# 高度情報転送における情報源符号化の効率化

**岡田　守 *　　福本昌弘 **　　浜崎真二　　藤村和人**

高知工科大学情報システム工学科
〒 782-8502 高知県香美郡土佐山田町宮ノ口 185

E-mail : * okada.mamoru@kochi-tech.ac.jp, ** fukumoto.masahiro@kochi-tech.ac.jp

**要約**：本研究では、転送される情報への新たな価値の付与や、必要な情報をいつでもどこでも思いのままに利用できる環境を実現するシステムの研究開発を目指している。そのため、情報転送時に物体・動作認識の技術を適用し、転送する情報を高度な価値を付加した情報として符号化する。また、転送される情報を所望の特性で再現するため、適応信号処理技術を用いた情報符号化時処理を施す。本稿では、これらのうち、両眼視差法を用いた立体画像の認識と、最も単純な構成でのステレオ音場再現システムの構成法を示している。

Abstract : Recently, a large amount of data is exchanged on the Internet. It is neces-sary to assignment high added value for transferred data. The purpose of our research is toestablish the encoding and signal processing technique for exible and comfortable informa-tion environment. In This paper, the experiment of recognizing the sign language and the
multi-channel sound eld reproduction system have been presented.

## 1．まえがき

情報通信技術の進歩に伴い、処理しきれないほどの大量のデータがやりとりされるようになりつつある。この氾濫する情報を的確に配信表示できるような仕組みが不可欠になっている。本研究では、柔軟で快適な情報ネットワークシステムの構築をめざし、転送される情報への新たな価値の付与や、必要な情報をいつでもどこでも思いのままに利用できる環境を実現するシステムの研究開発を行う。そのため、情報転送時に物体・動作認識の技術を適用し、転送する情報を単なる映像情報としてではなく高度な価値を付加した情報として符号化する。また、転送される情報を所望の特性で再現するため、適応信号処理技術を用いた情報符号化時処理を施す。更に、転送すべき情報を自動処理することにより転送の効率化をはかる。

本稿では、これらを実現するために必要な３時限画像情報表現法およびステレオ型音響再現技術を示す。まず、両眼視差法を用いた物体の３次元情報化と手話認識への応用について述べる。次いで、最も簡単な構成で実現可能なステレオ型音場再現システムを提案する。

## 2．両眼視差法を用いた手話の認識

本章では、両眼視差法による物体認識の応用として、手話認識を取り上げる。

聴覚障害者はコミュニケーションツールとして、手話や筆談などを利用している。特に手話は、生まれた時から、あるいは幼少等の早い時期から聴力を失った人達の主要なコミュニケーションツールとして使われている。しかし、現在国内において、手話を理解、使用できる手話人口は健常者を含めて 30 万人[1)] と非常に少ないのが実情である。

1) このうち手話母語話者は約 6 万人とされている

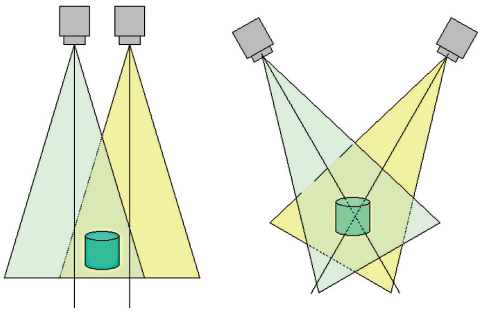


図 1　両眼視差法

手話により意思伝達される情報の 8 割は、腕の動きによるものである。また、腕の上下左右の動きだけでなく、その前後 ( 奥行き ) も含めた 3 次元の位置情報が手話を理解する上で必要となる。

本研究では、3 次元情報をステレオカメラより取得、隠れマルコフモデル (HMM)[1] を用いて手話単語の認識実験を行う。

### 2．1　両眼視差法 ( ステレオ視法 )

両眼視差法は、左右一定距離だけ離れた、異なる 2 点から観測した画像をもとに、対象物までの距離を求める 3 次元空間の計測法の一つである。3 次元空間座標は式（1）により算出することができる [2]。

$$\begin{cases} x = \dfrac{(X_L + X_R)}{2}\dfrac{L}{(X_L - X_R)} \\ y = Y_L\dfrac{L}{(X_L - X_R)} \\ z = f\dfrac{L}{(X_L - X_R)} \end{cases} \qquad (1)$$

対象座標（x、y、z）、撮影画像座標（$X_{[L、R]}$、$Y_{[L、R]}$）、焦点距離 f、光軸間距離 L、視差（$X_L - X_R$）、$Y_L = Y_R$ とする。

今回の実験では、実験機材として、市販のデジタルカメラ[2)] 2 台を 1 台の三脚に固定、カメラの光軸を平行に設定しステレオカメラとして利用した。認識特徴として入力する Z 軸座標については、f、L は定数とし、視差（$X_L - X_R$）を利用している。

2) Victor GR-DV700K：動画撮影有効画素数 62 万画素

表 1　撮影サンプル

| 被験者 | A | B | C | D | E | F | G | 合計 |
|---|---|---|---|---|---|---|---|---|
| OK | 1 | 2 | 3 | 1 | 2 | 2 | 1 | 12 |
| 飴 | 1 | 2 | 3 | 1 | 1 | 3 | 1 | 12 |

### 2．2　実験

両眼視差法による手話認識の有効性を確認するために実験を行う。

#### 2．2．1　撮影

男性 5 名、女性 2 名の計 7 名の被験者に、「OK」、「飴」の手話動画を見てもらい数回のリハーサル後撮影を行った。撮影した手話のサンプル数は表 1 に示す。

#### 2．2．2　特徴抽出

撮影した「OK」「飴」の手話動画を、右手第 1 指、右手第 2 指、右目を計測点とし、動画計測ソフト「Move-tr/2D」を用いて座標を計測する。計測した左右の座標系列から視差値（$X_L - X_R$）算出し、計測対象フレームと前後フレームとの均一重み平均値フィルタを用いて計測誤差によるノイズを除去する。右カメラの測定座標値と、ノイズを除去した視差値を、初期フレー

ムで検出された右目の XY 座標と視差値（$X_L$ — $X_R$）を原点とする座標系に変換し、認識特徴として利用する。

### 2.2.3 認識実験

HMM への記号登録は被験者 1 人につき 1 つ登録し、比較として視差（$X_L$ — $X_R$）情報を削除した計測データを作成し、HMM の出力記号列による認識実験を行う。表 2 は出力記号が、話者本人の登録記号であった場合を同一人物としている。誤認率は「OK」の手話において「飴」の登録記号を出力した場合とその逆である。

### 2.3 まとめ

現段階では「OK」と「飴」、2 種類の手話単語の識別が可能となった段階でしかない。しかし、手話「OK」は「飴」において、視差値ではそれぞれ集合が見られ、本実験の撮影システムでも、奥行き情報の取得は十分可能であるといえる . また、「OK」の識別率から見られるように、同一話者の同じ手話であっても、表出位置や体格、癖、タイミングなどが異なる。今回の認識特徴では、それらへの対応に限界があり、他の特徴抽出法の検討が今後の課題である。

表 2 OK-飴，識別結果

| | OK | | 飴 | |
|---|---|---|---|---|
| 出力記号列 | ST | MONO | ST | MONO |
| 複数 (同一人物含) | 66% | 16% | 92% | 42% |
| 非同一人物 | 17% | 0% | 8% | 8% |
| 同一人物のみ | 17% | 0% | 0% | 0% |
| 誤識別 | 0% | 84% | 0% | 50% |

## 3. クロストーク成分の相互相関に着目した音場再生システム

本章では、構成が最も単純で実現容易なステレオ型音響再現システムを提案する。

要求するような音響空間を再現させるにはスピーカ（2 次音源）からマイク（制御点）までの伝達特性の影響を打ち消す、すなわち伝達特性の逆特性を近似する必要がある。現在までに逆特性を近似する手法についてはシステム論的立場から様々な検討がなされてきた。現在最も有効とされている手法に MINT 理論に基づいた多チャネル多点制御系 [4] がある。これは制御点数 M に対し 2 次音源数を M +1 用意することで、逆特性の推定と同時にクロストーク現象を除去し音の再現を可能にしている。しかし多くのスピーカを用いることから制御系が複雑になってしまう。そこで、制御点数 2 に対し 2 次音源数 2 で構成される新たな制御系を提案する。

### 3.1 多入力信号補正システム

本節では提案する 2 チャネル {2 点制御系（多入力信号補正システム）について述べる。図 2 はその構成図である。適応フィルタ $H_{11}$ は右側入力信号 $d_1(t)$ とフィルタ通過後の信号 $s_{11}(t)$ から出力誤差を算出しているため $G_{11}$ の逆特性に近似させるフィルタとなる。ここで得られた係数を右側補正フィルタ係数 $c_1$ として与えることによって右側所望信号 $x_1(t)$ と右側観測信号 $y_1(t)$ を近似させることができる。しかしクロストー

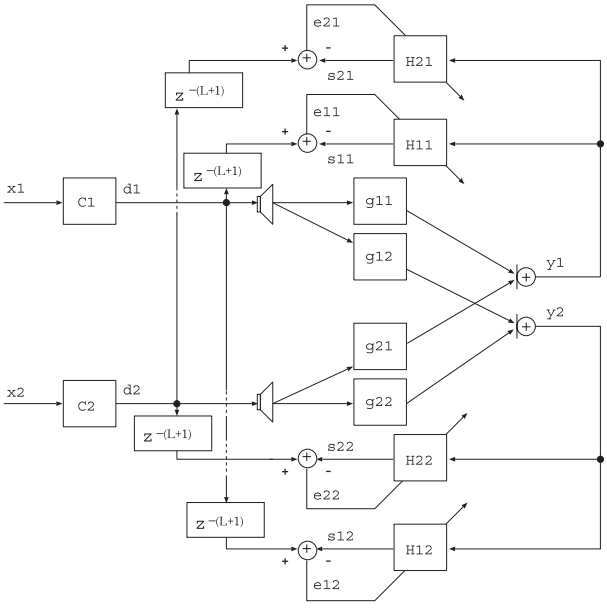


図２　多入力信号補正システムの構成

ク $G_{12}$ に対する補正処理により左側観測信号 $y_2(t)$ を右側所望信号 $x_1(t)$ に近似させるような動作をするため、これを防ぐ工夫が必要となる。適応フィルタ $H_{21}$ は左側入力信号 $d_2(t)$ と $s_{21}(t)$ から出力誤差を算出しているため $G_{21}$ の逆特性を近似させるフィルタとなる。通常ここで得られるフィルタ係数は右側補正フィルタ係数 $c_1$ として与えることはできないが、クロストークの伝達特性 $G_{12}$ と $G_{21}$ が互いに強い相関をもつことに着目すると、$H_{21}$ によって得られるフィルタ係数は $G_{12}$ の特性を予測するフィルタとして代用できる。これを右側補正フィルタ係数 $c_1$ として与えることで、$G_{12}$ を通過するクロストーク成分を左側所望信号 $x_2(t)$ に近似させることができる。したがって $H_{11}$ で得られたフィルタ係数と $H_{21}$ で得られたフィルタ係数を合成させたものを右側補正フィルタ係数として与えることによってクロストークの影響を軽減させる働きをもつ補正フィルタが設計できる。ここで左右の補正フィルタ係数はそれぞれ

$$c_1,L(t+1) = w_1 h_{11},L(t) + (1 - w_1) h_{21},L(t)$$
$$c_2,L(t+1) = w_2 h_{22},L(t) + (1 - w_2) h_{12},L(t)$$

として与えられる。wi は第 1 項と第 2 項に対する重み付けパラメータ（$0<wi \leq 1$）である。適応フィルタ $H_{ij}$ のフィルタ係数の更新には

$$\boldsymbol{h}_{ij,L}(t+1) = \boldsymbol{h}_{ij,L}(t) + \alpha \frac{\boldsymbol{y}_{j,L}(t)}{\|\boldsymbol{y}_{j,L}(t)\|^2} e_{ij}(t) \quad (2)$$

で示される学習同定法 [5] を用いる。ここではステップゲイン、$e_{ij}$ はフィルタの出力誤差を示す。

表 3 最大改善量と各パラメータおよび相互相関係数

| | $\alpha$ | $w_1$ | $w_2$ | $r_N$ | 改善量 [dB] |
|---|---|---|---|---|---|
| パターン 1 | 0.05 | 0.35 | 0.25 | 0.73 | 6.23 |
| パターン 2 | 0.05 | 0.35 | 0.50 | 0.55 | 5.40 |
| パターン 3 | 0.05 | 0.45 | 0.85 | 0.45 | 5.13 |

### 3.2 聴覚特性に基づく出力誤差の算出

場再生システムでは適応フィルタを可能な限り速く収束させることが望ましい。このことに対する一提案手法として、出力誤差を算出する際に人間の聴覚特性に基づき重み付けを行うことで余分な周波数帯域を縮小し計算効率の向上をはかる。人間の聴覚特性には騒音測定の指標として利用されている A 特性音圧レベル [6] を用いる。この特性から周波数毎の重み付け関数を

$$\phi(k) = \frac{Res(k) + |\min Res|}{|\min Res|} \quad (3)$$

により算出する。ここで Res(k)、minRes はそれぞれ A 特性における周波数応答およびその中の最小値を示す。入力信号 di(t) と適応フィルタの出力信号 sij(t) に対して周波数領域で (k) との積をとり、時間領域での差を新たな出力誤差 $e_{ij}$ として扱う。

### 3.3 計算機シミュレーション

提案手法の有効性を検証するために、計算機によるシミュレーションを行う。ステップゲインや重み付けパラメータ wi の違いによる収束特性を比較する。

#### 3.3.1 シミュレーション 1

所望信号 xi(t) として成人男性の声を 8kHz でサンプリングした音声信号を与える。また、適応フィルタのインパルス応答長 L = 512 とする。シミュレーションで用いる伝達関数は、反響性のある部屋においてスピーカとマイクロフォンの位置関係を変えたもの 3 パターンの実環境を想定し、原信号（白色雑音）と観測信号により実測したものを用いる。ここでは聴覚特性の重み付けは行わず、ステップゲインや重み付けパラメータ wi の違いによる収束特性を比較する。評価量には次に示す原音に対する再現音の再現精度（SNR）を用いる。

$$SNR\ [\mathrm{dB}] = 10\log_{10}\frac{E[x_j^2(t)]}{E[y_j^2(t) - x_j^2(t)]} \qquad (4)$$

ここでE[·] は期待値を表す。ただし、左右の所望信号に対して一切補正を行わずに算出したSNRと提案手法を適用し算出したSNRとの差を改善量とし評価を行う。

表3 に3パターンそれぞれの場合の左右チャネルに対する改善量の平均が最大となった条件を示す。これらの結果より、重み付けパラメータに大きなばらつきが見られるが、ステップゲイン = 0.05 付近で最も高い改善量が得られた。また2本のクロストーク成分の相互相関係数 $r_N$ が大きい値を示すほど改善量も高い値を示している。

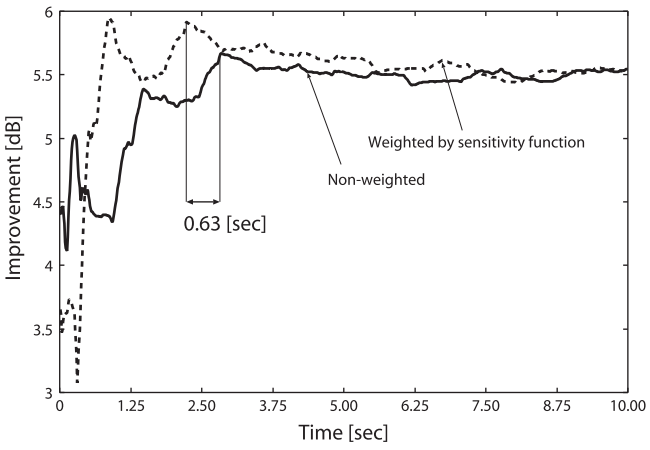


図3　収束特性の比較

### 3.3.2　シミュレーション2

出力誤差 eij を算出する際に聴覚特性の重み付けを行い、改善量の最も高い条件下で収束特性を比較する。ただし所望信号 $x_j(t)$ および観測信号 yj(t) のそれぞれの周波数帯域において $\phi$ (k) を重み付け、式（4）のSNR から改善量を算出する。

図3はパターン1 において $\alpha$ = 0.05, $w_1$=0.35, $w_2$= 0.25 を与え、改善量を比較したものである。この結果から聴覚特性の重み付けを行ったことで 0.63 秒程度の速度向上が確認された。

### 3.4　まとめ

本章ではクロストーク成分の相互相関に着目した新たな多入力信号補正法を提案した。計算機シミュレーションによりクロストーク成分の相互相関係数が大きいものほど、本手法が有効であることを示した。また適応フィルタ係数更新のための出力誤差を導出する過程で、聴覚特性に基づき重み付けを行うことにより収束速度を向上させることができた。今後は様々な環境の室内インパルス応答を与え、2本のクロストークにおける伝達特性の相互相関と最適な重み付けパラメータとの関連性を明らかにしていく必要がある。

## 4．むすび

本稿では、両眼視差法を用いた立体の3次元情報化と手話認識への応用の有効性を示した。また、最も簡単な構成で実現可能なステレオ型音場再現システムを提案した。

これらを活用することで、情報転送の効率化と情報資源の有効な活用が期待できる。

## 謝　辞

本研究は、文部科学省私立大学学術研究高度化推進事業の援助のもとで行われたものである。

## 文　献

[1] 山本拓，“手話単語認識における HMM の適用，” 平成 15 年度特別研究セミナー報告書 .

[2] 山本美子，“立体視を用いたロボット搭載ビジョンシステムの研究，”http://www.dsl.hiroshima-u.ac.jp/presen01/miko.pdf.

[3] K.Fujimura and M.Okada, “Recognition the sign language using the binocular parallax,” nternational Conference on Next Era Information Networking, NEINE-121,pp.441－445, Sept. 2004.

[4] P.A.Nelson, H.Hamada and S.J.Elliott,

“Inverse Filters for Multi-Channel Sound Reproduction,” IECE Trans. Fundamen-tals, vol.E75－A, no.11. pp.1468－1473, Nov. 1992.
[5] J.Nagumo and A.Noda, “A Learning Method for System Identication,” IEEE Trans. AC,vol.12, no.3, pp.282－287, 1967.
[6] “JIS C 1502-1990, 普通騒音計,” 日本工業規格 , 1997.
[7] S.Hamasaki and M.Fukumoto, “Stereo type sound reproduction systems using multi-input correction,” International Conference on Next Era Information Networking, NEINE-129, pp.468－475, Sept. 2004.
[8] 浜崎真二 , 福本昌弘 , “多入力信号補正によるステレオ型音場再生システム ,” 第 19 回信号処理シンポジウム講演論文集 , A4－3, Nov. 2004.