

# HyperAudio: 音を媒介にした マンマシンインタフェースの一手法

中山 彰 岩城 敏

NTTサイバーソリューション研究所

〒180-8585 東京都武蔵野市緑町3丁目9-11

nakayama.akira@lab.ntt.co.jp

iwaki.satoshi@lab.ntt.co.jp

## あらまし

複数の人間と機械が混在する環境において、人間と機械がコミュニケーションするための新しい、複合メディア「HyperAudio」のコンセプトを提案する。これは可聴音に機械指令信号を音声電子透かし技術を用いて重畳した音データである。HyperAudioは音信号に密かに情報を埋め込むといった点で、電子透かし技術を利用したものであるが、従来の電子透かしから安全性や、耐攻撃性、極限なまでの品質の保持を取り外し、積極的に情報伝達に使用するという視点にたっている。このためにリアルタイム動作性、空中伝播といった、電子透かしにはないあらたな技術的障壁が生じる。このHyperAudioの応用として、音楽と運動情報が一元化されたオーディオモーションメディアを提示する。

キーワード インタフェース, 電子透かし, 多重化, 音, 情報, 動作

## HyperAudio: A New Approach for Man-Machine Interface via Audio

Akira NAKAYAMA Satoshi IWAKI

NTT Cyber Solutions Laboratories

3-9-11, Midori-Cho, Musashino-Shi, Tokyo, 180-8585 JAPAN

## Abstract

We propose the concept of a new multiplexing medium "HyperAudio" for man-machine communication via audio. HyperAudio is an audio signal that is secretly embedded in machine-control signals by using digital watermarking technology. We also show the application of HyperAudio. It is the "Audio-motion" medium which enables us to send audio and motion information simultaneously.

key words interface, digital watermarks, multiplexing, audio, information, motion

## 1 はじめに

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われる。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。この観点から本報告では、音のメディアに着目し、人間と機械とが混在するコミュニケーション空間におけるシームレスなインタフェースメディアの一例として、HyperAudioを提案する。HyperAudioは、音声電子透かし技術を利用して、音楽や音声など可聴原音に対し、伝達すべき情報を密かに重畳した「音」である。

本報告ではまず、HyperAudioが提案された背景を述べた後、HAの概要とその特徴、通常の音声電子透かし技術との違い、について述べる。次に、HyperAudio実現の要であるHyperAudio Encoder/Decoderの基本原理を述べ、その原理確認実験後、応用システムを紹介する。

## 2 HyperAudioの提案

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役を演じ続けるであろう。また音楽（歌）も、人一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、事前に特定の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

現が一画的であったり、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[1]。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、モデムなどに使われている変調波、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っている」以上の情報を人間にもたらしものではないし、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、人間みでのコミュニケーションで例えれば、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。

我々はこのような考察から、人間に取ってなじみ易い音（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、「HyperAudio」を提案するに至った。HyperAudioは、近年著作権保護の用途で着目されている音声電子透かし技術を利用して、音楽や音声など可聴原音に対し、伝達すべき情報を密かに重畳した「音」である。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に、音声電子透かし技術を応用するのである。HyperAudioの通信メディアとしての特徴は次のようにまとめることができる。

表 1: 音声電子透かし技術とHyperAudioの比較

	電子透かし	HyperAudio
耐改ざん性	必要	不要
リアルタイム性	不要	必要
空中伝播	不要	必要
音質	極めて重要	多少劣化許容

[多値性] 音であると同時に信号が含まれており、しかもそれぞれの意味は完全に独立である。この点が音声認識技術と大きく異なるメリットである。

[柔軟な伝達性] 音波の性質から空中を伝搬し、その回折性から障害物を回りこんで情報が伝わっていく。また、水中など電波が伝わりにくい空間や、病院等電波の利用が制限されるような空間においても有効である。

[既存の設備との親和性] 音というメディアに情報を直接埋め込むため、音を蓄積および伝送する既存媒体すべてが、HyperAudioの伝送媒体となる。同時に、音を音波そのものあるいは電気信号として流通可能な既存NWインフラ全てが、有効に利用可能である。すなわち図1に示すように、HyperAudioは、あらゆる既存音メディア（CD、MD、電話、テープなど）で保存され、あらゆる既存ネットワーク（電話、ラジオ、空間、CDなど）で流通される。そして、この図における、出力および入力デバイス、伝送媒体組み合わせにより、様々なサービスおよびシステムが実現できる。図中HyperAudio Decoder /Encoderと記しているのは、詳細は後述するが、入力デバイスからの信号を音声もしくは音楽に情報を埋め込む装置、またHyperAudioに埋め込まれた情報を取り出す情報抽出装置を表している。

[簡便性] 通常のオフィスや家庭内にごくありふれているマイクやスピーカを利用して情報の受発信が可能である。

なお、音信号に密かに情報を埋めこむといった点で、HyperAudioは、電子透かし技術を利用したものであるが、従来の電子透かしから安全性や、耐攻撃性、極限なまでの品質の保持を取り外し、積極的に情報伝達に使用するという視点にたっている。このためにリアルタイム動作性、空中伝播といった、電子透かしにはないあらたな技術的障壁が生じる。両者の違いを表1にまとめておく。

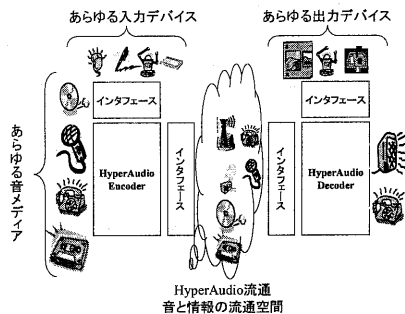


図 1: HyperAudio 流通システム概要図

### 3 HyperAudio Encoder / Decoder 方式

#### 3.1 情報埋め込み

情報の埋め込み部を、HyperAudio Encoderと呼ぶこととする。HyperAudio Encoderは、入力デバイスからの入力を、入力された音・音楽に埋め込む働きをする。

電子透かし技術における、情報の埋め込み方法は、周波数拡散による方法[6][8]、エコー成分を付加する方法[4]、位相を変化させる方法[3]、量子化雑音[7]を利用する方法などさまざまな方法が提案されている。しかし今回のように、AD/DA変換にロバストかつ受け取った信号のみから自己同期可能であり、高いビットレートをもつ方法は提案されていないように思われる。そこで今回は、HyperAudio Encoderに適した、人間の耳の周波数特性を考慮した情報埋め込み方法を考案した。

人間の耳の感度（最小可聴域）は周波数に対して、一様ではなく図2のような特性を持つことが知られている[9]。この図で、横軸は周波数、縦軸は、音圧レベルをあらわしており、一点鎖線は人間の最小可聴域（音の強度を弱くしていった場合の、聞きうる最小のレベル）をあらわしている。今回は、人間の耳に感じにくい16kHz付近に帯域雑音を付加することで情報の埋め込みを行っている。帯域雑音のエネルギーは、人間の最小可聴域と雑音のレベルを考慮し、適当なレベルを設定している。

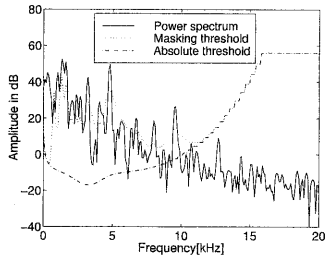


図 2: 人間の耳の感度

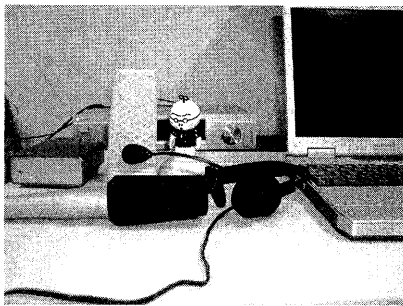


図 3: HyperAudio 検証システム写真

### 3.2 情報取り出し

情報の取り出し部を、HyperAudio Decoder と呼ぶこととする。

今回は、HyperAudio Encoder で雑音が付加された帯域を、狭帯域バンドパスフィルタに通して、そこからの出力エネルギーの変化を閾値と比較することで情報の抽出を行った。また回りの雑音レベルや、音楽のレベルを考慮し、閾値は動的に変化させるようにした。

## 4 原理確認実験

### 4.1 実験の目的

HyperAudio の実現可能性を検証する目的で以下に示すような実験を行った。本実験では特に技術的困難性が高くかつ実用上極めて重要と考えられる、HyperAudio の空中伝播の可能性、FM 電波上での伝送の可能性に着目し、検証実験を行った。

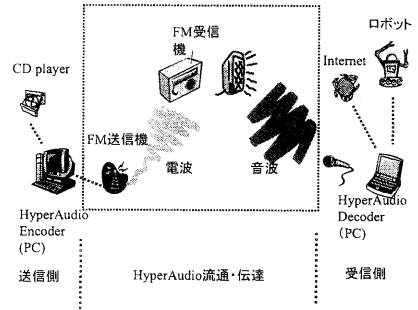


図 4: HyperAudio 検証システムイメージ図

### 4.2 システム構成

今回利用した実験した実験システムの写真を図3に示す。また概要を図4に示す。

システムに用いた機器は以下の通りである。

[HyperAudio Encoder-Decoder] 今回の実験では、HyperAudio Encoder-Decoder は、IBM PC-AT 互換機 (Pentium III 500MHz) に実装した。音の取り込み、および、出力は互換機付属の安価なサウンドカードを使用した。この試作 HyperAudio Encoder-Decoder では、40bits/sec のデータが、実時間で埋め込みかつ取り出し可能である。処理の高速化のために、MMX 命令を使用した。システムの実装は、Window98 上の VisualC++ で行った。

[伝送メディア (FM 送信機・受信機)] HyperAudio システムの伝送路として FM 放送を考えた。FM 送信機は、市販のミニ FM 放送局用送信機を用いた。通常の FM 放送に比べれば、多少雑音が多く、品質は劣る。伝送帯域は測定の結果、ほぼ 16kHz 帯域まで伝送していることを確認した。また FM 受信機は市販の FM・AM チューナを用いた。

[HyperAudio 出力デバイス (スピーカ)] パソコンに付属の安価なスピーカを使用した。低域と高域で特性が落ちるが、100Hz から 18KHz 付近まで、伝送帯域をもっていることを確認し、今回の情報埋め込み手法が適用できると判断した。

[HyperAudio 入力デバイス (マイク)] ヘッドセットマイクおよび、コンデンサマイクロホンを利用した。ヘッドセットマイクは、近接型のマイクのため、マイクとスピーカの距離が非常に短い場合有効であり、また雑音に強いことを確認している。またコンデンサマイクロホンは、マイクとスピーカの距離が多少離れている場合においてもスピーカから

の音を收音できる。

[出力インタフェース] 今回は市販の玩具ロボットを利用した。この玩具ロボットは、RS232Cにつながるされており、PCから制御できるようなライブラリが用意されている。このロボットの自由度は2（首と手）であり、LEDランプ（緑色と赤色）が用意されている。

### 4.3 実験条件

一般的な静寂なオフィス環境に（騒音 50 dBA）において、HyperAudio Encoder-Decoderの伝送実験を行った。CD音源に対して、情報（文字列）を埋め込み、FM電波に乗せて、それをFMチューナ受けたあと、増幅し、スピーカで放射する経路で実験を行った。HyperAudioの放射は上記のスピーカを用い、また受信は上述のコンデンサマイクロホンとヘッドセットを用いた。コンデンサマイクロホンの場合、マイクとスピーカの距離はおよそ1メートル、ヘッドセットマイクの場合数センチメートルであった。おもな雑音源は、エアコン、PC2台、人の話し声などがあつた。

### 4.4 実験結果

実験の結果、コンデンサマイクロホンの場合は、5%ないしは6%の符号誤りが発生した。また、ヘッドセットマイクロホンの場合は、1%程度であった。その原因としては、PCから発生し、FM受信機に入る高周波ノイズさらに、空気中のノイズなどの影響が考えられる。

また Encoder / Decoder のコンピュータ（Pentium III 500 MHz）に対する負荷は5%から7%程度であり、特別に高価なPCを用意しなくとも、HyperAudioシステムは十分実行可能なことを検証した。

HyperAudioによる、音質に与える影響はほとんどなく、20名の人に試聴してもらったところ、ほとんどの人から気付かないか気にならないレベルであるとの解答が得られた。

### 4.5 考察

実験結果より、HyperAudioの原理的な実現可能性が確認できた。また今回の実装においては業務用のハイスpekな機器ではなく、市販の安価な機器で成功した意義は大きい。しかしながら、まだ空気中の

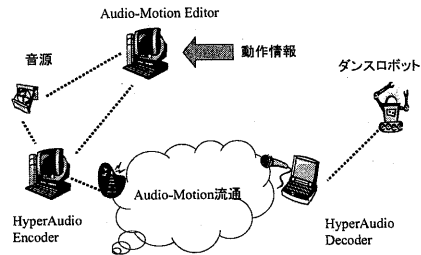


図5: ダンシングロボット システム概要図

ノイズに弱いなどの欠点があり、今後HyperAudio Encoder / Decoder アルゴリズムの検討や、エラー訂正技術を取り入れることが必要ある。

## 5 応用例

### 5.1 玩具への応用 (dancing robot)

音楽にロボット動作コマンドを重畳させたメディア、「オーディオモーションメディア」をHyperAudioを用いて実現することが可能となる。図5にその概要を示す。従来、音の強弱に反応するロボットがあつたが、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、エンタテイメント要素かつ鑑賞に値する動きを行わせることも可能になるとと思われる。CD、MDなどのメディアに動き情報を蓄積し、配布することも可能であろう。

## 6 まとめ

音声電子透かし技術を応用したメディア、HyperAudioの概念とその流通システムおよび応用例を提案した。実験によりその原理的な可能性を確認し、プロトタイプデモの作成および評価を通して、新しい通信メディアとしての高いポテンシャルを明らかにした。今後は、耐雑音性の向上や、情報伝送ビットレートの向上、音質の向上などのHyperAudio Encoder-Decoderの高度化を進めていきたい。また、ワンチップ化などを進め、新たなサービスイメージの創出に努めていきたい。モーションメディア用コンテンツ作成ツールの作成や、オーディオモーションメディアの心理的評価も行っていく必要がある。

## 謝辞

本研究を行うにあたり、有益な意見を頂いた、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プロジェクトおよびサイバースペース研究所センシング&ダイナミクスグループの皆様には感謝します。

## 参考文献

- [1] 鹿野, 中村, 伊勢: 「音声・音情報のデジタル信号処理」, 昭晃堂(1997).
- [2] 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).
- [3] Y. Yardimci, A. E. Cetin, R. Ansari, "Data Hiding in Speech Using Phase Coding," ESCA. Eurospeech97, Greece. pp.1679-1682(1997).
- [4] D. Gruhl, A. Lu, W. Bender: "Echo Hiding," pp.295-315, Lecture Notes in Computer Science, Vol.1174, Information Hiding, Springer(1996).
- [5] W. Bender, D. Gruhl, N. Morimoto, A. Lu, "Techniques for data hiding," IBM SYSTEMS JOURNAL, Vol. 35, NOS 3&4 (1996).
- [6] 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).
- [7] 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).
- [8] Laurence Boney, Ahmed H. Tewfik, Khaled N. Hamdy, "Digital Watermarks for Audio Signals," IEEE Intl. Conf. on Multimedia computing and Systems, Hiroshima, pp.473-480(1996).
- [9] 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).