695-700, 2006

特許

1. Iwaki Satoshi, Nakayama Akira, Machino Tamotsu, Kitagishi Ikuo and Nakamura Yukihiro: “Acoustic signal transmission method and acoustic signal transmission apparatus,” United States Patent Application 20060153390.
2. 岩城 敏, 中山 彰: “音響信号伝達方法および音響信号伝達装置,” 特開 2001-148670.
3. 岩城 敏, 中山 彰: “音声反応型機械,” 特開 2001-175271.
4. 岩城 敏, 中山 彰, 町野 保, 中村 幸博: “自動車情報提供方法及びカーナビゲーションシステム,” 特開 2001-184597.
5. 中山 彰, 岩城 敏: “情報伝達方法および情報伝達装置,” 特開 2001-186093.
6. 岩城 敏, 中山 彰, 町野 保, 北岸 郁雄: “音楽反応型ロボット, 音声反応型ロボットおよび通信・放送融合ロボットシステム,” 特開 2001-239485.
7. 岩城 敏, 中山 彰, 町野 保: “音響信号伝達方法および音響信号伝達装置および記憶媒体,” 特開 2001-320337.
8. 中山 彰, 岩城 敏, 町野 保, 北岸 郁雄: “ロボットシステム, ロボット制御信号生成装置, ロボット制御信号生成方法, 記録媒体, プログラムおよびロボット,” 特開 2002-127062.
9. 岩城 敏, 中山 彰: “商取引方法,” 特開 2002-150102.
10. 町野 保, 中山 彰, 北岸 郁雄, 岩城 敏: “制御機器, 機器制御信号生成方法, 機器制御方法, 記録媒体およびプログラム,” 特開 2002-305781.
11. 北岸 郁雄, 岩城 敏, 町野 保, 中山 彰, 中山 丈二: “動作入力再生方法, 装置及びプログラム,” 特開 2003-039358.

12. 中山 彰, 町野 保, 岩城 敏, 北岸 郁雄, 奥平 雅士: “ロボットシステム, ロボット制御方法, ロボット制御装置及びロボット制御プログラム,” 特開 2003-150186.
13. 町野 保, 北岸 郁雄, 中山 彰, 岩城 敏: “動作制御装置, 方法及びプログラム,” 特開 2003-173452.
14. 中山 彰, 岩城 敏, 奥平 雅士, 北岸 郁雄, 町野 保: “ロボットシステム, ロボット制御信号配信装置, ロボット制御装置, ロボット制御信号配信方法, ロボット制御方法, プログラム, およびプログラムを記録したコンピュータ読み取り可能な記録媒体,” 特開 2004-034250.
15. 北岸 郁雄, 町野 保, 中山 彰, 岩城 敏, 奥平 雅士: “ロボットの制御方法及び装置並びにプログラム及び記録媒体,” 特開 2004-054841.
16. 北岸 郁雄, 八木 貴史, 中山 彰, 岩城 敏, 小林 稔, 石橋 聡: “共有情報操作装置, 共有情報操作方法, 共有情報操作プログラム, 共有情報操作プログラムを記録したコンピュータ読み取り可能な記録媒体,” 特開 2004-171356.
17. 中山 彰, 岩城 敏, 八木 貴史, 小林 稔, 北岸 郁雄, 磯 和之, 石橋 聡: “会議情報可視化方法, 装置, プログラムおよび該プログラムを記録した記録媒体,” 特開 2004-173058.
18. 中山 彰, 岩城 敏, 北岸 郁雄, 磯 和之, 石橋 聡, 八木 貴史: “多地点電子会議システムにおける会議概要把握支援方法, 多地点電子会議システム用サーバ, 会議概要把握支援プログラム, および該プログラムを記録した記録媒体,” 特開 2004-350134.
19. 細田 真道, 中山 彰, 岩城 敏, 小林 稔, 犬童 拓也: “電子会議データ取得方法, 装置, プログラム, および記録媒体ならびに電子会議データ検索方法, 装置, プログラム, および記録媒体,” 特開 2005-117155.
20. 細田 真道, 中山 彰, 岩城 敏, 小林 稔: “会議状態検出方法, 装置, および会議状態検出プログラム,” 特開 2005-198086.

21. 細田 真道, 中山 彰, 小林 稔, 岩城 敏: “会議 / 談話状態推定方法, 装置, および会議 / 談話状態推定プログラム,” 特開 2006-065765.
22. 中山 彰, 細田 真道, 小林 稔: “会議構造把握支援方法, 装置, プログラム, 及び該プログラムを格納した記録媒体,” 特開 2006-081005.
23. 細田 真道, 中山 彰, 小林 稔: “存在・動作取得方法, 装置, およびプログラム,” 特開 2006-228104.

紀要

1. 中山 彰, 林 昭博: “音の $1/f$ ゆらぎに関する研究,” 神戸市立工業高等専門学校研究紀要, No. 34, pp. 43-48, 1996.

解説記事

1. 町野 保, 中山 彰, 岩城 敏: “音データでロボットを制御する技術「Cyber-Performer/Audio」,” NTT 東日本プラザ 2002Feb. Mar 号, 2002.
2. 中山 彰: “ネットワークロボティクス,” 日本機械学会誌, Vol. 106, No. 1017, p. 183, 2003.

展示会 (公開分のみ)

1. 計測自動制御学会 モーションメディア調査研究会: “モーションメディア,” ロボカップ共催行事キャラロボ展示ブース, 2005.
2. MeetBall 開発チーム: “ミーティングを支援するロボットメディア MeetBall,” NTT R&D フォーラム 2005, 2005.
3. MeetBall 開発チーム: “ミーティングを支援するロボットメディア MeetBall,” NTT R&D Expo, 2005.
4. MeetBall 開発チーム: “Information Snow,” NTT R&D Expo, 2005.

学位論文

1. 中山 彰: “音の $1/f$ ゆらぎに関する研究,” 神戸市立工業高等専門学校電子工学科卒業論文, 1995.
2. 中山 彰: “wavelet 変換を用いた音声変換,” 静岡大学工学部電子工学科卒業論文, 1997.
3. 中山 彰: “心理音響モデルに基づくオーディオ信号への電子透かしの研究,” 奈良先端科学技術大学院大学 情報科学研究科 修士論文, NAIST-IS-MT9751078, 1999.

メディア

1. “会話弾むヒントテーブルに投影 天井ロボ , NTT 開発,” 日本経済新聞 (2005年12月19日掲載)
2. “コミュニケーション支援ロボット ミートボール,” TV 東京 5時のニュース (2005年12月20日放映)
3. “「あなた,話し過ぎです」—会話をフォローする丸いロボ,” IT Media ニュース (2005年12月20日掲載)
4. “適切なタイミングで会話にツッコミを入れるコンダクター,” Internet Watch (2005年12月20日掲載)
5. “ボールで会話も弾みます!,” WBS トレンドたまご (2006年1月31日放映)
6. “MeetBall,” 月間ビジネスコミュニケーション, Vol. 43, No. 2, pp. 69–70, 2006.
7. “ロボットが会話の壁を除去 MeetBall,” 電経新聞 (2006年4月24日掲載)

表彰

1. 岩城 敏, 北岸 郁雄, 中山 彰, 町野 保, 奥平 雅士: “SI2002 ベストセッション賞,” 2002.
2. 町野 保, 北岸 郁雄, 中山 彰, 岩城 敏: “NTT サイバーコミュニケーション総合研究所 所長表彰 事業貢献賞,” 2002.(社内表彰)
3. 風インタフェースチーム: “NTT サイバーソリューション研究所 所長表彰 報道賞,” 2005. (社内表彰)
4. MeetBall 開発チーム: “NTT サイバーソリューション研究所 所長表彰 報道賞,” 2006. (社内表彰)
5. 中山 彰: “NTT サイバーソリューション研究所 所長表彰 特別賞年間 MVP,” 2006. (社内表彰)

学会・社会活動

1. NTT 武蔵野研究開発センタ主催 小学生科学体験教室 2001, 音と動きの不思議な世界, 講師, 2001.
2. 日本科学技術振興財団主催 サイエンスキャンプ 2004, 遠隔コミュニケーションを測る, 講師, 2004.
3. 計測自動制御学会 SI 部門 モーションメディア調査研究会, 委員, 2004~.
4. Collabtech2006, プログラム委員, 2006.
5. Collabtech2007, プログラム委員, 2007.