音響センサの知能化と視聴覚融合

高橋弘太

(1)

音響センサの知能化と視聴覚融合

1992年11月

髙橋弘太

目次

Table 0.1 記号表A（非時間関数）

種別	記号	意味　　実験やシミュレーションで採用した代表的な値
定数	J	独立な音源の個数　2～6
	M	マイクロホンの個数　6
	K	FIR フィルタのタップ数　32
	N	総タップ数＝線型結合器の入力数　$N=MK$　192
	L	$L=[K/2]$　16
	I	ブロック適応化の時のブロックあたりのサンプル数　220500
	T	ディジタル信号処理の時のサンプリング周期　22.7μs
	F	ディジタル信号処理の時のサンプリング周波数　44.1kHz
インデックス	j	音源の番号　$(j=0,1,2,\ldots,J-1)$
	m	マイクロホンの番号　$(m=1,2,3,\ldots,M)$
	n	FIR フィルタの通算タップ番号 $(n=1,2,3,\ldots,N)$
変数	f_n	FIR フィルタの係数　$(n=1,2,3,\ldots,N)$
	r_{nq}	$y_n(t)$ と $y_q(t)$ の相互相関　（$r_{y_n y_q}(0)$ のこと）
	p_n	$y_n(t)$ と $d(t)$ の相互相関　（$r_{y_n d}(0)$ のこと）
変数（行列など）	R	$y_n(t)$ の相関行列　$R=[r_{nq}]$
	$\boldsymbol{p}$	$y_n(t)$ と $d(t)$ の相関ベクトル　$\boldsymbol{p}=[p_n]$
	$\boldsymbol{f}$	FIR フィルタの係数ベクトル　$\boldsymbol{f}=[f_n]$
	$\boldsymbol{d}$	内部目標ベクトル　$\boldsymbol{f}=[d_i]$
	Y	タップ行列　$Y=[y_{ni}]$
演算記号等	$\mathrm{E}[\bullet]$	確率信号の集合平均
	$\overline{\bullet}$	確定信号の時間平均（平均サンプル数 $=I$）
	$\langle\bullet\rangle$	エルゴード性を有する確率信号の集合平均と， 確定信号の時間平均，の両方を指す記号
	$K[\bullet,\circ]$	$\langle\bullet(t)\circ(t)^2\rangle/\langle\circ(t)^2\rangle$
	$r_{\bullet\circ}(t)$	信号 $\bullet(t)$ と信号 $\circ(t)$ の相関（$r_{\bullet\circ}(\tau)=\langle\bullet(t)\circ(t+\tau)\rangle$）
	$R_{\bullet\circ}$	信号 $\bullet(t)$ と信号 $\circ(t)$ の相関値（$r_{\bullet\circ}(0)$ のこと）
	$B_\bullet$	帯域制限信号 $\bullet(\omega)$ の上限の周波数
	$\bullet_S(t)$	信号 $\bullet(t)$ の目的音成分
	$\bullet_N(t)$	信号 $\bullet(t)$ の妨害音成分
	$\bullet_C(t)$	信号 $\bullet(t)$ の搬送波成分
	$\otimes$	畳み込み積分の記号
物の記号	S_j	j 番目の音源　$(j=0,1,2,\ldots,J-1)$
	S	目的音源
	N_j	j 番目の妨害音源　$(j=1,2,\ldots,J-1)$
	M_m	m 番目のマイクロホン　$(m=1,2,3,\ldots,M)$

Table 0.2 記号表B（時間関数）

時間関数	時系列	周波数関数	意味
t	i	ω	
$s_j(t)$	s_{ji}	$S_j(\omega)$	各音源の信号 （$s(t), n_1(t), n_2(t), \ldots$ の総称）
$s(t)$	s_i	$S(\omega)$	目的音（target signal）
$n_0(t)$	n_{0i}	$N_0(\omega)$	$= s(t)$
$n_j(t)$	n_{ji}	$N_j(\omega)$	妨害音（disturbance signal） $(j = 1, 2, \ldots, J-1)$
$a(t)$	a_i	$A(\omega)$	目的音の強度エンベロープ［非定常信号］
$c(t)$	c_i	$C(\omega)$	目的音の搬送波成分［定常信号］ $(s(t) = a(t)c(t))$
$a_j(t)$	a_{ji}	$A_j(\omega)$	各音源の強度エンベロープ［非定常信号］
$c_j(t)$	c_{ji}	$C_j(\omega)$	各音源の搬送波成分［定常信号］ $(s(t) = a(t)c(t))$
$u_m(t)$	u_{mi}	$U_m(\omega)$	マイクロホン出力（離散系のときは AD 変換器の出力）
$y_n(t)$	y_{ni}	$Y_n(\omega)$	FIR フィルタのタップの信号 $(n = K(m-1)+k)$
$\psi(t)$	ψ_i	$\Psi(\omega)$	受信音を遅延した信号 （例えば $y_L(t)$）
$\theta(t)$	θ_i	$\Theta(\omega)$	Event Signal （事象生起推定量）
$\alpha(t)$	α_i	$\Pi(\omega)$	Cue Signal （手がかり量）
$\kappa(t)$	κ_i		Enable Signal （学習許可信号）
$d(t)$	d_i	$D(\omega)$	線型フィルタの適応目標 （内部目標: $d(t) = \alpha(t)\psi(t)$）
$\phi(t)$	ϕ_i	$\Phi(\omega)$	システム出力
$e(t)$	e_i	$E(\omega)$	推定誤差 $e(t) = \phi(t) - d(t)$
$h_{jm}(t)$	h_{jmi}	$H_{jm}(\omega)$	音源 S_j から マイクロホン M_m までのインパルス応答
$f_m(t)$	f_{mi}	$F_m(\omega)$	マイクロホン M_m に対するフィルタのインパルス応答
$g_{jn}(t)$	g_{jni}	$G_{jn}(\omega)$	音源 S_j から タップ y_n までのインパルス応答
$\boldsymbol{y}(t)$	$\boldsymbol{y}_i$		タップ信号のベクトル $\boldsymbol{y}(t) = [y_n(t)]$

第 1 章
はじめに

1.1 研究の概要

本研究は，マルチセンサを有するセンシングシステムに対して「新しい学習手法」を導入し,「センサの知能化」をはかろうとするものである．

ここで言う「センサの知能化」とは，複数の信号の加算された生のセンサデバイス出力から，目的とする信号のみを推定し，抽出することを指す．このための信号処理には線型フィルタを用いる．

また「新しい学習手法」とは，目的とする信号の事象生起推定量を元にした適応方法である．これは，具体的には目的とする音響信号の信号強度推定量である．本論文では，この推定量を事象が起こった度合を表わす量という意味で **Event Signal** と呼ぶ．

実際の適応では，Event Signal はそのまま用いられるのではなく，その時間平均を 0 にした信号が用いられる．本論文では，その信号を **Cue Signal** と呼ぶ．

本手法は，線型フィルタ適応のための参照信号（教師信号，目標信号）に，Cue Signal と生のセンサ出力の積信号を用いる．このように，適応目標をシステム内部で生成することで，学習のためのトレーニング期間等が不用となるという利点がある．このため，何らかの方法で Cue Signal さえ推定できれば，自律的な適応性をもったセンシングシステムを構成することができる．

筆者は，まず，Cue Signal を用いた適応型の音響センシングシステムについて研究した．このシステムは未知の音環境に置かれたとき，定常音（一定のパワーで放射される騒音）のなかから，非定常音（音声など）を取り出すなどの機能を有する．なお，実験は音響センサに限って行なっているが，本手法は，マルチセンサと線型フィルタを有するセンシングすべての適応化に有効な方法である．

5.2 節で述べるように，Cue Signal に課せられる条件は，目的音の強度に相関を持てばよいという統計的な条件である．Cue Signal に関する許容度は高い．そこで，Cue Signal を，

音響情報（聴覚情報）に限らず動画像（視覚情報）から生成したり，さらに知識ベースなどをも利用して生成することができる．

これは異種情報にまたがったセンサの知能化を意味する．または，最近注目を集めているセンサフュージョン（Sensor Fusion または Sensor Fusion and Integration）技術の一分野である視聴覚融合の具体例とみなすこともできる．

そこで筆者は，この新しい適応手法の応用として，視聴覚融合による知能化センシングシステムについて研究を行なった．このシステムでは，ビデオカメラで捉えた動画像をもとに，対象物の音強度を推定し，その対象物の音信号を出力するなどの動作を行なわせることが可能である．

また，本研究ではアルゴリズムの提案だけでなく，実際にリアルタイムで動作するセンシングシステムの試作を行なった．これにより，実験の効率化がはかれることはもちろん，信号形態の異なる異種情報を包括的にとりあつかうシステムを設計するときに問題点等を明らかにすることをめざした．

1.2 本論文の構成

本論文は，大きく2つの内容から構成される．これを第I部と第II部に分けた．

マルチセンサと線型フィルタからなるセンサシステムの自律的な適応手法，これを第I部とした．そして，この成果を応用した視聴覚融合型の知能化センサシステムについて，第II部で述べた．

第I部は以下のように構成した．

第2章は第1部の序章的役割を持った章である．ここでは，本研究の目的 — すなわち，どのようなセンサの知能化をめざしているのか — を明確にする．

次に第3章では，マルチセンサ情報の線型フィルタによる処理について考える．3.1節では，記号の定義と問題の定式化の準備をした後，平均2乗誤差最小の意味で最適な線型フィルタ係数についてまとめる．3.2節では，その最適な線型フィルタ特性を，与えられた学習信号から具体的に獲得するための手法のバリエーションを列挙する．この第3章は，以降の本論のための準備の章である．

第4章では，従来の研究についてサーベイを行なう．本論文で紹介する Cue Signal 法と比較するものとして，Widrow の適応ノイズキャンセラ，Flanagan のマイクロホンシステム，金田の AMNOR 受信方式の3つをとりあげ，それらの特徴や動作条件を本論文の手法とともに比較・検討する．

第5章と第6章が本論である．この2つの章で，本論文の主題である方法—Cue Signal による方法—について述べる

第5章では，5.1節で手法の概略を直感的に説明し，5.2節でなぜ Cue Signal で学習が可能かの証明とそのための条件を明確にする．5.3節では本手法の有効性を計算機シミュレーションで確かめる．さらに，5.4節で実際の音響信号を用いた実験で学習動作を確認する．

第6章では，第5章の結果に基づき，考察を行なう．まず，6.1節で Cue Signal 法の特徴をまとめる．次に，6.2節で，天下り的に導入された信号抽出の評価基準「平均2乗誤差最小の規範」の妥当性を再検討する．6.3節～6.5節の3つの節では，Cue Signal 法と各種適応アルゴリズム（LMS アルゴリズム，学習同定法，直接法）との相性を考察する．また，6.6節では音源モデルの解釈を拡張する．マイクロホンの本数，設置位置，フィルタの次数，等の最適化問題は，本手法特有の問題ではないが興味深い問題である．そこで，6.7節では，これ

らの問題に簡単に触れる．

これらの結果を，第 7 章でまとめる．以上が第 1 部である．

第 II 部の視聴覚融合は以下の構成をとる．

第 8 章は第 2 部の序章的役割を持った章である．ここでは，センサフュージョン技術と視聴覚融合について述べる．8.1 節でセンサフュージョン技術全般について紹介したあと，8.2 節で視覚情報と聴覚情報の場合のセンサフュージョン技術について，開発のキーポイントとなる事項を指摘する．また，ここでは視聴覚融合の他の具体的研究例として，視覚と聴覚による音声認識を紹介する．

第 9 章では，Cue Signal 法を視聴覚融合に応用する場合の具体的手法を述べる．9.1 節で概要を述べ，9.2 節では視覚情報から Cue Signal を生成するアルゴリズムについて記述する．また，9.3 節では実験の結果を記述する．

第 10 章では，視聴覚融合のために試作したリアルタイムシステムを紹介する．10.1 節で視覚部について，10.2 節で聴覚部について述べる．

第 11 章は，第 II 部の考察の章である．11.1 節では，視聴覚融合モデルを構成する立場から，Cue Signal 法による視聴覚融合が一体何をやっているのか改めて考え直してみる．11.2 節では，視覚センサで音源位置を計測してマイクロホンアレイの指向特性をむける方法に比較したとき，Cue Signal 法の長所は何であるのかを論じる．11.3 節では，さらなる発展問題として Event Signal の拡張を試みる．すなわち，従来のセンシング出力は，単に「値」を出力するか，あるいは，せいぜい「値・誤差の大きさ」を出力するものがほとんどだった．ここでは，Event Signal をそうした単純な「信号」から「メタ信号」とよぶべきものに一般化する．どのような一般化が考えうるのか，そしてそのメタ信号をどのようにして扱ったらよいのかを考察する．

最後に，第 12 章で 第 II 部のまとめを行なう．

第 I 部

音響センサの知能化

第 2 章
研究の目的

多数ある音源の発する音の中から特定の音源（以下**目的音源**と書く）の音（以下**目的音**と書く）を抽出することを考える．ここで，目的音源以外の音源を**妨害音源**と書き，その音を**妨害音**と書くことにする．妨害音の混入した生のセンサ出力（マイクロホン出力）から，目的音波形を線型フィルタで濾過するセンシングシステムを考える．

本研究では以下のような**自律的**な適応機能をもったセンシングシステムをめざす．

その第１は，トレーニング期間を必要としないことである．線型フィルタの学習においては，あるトレーニング時区間のあいだ外部から出力すべき信号（教師信号）が与えられれば，その教師信号と出力の間の２乗誤差を最小にすべく学習することはたやすい．

しかし，このような適応を必要とするシステムは自律的システムとは言い難い．このため，線型フィルタの適応に必要な信号はセンサシステム自らが内部で生成してやる必要がある．

その第２は，対象物（音源）の位置やスペクトル，音環境の条件（具体的には，部屋の寸法や，壁の反射率など），マイクロホンの位置などの先験情報を持たないことである．

これらの条件が明示されていれば，ある評価関数（例えば最小２乗誤差など）を定めることで最適なフィルタ特性は計算によって一意に導出することができる．しかし，このような先験情報は自律的適応性に反する．

以上２つをあわせて，「未知の世界に置かれたとき，自律的にその環境に適応して，目的とする仕事—すなわち音の抽出—を実行すること」，これを目標とする．

しかし，このように多数ある音源のなかから目的音源を選定する際に，出力すべき信号（教師信号）や先験情報を用いないという条件のもとでは，目的音源の選定に別の方法を用意しなければならない．これは，以下のようにする．

まず，生体の場合を考えてみよう．騒音のなかで音声を聞き取る場合である．人間は「話を聞こう」と意図しているだけで，話者の発声部位が前方 1200mm 左へ 15mm に位置するなどということは考えない．または，その音声の真の波形そのものを知って聴覚系をトレーニングしているわけでもない．音声を聞こうとしているだけである．

または，今，目の前でピンが落ちたとする．騒音の大きな環境でのこの小さな音であって

も，ビンが落ちる瞬間を目で捉えることが出来れば，その音を聞き分けることは，よりたやすいであろう．つまり，今見えた現象の音を聞き取ろうとしているのである．

このように，生体においては，音声を聞こうとか，目でみた特定の対象の音を聞きとろうなどという等という意思によって信号選択系が駆動されていると考えることが出来る．これは，先験情報や信号波形など単純物理量（数値）に比較して，より抽象的で意味レベルに近い規範であると言える．本研究では，目的音をこのような規範で指定することをめざす．

最後に，本論文でめざす知能化を他の手法と比較して整理すると **Table. 2.1** のようになる．最下段が本手法である．

Table 2.1 目的音指定の規範

規範	タイプ	事前に与える情報の具体例	適応方法
先験情報	数値	音源の位置，音源のスペクトル，マイクロホンの位置，音響伝達特性	評価関数に基づいた最小化問題
教師信号		トレーニングのための目的音波形，妨害音波形	トレーニング信号を出力するように学習
抽象的規範	意味	目的音は音声であるということ，異常音であるということ，あるいは視覚情報との関連で目的音を指示	Cue Signal を用いた方法

第 3 章

マルチセンサと線型フィルタによる信号抽出

この章では，マルチセンサ出力の中に線型の伝達関数で混入した特定の信号を，線型フィルタで推定する問題についてまとめておく．ここは以後の章のための準備の章である．

内容は，最適解の定式化（**3.1**節）とその解へ到達するための適応手法（**3.2**節）の 2 節より成る．

3.1 最適解

ここでは，以下の理論の準備として，線型フィルタによる最適フィルタリングについて整理しておく．本節は，後の議論のための理論的な前提をまとめるだけでなく，定式化の際の記号の定義を整理しておくために設けてある．

なお，本章の内容は音響センサに限定されるものではないが，簡潔に説明するために必要な場合は，音響センサの用語で記述した．

3.1.1 最適化の対象について

最適化とは，ある評価関数を定義して，その評価関数の最小化（最大化）によって行なわれることが一般的である．そして，信号の推定の場合は，この評価関数として，誤差の 2 乗平均などが用いられる．

この平均操作には，次の 2 とおりがある．

第一は，対象を確率的な信号として扱う方法である．この場合，最適化は集合平均された評価量に関して行なわれる．以下では，集合平均（期待値）を “$\mathrm{E}[\bullet]$” で表わす．対象がエルゴード的でない場合は，$\mathrm{E}[\bullet]$ には時間 t がパラメータとして残るが，対象がエルゴード的である場合は，$\mathrm{E}[\bullet]$ は時間 t に依存しない量となる．

第二は，対象を確定的な信号として扱う方法である．最適化は，時間平均された評価量に関して行なわれる．以下では，時間平均を “$\overline{\bullet}$” で表わすものとする．

さて，以下の最適化の理論では，対象がエルゴード的である場合の期待値 $\mathrm{E}[\bullet]$ による評価と，確定信号の時間平均 $\overline{\bullet}$ による評価のどちらでも成立する式が非常に多い．そこで，この 2 つの平均の両方で成立する式には平均の記号として "$\langle \bullet \rangle$" を用いることにする．

3.1.2 定式化

ここでは，計測対象である音響信号や，その他の信号の定義をしておく．また，それら基本信号の相互関係を定式化しておく．(記号表は，目次の次のページから 2 ページにわたって一覧にしてある)

まず，ある対象となる空間を考える．その空間には残響があってもなくてもよい．空間内に独立な音源が J 個あるとする．それぞれの音源の発する音を $s_0(t), s_1(t), s_2(t), \ldots, s_J(t)$ とする．これらのなかで，システムが抽出すべき音を**目的音**と呼び $s(t)$ と書く．また，目的音以外の音を**妨害音**と呼び，$n_1(t), n_2(t), \ldots, n_J(t)$ と書く．(目的音が単一の場合は，とくにことわらないかぎり，$s(t) = s_0(t)$, $n_j(t) = s_j(t)$ であるとする．)

目的音抽出のため，空間内に M 個のマイクロホンを置く．これを，$\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_M$ とする．また，それぞれの出力を $u_1(t), u_2(t), \ldots, u_M(t)$ とする．

以下では，特にことわらないかぎり，信号の伝達に非線型な変換は行なわれないものとする．

このとき，マイクロホン信号 $u_m(t)$ と，音源信号 $s_j(t)$ の関係は以下のように空間の伝達関数によって記述できる．

$$u_m(t) = \sum_{j=0}^{J} \int s_j(t-\tau)\, h_{jm}(\tau)\, d\tau \tag{3.1}$$

周波数領域では，

$$U_m(\omega) = \sum_{j=0}^{J} S_j(\omega)\, H_{jm}(\omega) \tag{3.2}$$

ここで，$h_{jm}(\tau)$ は，音源 S_j からマイクロホン M_m までのインパルス応答である．これには，音源の位置，マイクロホンの位置，部屋の壁の位置と反射係数などのすべての音響条件が反映されている．

次に，センサシステムの線型フィルタも含めて記述しておこう．

マイクロホン M_m に対する線型フィルタのインパルス応答を $f_m(\tau)$ とかけば，システム出力は，

$$\phi(t) = \sum_{m=1}^{M} \int u_m(t-\tau)\, f_m(\tau)\, d\tau \tag{3.3}$$

と書け，周波数領域では，

$$\Phi(\omega) = \sum_{m=1}^{M} U_m(\omega)\, F_m(\omega) \tag{3.4}$$

$$= \sum_{m=1}^{M} \sum_{j=0}^{J} S_j(\omega)\, H_{jm}(\omega)\, F_m(\omega) \tag{3.5}$$

と書ける．ここで，周波数をパラメータとした行列（ベクトル）を，

$$H(\omega) \stackrel{\mathrm{def}}{=} [H_{jm}(\omega)] \qquad (J \times M \text{ 行列}) \tag{3.6}$$

$$F(\omega) \stackrel{\mathrm{def}}{=} [F_m(\omega)] \qquad (M \times 1 \text{ 行列}) \tag{3.7}$$

$$U(\omega) \stackrel{\text{def}}{=} [U_m(\omega)] \qquad (1 \times M \text{行列}) \tag{3.8}$$

$$S(\omega) \stackrel{\text{def}}{=} [S_j(\omega)] \qquad (1 \times J \text{行列}) \tag{3.9}$$

$$\Phi(\omega) \stackrel{\text{def}}{=} [\Phi(\omega)] \qquad (1 \times 1 \text{行列}) \tag{3.10}$$

のように定義すれば，以下のように書くこともできる．

$$\Phi(\omega) = U(\omega)\,F(\omega) \tag{3.11}$$

$$= S(\omega)\,H(\omega)\,F(\omega) \tag{3.12}$$

3.1.3 音源信号の完全推定

音源信号をマイクロホン信号から完全に誤差ゼロで推定するための条件を述べる．

例えば，目的音 $S_0(\omega)$ の推定を考える．これを完全に推定するということは，式(3.12)より，

$$[S_0(\omega)] = S(\omega)\,H(\omega)\,F(\omega) \tag{3.13}$$

$$= [S_0(\omega)\ S_1(\omega)\ \cdots\ S_J(\omega)]\ H(\omega)\,F(\omega) \tag{3.14}$$

ということである．これが任意の $S_j(\omega)\,(j = 0, 1, \ldots, J)$ に対して成り立つためには，

$$H(\omega)\,F(\omega) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T \tag{3.15}$$

でなければならない．$F(\omega)$ を無制限としても，式(3.15)は，行列 $H(\omega)$ の列ベクトルの張る空間にベクトル $[1\,0\,0\cdots 0]^T$ が含まれていなければならないことを示している．

$H(\omega)$ は $J \times M$ 行列である．ベクトル $[1\,0\,0\cdots 0]^T$ は J 次ベクトルである．

$M < J$ のとき，$H(\omega)$ のランクは，J より小さい．$H(\omega)$ の列ベクトルは，J より小さい空間を張る．そこに，J 次のベクトル $[1\,0\,0\cdots 0]^T$ は一般には含まれない．

$M \geq J$ のとき，$H(\omega)$ が $M - J + 1$ 以上ランク落ちしないかぎり，その列ベクトルは，J 次元空間を張ることができる．よって，J 次のベクトル $[1\,0\,0\cdots 0]^T$ はそこに含まれる．しかし，ここで重要なのは，これがすべての ω に対して成立しなければならないということである．一般には，ω のいくつかの点で $H(\omega)$ がランク落ちして，ランクが J 以下になることは充分考えられる．特に，$J = M$ のときはその可能性は高い．

以上より，任意の広帯域信号のすべての周波数について完全推定を行なうためには，$M = J$ もしくは $M = J + 1$ が目安となると言える．もちろん，特定の音源配置や特定の音源スペクトルの場合には，$M < J$ でも完全推定ができることが有り得る．

また，音波伝達の遅延があるので，完全推定をするためのフィルタは一般には因果率に反するものとなる．ただし，遅延を許した推定ならば場合によっては実現可能である．

最後に，マイクロホンや，マイクロホンプリアンプ，A/D 変換器などのノイズに対する影響について述べる．

実は，これらのノイズが系に与える影響は，それが加算的なものであると仮定すれば，妨害音 $s_j(t)$ による影響と形式的には全く同じものである．

よって，これらのノイズは見かけ上音源の個数 J が増えたものとして作用する．各センサに独立なノイズと，共通なノイズ（電源等に起因するもの）があるので，増加する個数は最低でも，$M + 1$ 程度と考えられる．

形式的な音源数が $J + M + 1$ になってしまうため，音源信号の完全推定は，実際には不可能である．

しかし，これらシステムノイズの信号レベルは音源からの信号レベルに比較して微小であるので，さきに述べた $M \geq J$ または $M \geq J+1$ という目安は有効であることが多い．

3.1.4 最適な線型近似量

M チャネルの入力 $u_m(t)$ 〈 $i = 1, 2, \ldots, m$ 〉と，何らかの**適応目標** $d(t)$ が与えられたとき，各チャネルからのインパルス応答が $f_m(t)$ である線型演算

$$\phi(t) = \sum_{m=1}^{M} \int f_m(\tau)\, u_m(t-\tau)\, d\tau \tag{3.16}$$

によって，$d(t)$ の近似値 $\phi(t)$ を算出する場合を考える．このとき，近似誤差 $e(t) \stackrel{\text{def}}{=} d(t) - \phi(t)$ の 2 乗平均 $\langle e(t)^2 \rangle$ を最小にするインパルス応答 $f_m(t)$ は以下のように算出される．

まず，最小にすべき量 $\langle e(t)^2 \rangle$ は次式で表される．

$$\begin{aligned} \langle e(t)^2 \rangle &= \langle\, [d(t) - \phi(t)]^2 \,\rangle \\ &= r_{dd}(0) - 2 \sum_{m=1}^{M} \int f_m(\tau)\, r_{u_m d}(\tau)\, d\tau \\ &\quad + \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} \iint f_{m_1}(\tau_1)\, f_{m_2}(\tau_2)\, r_{u_{m_1} u_{m_2}}(\tau_1 - \tau_2)\, d\tau_1 d\tau_2 \end{aligned} \tag{3.17}$$

ここで，$r_{ab}(\tau)$ は，信号 a と b との相関を表わすものとする〈$r_{ab}(\tau) \stackrel{\text{def}}{=} \langle a(t)\, b(t+\tau) \rangle$〉．

式(3.17) を最小とする $f_m(\tau)$ を変分法で解くと[1]，次の Wiener-Hopf の積分方程式が得られる．

$$r_{u_k d}(t) = \sum_{m=1}^{M} \int r_{u_k u_m}(t-\tau)\, f_m(\tau)\, d\tau \qquad (k = 1, 2, \ldots, M) \tag{3.18}$$

これを，フーリエ変換して，周波数軸上で表現すれば，

$$R_{u_k d}(\omega) = \sum_{m=1}^{M} R_{u_k u_m}(\omega)\, F_m(\omega) \qquad (k = 1, 2, \ldots, M) \tag{3.19}$$

となる．ただし $R_{\bullet\bullet}(\omega), F_{\bullet\bullet}(\omega)$ は，それぞれ $r_{\bullet\bullet}(\tau), f_{\bullet\bullet}(\tau)$ のフーリエ変換を表す．

以上のように，すべての線型フィルタのなかで，2 乗誤差最小という意味で最適なフィルタは，方程式(3.18) または (3.19) で表わされる．

また，式(3.19) は行列形式で

$$\begin{bmatrix} R_{u_1 d}(\omega) \\ R_{u_2 d}(\omega) \\ \vdots \\ R_{u_M d}(\omega) \end{bmatrix} = \begin{bmatrix} R_{u_1 u_1}(\omega) & R_{u_1 u_2}(\omega) & \cdots & R_{u_1 u_M}(\omega) \\ R_{u_2 u_1}(\omega) & R_{u_2 u_2}(\omega) & \cdots & R_{u_2 u_M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ R_{u_M u_1}(\omega) & R_{u_M u_2}(\omega) & \cdots & R_{u_M u_M}(\omega) \end{bmatrix} \begin{bmatrix} F_1(\omega) \\ F_2(\omega) \\ \vdots \\ F_M(\omega) \end{bmatrix} \tag{3.20}$$

とも書ける．最適化をはかる全ての周波数 ω に対して上式が成り立たなければならない．

なお，上式は，式(3.12) の左辺 $\Phi(\omega)$ を，

$$D(\omega) \stackrel{\text{def}}{=} [\, D(\omega) \,] \qquad (1 \times 1 \text{行列}) \tag{3.21}$$

で置き換え，両辺について行列 $U(\omega)$ との相互相関をとったものに等しい．これは，$U(\omega)$ と相関のある成分のみで比較すれば，適応目標 $D(\omega)$ と出力 $\Phi(\omega)$ が等しいということである．

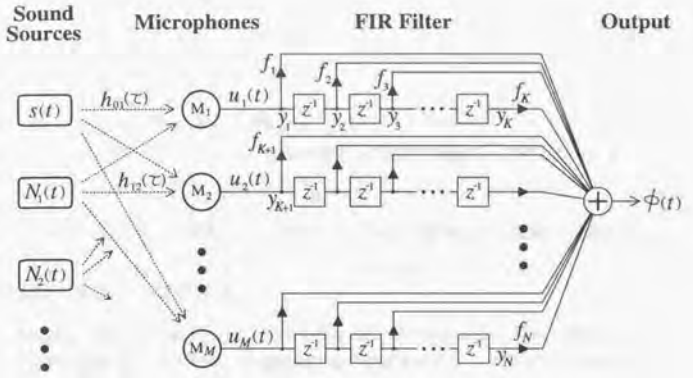


Fig. 3.1 FIR フィルタによる信号の流れ

3.1.5 FIR フィルタを用いた場合

特に，**Fig. 3.1** に示すように，式(3.19) の特性 $F_m(\omega)$ を，k 次の FIR フィルタで実現することを考える．(本研究では，すべてこの形式の線型フィルタを用いている．) 式(3.16) のインパルス応答 $f_m(\tau)$ は，一般には無限の長さを持つので，FIR フィルタでは実現不可能である．$F_i(\omega)$ を FIR フィルタで近似してもよいが，与えられた自由度の範囲内での最適性は保証されない．そこで，以下で改めて最適条件を求める．

Fig. 3.1 の積和をとる部分は，適応線型結合器（adaptive linear combiner）と呼ばれる．[2] 適応線型結合器の最適化を調べればよい．

適応線型結合器の $N = M \cdot K$ 個の入力を，$y_n(t)$ 〔 $i = 1, 2, \ldots, N$ 〕 とする．ここで，$y_n(t)$ は，$u_m(t)$ または，$u_m(t)$ を遅延させた信号である．出力を，

$$\phi(t) = \sum_{n=1}^{N} f_n y_n(t) \tag{3.22}$$

と，$y_n(t)$ の実係数の一次結合で生成するとすれば，近似誤差の 2 乗平均

$$\begin{aligned} \langle e(t)^2 \rangle &= \langle [d(t) - \phi(t)]^2 \rangle \\ &= \left\langle \left[d(t) - \sum_{n=1}^{N} f_n y_n(t) \right]^2 \right\rangle \end{aligned} \tag{3.23}$$

を最小にする f_i の満たすべき条件（式(3.18) に相当する方程式）は，次のようになる．

$$R_{y_k d} = \sum_{n=1}^{N} R_{y_k y_n} f_n \tag{3.24}$$

これを，行列形式で表示すると，

$$\begin{bmatrix} R_{y_1 d} \\ R_{y_2 d} \\ \vdots \\ R_{y_N d} \end{bmatrix} = \begin{bmatrix} R_{y_1 y_1} & R_{y_1 y_2} & \cdots & R_{y_1 y_N} \\ R_{y_2 y_1} & R_{y_2 y_2} & \cdots & R_{y_2 y_N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{y_N y_1} & R_{y_N y_2} & \cdots & R_{y_N y_N} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} \tag{3.25}$$

である．上式の行列・ベクトルを，左から順に $\boldsymbol{p}, R, \boldsymbol{f}$ で定義し，以下のように書く．

$$\boldsymbol{p} = R\,\boldsymbol{f} \tag{3.26}$$

最適なFIRフィルタの係数 f_n は式(3.24)（または式(3.25)，式(3.26)）で定められる．

3.2 適応アルゴリズム

ある適応目標が得られたとして，その目標とFIRフィルタ出力の平均2乗誤差を最小にするフィルタ係数 f_n（すなわち，方程式(3.25)で決定される f_n）を求める具体的手法ついて述べる．

ところで，本論文で述べるCue Signal法は，センサ情報から適応目標をどのように生成するかという点に主眼を置いたものである．それゆえ，適応目標からフィルタ係数を得る方法には，既存の手法をもちいることとした．

このようなわけで，本節は，簡単にまとめた．すなわち，既存の手法の中からどの手法を選択すればCue Signal法に適しているかを後に論じるのに十分な程度に，主要な手法を列挙するにとどめた．

ここでは，LMSアルゴリズム，学習同定法，RLSアルゴリズム，直接法の4つを取りあげる．なお，FIRフィルタに限定しないのであれば，周波数領域での適応フィルタ[3)4)]など，他にも多数の方法がある．

3.2.1 LMSアルゴリズム

FIRフィルタの誤差の2乗平均は，**3.1.5**の式(3.23)で示した．誤差の2乗平均 $\langle e(t)^2 \rangle$ の値による等高面を f_n 空間で書いてみれば，それは，式(3.23)より，N 次元の楕円面になっていることがわかる．そして，到達すべき誤差2乗平均最小の点はその中心にある．そこで，ある時点での f_n をもとに，最急勾配法で漸近的に f_n を修正していく方法が考えられる．すなわち，勾配は

$$\frac{\partial \langle e(t)^2 \rangle}{\partial f_n} = -2 \langle e(t)\, y_n(t) \rangle \qquad (n = 1, 2, \ldots, N) \tag{3.27}$$

であるので，

$$[f_n]_{\mathrm{NEW}} = [f_n]_{\mathrm{OLD}} + \mu \langle e(t)\, y_n(t) \rangle \qquad (n = 1, 2, \ldots, N) \tag{3.28}$$

と修正するのが**最急勾配法**である．ここで定数 μ は修正係数である．

最急勾配法の欠点は，平均操作 $\langle \bullet \rangle$ を必要とする点である．修正の1ステップに長い時間平均を要するのでは，実時間で適応させる場合には適さない．そこで，最急勾配法の平均操作を省いてしまったのが，確率勾配法である**LMSアルゴリズム**（最小2乗平均アルゴリズム，Least-Mean-Square algorithm）である．1960年にWidrowtとHoff[5)]により提案された．

$$[f_n]_{\mathrm{NEW}} = [f_n]_{\mathrm{OLD}} + \mu\, e(t)\, y_n(t) \qquad (n = 1, 2, \ldots, N) \tag{3.29}$$

LMS アルゴリズムには，修正量 μ を可変にしたり，直接修正するのではなく，ローパスフィルタをかけて修正するなど，数々のバリエーションがある．

また，係数の収束条件についても様々な議論がある．しかし，非定常信号の場合については，収束を保証する統一的な理論はない．本研究の対象は，まさに非定常信号であるので，LMS アルゴリズムが有効であるか否かは実験で確かめることとする．

なお，定常音の場合の係数の期待値の収束条件は，（修正繰り返しの間隔が，異なる時間の y がすべて統計的に独立となる程度に十分長い，と仮定した場合）

$$0 < \mu < 2/\lambda_{\max} \tag{3.30}$$

である．[6][7] ここで $\lambda_{\max}$ は，相関行列 R（式(3.26) 参照）の最大固有値である．

LMS アルゴリズムの長所は計算量が非常に少ない（N の 1 次のオーダー）ことである．一方，欠点は，修正係数 μ の最適値が信号強度に依存するということ，他の方法に比較して収束が遅いということ，特に非定常信号の場合には収束の保証が得られないこと，などである．

3.2.2 学習同定法

南雲・野田の**学習同定法**[8][9]は，信号強度で修正係数 μ を変えるタイプの LMS アルゴリズムと見ることもできる．このため Bitmead と Anderson[10]は，これを正規化 LMS アルゴリズム（NLMS algorithm）とよんだ．学習同定法では，修正係数はタップベクトル $\boldsymbol{y}$ のノルムの 2 乗に反比例させられる．

$$[f_n]_{\mathrm{NEW}} = [f_n]_{\mathrm{OLD}} + \frac{\nu}{\sum y_n(t)}\, e(t)\, y_n(t) \qquad (n = 1, 2, \ldots, N) \tag{3.31}$$

このようにすると，収束条件が修正係数 ν に関して

$$0 < \nu < 2 \tag{3.32}$$

と，受信信号の強度に依存しなくなる．これが大きな特徴である．

なお，$\nu = 1$ の場合の修正を $\boldsymbol{f}$ 空間で幾何学的に見れば，式(3.31) は，$[\boldsymbol{f}]_{\mathrm{OLD}}$ を，$\boldsymbol{d} = \boldsymbol{yf}$ 平面（現在のサンプル値 $\boldsymbol{y}$ について誤差を 0 にする $\boldsymbol{f}$ の作る平面）へ降ろした垂線の足 $[\boldsymbol{f}]_{\mathrm{NEW}}$ へ移動させることに他ならない．

学習同定法の，LMS アルゴリズムに比較した欠点は，計算量が僅かに多くなることであろう（しかし，依然 N の 1 次のオーダー）．また，他の利点として，普通の LMS アルゴリズムに比較して速く収束することが示されている．[11]

3.2.3 RLS アルゴリズム

ある時間区間について受信したデータをもとに式(3.25) を解けば，実行可能ななかでは最適な解 f_n を得ることができる．しかし，それには連立 1 次方程式を解かなければならない．これは，センサ数やフィルタ次数が増加すると膨大な計算量を必要とする．しかも，実時間で適応させるためには，新しいデータ y_n が入ってくるたびに f_n を更新しなければならない．

しかし，もし時点 i において，相関行列 R_i の逆行列 Q_i が得られているならば，時点 $i+1$ での Q_{i+1} は N の 2 乗オーダーの計算で算出することができる．これを利用して，1 ステップごとに f_n を更新するのが**RLS アルゴリズム**（再帰最小 2 乗アルゴリズム，Recursive Least Square algorithm）である．この手法は，1950 年に Plackett[12]により提案された．具体的に

は，以下の操作を繰り返すことで，次々と新しい $\boldsymbol{f}_i$ を得ることができる．

$$\boldsymbol{k}_i = \frac{Q_{i-1}\,\boldsymbol{y}_i}{1 + \boldsymbol{y}_i^T Q_{i-1}\,\boldsymbol{y}_i} \tag{3.33}$$

$$e_i = d_i - \boldsymbol{f}_{i-1}\,\boldsymbol{y}_i \tag{3.34}$$

$$\boldsymbol{f}_i = \boldsymbol{f}_{i-1} + \boldsymbol{k}_i e_i \tag{3.35}$$

$$Q_i = Q_{i-1} - \boldsymbol{k}_i \boldsymbol{y}_i Q_{i-1} \tag{3.36}$$

ここで，時点 i の タップ信号である $y_{ni}\ (n = 1, 2, \ldots, N)$ を，$\boldsymbol{y}_i$ などとベクトルで書いた．

その特徴は，現在までに知っている情報を用いた最適解そのものが求まるということである．なお，RLS アルゴリズムは，カルマンフィルタ理論を FIR フィルタに当てはめたものと見ることもできるので，カルマンアルゴリズムともよばれる．

しかし，逆行列計算が不用になったとは言っても，計算量は N の 2 次のオーダーはあるわけで，センサ数やタップ数が増加すると実時間動作は困難になってくる．

3.2.4 直接法

式(3.25) をそのまま解く方法を，本論文では直接法とよぶことにする．もちろん，有る程度のセンサ数とタップ数を有するシステムになると，各サンプリング時点ごとに f_n を求めるのは実際上不可能である．

第 4 章
従来の研究

本章では，従来提案された手法のうち，最小限の先験的知識（信号源の位置，信号源とセンサ間の伝達関数，信号の統計的性質）で適応可能な手法をまとめる．これは，言い替えれば，できるだけ受信信号のみから適応可能な手法である．いずれも，複数のセンサと線型フィルタで信号の抽出を行なうものである．

また，本章の最終節では，それらの方法を，「必要とする先験的知識」と「目的音指定の規範」の２点で整理し，比較・検討を行なう．

4.1 適応ノイズキャンセラ

適応ノイズキャンセラ[13]の概念図を **Fig. 4.1** に示す．

適応ノイズキャンセラでは，各センサは対等ではなく，主センサ（primary input）と参照センサ（reference input）に分けられる．Fig. 4.1では，参照センサは１個であるが，複数の参照センサを用いることもできる．

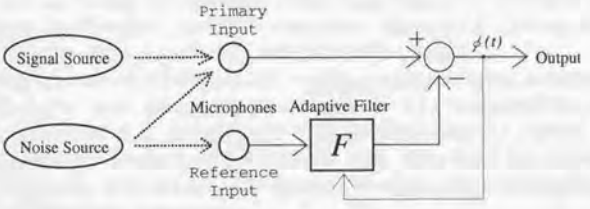


Fig. 4.1 適応ノイズキャンセラ

適応のノイズキャンセラの動作の条件は，目的信号源と参照センサ間の伝達関数が０である（または非常に小さい）ことである．

この条件のもとで，適応ノイズキャンセラは，参照センサ出力のみを適応フィルタにかけて主センサ出力との差の２乗 $\phi(t)^2$ を最小にすべく学習を行なう．この結果，システム出力 $\phi(t)$（主センサ出力と適応フィルタ出力の差）における妨害信号成分の強度が最小化される．一方，目的信号に関しては，その伝達特性を変化させる機構が存在しないので，主センサで受信された目的信号成分はそのまま出力される．これによって，目的信号の抽出が行なわれるのである．

実際の応用では，目的信号源と参照センサ間の伝達関数が完全に０になることは少ない．Widrow は，これについても解析を行い，参照センサの SN 比が０から増加していったとき，システム出力の SN 比や目的信号劣化比がどのように劣化していくかを示している．[13]

いずれにしても，本手法は強大・大域的な妨害信号のなかから微弱・局所的な目的信号を抽出する場合には有効な方法である．例えば，商用周波数のノイズの存在する環境下で生体の微弱な信号を検出する場合など，このような状況は数多く存在するので，適応ノイズキャンセラの応用範囲は広い．

適応ノイズキャンセラの目的信号指定の規範は，「参照センサへの信号経路が遮断され，しかも主センサには信号が伝達されている信号」である．言い替えれば，信号源－センサ間の伝達特性で目的信号と妨害信号を分割する手法である．

なお，Widrow らと初期においては独立に発展したアダプティブアンテナ技術は，本節で紹介したものと本質的には同じものである．[14] 主センサに相当するアンテナに指向特性の鋭い物を利用し，指向特性のピークの方向にある信号源がターゲットとして規定されるのである．

4.2 Flanagan の手法

Flanagan ら[15]は，**Fig. 4.2** に示すように計算機で指向特性を自動走査し，音声を集音するシステムを提案している．

マイクロホンを縦７×横９の６３個，格子型に配列し，各マイクロホンに対して遅延を調節してやりその和をとることで，任意の方向の指向特性が実現できる．この６３個の遅延量を計算機で制御してやれば，自律的な集音システムを実現することが可能である．

指向特性の選択は，以下のように行なう．まず，全体をくまなく走査する．次に，音声を検出したかどうかを判定する．これは，基音の周波数，全体の周波数包絡（約５００ Hz 以上で８～１０ dB/oct で減衰），時間軸上でのエネルギーの集中などの規範で，バックグラウンドノイズと区別することが可能である．ただし，文献[15]では計算速度の制約から，３０ ms と１５０ ms での平均パワーの比が前もって決めた閾値を超えたかどうかで音声の検出の有無を判定している．最後に，音声が有りと判定された（閾値を超えた）方向のなかで３０ ms での平均パワーが最も大きいものを，次の集音の方向として採用する．

実際には，２つのフィルタを用意しておき，一方をシステム出力用（抽出した現在の発言者の音声用）にあて，もう一方は次の発言者を探すために用いるという方法が提案されている．

さて，この方法の欠点は，周波数が低下するほど指向特性が悪化することである．また，$M = 63$ 個のマイクロホンは前もってその周波数特性（利得，位相）を揃えておく必要がある．

文献[15]によれば，マイクロホンは，800～2000Hz において± 1.0dB 以内に周波数特性（利得）の揃った物を選別して用いている．

また，マイクロホン１個あたりの適応の自由度が僅かに１（遅延時間）しかないという問題もある．これは，マイクロホン１個につきパラメータが数個から数十個ある線型フィルタ

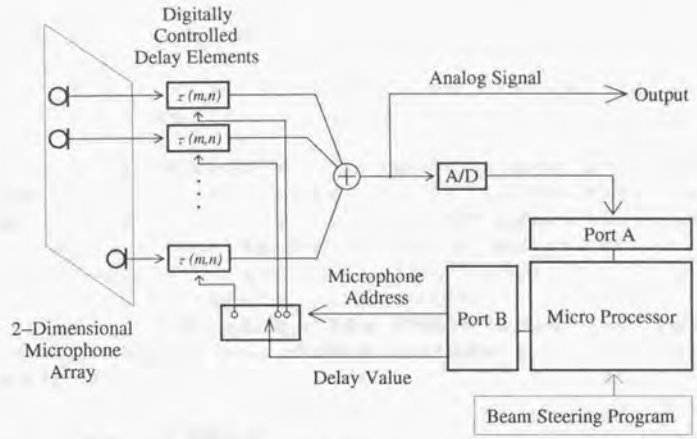


Fig. 4.2 自動走査型マイクロホンアレイ

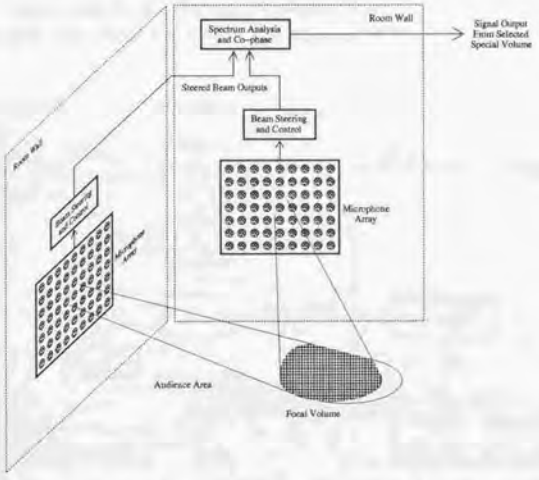


Fig. 4.3 3次元選択性を持たせたアレイマイクロホン

を用意するのと比較すれば，かなり少ない値である．さらに，その M 個の自由度を，上下角と左右角の2個の自由度に絞っている．このため，目的音に対する適応のみが行なわれ，ノイズ方向にシステム伝達特性の零点を持って来るような適応は含まれない．

一方，以下のような利点がある．第一に，目的音の劣化は，指向特性のピークを正確に音源に合わせれば原理上生じない．第二に，目的音自身の残響（反射音）も抑制できる．これらは，マイクロホン位置とマイクロホン特性（相対的特性）を既知としたために生じた利点である．

有効に応用できる条件を考える．低い周波数の選択性を向上させるためには，マイクロホンアレイ全体の大きさを大きくする必要がある．また，本システムには上で述べたように残響を取り除く効果がある．以上2点より，本システムは，講堂や会議場のように（可聴音の波長に比較して）広い部屋で，残響を抑制した集音を行ないたい場合に有用であると言える．

さらに，Flanagan らは **Fig. 4.3** に示すように，2次元アレイのマイクロホンシステムを2セット用いて3次元の走査を行なうことも提案している．[15]

いずれにしても，本手法の目的音指定の規範は，目的信号の性質である．これは，本論文の趣旨に近い．具体的には，「150ms の時区間で信号強度の変動がある」が目的音としての音声の規範である．

4.3 AMNOR受信方式

金田らによって提案された AMNOR 受信方式[16]は，内部で生成した仮想的な目的信号を用いた適応処方である．このような目的信号は Widrow ら[17]が適応アンテナのテクニックのひとつとして用いたパイロット信号と本質的には同じものである．しかし，金田らの方法は，その仮想目的信号の信号レベルを適応的に変化させることで，SN 比と目的信号劣化量のトレードオフを自由に調節できるようにしたことが特徴である．

AMNOR 受信方式のブロックダイアグラムを **Fig. 4.4** に示す．

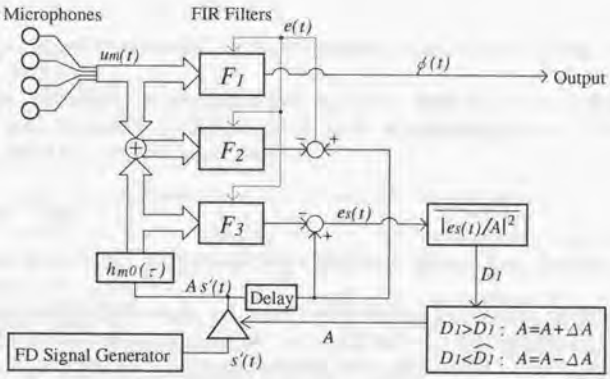


Fig. 4.4 AMNOR受信方式

本手法は次の2条件が満たされたとき動作する．

1. 目的信号の到来方向が既知であること．
2. 目的信号には無音時間区間があり，そのとき抑圧すべき音のみが受信されること．

Fig. 4.4のブロックダイアグラムを簡単に説明する．

システムは3つの適応フィルタ F_1, F_2, F_3 を持つ．それらの係数（特性）はすべて連動している．マイクロホン信号は，F_1 と F_2 に加えられている．一方，仮想目的信号は既知情報（目的信号の到来方向）に基づいた目的音源－マイクロホン間の伝達特性（$h_{m0}(\tau)$）をかけられて，F_2 と F_3 に加えられている．

F_1 は，システム出力用である．抽出された目的音を出力する．F_2 は，適応用である．これは，Widrow[17]の'Auxiliary Adaptive Processor'に相当する．F_2 では，目的音無音の時間区間で妨害音と仮想目的信号を入力し，仮想目的信号を最小2乗誤差で推定するフィルタ係数を学習する．こうして得られたフィルタの係数は，F_1 と F_3 にもコピーされる．

F_3 は目的音劣化量の評価用である．仮想目的信号のみを現在のフィルタに通して，仮想目的信号と引算することで目的音の劣化量を算出する．こうして得られた目的音劣化量 D_1 を，前もってきめておいた目的音劣化量の限界値 $\widehat{D_1}$ と比較し，限界を超えていれば，仮想目的信号のレベルを上げ，限界より小さければ，仮想目的信号のレベルを下げてより高いSN比を狙う．

仮想目的信号のレベルで，SN比・目的信号劣化比を制御できることは，以下のように説明できる．

適応フィルタ F_2 のフィルタリング誤差の2乗平均 $\overline{[e(t)]^2}$ は，仮想目的信号と妨害音が無相関な定常確率過程であるとすれば，

$$\langle e(t)^2 \rangle = A^2 \langle [s'(t) - \sum_m f_m(t) \otimes h_{m0}(t) \otimes s'(t)]^2 \rangle \tag{4.1}$$

$$+ \langle [\sum_j \sum_m f_m(t) \otimes h_{mj}(t) \otimes n_j(t)]^2 \rangle \tag{4.2}$$

$$\stackrel{\text{def}}{=} A^2 D_1 + D_2 \tag{4.3}$$

のように，目的信号の劣化に関する項 D_1 と，妨害音の混入に関する項 D_2 とに分離して書き表すことができる．

最適化の評価関数が，2つの評価関数の和になっており，仮想目的音レベル A が両者の加重比を表わしているため，A を調節することで，SN比・目的信号劣化比とのトレードオフを自由に調節することが可能となるのである．

4.4 比較

本章で述べた手法と，本研究でめざす手法を比較すれば，**Table. 4.1** のようになる．項目別に見ていこう．

最初の比較項目「動作の条件」は，各方法が使用可能となるための制約や，先験的に必要な知識である．Widrowの適応ノイズキャンセラでは目的音－センサ間の伝達関数 $h_{0m}(\tau)$ が0であることが条件である．その条件を満たすセンサ m が参照センサとなる．Flanaganのアレイマイクシステムでは，指向特性を制御するためにマイクロホンの位置と特性という先験的知識が必要である．金田のAMNOR受信方式では，仮想目的信号から各マイクロホンでの仮想目的音を生成するために目的音－センサ間の伝達関数 $h_{0m}(\tau)$ が必要である．具体的

Table 4.1 従来研究と本研究でめざす手法の比較（i）

手法	動作の条件	目的音の規範
適応ノイズキャンセラ (Widrow)	$\exists m, h_{0m}(t) \equiv 0$	伝達関数 主センサはSN比が高く， 参照センサはSN比が低い．
アレイマイクシステム (Flanagan)	マイク配置・特性が既知	信号の性質 150msの時区間で，強度が 時間変化する信号
AMNOR (Kaneda)	$s(t)=0$ の時区間の存在， h_{0m}がすべて既知	音源位置，特定時区間で $s(t)=0$
本研究でめざす手法	目的音と妨害音の強度変化が同一でないこと	意味レベルの規範（音声，異常音，視覚的変化を伴う音，など）

にはマイクロホンの位置と目的音源の位置を先験的知識として用いている．また，学習は目的音がゼロ ($s(t)=0$) の時区間でおこなわれるので，その時区間を用意してやらなければならない．

本論文ではこのような伝達関数に対する先験的知識を不用とする手法をめざしている．なお，あとで（第5章）述べるが，本論文で述べる方法（Cue Signal 法）での動作条件は「目的音と妨害音の強度変化が全く同一ではないこと」というものである．

次の比較項目「目的音の規範」は，混合して受信される多数の音信号の中でどの信号が目的信号であるかをどのような規範で定義するかである．Widrow の適応ノイズキャンセラでは，参照センサへの伝達関数がほぼ0であることをもって目的音とする．Flanagan のアレイマイクシステムでは，受信した信号の性質から目的音強度が強いかどうかを決定する．具体的には，音声の強度変化的特徴・スペクトル的特徴などを利用して音声の識別を行なっている．金田の AMNOR 受信方式では，マイクロホンからの相対的位置で定め，さらに学習時区間で無音であることで目的音であると定めている．

これらに対して，本論文では，意味レベルの規範で目的音を定義することが目標である．意味レベルの多様な情報に柔軟に対応できるような適応方法をめざす．

以上4つの手法は，その他にも多くの本質的な違いがある．それらを **Table. 4.2** にまとめる．

「個々のセンサ特性」を揃える必要がないのは，Widrow のノイズキャンセラと本論文の手法である．両者は，「残響」を抑制せずそのまま出力するという点においても共通している．

この理由は以下のように考えられる．それは，両者の手法が受信信号のみから適応するということである．すなわち，空間の伝達関数と，反射による影響と，マイクロホン特性による影響などが不可分に取り扱われているのである．このため，センサの特性のバラツキなどは伝達関数への適応化として一括して処理できる一方，残響の抑制は不可能となるのである．

「目的音の劣化」も重要な観点である．Widrow の適応ノイズキャンセラは，目的音成分の全く含まれない参照センサ信号が得られたときは目的音の劣化がない．また，Flanagan のアレイマイクシステムでは，目的音の劣化がない．しかし，Flanagan の手法は，「適応のための評価関数」が誤差の2乗平均ではないので，多数のマイクロホンを用いる割には，高い SN

Table 4.2 従来研究と本研究でめざす手法の比較（2）

手法	個々のセンサ特性	残響	目的音劣化	適応の評価
適応ノイズキャンセラ (Widrow)	揃える必要なし	保存	無	誤差の2乗平均
アレイマイクシステム (Flanagan)	揃える必要あり	抑制	無	目的信号の感度
AMNOR (Kaneda)	揃える必要あり	抑制	有（調節可）	誤差の2乗平均
本研究でめざす手法	揃える必要なし	保存	有	誤差の2乗平均

比が得られないという欠点もある．その他の手法は，最適性の規範誤差の2乗平均を採用している．

第 5 章

Cue Signal による方法

第 I 部の中心をなすのが本章である．本章では，まず **5.1** 節で，Cue Signal による適応方法の概略を説明する．ここでは，記述の厳密さは以下の節に譲り，直観的な理解のしやすさに重点をおいて記述する．次に **5.2** 節で Cue Signal による適応の条件を述べ，その条件のもとで内部学習信号による適応が可能であることを証明する．ここでの証明を見ればわかるように，内部学習信号の正当性は統計的に（すなわち，「時間平均が」とか「期待値が」とか「極限において」などの形式で）示される．それゆえ，実際に本手法で適応が可能であるかどうかは，シミュレーションで確かめる必要がある．この結果を **5.3** 節で示す．しかし，シミュレーションだけでは，理論で仮定したモデル化された目的音とと実際の目的音の違いによる影響や，実際のセンサノイズや量子化誤差による影響がでてこない．また，係数算出のアルゴリズムとして LMS アルゴリズムを用いた場合の収束性も確認しておく必要がある．そこで，**5.4** 節では，実際の音響信号とマイクロホンを用いた実験結果について述べる．

5.1 手法の概要

第 2 章の研究の目的でも述べたように，本研究では，できる限り自律的な適応センシングをめざす．すなわち，以下のような適応方法はとらない．

- 特定の時間区間で教師信号を与える適応法．いわゆるトレーニング信号による学習．
- 対象物やセンサの位置，信号のスペクトルなど，先験的知識を用いた環境への適応．

めざすのは，未知の信号と未知の伝達関数を相手にした適応化である．センサ情報のみから適応しなければならない．このため，線型フィルタの学習信号（教師信号，適応目標信号）を，システム内部で生成する．これを，本論文では，**内部目標**と呼ぶことにする．

内部目標は，できるだけセンサ信号のみから生成できるものがのぞましい．**Cue Signal 法**は，その内部目標として，**Cue Signal** $\alpha(t)$ と 受信音響信号（当然妨害音を含んでいる）$\psi(t)$ を乗算した信号 $d(t) = \alpha(t)\psi(t)$ を用いる方法である．[18)-36)]

5.2 節で示されるように，内部目標として $d(t) = \alpha(t)\,\psi(t)$ を用いる適応は，ある条件のもとで，理想的なトレーニング信号を用いた適応と全く同じフィルタ特性に収束することができる．その条件とは，Cue Signal が，目的音の強度変動のみに相関を有し，妨害音の強度変動に相関を持たないことである．また，これも詳しくは **5.2** 節で述べるが，Cue Signal は直流成分が 0 でなければならない．

Fig. 5.1 の模式図で Cue Signal を説明する．上段の 3 つは，左から，目的音波形，妨害音 1 の波形，妨害音 2 の波形である．マイクロホンでは，この 3 つの信号の和信号が受信される．ここで，なんらかの方法で目的音の強度変動を推定できたとして，下段の 3 つの信号が得られたとする．これらはどれも，目的音の強度変動に相関があり，妨害音の強度変動には相関がなく，かつ平均 0 の信号である．よって，これら 3 つはすべて Cue Signal として使用可能である．

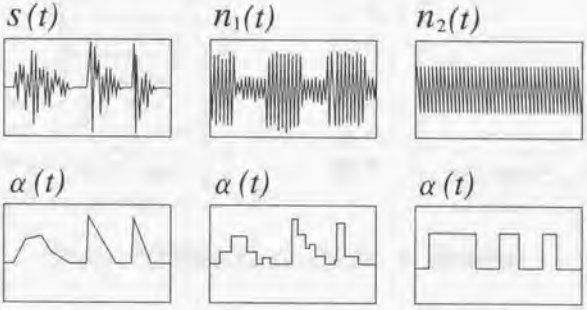


Fig. 5.1 目的音 $s(t)$，妨害音 $n_j(t)$，Cue Signal $\alpha(t)$ の模式図

まとめると，Cue Signal に対する条件は以下のようなものである．

- 目的音の強度変動に相関を有し，妨害音の強度変動に無相関であること．
- 平均が 0 であること．

このように Cue Signal に対する要求は，平均操作をした上で規定されており，柔軟な条件であると言える．すなわち，多様な Cue Signal を受け入れることができるのである．

目的音強度の変動を大ざっぱに推定すればいいのであるから，Cue Signal は受信音のみから生成することもできるし，視覚センサ（ビデオカメラ）など別の種類のセンサを用いて生成することもできる．

両者による基本的なブロックダイアグラムを，**Fig. 5.2** に示す．多チャネルのセンサ信号からＦＩＲフィルタによって目的とする信号を抽出する．このフィルタの係数を学習によって適応させる．適応のための内部目標 $d(t)$ は，Cue Signal $\alpha(t)$ と受信音 $\psi(t)$ の積信号を用いる．学習信号からフィルタの係数を得るためのアルゴリズムとしては，**3.2** 節で述べたさまざまな方法のいずれかを用いる．なお，$\psi(t)$ を生成するにあたって，遅延（Delay）をかけるのは，ＦＩＲフィルタによる遅延を補正するためである．

平均を 0 にする前の Cue Signal を，**Event Signal** $\theta(t)$ と呼ぶことにする．すなわち，目的信号強度の推定値である．音が発生しているという事象（Event）の度合を推定した信号であるから，Event Signel という用語を用いた．ただし，本手法で最小限必要なのは Cue

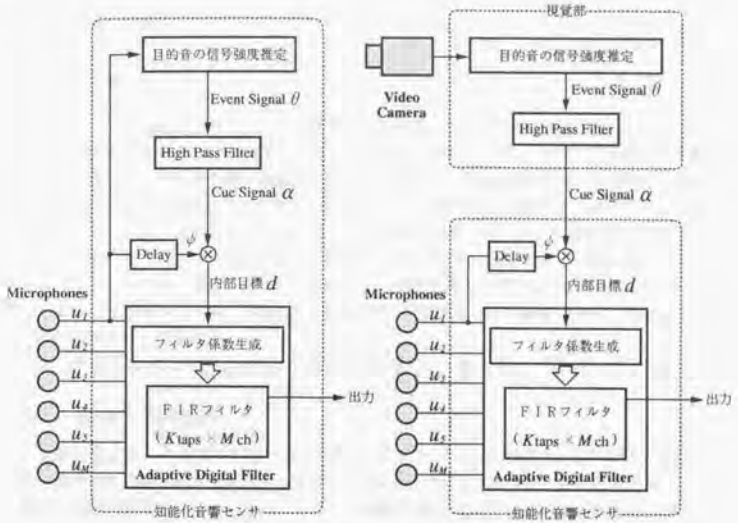


Fig. 5.2 知能化音響センサ（左：基本型，右：視聴覚融合型）

Signal のみであるので，必ずしもこの Event Signal を中間信号として生成する必要はない．

Cue Signal は，マイクロホン信号のみから生成する（Fig. 5.2 左図）こともできるし，異種センサからの情報をもとに生成する（Fig. 5.2 右図）こともできる．前者は，例えば定常的な妨害音のなかから強度変動のある目的音を抽出する場合に，マイクロホン信号の強度変化から目的音の強度変化を推定するなどの方法である．後者は，例えばビデオカメラを用いて，視覚的に目的音の強度を推定する場合などである．これは視聴覚融合型のセンシングともみなせる．後者のタイプの知能化センサについては，第 II 部で述べることとし，第 I 部では，Fig. 5.2 左図のタイプの音響センサのみをとりあつかう．

概要の最後として，Cue Signal を用いた内部目標で学習が可能である理由を，直感的に説明しておく．

まず，妨害音に関する学習がどのように行なわれるかを考える．妨害音は，正負値をとる平均 0 の Cue Signal とかけ合わされて内部目標となる．すなわち，ある時は妨害音そのものが目標となり，またある時は妨害音の位相反転信号が目標となる．これが平均化されると，この反転は妨害音強度とは無相関に行なわれるため，結局は妨害音の 0 倍信号を目標とした学習が行なわれることになるのである．これは，妨害音を抑制することである．

一方，目的音を考えてみよう．これも，妨害音のときと同じく，ある時は目的音そのものが目標となり，またある時は目的音の位相反転信号が目標となる．しかし，目的音の場合は Cue Signal と目的音信号変動のあいだに相関がある．すなわち，目的音が強いときには目的音そのものが目標となり，目的音が弱いときには目的音の位相反転信号が目標となる．これ

が平均化されると，こんどは目的音の正相信号を目標とした学習となる．すなわち，目的音は抽出されるのである．

5.2 理論

ここでは，Cue Signal から生成した内部目標による適応フィルタの理論を説明する．**5.2.1** と **5.2.2** は準備である．**5.2.3** ～**5.2.5** の 3 小節が適応の理論である．このうち，**5.2.3** では最も典型的な場合，すなわち 1 個の目的音と複数の妨害音が存在する場合をとりあげる．**5.2.4** では，音源の条件を一般化して，複数の目的音が存在する場合をとりあげる．以上の 2 つ小節では，線型フィルタとして FIR フィルタを用いた場合に限定し述べる．本研究では FIR フィルタのみを用いているので，理論的にはこれで十分であるが，**5.2.5** では，一般の線型フィルタの場合をとりあげる．線型フィルタにどのような制限を加えれば Cue Signal による学習が可能となるかを明らかにする．最後に，**5.2.6** では，理想的な Cue Signal が得られなかった場合の対応法について述べる．

なお，本節では，記号表（p.4～p.5）および第 3 章での表記に従って記述する．

5.2.1 等価な適応目標

この節の目的は，ある適応目標 $d_1(t)$ で学習した FIR フィルタ係数 f_n と同じフィルタ係数を，別の適応目標 $d_2(t)$ で獲得するための，$d_2(t)$ に対する条件を述べることである．

まず，用語の定義をする．今，ここに 2 つの適応目標 $d_1(t), d_2(t)$ があるとする．このとき，あるフィルタ係数 $f_n (n = 1, 2, \ldots, N)$ が存在してそれが，$d_1(t)$ と出力との平均 2 乗誤差と，$d_2(t)$ と出力との平均 2 乗誤差の両方を最小化するとき，$d_1(t)$ と $d_2(t)$ は同等の適応を与える適応目標であると言える．

さて，平均 2 乗誤差最小の規範の場合，適応目標とフィルタ係数を結びつけているのは式 (3.25)，再掲すれば，

$$\begin{bmatrix} R_{y_1 d} \\ R_{y_2 d} \\ \vdots \\ R_{y_N d} \end{bmatrix} = \begin{bmatrix} R_{y_1 y_1} & R_{y_1 y_2} & \cdots & R_{y_1 y_N} \\ R_{y_2 y_1} & R_{y_2 y_2} & \cdots & R_{y_2 y_N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{y_N y_1} & R_{y_N y_2} & \cdots & R_{y_N y_N} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} \tag{3.25}$$

である．上式は，単に「任意の信号 $d(t)$ について，ある時間区間で，$d(t)$ との 2 乗誤差を最小にする線型結合係数 f_n を決定する方程式」と見なしてよい．上式を観察すれば，右辺の行列 R は，タップ信号 y_n のみに依存しており，適応目標 $d(t)$ は左辺のみに影響することがわかる．すなわち，

$$R_{y_i d_1} = R_{y_i d_2} \qquad (i = 1, \ldots, n) \tag{5.1}$$

であれば，$d_1(t)$ と $d_2(t)$ について，式(3.25) は全く同一になる．よって，2 乗誤差を最小にする解 f_n の作る空間も同一である．そこで，式(5.1) の成立する 2 つの適応目標 $d_1(t)$ と $d_2(t)$ を，**等価な適応目標**と呼ぶことにする．以下では，この条件を用いて適応目標の評価を行なう．

5.2.2 問題のモデル化（理論のための仮定）

次に，目的信号をモデル化しておく．このモデル化は，本手法の動作条件を明確に記述するためにも必要である．

まず，実際の状況を考えてみる．目的音は，音声や楽音，または異常音などのような音を考えている．これらは，みな非定常音である．そして，その非定常音は，通常は強度変化を伴う．そこで，非定常な目的信号 $s(t)$ が，定常確率過程によって生成される**搬送波** $c(t)$ と，振幅の変動を表す非定常なエンベロープ $a(t)$ の積によって表現できると仮定する．

$$s(t) = a(t)\,c(t) \tag{5.2}$$

ここで，エンベロープおよび搬送波の意味から，$a(t) \geq 0$ および $\langle c(t) = 0 \rangle$ が成り立つものとする．ちなみに，非定常な信号を式(5.2)のように分解して記述する手法は，地震波の分析の際にもしばしば用いられる．その場合は，$s(t)$ を**分離可能**（separable）な過程[37]と呼ぶこともある．

また，エンベロープ $a(t)$ の2乗信号 $a(t)^2$ を，**強度エンベロープ**とよぶことにする．さらに，$a(t)^2$ や $a(t)\,a(t)$ や $a(t)\,a(t)^2$ など，**包絡類**のみの関数として得られる信号は，音響信号（搬送波帯域を占有する）のみの関数として得られる信号とは独立であるとする．すなわち，

$$\begin{aligned}&\langle\,[a(t)\text{と}\alpha(t)\text{から作られる信号}]\cdot[\text{音響信号から作られる信号}]\,\rangle\\ &\quad=\langle\,[a(t)\text{と}\alpha(t)\text{から作られる信号}]\,\rangle\,\langle\,[\text{音響信号から作られる信号}]\,\rangle\end{aligned} \tag{5.3}$$

なお，妨害音については，特にことわらないかぎり，定常確率過程によって生成される定常信号であるとする．

以上述べたモデル化の最も厳しい制約は，搬送波成分 $c(t)$ を定常としている点である．音声などを例にとってみても，これは一般には成立しない．この仮定を緩めた理論は，本章の考察で若干ふれるが，モデルの不適合による問題が起こるか否かは最終的には実験で確認することにする．

さて，以上のモデル化を行なったあとで，タップ信号 $y_n(t)$ を計算する．ここでは，簡単のため，目的音源は1個とする．適応線型結合器の n 番目の入力信号であるタップ信号 $y_n(t)$ は，目的音源 S_0 および妨害音源 S_j から各タップ y_n までの経路のインパルス応答をそれぞれ $g_{0n}(t), g_{jn}(t)$ と書くことにすると，妨害音源の総数を J として，

$$y_n(t) = g_{0n}(t) \otimes s(t) + \sum_{j=1}^{J} g_{jn}(t) \otimes n_j(t) \tag{5.4}$$

$$= g_{0n}(t) \otimes [a(t)c(t)] + \sum_{j} g_{jn} \otimes n_j(t) \tag{5.5}$$

で表わされる．ここで，$g_{0n}(t)$ の時間軸方向の変化に比較して $a(t)$ の時間変化がゆっくりとしたものであるとすれば，$g_{0n}(t)$ が有効な値（0でない値）をとる時間範囲において $a(t)$ は一定とみなせるので，$a(t)$ を畳み込みの前に出すことができて，

$$y_n(t) = a(t)\,[g_{0n}(t) \otimes c(t)] + \sum_{j} g_{jn}(t) \otimes n_j(t) \tag{5.6}$$

と近似できる．$y_{nC}(t) \stackrel{\mathrm{def}}{=} g_{0n}(t) \otimes c(t)$ と定義し，上式の第1項（目的音成分）を $y_{nS}(t)$ と書き，上式の第2項（妨害音成分）を $y_{nN}(t)$ と書くことにすれば，タップ信号 $y_n(t)$ は，

$$y_n(t) = a(t)\,y_{nC}(t) + y_{nN}(t) \tag{5.7}$$

$$= y_{nS}(t) + y_{nN}(t) \tag{5.8}$$

のように，目的音を要因とする項と妨害音を要因とする項の和として表現することができる．以下，式(5.7)と式(5.8)の形式を用いて計算を進める．

5.2.3 内部目標による学習の正当性（FIR フィルタ・単一目的信号）

本手法（Cue Signal 法）における内部目標は，Fig. 5.2 に示したように，Cue Signal $\alpha(t)$ と妨害音を含んだ受信信号 $\psi(t)$ を掛け合わせて生成する．

$$d(t) = \alpha(t)\,\psi(t) \tag{5.9}$$

実際には $\psi(t)$ としては，タップ信号 $y_n(t)$ のいずれかひとつが使われる．よって $\psi(t)$ は式(5.7) と式(5.8) と同じように

$$\begin{aligned} \psi(t) &= a(t)\,\psi_C(t) + \psi_N(t) && (5.10)\\ &= \psi_S(t) + \psi_N(t) && (5.11) \end{aligned}$$

と，目的音成分と妨害音成分の和として記述できる．

さて，$\psi_S(t)$ は受信音から妨害音を完全に抑圧した信号である．今，われわれは $\psi_S(t)$ を得ることはできないが，仮に $\psi_S(t)$ を用いて適応化を行なうことができたならば，妨害音除去の観点では最良であるといえる．

そこで，$\psi_S(t)$ を内部目標 $d(t)$ に対比する意味で，**最適目標**と呼ぶことにする．

本節では，内部目標による学習の正当性を，内部目標が最適目標と等価な適応目標であることによって示す．これが示されれば，内部目標による適応によって，妨害音を全く含まない適応目標でトレーニングしたのと同じ性能の信号選択性が得られことになる．

これを示す準備として，以下の量を定義する．

$$K[\alpha, a] \stackrel{\text{def}}{=} \frac{\langle \alpha(t) a(t)^2 \rangle}{\langle a(t)^2 \rangle} \tag{5.12}$$

これは，Cue Signal と強度エンベロープの相関の大きさを強度エンベロープで正規化した量である．

【定理 5.1】 内部目標 $d(t) = \alpha(t)\,\psi(t)$ と最適目標 $\psi_S(t)$ は，以下の 4 つの条件が満たされるならば，等価な適応目標である．

1. 目的信号 $s(t)$ が，エンベロープと搬送波に分離可能であること． $s(t) = a(t)\,c(t)$
2. 目的音 $s(t)$ と定常妨害音 $n_j(t)$ は，独立な信号であること． $\langle y_{n_i N}(t)\, y_{n_j C}(t) \rangle = 0$
3. Cue Signal $\alpha(t)$ が，強度エンベロープ $a(t)^2$ と相関をもつこと． $K[\alpha, a] = 1$
4. Cue Signal $\alpha(t)$ が，平均 0 の信号であること． $\langle \alpha(t) \rangle = 0$

（定理おわり）

［証明］ $R_{y_n \psi_S}$ と $R_{y_n d}$ を比較すればよい．最適目標を使った場合は，

$$\begin{aligned} R_{y_n \psi_S} &= \langle y_n(t)\,\psi_S(t) \rangle && (5.13)\\ &= \langle [a(t)\,y_{nC}(t) + y_{nN}(t)]\; a(t)\,\psi_C(t) \rangle && (5.14)\\ &= \langle a(t)^2 y_{nC}(t)\,\psi_C(t) \rangle + \langle a(t)\,y_{nN}(t)\,\psi_C(t) \rangle && (5.15)\\ &= \langle a(t)^2 \rangle \langle y_{nC}(t)\,\psi_C(t) \rangle + \langle a(t) \rangle \langle y_{nN}(t)\,\psi_C(t) \rangle && (5.16)\\ &= \langle a(t)^2 \rangle \langle y_{nC}(t)\,\psi_C(t) \rangle && (5.17) \end{aligned}$$

となる．ここで，式(5.13) から式(5.14) では定理の条件 1，式(5.14) から式(5.15) では和と平均の順序の可換性，式(5.15) から式(5.16) では式(5.3) の性質，式(5.16) から式(5.17) では定理の 条件 2 を用いた．

一方，内部目標を使った場合は，

$$
\begin{aligned}
R_{y_n d} &= \langle\, y_n(t)\, d(t) \,\rangle && (5.18)\\
&= \langle\, [a(t)\, y_{nC}(t) + y_{nN}(t)]\, \alpha(t)\, [a(t)\, \psi_C(t) + \psi_N(t)] \,\rangle && (5.19)\\
&= \langle\, \alpha(t)\, a(t)^2\, y_{nC}(t)\, \psi_C(t) \,\rangle + \langle\, \alpha(t)\, y_{nN}(t)\, \psi_N(t) \,\rangle \\
&\quad + \langle\, \alpha(t)\, a(t)\, y_{nC}(t)\, \psi_N(t) \,\rangle + \langle\, \alpha(t)\, a(t)\, y_{nN}(t)\, \psi_C(t) \,\rangle && (5.20)\\
&= \langle\, \alpha(t)\, a(t)^2 \,\rangle \langle\, y_{nC}(t)\, \psi_C(t) \,\rangle + \langle\, \alpha(t) \,\rangle \langle\, y_{nN}(t)\, \psi_N(t) \,\rangle \\
&\quad + \langle\, \alpha(t)\, a(t) \,\rangle \langle\, y_{nC}(t)\, \psi_N(t) \,\rangle + \langle\, \alpha(t)\, a(t) \,\rangle \langle\, y_{nN}(t)\, \psi_C(t) \,\rangle && (5.21)\\
&= \langle\, \alpha(t)\, a(t)^2 \,\rangle \langle\, y_{nC}(t)\, \psi_C(t) \,\rangle && (5.22)\\
&= K[\alpha, a] \langle\, a(t)^2 \,\rangle \langle\, y_{nC}(t)\, \psi_C(t) \,\rangle && (5.23)
\end{aligned}
$$

となる．ここで，式(5.18) から式(5.19) では定理の条件 1，式(5.19) から式(5.20) では和と平均の順序の可換性，式(5.20) から式(5.21) では式(5.3) の性質，式(5.21) から式(5.22) では定理の 条件 2 と 条件 4 を用いた．

よって，式(5.17)，式(5.23)，および $K[\alpha, a]$ の条件（定理の条件 3）より，

$$
R_{y_n d} = R_{y_n \psi_S} \tag{5.24}
$$

となり，内部目標と最適目標は等価な適応目標である． （証明終り）

さらに，$K > 0$ （K が 1 でない場合）を考える．このとき，式(5.17) と式(5.23) より

$$
R_{y_n d} = K[\alpha, a]\, R_{y_n \psi_S} \tag{5.25}
$$

となる．これを式(3.25) に代入すれば，

$$
[f_i]_d = K[\alpha, a] \cdot [f_n]_{\psi_S} \qquad (n = 1, \ldots, N) \tag{5.26}
$$

となるので，K は 1 でなくても 0 でさえなければ，フィルタの係数がすべて等しく定数倍されるにすぎないことがわかる．すなわち，出力が $K[\alpha, a]$ 倍になることを除けば最適目標に等価な適応目標を使用したのと同じ効果が得られる．そこで，定理 5.1 の 条件 3 のかわりに，

3a. $K[\alpha, a] > 1$ であること．

の場合，「定数倍を除いて等価な適応目標」ということにする．

5.2.4 信号モデルの一般化（FIR フィルタ・複数目的信号）

前節では，典型的な場合を扱った．本節では，それを一般化して，Cue Signal によっては，抽出と抑制の中間的なフィルタリングをうける音信号が存在することを示す．また，その抽出の度合と Cue Signal との関係を定式化する．

まず，$s_0(t) \stackrel{\text{def}}{=} s(t)$，および $s_j(t) \stackrel{\text{def}}{=} n_j(t)$ $(j = 1, 2, \ldots)$ として，**3.1.2**（p.13参照）でやったようにすべての音源の波形を統一的に $s_j(t)$ $(j = 0, 1, 2, \ldots)$ であらわしておく．

次に，各音源はエンベロープと搬送波（定常確率過程で生成される信号）の積で

$$
s_j(t) = a_j(t)\, c_j(t) \quad (j = 0, 1, 2, \ldots) \tag{5.27}
$$

と記述できると仮定する．（定常信号の場合は $a_j(t)$ を定数とおけばよい）

エンベロープの変化は，音源ーマイク間の伝播時間やＦＩＲフィルタの遅延時間に比較して大きな時定数で変化するので，式(5.6)を求めたのと同じようにして，タップ信号 $y_n(t)$ は，

$$y_n(t) = \sum_j a_j(t)\,[g_{jn}(t) \otimes c_j(t)] \tag{5.28}$$

と近似できる．ここで，$y_{nC_j}(t) \stackrel{\text{def}}{=} g_{jn}(t) \otimes c_j(t)$ と定義すれば，タップ信号 $y_n(t)$ と $\psi(t)$ は，エンベロープと搬送波成分の積和で

$$y_n(t) = \sum_j a_j(t)\, y_{nC_j}(t) \qquad (j = 0, 1, 2, \ldots) \tag{5.29}$$

$$\psi(t) = \sum_j a_j(t)\, \psi_{C_j}(t) \tag{5.30}$$

と書くことができる．

ここで，各音源 S_j の強度エンベロープと Cue Signal のあいだの相関の大きさをあらわす量 $K[\alpha, a_j]$

$$K[\alpha, a_j] \stackrel{\text{def}}{=} \frac{\langle\, \alpha(t)\, a_j(t)^2 \,\rangle}{\langle\, a_j(t)^2 \,\rangle} \tag{5.31}$$

を定義する．次に，受信音 $\psi(t)$ に含まれる音源 i に起因する成分を $K[\alpha, a_j]$ 倍した信号 $d_W(t)$

$$d_W(t) \stackrel{\text{def}}{=} \sum_j K[\alpha, a_j]\; a_j(t)\; \psi_{C_j}(t) \tag{5.32}$$

を定義する．$d_W(t)$ を内部目標 $d(t)$ に対比する意味で，**加重目標**と呼ぶことにする．この加重目標は最適目標を一般化したものである．なぜならば，$K[\alpha, a_0] = 1$ かつ $K[\alpha, a_j] = 0\,(j \neq 0)$ のとき，加重目標は最適目標に等しいからである．

【定理 5.2】 内部目標 $d(t) = \alpha(t)\,\psi(t)$ と加重目標 $d_W(t)$ は，以下の条件が満たされるならば，等価な適応目標である．

1. 音源信号 $s_j(t)$ が，エンベロープと搬送波に分離可能であること．　$s_j(t) = a_j(t)\, c_j(t)$
2. 音源 j_1 と 音源 j_2 ($j_1 \neq j_2$) が独立な信号であること．　$\langle\, y_{n_3 C_{j_1}}(t)\; y_{n_4 C_{j_2}}(t)\,\rangle = 0$

（定理おわり）

［証明］定理 5.1 のときと同じく，$R_{y_n d}$ と $R_{y_n d_W}$ を比較すればよい．まず，$R_{y_n d_W}$ を計算すると，

$$R_{y_n d_W} = \langle\, y_n(t)\; d_W(t)\,\rangle \tag{5.33}$$

$$= \langle\, \{\sum_j a_j(t)\, y_{nC_j}(t)\}\; \{\sum_j K[\alpha, a_j]\, a_j(t)\, \psi_{C_j}(t)\}\,\rangle \tag{5.34}$$

$$= \sum_{j_1} \sum_{j_2} K[\alpha, a_{j_1}]\, \langle\, a_{j_1}(t)\, a_{j_2}(t)\, \psi_{C_{j_1}}(t)\, y_{nC_{j_2}}(t)\,\rangle \tag{5.35}$$

$$= \sum_{j_1} \sum_{j_2} K[\alpha, a_{j_1}]\, \langle\, a_{j_1}(t)\, a_{j_2}(t)\,\rangle\, \langle\, \psi_{C_{j_1}}(t)\, y_{nC_{j_2}}(t)\,\rangle \tag{5.36}$$

$$= \sum_j K[\alpha, a_j]\, \langle\, a_j(t)^2 \rangle\, \langle\, \psi_{C_j}(t)\, y_{nC_j}(t)\,\rangle \tag{5.37}$$

$$= \sum_j \langle\, \alpha(t)\, a_j(t)^2 \rangle\, \langle\, \psi_{C_j}(t)\, y_{nC_j}(t)\,\rangle \tag{5.38}$$

となる．ここで，式(5.33) から式(5.34) では定理の条件1，式(5.34) から式(5.35) では和と平均の順序の可換性，式(5.35) から式(5.36) では式(5.3) の性質，式(5.36) から式(5.37) では定理の条件2，式(5.37) から式(5.38) では式(5.31) の定義を用いた．一方，$R_{y_n d}$ を計算すると，

$$
\begin{aligned}
R_{y_n d} &= \langle y_n(t)\, d(t) \rangle && (5.39)\\
&= \langle \{\sum_j a_j(t)\, y_{nC_j}(t)\}\ \alpha(t)\ \{\sum_j a_j(t)\, \psi_{C_j}(t)\} \rangle && (5.40)\\
&= \sum_{j_1}\sum_{j_2} \langle \alpha(t)\, a_{j_1}(t)\, a_{j_2}(t)\, \psi_{C_{j_1}}(t)\, y_{nC_{j_2}}(t) \rangle && (5.41)\\
&= \sum_{j_1}\sum_{j_2} \langle \alpha(t)\, a_{j_1}(t)\, a_{j_2}(t) \rangle \langle \psi_{C_{j_1}}(t)\, y_{nC_{j_2}}(t) \rangle && (5.42)\\
&= \sum_{j} \langle \alpha(t)\, a_j(t)^2 \rangle \langle \psi_{C_j}(t)\, y_{nC_j}(t) \rangle && (5.43)
\end{aligned}
$$

となる．ここで，式(5.39) から式(5.40) では定理の条件1，式(5.40) から式(5.41) では和と平均の順序の可換性，式(5.41) から式(5.42) では式(5.3) の性質，式(5.42) から式(5.43) では定理の条件2を用いた．以上式(5.38) と式(5.43) より

$$
R_{y_n d} = R_{y_n d_W} \qquad (5.44)
$$

である．これより，$d(t)$ は $d_W(t)$ に等価な適応目標である．　（証明おわり）

この定理は，「Cue Signal $\alpha(t)$ を採用した知能化音響センサは，出力中の音 j 成分を $K[\alpha, a_j]$ 倍にするように適応化が行なわれる」ことを意味する．

すなわち，$K[\alpha, a_j]$ が音源 S_j の抽出の度合をあらわしており，その極端な場合として，$K[\alpha, a_j] = 0$ である音 j は抑制され，$\left|K[\alpha, a_j]\right| \gg 0$ である音 j は抽出される．

定理5.1と定理5.2を比較してみよう．定理5.1の場合は妨害音はすべて定常という仮定であった．これは，$a_j(t) = 1$ $(j = 1, 2, \ldots)$ ということである．すなわち，定理5.1の条件4は，定理5.2の言葉で書けば，$K[\alpha, a_j] = K[\alpha, 1] = \langle \alpha \rangle = 0$ $(j \neq 0)$ であり，抑制条件に対応している．また，定理5.1の条件3は，定理5.2の言葉で書けば，$K[\alpha, a_0] = 1$ であり，抽出条件に対応している．このように，定理5.1は定理5.2の特殊な場合になっている．

5.2.5　一般の線型フィルタを用いた場合（単一目的信号）

以上の2つの節では，内部目標で適応可能であることを，FIRフィルタを用いた場合に限定して記述した．もちろん，内部目標による学習はFIRフィルタだけでなく，すべての線型フィルタで有効なものである．

本節では，一般の線型フィルタの場合について，内部目標の有効性を示す．なお，フィルタを一般化したことで記述がより複雑になるため，目的音，妨害音がともに1個である場合を計算した．音源の数を増加した場合への拡張が容易であることは，途中の式の変形をたどれば明らかであろう．

FIRフィルタの場合は，式(3.25) をもとに，計算を進めた．一般の線型フィルタの場合は，式(3.20) を出発点とすればよい．再掲すれば，

$$
\begin{bmatrix} R_{u_1 d}(\omega) \\ R_{u_2 d}(\omega) \\ \vdots \\ R_{u_M d}(\omega) \end{bmatrix} = \begin{bmatrix} R_{u_1 u_1}(\omega) & R_{u_1 u_2}(\omega) & \cdots & R_{u_1 u_M}(\omega) \\ R_{u_2 u_1}(\omega) & R_{u_2 u_2}(\omega) & \cdots & R_{u_2 u_M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ R_{u_M u_1}(\omega) & R_{u_M u_2}(\omega) & \cdots & R_{u_M u_M}(\omega) \end{bmatrix} \begin{bmatrix} F_1(\omega) \\ F_2(\omega) \\ \vdots \\ F_M(\omega) \end{bmatrix} \qquad (3.20)
$$

である．これを，「任意の信号 $d(t)$ について，ある時間区間で，$d(t)$ との2乗誤差を最小にする線型フィルタ特性 $F_m(\omega)$ を決定する方程式」と見なす．適応目標 $d(t)$ は左辺のみに影響するので，式(5.1) に相当する「等価な適応目標の条件」は，

$$R_{u_m d_1}(\omega) = R_{u_m d_2}(\omega) \qquad (m = 1, \ldots, M) \tag{5.45}$$

となる．なお，式(5.45) は，式(5.1) に比較して，式の数が $N = MK$ 個から M 個に減少した代わりに，新たにパラメータ ω が増えている．以下，この条件を用いて適応目標の評価を行なう．パラメータ ω が入ったため，周波数領域での計算が主体となる．

まず，目的信号のモデル化（式(5.2)）を周波数領域で書けば，エンベロープ $A(\omega)$ と包絡 $C(\omega)$ の畳み込みで

$$S(\omega) = A(\omega) \otimes C(\omega) \tag{5.46}$$

となる．また，Cue Signal $\alpha(t)$ のフーリエ変換を $\Pi(\omega)$ と書くことにする．ここで，エンベロープ $A(\omega)$ や Cue Signal $\Pi(\omega)$ は包絡類であるので，超低周波に帯域制限された信号である．その帯域をそれぞれ B_A, B_Π とする．

$$A(\omega) = 0 \qquad |\omega| > B_A \tag{5.47}$$

$$\Pi(\omega) = 0 \qquad |\omega| > B_\Pi \tag{5.48}$$

適応効果を示す方針は，FIR フィルタのときと同じく，$R_{u_m \psi_S}(\omega)$ と $R_{u_m d}(\omega)$ が等価な適応目標であるかを調べる．$u_m(t), \psi_S,(t), d(t)$ の周波数領域での表現である $U_m(\omega), \Psi_S,(\omega), D(\omega)$ を計算する．以下，特にことわらない変数の定義は，記号表を参照のこと．

$$U_m(\omega) = S(\omega)\, H_{0m}(\omega) + N(\omega)\, H_{1m}(\omega) \tag{5.49}$$

$$= \big(A(\omega) \otimes C(\omega)\big)\, H_{0m}(\omega) + N(\omega)\, H_{1m}(\omega) \tag{5.50}$$

$$\Psi(\omega) = S(\omega)\, H_{0\psi}(\omega) + N(\omega)\, H_{1\psi}(\omega) \tag{5.51}$$

$$= \big(A(\omega) \otimes C(\omega)\big)\, H_{0\psi}(\omega) + N(\omega)\, H_{1\psi}(\omega) \tag{5.52}$$

$$\Psi_S(\omega) = [\Psi(\omega) \text{ の目的音成分}] \tag{5.53}$$

$$= [\text{式(5.52) の第1項}] \tag{5.54}$$

$$= \big(A(\omega) \otimes C(\omega)\big)\, H_{0\psi}(\omega) \tag{5.55}$$

内部目標は，$d(t) = \alpha(t)\, \psi(t)$ を周波数領域で書けば以下のようになる．

$$D(\omega) = \Pi(\omega) \otimes \Psi(\omega) \tag{5.56}$$

$$= \Pi(\omega) \otimes \big[\big(A(\omega) \otimes C(\omega)\big)\, H_{0\psi}(\omega)\big] + \Pi(\omega) \otimes \big[N(\omega)\, H_{1\psi}(\omega)\big] \tag{5.57}$$

$$= \big[\Pi(\omega) \otimes \big(A(\omega) \otimes C(\omega)\big)\big]\, H_{0\psi}(\omega) + \big[\Pi(\omega) \otimes N(\omega)\big]\, H_{1\psi}(\omega) \tag{5.58}$$

ここで，式(5.57) から式(5.58) への変形には次式の性質を用いた．

$$[\text{包絡類}] \otimes \Big([\text{音響信号}] \cdot [\text{伝達関数類}]\Big)$$

$$= \Big([\text{包絡類}] \otimes [\text{音響信号}]\Big) \cdot [\text{伝達関数類}] \tag{5.59}$$

これは，包絡類信号（$\Pi(\omega)$ や $A(\omega)$）が，ω が0にごく近い部分（すなわち B_Π, B_A 以内）でしか非0の値を持たないことと，伝達関数類（$H_{\bullet\bullet}(\omega)$）が，$\omega$ 軸方向でなだらかに変化す

ることの2条件から可能な近似である．なお，伝達関数類が，ω 軸方向でなだらかに変化することは，極端に大きな遅延がないことを意味している．すなわち，包絡類信号の時間変化のオーダーよりも残響等の時間のオーダーがずっと小さいことがこの変形の条件である．音源の強度変化時間より残響が長ければ Cue Signal 法が無効であることは明らかであるので，この仮定は妥当であろう．

以上で，$U_m(\omega), \Psi_S,(\omega), D(\omega)$ の準備ができたので，$R_{u_m\psi_S}(\omega)$ と $R_{u_md}(\omega)$ を計算する．

$$
\begin{aligned}
R_{u_m\psi_S}(\omega) &= U_m(\omega)\Psi_S(\omega)^* && (5.60)\\
&= \left[A(\omega)\otimes C(\omega)\right]\left[A(\omega)^*\otimes C(\omega)^*\right] H_{0m}(\omega)\, H_{0\psi}(\omega)^* && (5.61)\\
&\quad + N(\omega)\left[A(\omega)^*\otimes C(\omega)^*\right] H_{1m}(\omega)\, H_{0\psi}(\omega)^* && (5.62)
\end{aligned}
$$

$$
\begin{aligned}
R_{u_md}(\omega) &= U_m(\omega)\, D(\omega)^* && (5.63)\\
&= \left[A(\omega)\otimes C(\omega)\right]\left[\Pi(\omega)^*\otimes A(\omega)^*\otimes C(\omega)^*\right] H_{0m}(\omega)\, H_{0\psi}(\omega)^* && (5.64)\\
&\quad + N(\omega)\left[\Pi(\omega)^*\otimes A(\omega)^*\otimes C(\omega)^*\right] H_{1m}(\omega)\, H_{0\psi}(\omega)^* && (5.65)\\
&\quad + \left[A(\omega)\otimes C(\omega)\right]\left[\Pi(\omega)^*\otimes N(\omega)^*\right] H_{0m}(\omega)\, H_{1\psi}(\omega)^* && (5.66)\\
&\quad + N(\omega)\left[\Pi(\omega)^*\otimes N(\omega)^*\right] H_{1m}(\omega)\, H_{1\psi}(\omega)^* && (5.67)
\end{aligned}
$$

$D(\omega)$ と $\Psi_S,(\omega)$ が「等価な適応目標」であるためには，上の $R_{u_m\psi_S}(\omega)$ と $R_{u_md}(\omega)$ が等しくなければならない．このためには，項(5.61)=項(5.64)，項(5.62)=0，項(5.65)=0，項(5.66)=0，項(5.67)=0 の5つの条件が成り立てばよい．さらに，音響環境に依存せずに，常に「等価な適応目標」であるためには，$H_{\bullet\bullet}(\omega)$ に依存せずに $R_{u_m\psi_S}(\omega)$ と $R_{u_md}(\omega)$ が等しくなければならない．すなわち，以下の5つの式が成立すればよい．

$$
\begin{aligned}
\left[A(\omega)\otimes C(\omega)\right]\left[A(\omega)^*\otimes C(\omega)^*\right] &= \left[A(\omega)\otimes C(\omega)\right]\left[\Pi(\omega)^*\otimes A(\omega)^*\otimes C(\omega)^*\right] && (5.68)\\
N(\omega)\left[A(\omega)^*\otimes C(\omega)^*\right] &= 0 && (5.69)\\
N(\omega)\left[\Pi(\omega)^*\otimes A(\omega)^*\otimes C(\omega)^*\right] &= 0 && (5.70)\\
\left[A(\omega)\otimes C(\omega)\right]\left[\Pi(\omega)^*\otimes N(\omega)^*\right] &= 0 && (5.71)\\
N(\omega)\left[\Pi(\omega)^*\otimes N(\omega)^*\right] &= 0 && (5.72)
\end{aligned}
$$

しかし，残念ながら式(5.59)の仮定だけでは，以上5つの式を示すことはできない．定理5.1の時との違いは，2乗誤差最小の土台となる平均操作 $\langle\bullet\rangle$（時間平均 $\overline{\bullet}$ または集合平均 $\mathrm{E}[\bullet]$）に対して何も言及していないということである．そこで，平均操作 $\langle\bullet\rangle$ が時間平均 $\overline{\bullet}$ であるときは，その観測時間は十分長いものと仮定する．（実は，定理5.1の証明では，信号の独立性を 条件2 や 式(5.3) で用いたが，その独立性の定義 $\langle\bullet\bullet\rangle = \langle\bullet\rangle\langle\bullet\rangle$ の中に「時間平均の場合はその観測時間は十分長いものとする」という仮定が暗に埋め込まれていたわけである）．さて，観測時間が長くなればなるほど，$C(\omega)$ や $N(\omega)$ は，周波数軸上で激しく変化するようになる．（周波数軸方向で見たときの空間周波数が広帯域になる）．これによって，式(3.20) の $R_{u_m\psi_S}(\omega)$ や $R_{u_md}(\omega)$ や $R_{u_{m_1}u_{m_2}}(\omega)$ が周波数軸方向に激しく変動する．よって，式(3.20) を解くと，解である線型フィルタの周波数特性 $F_m(\omega)$ も一般に周波数軸方向に激しく変動するようになる．これは，線型フィルタの時間特性で言い替えれば，インパルス応答が（観測時間と同程度の長さで）伸びなければならないという要請に対応する．しかし，このように長いインパルス応答は現実的ではない．

これは，言い替えれば次のようなことである．長時間の確定信号を取り扱い，しかも線型フィルタ $F_m(\omega)$ に全く制限を加えない（すなわち，数秒以上のインパルス応答を許す）ので

あれば，フィルタの自由度が高すぎてアドホックなフィルタができてしまう．これは，音響信号の統計的性質や伝達関数に適応したのではなくて，確定的な波形そのものに適応したにすぎない．すなわち，ある時間区間で適応した結果を，次の未知の時間区間に適用することはできない．

音響環境の統計的性質に適応するためには，線型フィルタは自由度の小さな短いインパルス応答のものでなければならない．すなわち，線型フィルタの周波数特性 $F_m(\omega)$ は周波数軸方向にある程度なだらかに変化するものでなければならない．すなわち，式(3.20) をそのまま解くのではなく，ある程度 ω 方向にボケをかけて解くべきである．そこで，この操作を，期待値 $\mathrm{E}[\bullet]$ をとることで表現することにする．

これは，次の様に説明してもよい．すなわち，十分長い時間区間で式(3.20) を解くのであるが，フィルタとして短いインパルス応答しか用意しない．そして，長い時間区間をフィルタのインパルス応答程度の細かい区間に分割して多数の式(3.20) をたてる．それらの平均をとることを，集合平均 $\mathrm{E}[\bullet]$ と見るのである．

もちろん，2 乗誤差最小のベースとなる平均操作 $\langle\bullet\rangle$ が時間平均 $\bar{\bullet}$ でなくて集合平均 $\mathrm{E}[\bullet]$ である場合は，もともと $\mathrm{E}[\bullet]$ をとって評価する必要がある．

以上の理由から，$R_{u_m\psi_S}(\omega)$ と $R_{u_m d}(\omega)$ をそのまま比較するのではなく，$\mathrm{E}[R_{u_m\psi_S}(\omega)]$ と $\mathrm{E}[R_{u_m d}(\omega)]$ を比較することとする．よって，式(5.68)～式(5.72) の 5 つの式もその期待値で評価する．

まず，式(5.69)～式(5.71) の 3 つの式は，どれも，目的音成分と妨害音成分の積である．目的音と妨害音が無相関な信号であるとすれば，すべての ω に対して目的音成分の偏角と妨害音成分の偏角は独立であるので，両成分の積の期待値は 0 になる．

$$\mathrm{E}\left[N(\omega)\left[A(\omega)^*\otimes C(\omega)^*\right]\right] = 0 \tag{5.73}$$

$$\mathrm{E}\left[N(\omega)\left[\Pi(\omega)^*\otimes A(\omega)^*\otimes C(\omega)^*\right]\right] = 0 \tag{5.74}$$

$$\mathrm{E}\left[\left[A(\omega)\otimes C(\omega)\right]\left[\Pi(\omega)^*\otimes N(\omega)^*\right]\right] = 0 \tag{5.75}$$

次に，式(5.72) の左辺の期待値をとってみよう．

$$\mathrm{E}\left[N(\omega)\left[\Pi(\omega)^*\otimes N(\omega)^*\right]\right] = \mathrm{E}\left[N(\omega)\int\Pi(\xi)^*N(\omega-\xi)^*\,d\xi\right] \tag{5.76}$$

$$= \mathrm{E}\left[\int\Pi(\xi)^*N(\omega)\,N(\omega-\xi)^*\,d\xi\right] \tag{5.77}$$

$$= \int\Pi(\xi)^*\,\mathrm{E}\left[N(\omega)\,N(\omega-\xi)^*\right]d\xi \tag{5.78}$$

ここで，式(5.78) 内の $\mathrm{E}[N(\omega)\,N(\omega-\xi)^*]$ について，ω と ξ $(\xi\neq 0)$ を固定して考える．もし，$\mathrm{E}[N(\omega)\,N(\omega-\xi)^*]\neq 0$ であれば，妨害音の ω 成分と $\omega-\xi$ 成分の両者の位相には一定の関係が存在する．これは，もし ξ が小さければ，この両成分の和によって周波数 ξ のエンベロープが（期待値で平均化してもなお）発生することを意味している．つまり，妨害音の強度が緩やかに変動するわけで，これでは定常とは言えなくなってしまうのである．

そこで，$\Pi(\xi)$ が非 0 の値を持つような微小の ξ（すなわち $|\xi|\leq B_\Pi$，ただし $\xi\neq 0$ とする）に対しては，$\mathrm{E}[N(\omega)\,N(\omega-\xi)^*]=0$ が成立するものと仮定する．これは，Cue Signal $\Pi(\xi)$ が変動する程度の時間よりも長い時間で観察すれば，妨害音が定常であるという仮定であり，意味的にも妥当であろう．

この仮定が認められれば，式(5.78) が 0 になるための残る条件は，$\xi=0$ の一点のみである．よって，

$$\Pi(0)=0 \tag{5.79}$$

であれば，

$$\mathrm{E}\left[\, N(\omega)\,[\Pi(\omega)^{*}\otimes N(\omega)^{*}]\,\right]=0 \tag{5.80}$$

となる．なお，式(5.79)は時間領域で書き直せば，定理5.1の条件4と同じ

$$\langle \alpha(t)\rangle = 0 \tag{5.81}$$

である．すなわち，Cue Signalの平均が0という要請に他ならない．

5つの式の最後は式(5.68)である．これも，期待値をとる．なお，式(5.68)は共通の因子$[A(\omega)\otimes C(\omega)]$を持つが，期待値をとるので，この因子を約分して消すことはできない．まず，左辺の期待値は，

$$\begin{aligned}
&\mathrm{E}\left[\,[A(\omega)\otimes C(\omega)]\,[A(\omega)^{*}\otimes C(\omega)^{*}]\,\right]\\
&\qquad = \mathrm{E}\left[\,\left\{\int A(\omega-\xi)\,C(\xi)\,d\xi\right\}\left\{\int A(\omega-\eta)^{*}\,C(\eta)^{*}\,d\eta\right\}\,\right] &(5.82)\\
&\qquad = \iint A(\omega-\xi)\,A(\omega-\eta)^{*}\,\mathrm{E}\left[\,C(\xi)\,C(\eta)^{*}\,\right]\,d\xi\,d\eta &(5.83)
\end{aligned}$$

である．次に，右辺の期待値を計算すると，

$$\begin{aligned}
&\mathrm{E}\left[\,[A(\omega)\otimes C(\omega)]\,[\Pi(\omega)^{*}\otimes A(\omega)^{*}\otimes C(\omega)^{*}]\,\right]\\
&\qquad = \mathrm{E}\left[\,\left\{\int A(\omega-\xi)\,C(\xi)\,d\xi\right\}\left\{\iint \Pi(\zeta)^{*}A(\omega-\eta-\zeta)^{*}C(\eta)^{*}\,d\eta\,d\zeta\right\}\,\right] &(5.84)\\
&\qquad = \iiint \Pi(\zeta)^{*}A(\omega-\xi)\,A(\omega-\eta-\zeta)^{*}\,\mathrm{E}\left[\,C(\xi)\,C(\eta)^{*}\,\right]\,d\xi\,d\eta\,d\zeta &(5.85)
\end{aligned}$$

ここで，$\xi\eta$空間で式(5.83)の$A(\omega-\xi)\,A(\omega-\eta)^{*}$が非0の値をもつ範囲は，$A(\omega)$が帯域$B_A$に制限されているので，**Fig. 5.3**左に示す部分となる．

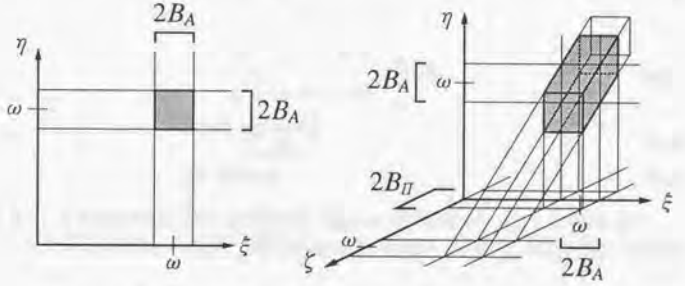


Fig. 5.3　式(5.83)と式(5.85)の包絡類部分の非0領域

一方，$\xi\eta\zeta$空間で式(5.85)の$\Pi(\zeta)^{*}A(\omega-\xi)\,A(\omega-\eta-\zeta)^{*}$が非0の値をもつ範囲は，$\Pi(\omega), A(\omega)$がそれぞれ帯域$B_\Pi, B_A$に制限されているので，**Fig. 5.3**右に示す部分となる．

そこで，妨害音のときの同じように長い時間で観察したときの定常性，

$$\mathrm{E}\left[\,C(\omega)\,C(\omega-\xi)^{*}\,\right]=0 \qquad |\xi|\le B_\Pi+2B_A \quad \text{かつ} \quad \xi\neq 0 \tag{5.86}$$

を仮定する．この仮定によって，式(5.83) と式(5.85) 内の $\mathrm{E}[C(\xi)\,C(\eta)^*]$ が 非 0 値をもつ領域は，$\xi=\eta$ 平面（直線）を除けば，Fig. 5.3左右両図の領域の外側に追い出すことができる．それゆえ，式(5.83) と式(5.85) の積分範囲を Fig. 5.3 両図の中でも $\xi=\eta$ 上のみに限定することができる．

さらに，さきに触れたように，アドホックはフィルタを作らないためには，フィルタのインパルス応答長に制限を加えることが必要である．そこで，Cue Signal 成分の周期のなかで最も短い周期よりもフィルタのインパルス応答長がさらに短いことを仮定する．すなわち，フィルタの周波数分解能が，B_A に比較してずっと粗いものであると仮定する．このとき，フィルタは $C(\omega)$ や $N(\omega)$ や $h_{\bullet\bullet}(t)$ の性質に対応して動作するので，$\mathrm{E}[C(\xi)\,C(\xi)^*]$ が，$\omega-B_A\le\xi\le\omega+B_A$ の範囲で一定であるとしても良いだろう．なぜなら，それ以上の $C(\omega)$ の微細構造への適応は Cue Signal の周期に匹敵する長さのインパルス応答を必要としてしまうからである．

以上の仮定をすると，式(5.83) と式(5.85) の比（すなわち，式(5.68) の両辺の比の期待値）は $C(\bullet)$ と ω によらない量となる．これを計算すると，

$$\frac{\text{式(5.85)}}{\text{式(5.83)}} = \frac{\mathrm{E}\left[\left[A(\omega)\otimes C(\omega)\right]\left[\Pi(\omega)^*\otimes A(\omega)^*\otimes C(\omega)^*\right]\right]}{\mathrm{E}\left[\left[A(\omega)\otimes C(\omega)\right]\left[A(\omega)^*\otimes C(\omega)^*\right]\right]} \tag{5.87}$$

$$= \frac{\displaystyle\iint \Pi(\zeta)^*A(\omega-\eta)\,A(\omega-\eta-\zeta)^*\,d\eta\,d\zeta}{\displaystyle\int A(\omega-\eta)\,A(\omega-\eta)^*\,d\eta} \tag{5.88}$$

$$= \frac{\displaystyle\iint \Pi(\zeta)^*A(\eta)\,A(\eta-\zeta)^*\,d\eta\,d\zeta}{\displaystyle\int A(\eta)\,A(\eta)^*\,d\eta} \tag{5.89}$$

$$= \frac{\displaystyle\iint \Pi(-\zeta)\,A(\eta)\,A(\zeta-\eta)\,d\eta\,d\zeta}{\displaystyle\int A(\eta)\,A(-\eta)\,d\eta} \tag{5.90}$$

$$= \frac{\left[\Pi(\omega)\otimes A(\omega)\otimes A(\omega)\right]\Big|_{\omega=0}}{\left[A(\omega)\otimes A(\omega)\right]\Big|_{\omega=0}} \tag{5.91}$$

$$= \frac{\langle\,\alpha(t)\,a(t)^2\,\rangle}{\langle\,a(t)^2\,\rangle} \tag{5.92}$$

$$= K[\alpha,a] \tag{5.93}$$

となる．すなわち，$K[\alpha,a]=1$ であれば，式(5.68) は期待値をとったうえで成り立つ．

以上をまとめると，式(5.68)～式(5.72) の条件は期待値をとれば，成立したことになる．よって，

$$\mathrm{E}[R_{u_m d}(\omega)] = \mathrm{E}[R_{u_m \psi_S}(\omega)] \qquad (m=1,\ldots,M) \tag{5.94}$$

であり，内部目標 d は最適目標 ψ_S に（期待値において）等価な適応目標であることが示された．なお，確定的な信号に対してまで期待値をとったのは，フィルタの自由度を制限するための措置であった．

式(5.94) を示すにあたり，数ページにわたって様々な条件を設定してきた．ここでそれらの条件を整理する意味もかねて，一般の線型フィルタの場合の Cue Signal 法の適応条件を定理にまとめておこう．

【定理 5.3】 一般の線型のフィルタを用いた場合，内部目標 $d(t) = \alpha(t)\,\psi(t)$ と最適目標 $\psi_S(t)$ は，以下の条件が満たされるならば，等価な適応目標である．

1. 目的信号が，エンベロープと搬送波に分離可能であること．
$$s(t) = a(t)\,c(t)$$
2. 包絡類（Cue Signal とエンベロープ）が，上限 B_A と B_Π の帯域制限信号でること．
$$A(\omega) = 0 \qquad \text{ただし } |\omega| > B_A$$
$$\Pi(\omega) = 0 \qquad \text{ただし } |\omega| > B_\Pi$$
3. 空間伝達のインパルス応答 $h_{\bullet\bullet}(t)$ が，$1/B_\Pi$ よりずっと短い時間で終了すること．
$$h_{\bullet\bullet}(t) = 0 \qquad \text{ただし } |t| > {}^\exists t_{max} \ll 1/B_\Pi$$
4. 線型フィルタのインパルス応答 $f_m(t)$ が，$1/B_A$ よりずっと短い時間で終了すること．
$$f_m(t) = 0 \qquad \text{ただし } |t| > {}^\exists t_{max} \ll 1/B_A$$
5. 目的音の搬送波成分が $1/(B_\Pi + 2B_A)$ 以上の観察では定常とみなせること．
$$\mathrm{E}[\,C(\omega)\,C(\omega-\xi)^*\,] = 0 \qquad \text{ただし } |\xi| \le B_\Pi + 2B_A \text{ かつ } \xi \neq 0$$
6. 妨害音が $1/B_\Pi$ 以上の観察では定常とみなせること．
$$\mathrm{E}[\,N(\omega)\,N(\omega-\xi)^*\,] = 0 \qquad \text{ただし } |\xi| \le B_\Pi \text{ かつ } \xi \neq 0$$
7. 目的信号と妨害信号が独立な信号であること．
$$\mathrm{E}[\,S(\omega_1)\,N(\omega_2)^*\,] = 0$$
8. Cue Signal が，強度エンベロープと相関をもつこと．
$$K[\alpha, a] = 1$$
9. Cue Signal が，平均 0 の信号であること．
$$\langle \alpha \rangle = 0$$

（定理おわり）

5.2.6 Cue Signal の直交化

ここまでの理論では，Cue Signal はすべての妨害音の強度エンベロープと無相関でなければならなかった．しかし，各音源の強度エンベロープがすべて無相関であるということは，実際の問題において少ない．たとえば，オーケストラ演奏の特定の楽器音を抽出することを考えてみれば，各音源の音の強度間には，かなり大きな相関関係があることがわかるだろう．したがって，目的音の強度エンベロープを推定し生成した Cue Signal は妨害音の強度エンベロープとは無相関にはならない，すなわち，定理 5.1 が当てはまる理想的な条件ではなく，定理 5.2 が述べるような目的音と妨害音の中間の取り扱いを受けるような音源が実際には存在するであろう．

そういうわけで，もし目的音の強度エンベロープを推定した信号をそのまま Cue Signal として用いたとすると，それは多くの場合，妨害音の強度エンベロープとも相関をもっているので，妨害音が抑制されないことになる．

そのような場合に，妨害音の強度エンベロープをも推定してやる必要がでてくる．以下，妨害音の強度エンベロープの推定値をどのように活用するかについて述べる．

一般に，2 つの信号 $A(t), B(t)$ があったとき，$A(t)$ は $B(t)$ に比例する成分と $B(t)$ に無相関な成分の和に分解できる．そこで，妨害音 j の強度エンベロープを，目的音の強度エンベロープに比例する項（比例定数 ϵ_j ）と，目的音の強度エンベロープに無相関な項（ $b_j(t)$ とする）の和として次のように記述する．

$$a_j(t)^2 = \epsilon_j\, a_0(t)^2 + b_j(t) \qquad (j = 1, 2, \ldots, J) \tag{5.95}$$

また，各音源の強度エンベロープを推定したときには，推定値は，各音源の強度エンベロー

ブに比例する項（比例定数 $\gamma_j > 0$ ）と，各音源の強度エンベロープに無相関な項（推定誤差）の和として次のように記述できる．

$$\beta_j(t) = \gamma_j\, a_j(t)^2 + c_j(t) \qquad (j = 0, 1, 2, \ldots, J) \tag{5.96}$$

ここで，$\beta_j(t)$ $(j = 0, 1, 2, \ldots, J)$ の推定誤差 $c_j(t)$ は，ランダムであり，他の推定誤差 $c_i(t)\,(i \neq j)$ および他の音の強度エンベロープ $a_i(t)^2\,(i \neq j)$ とは無相関であるとする．

Cue Signal $\alpha(t)$ として $\beta_0(t)$ を採用するのが最も単純な方法である．しかし，この方法では妨害音 j について

$$\begin{aligned} K[\beta_0, a_j] &= \left\langle \left[\gamma_0 a_0(t)^2 + c_0(t)\right] \left[\epsilon_j a_0(t)^2 + b_j(t)\right] \right\rangle \Big/ \left\langle a_j(t)^2 \right\rangle \\ &= \epsilon_j \gamma_0 \left\langle a_0(t)^4 \right\rangle \Big/ \left\langle a_j(t)^2 \right\rangle \end{aligned} \tag{5.97}$$

となって，$\epsilon_j \neq 0$ のとき（すなわち目的音強度に依存する強度変化をする妨害音 j に対しては），$K[\beta_0, a_j] \neq 0$ であるので，その妨害音は除去できないことになる．

しかし，強度エンベロープは常に非負である，つまり，妨害音 j の強度が恒等的に 0 でないかぎり $\epsilon_j \neq 0$ となってしまう．これをとりあえず避けるのが Cue Signal の平均を 0 にするという要請だったわけである．妨害音 j の強度エンベロープが未知の状況では，これが精いっぱいの対処だった．

これに対して，本節で述べる妨害音 j の強度エンベロープをも推定するアルゴリズムでは，Cue Signal $\beta_0(t)$ を $\beta_j(t)\,(j = 1, 2, \ldots, J)$ に直交させて妨害音 j を除去することにする．（複数の妨害音源に関して推定する場合には，それらすべてと順次直交させる．）この方法で特性が改善されることを以下で示す．

今，妨害音 j に関する強度推定から Cue Signal を改善するとすれば，直交化された新しい Cue Signal は，β_0 から β_j に相関をもつ成分を取り除いて，

$$\alpha(t) = \beta_0(t) - \frac{\langle \beta_j \beta_0 \rangle}{\langle \beta_j{}^2 \rangle} \beta_j(t) \tag{5.98}$$

で与えられる．

さて，Cue Signal $\beta_0(t)$ による適応と Cue Signal $\alpha(t)$ による適応を比較してみよう．従来の Cue Signal $\beta_0(t)$ を採用した場合は，前にも述べたように，目的音を $K[\beta_0, a_0]$ 倍にし，妨害音 j を $K[\beta_0, a_j]$ 倍にすべく適応化が行なわれる．一方，新しい Cue Signal $\alpha(t)$ を採用した場合には，目的音を $K[\alpha, a_0]$ 倍にし，妨害音 j を $K[\alpha, a_j]$ 倍にすべく適応化が行なわれる．以上より，次式の左辺が 1 より大きければ，後者のほうが妨害音 j を抑制できることになる．これを式(5.95)～式(5.98) を使って計算すると，

$$\frac{K[\alpha, a_0] \Big/ K[\alpha, a_j]}{K[\beta_0, a_0] \Big/ K[\beta_0, a_j]} = 1 + \frac{\gamma_j{}^2 \langle b_j{}^2 \rangle}{\langle c_j{}^2 \rangle} > 1 \tag{5.99}$$

となって，妨害音 j の強度エンベロープを推定して Cue Signal をそれと直交化したほうが，常に妨害音 j の抑制が良好となることがわかる．特に，妨害音の強度エンベロープを完全に推定できた場合（すなわち $\langle c_j{}^2 \rangle = 0$ ）には，妨害音 j の成分を 0 とすべく適応が行なわれることになる．

ただし，この方法で妨害音 j を抑制したときの他の妨害音 $i\,(i \neq j)$ に対する影響を同じように計算すると，

$$\frac{K[\alpha, a_0] \Big/ K[\alpha, a_j]}{K[\beta_0, a_0] \Big/ K[\beta_0, a_j]} = \left(1 - \frac{\epsilon_j \gamma_j{}^2 \langle b_j b_i \rangle}{\epsilon_i \left(\gamma_j{}^2 \langle b_j{}^2 \rangle + \langle c_j{}^2 \rangle\right)}\right)^{-1} \tag{5.100}$$

であるので，i については特性が良好になる場合も悪化する場合もある．しかし，もし $\epsilon_i \neq 0$ かつ $\langle b_j b_i \rangle \simeq 0$ であれば上式は約 1 となるので，妨害音 j で直交化することは，妨害音 $i\,(i \neq j)$ の抑制には影響しない．

以上のことから，妨害音のなかで相対的に大音量を出している妨害音源については，その音源の強度エンベロープをも推定して，積極的に抑制したほうがよいと結論できる．

5.3 シミュレーション

5.3.1 シミュレーション条件

まず，Cue Signal による適応を実証するために計算機シミュレーションを行なった．

音源数 J は 3 とし，マイクロホン数 M は 6 とした．音源は白色雑音を球面波で発する（いわいる呼吸球，あるいは点音源）ものとし，そのピークの信号強度が等しくなるようにした．また，マイクロホンは無指向性であるとした．

音源とマイクロホンの配置を **Fig. 5.4** に示す．マイクは正 8 面体の頂点に配置した．また，残響のある音空間をシミュレートするために，Fig. 5.4 の x-y 平面，y-z 平面，z-x 平面はすべて反射率 1 の反射壁とし，音は周波数によらず幾何光学的に反射すると仮定した．なお，3 方のみに反射壁があるので，音源からマイクロホンまでのインパルス応答はすべて有限時間で完了する．

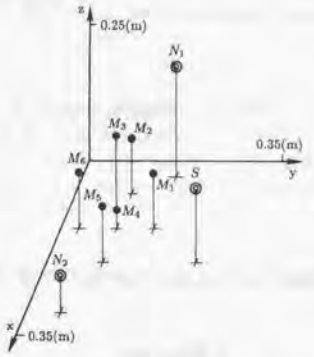


Fig. 5.4 計算機シミュレーション時のマイクと音源の位置

シミュレーションは，サンプリング周期は $T = 0.1$ms（周波数 10kHz，空間上で約 3cm 間隔のサンプリングに相当する）の離散時間系で行なった．0～5kHz の帯域の白色雑音は，乱数で代用した．各音源間の乱数は，無相関である．

なお，Fig. 5.4 に示す音源の位置では，0.1ms より細かい遅延差が生じる．厳密には，デジタルフィルタによる補間で対処するべきであるが，ここでは簡単のため，すべての遅延時間は 0.1ms の正数倍で近似して計算を行なった．このようにしても，本質的な差はないものと考えられる．

その他のシミュレーション条件もまとめて，**Table. 5.1** に示す．

Table 5.1 シミュレーションの設定

項目	シミュレーションで用いた値，条件	備考
音源数 J	3	$J < M$
センサ数 M	6	
目的音源 $s(t)$	白色雑音を振幅変調したもの	包絡は Fig. 5.5 参照
妨害音源 $n_1(t), n_2(t)$	白色雑音	
目的音源 S の位置	x=20cm, y=27cm, z=14cm	
妨害音源 N_1 の位置	x=3cm, y=17cm, z=20cm	
妨害音源 N_2 の位置	x=30cm, y=6cm, z=7cm	
マイクロホンアレイ	中心：x=14cm, y=10cm, z=10cm	M_1–M_1 間は 14cm
反射壁	x-y 平面，y-z 平面，z-x 平面	反射率= 1.0
目的音の規範	非定常音（強度の変化する音）	
サンプリング周期 T	0.1ms	F = 10kHz
FIR フィルタタップ数 K	32	
内部目標の被乗算量 $\psi(t)$	$y_{16}(t)$	

5.3.2 評価基準

本節では，シミュレーションに用いた評価関数について述べる．

本システムの適応は，目的音と出力の差の 2 乗 $\langle e(t)^2 \rangle$ を最小にするよう行なわれる．フィルタリングが線型に行なわれることから，この誤差は，SN 比（妨害音の混入量）と，目的信号劣化比（目的音の劣化量）の 2 つに分離して評価することができる．

5.3.2.1 SN 比

システムの出力 $\phi(t)$ は，目的音を要因とする項 $\phi_S(t)$ と，妨害音を要因とする項 $\phi_N(t)$ の和に分離できる．

そこで，SN 比 E_{SN} を，

$$E_{SN} \stackrel{\text{def}}{=} \frac{\langle \phi_S(t)^2 \rangle}{\langle \phi_N(t)^2 \rangle} \tag{5.101}$$

で定義する．

シミュレーションでは，各音源からシステム出力までのインパルス応答と各音源の自己相関関数が既知であるので，両者を用いてシステム出力での各音源成分の強度 $\phi_S(t)^2$ と $\phi_N(t)^2$ を算出している．

5.3.2.2 目的信号劣化比

受信音の特定の周波数帯域において妨害音の比率が高いとき，ウイナーフィルタは，その帯域のゲインを低下させるように動作する．

よって，目的音源からシステム出力までの伝達関数は 1 にはならない．これは，妨害音が音を発生していない瞬間においても出力が歪む（線型歪）ということを意味している．

そこで，目的信号劣化比 E_D を，

$$E_D \stackrel{\text{def}}{=} \min_K \frac{\langle (\phi_S(t) - K\psi_S(t))^2 \rangle}{\langle (K\psi_S(t))^2 \rangle} \tag{5.102}$$

で定義する．K に関する最小をとるのは，式(5.26) のところで述べたように，$\alpha(t)$ の強度によって出力レベルが変化するためである．

5.3.3 学習方法

内部目標 $d(t)$ と，出力 $\phi(t)$ の 2 乗誤差を最小にするための学習方法として，シミュレーションでは以下の 2 方法を用いた．

第一の方法は，正規方程式(3.25) を逆行列演算によって解く**直接法**である．ここでは，0 s からその時点までのデータをもとに正規方程式を解いてフィルタの更新を行なっている．また，このフィルタの更新間隔は，10ms である．

第二の方法は，**LMS アルゴリズム**である．LMS アルゴリズムには様々なバリエーションがあるが，今回ここで用いたのは最も単純な方法（各サンプルタイム毎に最急降下で修正）である．

なお，修正係数は，学習速度と学習完了後の特性のトレードオフを観察しながら，適当と思われる値に調節してある．よって，以下で示す結果は絶対的なものではなく，学習完了後の特性を犠牲にすれば，より迅速な学習速度が得られ，また，学習速度を犠牲にすれば，より良好な学習完了後の特性が得られることを付記しておく．

5.3.4 シミュレーション結果 1：一例

Fig. 5.5 にシミュレーションでの各信号波形の例を示す．

目的音は白色雑音を Fig. 5.5 (a) に示す波形で振幅変調して生成した断続的な音である（Fig. 5.5 (b)）．これに，一定振幅の白色雑音である妨害音 2 つを加えて作成された Fig. 5.5 (c)～(h) が 6 本のマイクロホン信号 $u_1(t)$～$u_6(t)$ である．目的音の包絡が見えるか見えないか程度に妨害音が加算されている．

Fig. 5.5 (i) が，Cue Signal $\alpha(t)$ の一例である．これは人為的に作った Cue Signal である．平均が 0 で，かつ Fig. 5.5 (a) との間に正の相関を有するので前に述べた Cue Signal の条件を満たしている．

最後に，Fig. 5.5 (j) が，この Cue Signal を用いたときのシステム出力波形 $\phi(t)$ である．この例では，2 乗誤差最小の学習法として，直接法を用いている．Fig. 5.5 (c)～(i) のみから，目的音波形 Fig. 5.5 (b) がほぼ再現できていることがわかる．また，Cue Signal として方形波的波形を入れていても，出力のエンベロープは正しく 3 角波的波形になっていることがも確認された．

5.3.5 シミュレーション結果 2：種々の Cue Signal

理論（**5.2** 節）で述べたように，Cue Signal は，定理の条件を満たせば，すべて適応可能である．しかし，それは，条件を満たす Cue Signal がすべて同じ特性へ収束するといっているだけである．よって，Cue Signal によって学習の効率（収束速度）には差が生じる．

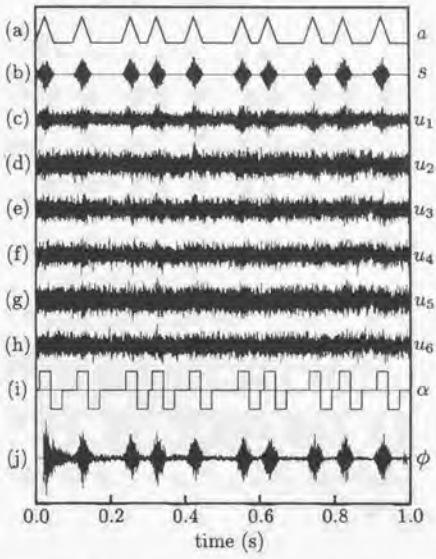


Fig. 5.5 シミュレーションでの波形の例

また，実際には，条件が完全に成立するとはかぎらない．条件からのズレが結果におよぼす影響も調べる必要がある．

以上2点をシミュレーションによって検討した．

Fig. 5.6 (A)~(G) にシミュレーションで採用した7種類の異なる Cue Signal を，Fig. 5.5 と同一の時区間で示す．

このうち，(A)~(C) の3種類は，Fig. 5.6 (a) から適当に Cue Signal として典型的な波形を作ったものである．言うなれば，人為的な Cue Signal である．また，残りの (C)~(G) の4種類は，マイクロホン信号（Fig. 5.5 (c)）を信号処理することで生成した Cue Signal である．こちらは，センサ情報のみから生成した Cue Signal である．

7種類の Cue Signal の生成方法と性質を **Table. 5.2** にまとめておく．

5.3.5.1 直接法による結果と考察

直接法による適応結果を **Fig. 5.7** に示す．上段が SN 比（E_{SN}）で評価した学習曲線で，下段が目的信号劣化比（E_D）で評価した学習曲線である．

Fig. 5.7 より以下のことがわかった．

まず，どの Cue Signal を用いても，2 s 以下の受信により SN 比で 20 dB 以上の特性改善が得られることが確かめられた．

また，学習は40秒程度でほぼ完了しているが，(A)~(C) に比較して，それ以外の4つ

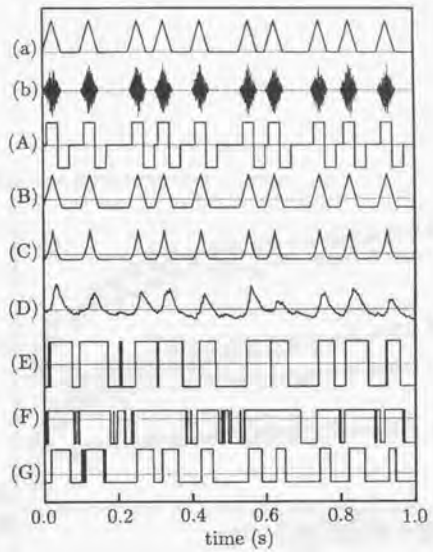


Fig. 5.6 シミュレーションで用いた Cue Signal の波形

〈(C)~(G)〉の SN 比の収束値は低い．これは，(C)~(G) が理想的な Cue Signal ではないこと—すなわち，妨害音の信号強度〈一定振幅の白色雑音だが，実際には信号強度のゆらぎが存在する〉にも若干の相関を持った Cue Signal であること — に起因すると考えられる．

また，(A)~(C) の SN 比がほぼ等しい収束値を有することは，理論（**5.2** 節）で述べた「条件さえ満たせばフィルタ係数の収束値が同じになる」ということを裏付けている．

次に，受信音がら Cue Signal を作る方法〈(D)~(G)〉に限定して，その比較を行なってみる．SN 比〈E_{SN}〉と目的信号劣化比〈E_D〉両方を勘案すると，最も好ましい Cue Signal は，(E) である．これは，(D) に比較して 2 値化処理が入るため妨害音強度の微少なゆらぎの影響をうけにくいこと，(F),(G) に比較して情報量が大きい（正負の確率が 1/2 だから）ためと考えられる．

5.3.5.2 LMS アルゴリズムによる結果と考察

LMS アルゴリズムによる適応結果を **Fig. 5.8** に示す．上段が SN 比（E_{SN}）で評価した学習曲線で，下段が目的信号劣化比（E_D）で評価した学習曲線である．

また，記号「+」は，目標信号として内部目標 $d(t) = \alpha(t)\,\psi(t)$ のかわりに最適目標 $\psi_S(t)$ を用いたときの学習曲線である．これは，教師つきのトレーニングであり，比較のために示してある．

Fig. 5.8 より以下のことがわかった．

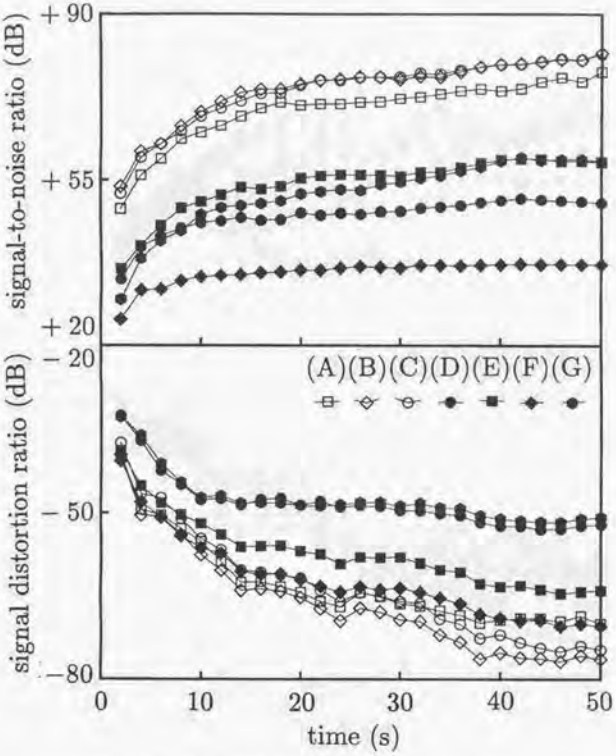


Fig. 5.7　直接法による学習曲線（シミュレーション）

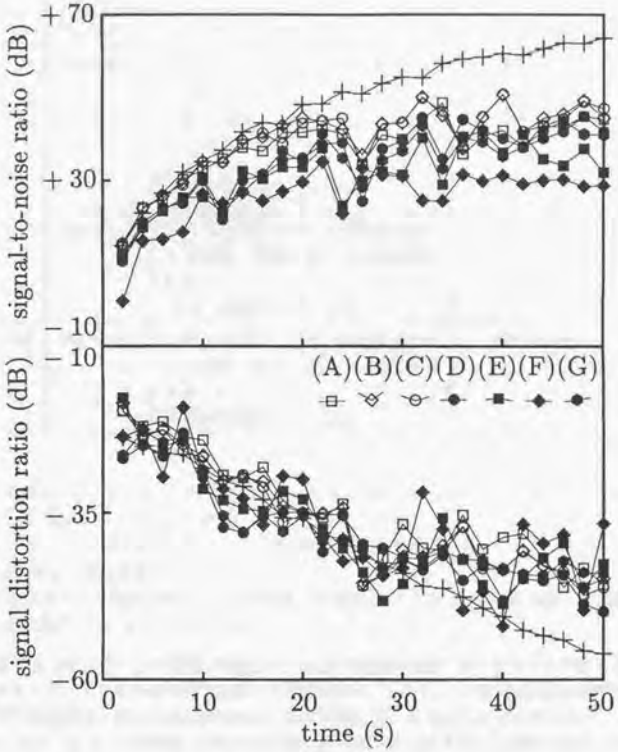


Fig. 5.8 LMS アルゴリズムによる学習曲線（シミュレーション）

Table 5.2 シミュレーションで用いた Cue Signal の生成方法と性質

	処理対象	処理方法	妨害音強度との相関
(A)	Fig. 5.5 (a)	→ 正の相関を持つ適当な 3 値関数で置き換える	完全にない
(B)	Fig. 5.5 (a)	→ 直流成分をカット	完全にない
(C)	Fig. 5.5 (a)	→ 2 乗 → 直流成分をカット	完全にない
(D)	Fig. 5.5 (c)	2 乗 → 低域フィルタ（時定数 20ms）→ 高域フィルタ（時定数 140ms）	僅かに有り
(E)	Fig. 5.5 (c)	2 乗 → 低域フィルタ（時定数 20ms）→ 2 値化（閾値：中）→ 直流成分をカット [正負の時間比＝ 1 ： 1]	僅かに有り
(F)	Fig. 5.5 (c)	2 乗 → 低域フィルタ（時定数 20ms）→ 2 値化（閾値：低）→ 直流成分をカット [正負の時間比＝ 3 ： 1]	僅かに有り
(G)	Fig. 5.5 (c)	2 乗 → 低域フィルタ（時定数 20ms）→ 2 値化（閾値：高）→ 直流成分をカット [正負の時間比＝ 1 ： 3]	僅かに有り

- 当然ながら，SN 比，目的信号劣化比，ともに直接法より劣化した．
- Cue Signal による差は直接法のときほど顕著ではない．
- 約 20 s までは，最適目標による学習曲線に準じた特性をたどる．
- 約 30 s 程度で学習が完了する．
- 学習完了後の最適目標による学習曲線との差は，SN 比で 約 20 dB の低下，目的信号劣化比で 約 10 dB の劣化である．

さきに述べたように，学習完了時間やそのときの特性は LMS アルゴリズムの修正係数に依存するので，それらの絶対的な数値には意味はない．しかし，上の結果は約 30 秒で学習が完了する必要がある場合の最終的特性の上限を表わしているということはできる．

Cue Signal による差が顕著でなかった理由は，LMS アルゴリズムを採用したことによる特性の劣化が大きく，Cue Signal の良し悪しが表面化しなかったためと考えられる．

また，学習完了後の特性が，最適目標のそれに及ばなかった理由は，勾配雑音が非常に大きく，しかもそれがゆっくりと変動することによる影響と考えられる．

これは，具体的には，以下のような意味である．

すなわち，もし最適目標を用いたときは，出力すべき波形が最適目標に全く同一であるため，学習完了後はエラー $e(t)$ がほとんど 0 になる．よって $e(t)^2$ の勾配もほとんど 0 になる．

しかし，内部目標を用いたときには，出力すべき波形と内部目標が似ても似つかない波形であるため，学習完了後もエラー $e(t)$ が小さくならない．よって $e(t)^2$ の勾配も小さくなら

ない．

LMS アルゴリズムの急所は，$\langle e(t)^2 \rangle$ を $e(t)^2$ で置き換えたところにある．そして，学習完了は，$\langle e(t)^2 \rangle$ の勾配が 0 ということである．このとき，最適目標に関してはこの置き換えは妥当であるが，内部目標に関しては $\langle e(t)^2 \rangle$ の勾配は 0 にもかかわらず，瞬時瞬時の $e(t)^2$ は大きな値を持ってしまうため妥当ではない．すなわち，仮に最適特性に到達したとしても，$e(t)^2$ によって不用な修正が施され，フィルタ特性は最適特性のまわりに振動させられてしまう．

以上の理由により，LMS アルゴリズムにおける学習完了特性では，最適目標と内部目標による差が生じると考えられる．

5.3.6 シミュレーション結果 3：インパルス応答でみた適応特性

5.2 節で述べたように，Cue Signal 法の学習は目的音そのものを抽出するのではなく，受信音の中の目的音成分のみを抽出すべく行なわれる．すなわち，システム出力 $\phi(t)$ は，$s(t)$ の推定値ではなく，$\psi(t)$ 中の目的音成分 $\psi_S(t)$ の推定値である．

このため，音響空間に反射や残響があれば，それはそのまま保存されて出力されることになる．これをシュミュレーションで示す．

Fig. 5.9 は，音源 $(s(t), n_1(t), n_2(t))$ とシステム出力 $(\phi(t))$ 間のインパルス応答が，学習の進行とともにどのように変化するかを示したものである．時間は，学習開始から，0.1s，1s，10s の 3 つを上から順に示した．左右方向には，左に目的音–出力間をとり，中央と右に妨害音–出力間をとった．このようにして定義した 9 個の升目にインパルス応答をプロットしてある．なお，シミュレーション条件は，Table. 5.2 である．

これより，時間の経過とともに妨害音–出力間のインパルス応答は抑制されていき，目的音–出力間の伝達のみが残されていく様子がわかる．

また，左下のインパルス応答を見ればわかるように，反射に起因する成分は保存されて学習が進んでいる．

一方，反射壁を取り除いた場合のインパルス応答の変化を **Fig. 5.10** に示す．シミュレーション条件は，壁の反射率を 0 にしたことを除いて，Table. 5.2 と同じである．

Fig. 5.10 の左下の応答を見ればわかるように，反射なしの場合は，目的音（を遅延したもの）を抽出するよう動作していることがわかる．

なお，Fig. 5.10 において，上の 3 つの応答は，すべてインパルス的になっている．これは，短時間の適応では，Cue Signal の平均が 0 になっていないため，妨害音も目的音とみなされて学習が行なわれていることを示すものである．

5.4 実験

本論文の研究における実験は，複数の DSP(Digital Signal Processor) を複数用いて製作した実時間処理系と，ワークステーションの適応化プログラムによる非実時間処理系の 2 種類の方法で行なっている．

このうち，非実時間の実験では，各音源（目的音，妨害音）の信号を別個にひとつづつ鳴らして採取し，各マイクロホンの出力信号は，ワークステーション内でそれら別個の信号の和をとることによって作成する．この利点は，

- 全く同じ波形で繰り返し実験できるので，種々の適応手法についてその適応能力を厳密に比較することができる．

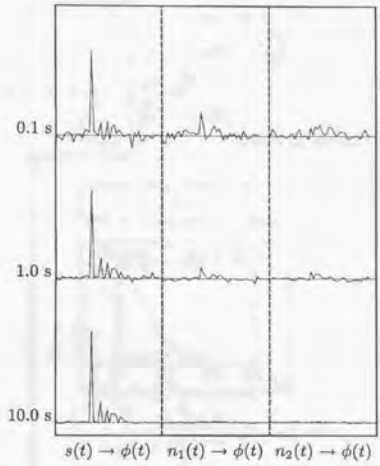


Fig. 5.9 音源とシステム出力間のインパルス応答の推移（反射壁あり）

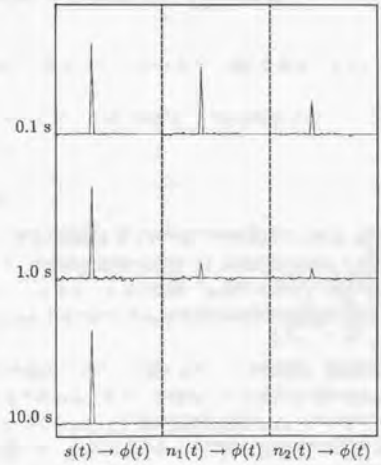


Fig. 5.10 音源とシステム出力間のインパルス応答の推移（反射壁なし）

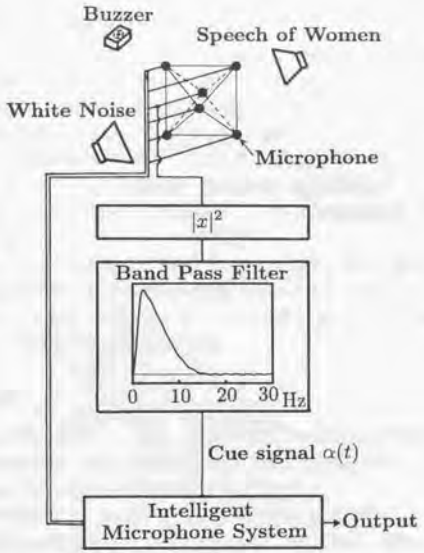


Fig. 5.11 非定常音を抽出する適応型音響センシングシステム

- 各音源ごとの真の波形がわかっているので，適応実験後に SN 比などの評価量を算出することができる．

である．この理由から，本節では非実時間系での実験結果を述べ，実時間系については，第 II 部で述べることとする．

5.4.1 実験方法

実験は，無響室にさざまな音源とマイクロホン群を設置して行なった．なお，無響室は，条件を単純にするためと，外部の雑音から遮断された環境を必要としたために使用した．内壁の反射による影響は，シミュレーションの節（p.51の **5.3.6**）で述べたとおりで，反射音が無いということは（**5.2.5** で述べたように秒単位の残響を問題にするのでなければ）それほど本質的なことではない．

Fig. 5.11 に実験の理論的なブロック図を示す．この実験は，定常妨害音（ブザー，白色雑音）のなかから，非定常目的音（音声）を抽出しようとするものである．

無響室には 6 個のマイクロホンと 3 個の音源を設置した．マイクロホンは一辺 155mm の正 8 面体の頂点に配置した．ちなみに，センサ配置の最適性については，**6.7** 節で述べる．

マイクロホン出力は，プリアンプ，アンチエイリアスフィルタを経て 44.1kHz の周波数で AD 変換されパーソナルコンピュータの RAM（16M バイト）に一旦蓄えられる．6 チャネル入力では 30 秒程度の実験が可能である．記憶された音響信号は，LAN によって別室のワーク

ステーションに転送される．そして，ワークステーション内のソフトウエアにより Cue Signal 法による適応的信号抽出の実験を行なうわけである．

Fig. 5.11 で，Cue Signal 生成について説明する．ここで与えられた音源識別の規範は，「目的音が非定常音（強度が変動する音響信号）であり，妨害音が定常音（一定強度の音響信号）である」という最も易しいものである．Cue Signal は，受信音を 2 乗した信号を，音声のエンベロープを抽出するためのバンドパスフィルタで濾過することによって生成できる．

なお，この Cue Signal 用バンドパスフィルタは，ワークステーション内の FIR デジタルフィルタで実現している．フィルタ係数は，通過帯域と阻止帯域を決めて，反復法による計算機設計で得た等リップルフィルタ[38]である．フィルタの周波数特性は，Fig. 5.11 の中央のワクの中に示してある．

ワークステーション上のソフトウエアは，フィルタの係数の更新と推定出力の算出をするだけでなく，各時点でのシステムの特性の評価値を出力するよう製作した．また，得られた出力信号は LAN によってパーソナルコンピュータに再度転送され，DA 変換器を介して学習結果をモニタリングできるようにもなっている．

5.4.2 実験結果

実験結果を **Fig. 5.12** に示す．これは，LMS アルゴリズムによる学習結果である．なお，ここでの評価量（SNR）は，信号と推定誤差の比で表わしている．すなわち，シミュレーションで述べた式(5.101) と式(5.102) の両者を加味した量である．

Fig. 5.12 について説明する．(a) は，マイクロホン信号 $u_1(t)$ である．これに，Fig. 5.11 に示した処理をして得られたのが (b) に示す Cue Signal $\alpha(t)$ である．実際の音声の波形は (d) に示されている．(d) の強度を (b) がある程度推定できていることがわかる．(c) がシステム出力である．(d) と見比べてみて，5 秒程度の学習で，出力が安定してきていることがわかる．なお，FIR フィルタ係数の初期値はすべて 0 とした．

Fig. 5.12 (e) に信号・誤差比で見た学習曲線を示す．太線が内部目標による（すなわち Cue Signal 法による）学習曲線を表わし，細線が最適目標による（すなわち 理想的なトレーニング信号による）学習曲線をあらわしている．それぞれの 3 本の学習曲線は LMS アルゴリズムにおける修正係数 μ を変化させて実験したものである．(四角，丸，菱形の順に値が約 2 倍になっている)．ちなみに，Cue Signal 法の場合は，中間の値である丸印が最も好ましい結果を与えている．(e) は，その値での出力波形である．また，最も上にある破線は，最適目標を用い，LMS アルゴリズムではなく直接法により式(3.25) の方程式を 0.22 s 間隔で解いていったものである．これは，未来のデータを使わないという制限内での最適解を示すために表示した．

以上の結果より，マイクロホン 1 個の場合に比較して 約 18 dB の信号・誤差比の向上が見られ，それは最適解に比較しても 約 4 dB の劣化にすぎないことがわかる．

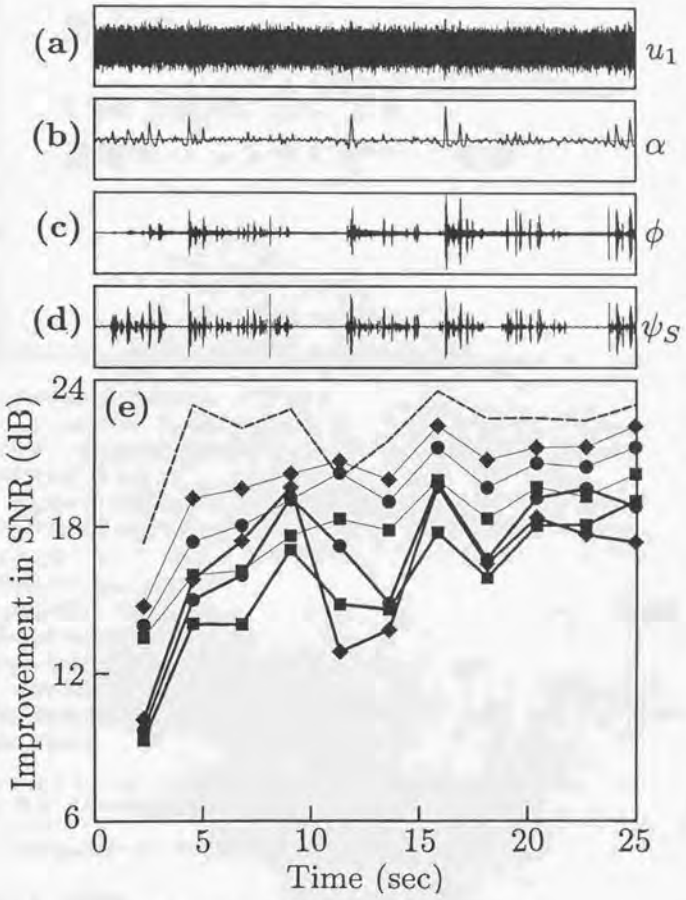


Fig. 5.12 非定常音（音声）の抽出実験の結果

第 6 章

Cue Signal 法による音響センシングに関する考察

この章は，第 I 部における考察の章である．

まず，**6.1** 節では，Cue Signal 法の特徴をまとめる．次に，**6.2** 節では，これまで天下り的に採用してきた信号抽出の評価基準である「平均 2 乗誤差最小の規範」が適当であるのか否かを考え直してみる．

その次の 3 つの節では，Cue Signal 法と各種適応アルゴリズムとの相性を考察する．LMS アルゴリズムとの適合性を **6.3** 節で，学習同定法との適合性を **6.4** 節で，直接法との適合性を **6.5** 節で，それぞれ検討する．なお，直接法に関する検討では，仮 Cue Signal を用いた発展型の Cue Signal 法を提案する．

理論の部分でも述べたように，Cue Signal 法が運用できるためには目的信号が分離可能な形にモデル化されなければならない．それでは，この基準を満たさない音源に対して，Cue Signal 法がどのような動作をするかを検討したのが **6.6** 節である．

マイクロホンの本数，その設置位置，FIR フィルタの次数，等の最適化問題は，Cue Signal 法特有の問題ではないが，興味深い問題である．本章の最後（**6.7** 節）では，これらの問題に簡単に触れる．

6.1 Cue Signal 法の特徴

Cue Signal 法には以下のような特徴がある．

6.1.1 柔軟性

Cue Signal の満足すべき条件は，平均値をとったうえで規定される．(p.32 定理 5.1の条件 3,4 または，p.41 定理 5.3の条件 8,9) このため，Cue Signal として，Fig. 5.6 (A)～(G) に示したようなさまざまな波形を波形を用いることが可能となった．すなわち，多様な Cue Signal に関して柔軟に対応することができるわけである．

多数の音源の中から何が目的音かを定義するのに，従来は単純物理量レベルの規範を用いていた．この規範を，意味レベルに近づけるのが我々の当初の目的であった．この目的に，Cue Signal の柔軟性が重要である．また，視聴覚融合など異種情報による適応や，センサシステム内部の知識ベースを用いた適応に発展させる際にも，Cue Signal の柔軟性は不可欠な特徴であるといえる．

6.1.2 システムノイズ

理論の章でも述べたように（p.14 参照），マイクロホンや，プリアンプ，A/D 変換器などで発生するシステムノイズは，妨害音と見かけ上全く同じように扱われる．すなわち，妨害音の数が増えた程度の影響しか及ぼさない．

例えば，あるマイクロホンのプリアンプが個体差により若干ノイズレベルが高かったとする．そのような場合には，システムは自動的にそのマイクロホンの加重を低めるように適応する．個々のシステムノイズを事前に評価する必要はない．

この利点は，音源位置などの知識を用いず，受信データのみからの適応することによってもたらされた．

6.1.3 センサ特性のバラツキ

マルチセンサを用いたセンシングシステムでは，特性の揃ったセンサデバイスを用意するのに労力を強いられることがある．また，センサの選別や調整が不可能な場合は，手間をかけてそれらのバラツキを補償する処理をしなければならないことも多い．

Cue Signal 法では，センサ特性のバラツキは問題にならない．それは，センサの周波数特性をも伝達関数 $H_{jm}(\omega)$ に繰り込んでしまって考えることができるからである．

これも，受信データのみからの適応の結果であると言える．

ただし，内部目標の被乗算信号 $\psi(t)$ をとりだすためのマイクロホンの周波数特性は，そのまま，出力の周波数特性に影響する．

また，センサが非線型な歪を発生する場合は，やはり影響がある．しかし，それにもある程度は自律的に対処することが可能である．例えば，多数のセンサのうち，ある1個のセンサが強烈な歪を発生していたとしよう．非線型歪は本システムでは妨害音が加算されたとして扱われるから，そのセンサのゲインは自動的に下げられるであろう．ここで言う自律的な対処（適応）とは，そのような意味である．しかし，その対処の能力を定量的に解析するのは，現在のところ難題である．

6.2 平均2乗誤差最小の規範の是非

われわれは，これまで，適応の規範として天下り的に平均2乗誤差最小の規範を導入してきた．しかし，これには何らかの必然性があるわけではない．ここでは，平均2乗誤差最小規範の問題点とその解決法について考える．

6.2.1 平均2乗誤差最小の規範の問題点

平均2乗誤差最小の規範で線型フィルタを適応化するということは，言うなれば，ウイナーフィルターを作るということである．

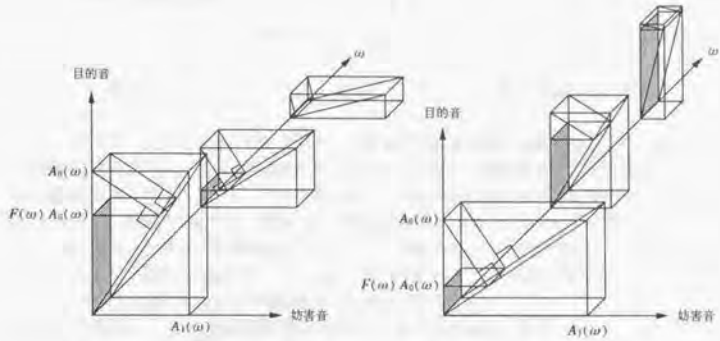


Fig. 6.1 平均2乗誤差を最小にする規範で生じる線型歪み

たとえば，1入力1出力の線型フィルタに目的音と妨害音の混合信号を入れてみよう．ここで両信号は無相関とする．今，ある周波数 ω での目的音強度が $A_0(\omega)^2$ で 妨害音強度が $A_1(\omega)^2$ と すると，平均2乗誤差最小で目的音波形を推定するためのその周波数のゲイン $F(\omega)$ は，

$$F(\omega) = \frac{A_0(\omega)^2}{A_0(\omega)^2 + A_1(\omega)^2} \tag{6.1}$$

となる．この様子を **Fig. 6.1** に図示する．左は妨害音がスペクトル一定（白色）の場合で，右は目的音がスペクトル一定の場合である．どちらを見ても，妨害音が目的音に比較して強い帯域ほど，出力の目的音強度が低下させられてしまう様子がわかるであろう．

これは，複数入力（複数マイクロホン）の場合でも当然起こりうる．すなわち，複数の入力を干渉させただけで妨害音が消去できれば問題ないが，そうでない場合は，平均2乗誤差最小の要望に沿うように周波数特性が操作される．具体的には，妨害音の比率が相対的に高い帯域が抑圧される．これは，目的音から見れば線型歪である．

この問題は，実際の音声と白色雑音を使った実験（p.51の **5.4** 節）でも如実にあらわれた．音声は 100～1kHz にエネルギーを集中させた信号[39]である．一方，実験で用いた白色雑音は，スピーカの特性の影響を受けるものの，0～20kHz に一様にエネルギーを有する信号である．このため，1kHz 以上の高い周波数成分が抑圧されてしまう．具体的には，音声が「こもったような音」になるという現象が認められた．特に音声の場合は，その認識に必要な情報（ホルマント）が，基音部分ではなくて，エネルギーの落ち込んでいる高い周波数帯域に存在するので，その影響は無視できない．

この現象は，「線型フィルタ＋平均2乗誤差最小規範」を使うかぎり避けられない問題である．Cue Signal 法とは関係がない．高い音を抑制したほうが平均2乗誤差が小さくなるのだから仕方がない．

6.2.2 平均2乗誤差最小規範による不具合の対処法

そういうわけで，システムの性能を保ったままシステムに改良を行ない対処するのは不可能であるが，この現象をさける方法として，次の2つの方法が考えられる．第一に，平均2乗誤差最小の規範をはずしてしまう方法，第二に，システム出力後に補償フィルタを入れる方法，である．

第一の方法は，具体的には，FIR フィルタの係数調整の自由度を制限したうえで平均2乗誤差最小に適応化するのが最も現実的と思われる．もちろん，制限をつけたわけだから，平均2乗誤差は増大する．しかし，それでも目的音を劣化させないことを優先させたい場合もあるだろう．この第一の方法は，実際には，拘束条件つきの LMS アルゴリズム[40]などを用いれば実現できる．ただし，その拘束条件の与え方が問題である．第4章で紹介した Flanagan の方法のように，適応の自由度を，上下左右の指向特性という2自由度に制限するためには，マイクロホンの位置や特性などの知識を与えなければならない．

第二の方法は，システム自体には手をつけずに，出力の周波数特性を1入力1出力の線型フィルタで補償しようというものである．しかし，問題がある．目的音の周波数特性がどのように劣化したかをどうやって知るかということである．これには，Cue Signal（または Event Signal）を再利用すればよいであろう．すなわち，入力と出力の各周波数帯域で出力変動を監視しておき，その変動を Cue Signal の変動と比較して見ることで，各帯域での目的音成分の減衰量を見積ることができそうである．

以上，平均2乗誤差最小を犠牲にして，システムを改造する方法を述べたが，つまるところ，これから先の価値判断は，本システムの出力を何に使うのかということに依存するのではないか．すなわち，本システム出力を人間が音声を聴くための用いるのであれば，現状は多少具合いが悪いだろう．平均2乗誤差最小は捨てたほうがいいかもしれない．しかし，音声認識装置などへの入力として用いるのであれば，ホルマントや極の位置の推定を行なうのに，高域が下がった特性であったとしても問題にはならないだろう．

また，人間の聴覚は定常音に対してはその位相特性をほとんど聞き分けられないことが知られている．本システムの出力を人間が聴いたり，人間と同じ機能を実現したシステムが聴くのであれば，位相特性は重要ではないだろう．この点において，平均2乗誤差最小規範は，やはりオーバースペックである．位相がずれれば，平均2乗誤差は増大するからである．ただし，最終出力の位相を周波数ごとに変化させても，目的音と妨害音の強度比は変化しないので，フィルタの自由度が十分高ければこのオーバースペックは実質的には問題にならないと考えられる．

6.3 LMSアルゴリズムとの適合性

Cue Signal 法と LMS アルゴリズムとの相性を検討する．

p.48の Fig. 5.7 と，つぎのページの Fig. 5.8 を比較してみればわかるように，LMS アルゴリズムでの適応結果は，直接法のそれと比較してかなり開きがある．これは，シミュレーション結果の部分（p.50）でも述べたように，内部目標 $d(t)$ が，目的音波形とは似ても似つかない波形をしていることに起因している．すなわち，瞬時ごとの修正では誤った修正が行なわれてしまうである．

我々は，LMSによって得られた修正量を，すぐに係数に加算するのでなく，平滑化（正確には1次おくれ系）をしてから加する方法を試みたりもした．しかし，平滑化の時定数を効果があるほど長くすると，今度は環境変化への適応速度が劣化してしまい，好ましい結果

は得られなかった．

本手法は LMS アルゴリズムとは，あまり相性が良くないと言える．

このため，第 II 部の後半では，実時間動作でも LMS アルゴリズムのお世話にならずにすむように，高速プロセッサの製作を行なっている．LMS アルゴリズムの利点は，計算量が著しく少ないということのみだからである．

なお，LMS アルゴリズムを Cue Signal 法に適用する場合，適応目的を内部目標にするだけなので，その収束条件は同じであることを付記しておく．なぜならば，収束条件は行列 R のみで決まるからである．

6.4 学習同定法との適合性

次に，Cue Signal 法と学習同定法との相性を検討する．本研究では，学習同定法での実験は行なっていない．それは，以下の理由による．

学習同定法は，LMS アルゴリズムに比較して修正係数が可変であるという点が異なる．そしてその変化は，タップベクトルのノルムの 2 乗に反比例する．すなわち，LMS に比較して，総受信強度が強いときに修正を弱めることになる．これが，Cue Signal 法と組み合わせたときに，重大な問題を引き起こす．

この問題を考えるために，非常に簡単な状況を設定してみよう．目的音は，ON/OFF の 2 状態のみを等時間づつとり，それぞれの状態では定常とする．また妨害音は定常音，という設定である．

目的音が ON のとき，Cue Signal は常に 1 とする．この状態での，平均 2 乗誤差の（フィルタ係数空間での）等高線は **Fig. 6.2** (a) のようになるだろう．もちろん中心ほど誤差が小さい．また，目的音が OFF のとき，Cue Signal は常に -1 とする．この状態での，平均 2 乗誤差の等高線は **Fig. 6.2** (b) のようになって，(a) とは異なる中心を持つ．実際には **Fig. 6.2** (c) のように，(a) と (b) が交互に繰り返される．(a) と (b) を平均したものが，Cue Signal 法の平均 2 乗誤差である．この等高線を **Fig. 6.2** (d) に示す．

直接法では，(c) をもとに適応するわけである．よって，最適な点（×印）を獲得することができる．

LMS アルゴリズムでは，(a) と (b) の 2 つの傾斜で交互に揺すられる．（もちろん，(a) も各時点での多数の誤差曲面を平均しともものであるので，各時点で正確に (a) の等高線の法線方向に揺すられるわけではない），その結果，最終的には (a) と (b) の傾斜が「正反対・同傾斜」の点で振動することになる．これは，(c) の×点であり，明らかに (d) の×点に一致する．ゆえに（振動があるのは好ましくないものの）問題ない．

しかし，学習同定法では，そうはいかない．(a) のときは，受信強度が強く，(b) のときは弱いからである．これは，(a) のときは小さく修正され，(b) のときは大きく修正されることを意味する．従って，学習を繰り返しても×点よりも，(b) の中心近くにズレた点での振動に落ちついてしまうだろう．

このように，学習同定法は Cue Signal 法とは相性が悪いことは明らかである．

学習同定法は，LMS アルゴリズムと違って，収束の保証に行列 R に関する先験的知識が不用である．これは受信信号のみから適応させたいという本研究の主旨には適した性質である．しかし，以上述べた不具合があるので，今回のシミュレーションや実験などでは，学習同定法は用いなかった．

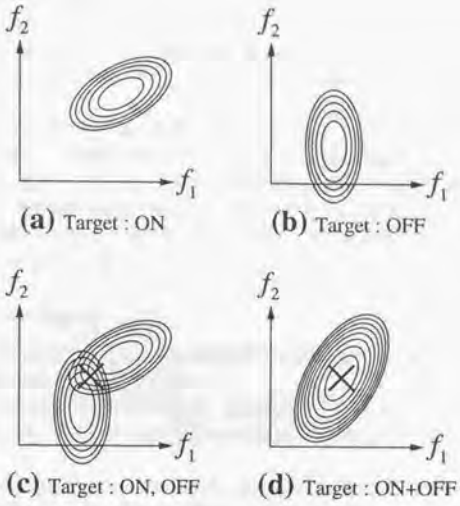


Fig. 6.2 平均2乗誤差の等高線

6.5 直接法との適合性

直接法は得られたデータの範囲内での最適なフィルタ係数が得られるのが特徴である．しかし，p.19 の **3.2.4** で述べたように，直接法は，連立一次方程式を解かなければならないので，実時間動作には向かない．

本論文では，次の2つの工夫によって Cue Signal 法の実時間動作を直接法で行なうことが可能であることを述べる．

- ブロック化
- 仮 Cue Signal

以下，この2つについて順に述べる．また，この節の最後では，直接法で連立1次方程式を解くときに一般化逆行列を用いる場合について若干のコメントをする．

6.5.1 ブロック化

Fig. 6.2の説明にもあったように，Cue Signal 法は短かい時間区間で適応してもあまり意味がない．すなわち，Cue Signal の平均が0になる程度の長さで適応する必要がある．

それならば，受信データを数秒程度の時間区間（ここではブロックとよぼう）に区切って，そのブロックごとに直接法で係数を更新すれば十分であると言えるだろう．

ブロック化したことで，新たな利点も生まれる．それは，Cue Signal の平均を0にするなどの操作を，完璧に行なうことができるということである．すなわち，ブロック長を可変にしておき，各ブロックではブロック開始からの Cue Signal を積分する．そして，最低ブロッ

ク長時間を経過後に積分値が0になったら，そのブロックを終了すれば Cue Signal の平均を厳密に0にすることができる．

ブロック化した直接法でフィルタ係数を算出するのに必要な仕事は，以下の2つである．

[**仕事1**] タップ信号間の相関行列 R と，タップ信号と内部目標の相関ベクトル $\boldsymbol{p}$ を計算する．

[**仕事2**] 連立1次方程式 $\boldsymbol{p} = R\boldsymbol{f}$ を解く．

このうち，仕事1は，音響信号のサンプリングごと（$T = 22.7\mu s$ ごと）に行なう必要があるが，仕事2は，数秒に1回でよい．また，仕事1の計算量は，R 算出に N^2 回の積和演算，$\boldsymbol{p}$ 算出に N 回の積和演算である．なお，本システムのように，多入力 FIR フィルタの場合，この演算量は大幅に落とすことができるが，それについては第II部のリアルタイムシステムの製作の部分（**10.2.2**）で述べる．

6.5.2 仮 Cue Signal

6.5.1 で述べた方法は，各ブロックの時間的長さを可変にして，「Cue Signal の平均0」という条件を満たそうとする方法であった．

しかし，ハードウエアを製作することや，実際の応用を考えると，この可変長ブロックは都合が悪い．そこで，固定長ブロックで「Cue Signal の平均0」という条件を満たす方法はないだろうか．

6.5.1 で述べた仕事1は，1サンプリングごとにすべて処理する完全な実時間計算である．そして，その計算に Cue Signal を用いている．しかるに，平均0の Cue Signal 生成は，そのブロックが完結しなければできない処理である．もし，1ブロックのデータを蓄えておき，完全な Cue Signal が得られてから仕事1を始めようとすれば，大きな記憶容量が必要であるし，仕事1の計算に要する時間がそのまま適応動作の遅れにつながってしまう．つまり，仕事1はデータが入った瞬間に処理してしまうことが望まれる．

しかし，幸運にもこの問題は容易に解決される．幸運の原因は，相関ベクトル $\boldsymbol{p}$ が Cue Signal に対して線型であるということである．すなわち，$\boldsymbol{p} = (R_{y_n d}) = \left(R_{y_n(\alpha\psi)}\right)$ なので，Cue Signal に対する線型変換は，相関ベクトル $\boldsymbol{p}$ に対する同じ線型変換で書ける．

そこで，以下のようにすればよい．まず，相関ベクトル $\boldsymbol{p}$ を計算するかわりに，2つの相関ベクトル $\boldsymbol{p}_\theta$ と $\boldsymbol{p}_1$ を実時間の積和で求めておく．ここで，

$$\boldsymbol{p}_\theta \stackrel{\text{def}}{=} \left(R_{y_n(\theta\psi)}\right) \tag{6.2}$$

$$\boldsymbol{p}_1 \stackrel{\text{def}}{=} \left(R_{y_n\psi}\right) \tag{6.3}$$

である．つまり，$\boldsymbol{p}_\theta$ は平均0にする前の Cue Signal（すなわち Event Signal）を Cue Signal として用いた相関ベクトルである．また，$\boldsymbol{p}_1$ は，定数1を Cue Signal として用いた相関ベクトルである．

そして，仕事1の終了時に，Event Signal の平均 $\overline{\theta}$ があきらかになるので，そのとき初めて正しい $\boldsymbol{p}$ を以下により求めればよい．

$$\boldsymbol{p} = \boldsymbol{p}_\theta - \overline{\theta}\,\boldsymbol{p}_1 \tag{6.4}$$

これは，すぐに終わる計算である．また，実時間の積和時に相関ベクトルの計算が2倍になってしまうが，相関行列 R の計算に比較すれば相関ベクトルの計算は軽いのでこれも問題にはならない．よって，Event Signal と 定数1を仮の Cue Signal として仕事1の積和演算を行

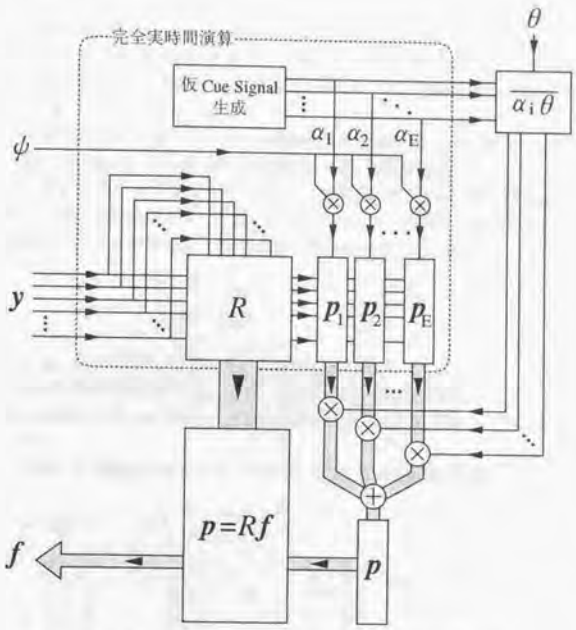


Fig. 6.3 ブロック化した直接法における 仮 Cue Signal

なっておけば，仕事2に移るときに簡単な補正をするだけで真の Cue Signal に対する相関ベクトル $\boldsymbol{p}$ が得られることがわかった．

さて，この技法はもっと応用できそうである．今述べたのは，Cue Signal に対する定数の加減算分が積和時に未知の場合の対処法であった．しかし，積和時に Cue Signal が全く未知の場合もありえよう．例えば，第II部で述べるように視覚センサから Cue Signal を生成する場合は，凝った処理をしようとすれば，1秒以上の時間遅れが起こるかもしれない．また，システム出力を見ながら Cue Signal を試行錯誤的に調節してみたい場合もあろう．

これらの場合，仮の Cue Signal を，(フーリエ級数やウォルシュ関数系などの) 完備直交な関数系 $\alpha_1(t), \alpha_2(t), \ldots$ で構成しておけば任意の Cue Signal について，積和演算をやり直すことなくフィルタ係数を求めることができる．ここで Cue Signal の帯域が非常に狭い (エンベロープ類だからゆっくり変動する信号) ということが活きてくる．少ない数の関数系で展開できれば，少数の仮 Cue Signal で済むからである．

このように，仮の Cue Signal を用いて，ブロック化した直接法を実現する方法を **Fig. 6.3** に示す．この図のうち，点線内のみが，音響信号帯域用のサンプリング周波数 T での動作を必要とする部分である．Event Signal を完備直交関数系に分離する演算 $\langle\ \overline{\alpha_i(t)\theta(t)}\ \rangle$ も，その処理時間は僅々たるものであるので，ブロック終了直後に行なえばよい．

6.5.3 一般化逆行列の選択

直接法では比較的大きな連立1次方程式を解かなければならない．特に，Fig.6.3のような方法を考えたときは，相関行列 R の逆行列を求めたておいた方がよい．正しい $\boldsymbol{p}$ が得られたとき，$\boldsymbol{f}=R^{-1}\boldsymbol{p}$ で素早く係数 $\boldsymbol{f}$ を求めることができるし，複数の Cue Signal による 複数の係数 を得ることも早いからである．

しかし，相関行列 R が正則であるという保証はなにもない．そこで，実際には逆行列 R^{-1} ではなく一般化逆行列 R^{-} を求められるようにしておく必要がある．

ところで，非正則行列 R が与えられたとき，一般化逆行列 R^{-} は一意には定まらない．そこで，どのような一般化逆行列 R^{-} を選んだらよいかについて簡単に述べておく．

まず，準備としてタップ行列 Y と内部目標ベクトル $\boldsymbol{d}$ を定義する．

$$Y \stackrel{\text{def}}{=} \left[y_{ni}\right] \qquad (N\times I) \tag{6.5}$$

$$\boldsymbol{d} \stackrel{\text{def}}{=} \left[d_i\right] \qquad (I\times 1) \tag{6.6}$$

すなわち，Y は，縦方向にタップ番号を，横方向に時間軸をとって，すべての受信した生データを並べた並べた巨大な行列である．また，$\boldsymbol{d}$ は，縦方向に時間軸をとって，内部目標をならべた非常に縦長のベクトルである．連立1次方程式を解く部分では，この両者が入力情報のすべてである．

いかなる行列でも特異値分解できる．そこで，まず，Y を特異値分解する．

$$Y = U\Delta V^T = U\begin{bmatrix}\Delta_r & O\\ O & O\end{bmatrix}V^T \tag{6.7}$$

$$= \begin{bmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots\cdots & \boldsymbol{u}_N \\ | & | & & | \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \sigma_2 & & O \\ & & \ddots & \\ & & & \sigma_r \\ \hline & O & & O \end{array}\right] \begin{bmatrix} \text{——} & \boldsymbol{v}_1^T & \text{——} \\ \text{——} & \boldsymbol{v}_2^T & \text{——} \\ & \vdots & \\ \text{——} & \boldsymbol{v}_N^T & \text{——} \end{bmatrix} \tag{6.8}$$

ここで，U と V は，それぞれ $(N\times N)$,$(N\times I)$ の行列で，構成する列ベクトル（$\boldsymbol{v}_\bullet$ どうし，または $\boldsymbol{v}_\bullet$ どうし）は，お互いにすべて直交している．特に，U は正方行列だから直交行列である．また，Δ_r は対角行列で，その対角成分を $\sigma_1,\sigma_2,\ldots,\sigma_r$ とする．ここで，r は Y のランクである．具体的に，センサ数 $M=6$，フィルタ次数 $K=32$，サンプリング周期 $T=22.7\mu$s，1ブロックの観測時間 $=5$ s とすると，$N=192$, $I=220500$ となるので，Y や V^T は実際には式(6.8) で見るよりずっと横長の行列である．

式(6.8) の意味するところを考えよう．

Y の任意の行ベクトル（任意のタップでの時系列）は，U^T を左から掛けることで座標軸を回転して観察すればすれば，式(6.8) から

$$U^TY = \begin{bmatrix} \text{——} & \sigma_1\boldsymbol{v}_1^T & \text{——} \\ \text{——} & \sigma_2\boldsymbol{v}_2^T & \text{——} \\ & \vdots & \\ \text{——} & \sigma_r\boldsymbol{v}_r^T & \text{——} \\ \text{——} & \boldsymbol{0} & \text{——} \\ & \vdots & \end{bmatrix} \tag{6.9}$$

と書けることからわかるように，r 個の時系列 $\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_r$ を各々 $\sigma_1,\sigma_2,\ldots,\sigma_r$ 倍したものと，$N-r$ 個の零ベクトルから成っている．

すなわち，Y を U^T で座標変換（回転）すれば，r 個の正規直交した信号 $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$ を定数倍したものになるか，あるいは 0 になる．そして，その定数倍の大きさが，$\sigma_1, \sigma_2, \ldots, \sigma_r$ である．また，$N-r$ 個の $\boldsymbol{v}_{r+1}, \boldsymbol{v}_r+2, \ldots, \boldsymbol{v}_N$ は，すべてのタップ時系列と直交する時系列である．

そこで，ここでは，N 個の時系列（長さ I） $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$ を**基底信号**とよぶことにしよう．Y の任意の行ベクトル（任意のタップでの時系列）は，r 個の基底信号の線型和で表現できる．

以上の準備の後，相関行列 R を式(6.8) を用いて表現すると次のように書ける．

$$
\begin{aligned}
R &= YY^T && (6.10)\\
&= U\Delta V^T V \Delta U^T && (6.11)\\
&= U\Delta^2 U^T && (6.12)\\
&= U\begin{bmatrix} \Delta_r^2 & O \\ O & O \end{bmatrix} U^T && (6.13)
\end{aligned}
$$

これは行列 R を特異値分解したことになっているので，R の一般化逆行列の一般形は以下のように[41]書くことができる．

$$
R^- = U\begin{bmatrix} \Delta_r^{-2} & S_1 \\ S_2 & S_3 \end{bmatrix} U^T \qquad (6.14)
$$

ここで，小行列 S_1, S_2, S_3 は任意の行列である．この小行列の選び方次第で，様々な一般化逆行列になる．例えば，$S_3 = S_2\Delta_r S_1$ となるのが「反射型一般化逆行列」，$S_2 = O$ が「ノルム最小一般化逆行列」，$S_1 = O$ が「最小 2 乗一般化逆行列」，$S_1 = S_2 = S_3 = O$ が「ムーアペンローズ一般化逆行列」である．

ここで求めた R^- の一般形を用いて，フィルタ係数ベクトル $\boldsymbol{f}$ を U^T で座標回転した世界で書くと以下のようになる．

$$
\begin{aligned}
U^T\boldsymbol{f} &= U^T R^- \boldsymbol{p} && (6.15)\\
&= U^T U\begin{bmatrix} \Delta_r^{-2} & S_1 \\ S_2 & S_3 \end{bmatrix} U^T U\begin{bmatrix} \Delta_r & O \\ O & O \end{bmatrix} V^T\boldsymbol{d} && (6.16)\\
&= \begin{bmatrix} \Delta_r^{-1} & O \\ S_2\Delta_r & O \end{bmatrix} V^T\boldsymbol{d} && (6.17)\\
&= \left[\begin{array}{ccc|c} \sigma_1^{-1} & & & \\ & \sigma_2^{-1} & & O \\ & & \ddots & \\ & & & \sigma_r^{-1} \\ \hline \multicolumn{3}{c|}{S_2\ \Delta_r} & O \end{array}\right] \left[\begin{array}{c} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \vdots \\ \boldsymbol{v}_r^T \\ \boldsymbol{0} \\ \vdots \end{array}\right] \boldsymbol{d} && (6.18)
\end{aligned}
$$

この式は，$N\times 1$ の列ベクトクルでの等式であるが，その要素のうちの上から r 個（Δ_r^{-1} に関する部分）の意味は明らかである．内部目標 $\boldsymbol{d}$ と基底信号 $\boldsymbol{v}_\bullet^T$ の内積をとり基底信号の大きさ $\sigma_\bullet$ で割ったものを，その基底信号の重さを決めるフィルタ係数にするということである．これは，意味的にも理解しやすい．

しかし，残りの $N-r$ 個（$S_2\Delta_r$ に関する部分）は，基底信号のうち，Y の行ベクトルに直交する成分に関するフィルタ係数を決めるものである．Y には含まれない成分の重さであ

るから，同一の Y を使うかぎり ここは任意においてかまわない．すなわち，S_1 や S_3 に限らず S_2 も何でもよい．出力は同じである．

ところが，ある時間区間での学習結果を，別の時間区間でのフィルタリングに利用しようとする場合を考えると事態は違ってくる．このときは，残りの $N-r$ 個の成分が入力される可能性がある．これらに対するフィルタ係数は，全く未知の信号成分に対する重みであるから（$\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \ldots$ の作る空間は既知だが，個々の $\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \ldots$ や，それを座標変換するための $\boldsymbol{u}_{r+1}, \boldsymbol{u}_{r+2}, \ldots$ が未知），平均2乗誤差最小の規範からは，明らかに0とおくべきである．すなわち，$S_2 = O$（ノルム最小一般化逆行列）とすべきである．

言いかえれば，残された自由度の中で係数ベクトルのノルムは最小にすべきであって，それは，ブロック中で受信されたのすべての時系列とは無相関の成分（各タップにどんな相対関係で入ってくるかが未知の信号）に対するゲインを0にすることなのである．

この節での結論は以下のとおりである．直接法を一般化逆行列を用いて実行する場合は，

- 係数学習と信号抽出が同一の時間ブロック
  $\longrightarrow$ 任意の一般化逆行列を用いてかまわない．
- 係数学習と信号抽出が異なる時間ブロック
  $\longrightarrow$ ノルム最小一般化逆行列を用いるべきである．

なお，ここでは，集合平均や期待値などの操作を一切行なわなかった．すなわち，ここでの話は，確定的な受信情報 $Y, \boldsymbol{d}$ について厳密に成り立つ議論である．

6.6 音源モデルの改良

5.2.2 で述べたように，Cue Signal 法が適用できるためには目的信号 $s(t)$ が分離可能な形にモデル化されなければならない．再掲すれば，搬送波 $c(t)$ と エンベロープ $a(t)$ の積で

$$s(t) = a(t)\, c(t) \tag{5.2}$$

と書けなければならなかった．そして，搬送波 $c(t)$ は定常信号であることが必要だった．

ところで，実際の状況を考えてみると，このようなモデル化が当てはまらないことも多い．例えば，音声を例にとってみても，各音韻ごとに統計的性質はすべて違う．2次の統計量で見て既に違っている．例えば母音の識別は，スペクトル包絡の差異があるからこそ可能なのである．それでは，Cue Signal 法で非定常搬送波を持つ目的音の抽出をすることは根本的に間違っているのだろうか．

この問題を，音声の場合で考えてみよう．搬送波は確かに非定常信号であるが，それらを有限個の定常信号の和で近似的に書き表せたとしてみよう．すなわち，ひとつひとつの母音ごとに定常信号と考えるわけである．

$$s(t) = \sum s_j(t) = \sum a_j(t)\, c_j(t) \tag{6.19}$$

このように書けば，**5.2.4** で述べた複数目的信号の理論がそのまま使えることが明らかであろう．すなわち，すべての音韻の強度変化に満遍無く相関を持つ Cue Signal を用意すればこの目的音を抽出することができると考えられる．

また，逆に，同一対象物がら発せられる音でも，抽出する音と抽出しない音を区別することが可能である．ただ，Cue Signal 法自体には何等の問題もないが，マルチセンサによる集音を考えたときには，同一音源からの音を選択的に取り分けるのはスペクトルの違いによるしかないので，選択性はそれほど期待できないであろう．

6.7 マイクロホン配置とフィルタ次数

マイクロホンの本数や，その設置位置，あるいはFIRフィルタの次数，等の最適化問題を考察するのは興味深い．

ただし，これらの問題はCue Signal法に固有な問題ではなく，マルチセンサと線型フィルタによる信号抽出・信号推定の全てに共通する一般的な問題である．このため，本論文では，これらの最適値問題については，本節で簡単に触れるにとどめる．

今，興味のある変量は以下のものである

- マイクロホンの数　M
- マイクロホン間の距離　D
- FIRフィルタの次数　K
- マイクロホン配置の形

ここでは，これら変量相互の最適な関係，および変量と信号選択特性の関係について検討する．

なお，評価の方法には，学習終了後の特性で比較する方法と，学習速度まで含めて評価する方法の2とおりある．ここでは，主に前者による検討について述べ，後者による検討は，**6.7.3**の後半で少し述べる．

評価の方法は数値計算によった．具体的な方法は以下のようなものである．

まず，評価基準である．学習終了後の最適特性は，式(3.25)で示されるフィルタ係数による特性であるので，平均2乗誤差は以下のように書きあらわされる．

$$\langle e(t)^2 \rangle = \langle d(t)^2 \rangle - \boldsymbol{f}^T \boldsymbol{p} \tag{6.20}$$

そこで，この数値と，適応前の妨害音の2乗和の比をとって評価をすることにする．

次に，具体的な計算方法である．サンプリング周波数 $F = 44.1$kHz とし，目的音，妨害音ともに0-20kHzの白色のスペクトルを持つとした．また，空間上には反射壁がないものと仮定した．音源は，目的音源1個，妨害音源2個として，音源の配置の特殊性が結果に表れないように，複数の異なった配置で計算を行い，結果を平均した．

以下，**6.7.1**では，K と D の関係を，**6.7.2**では，M と D の関係を，**6.7.3**では，M と K の関係を，それぞれ示す．また，**6.7.4**では，マイクロホン配置の形の問題に軽く触れる．

6.7.1 フィルタ次数とマイクロホン間隔の関係

マイクロホン4個を正4面体の頂点に配置した時の，FIRフィルタの次数 K と推定誤差特性の関係を，種々のマイクロホン間隔 D（中心からの距離）について求めたものを**Fig. 6.4**に示す．縦軸は下ほど特性が良いことを表している．音源配置の平均数は3である．

マイクロホン間隔 D を一定としてフィルタ次数 K を増加させてみれば，特性は「く」の字型になっている．すなわちあるマイクロホン間隔に対応して最低限必要なフィルタ次数 K が存在する．それが，「く」の字の曲り角にあたる．

逆に，フィルタ次数を固定して見れば，次のように言える．一般にマイクロホン間隔は広げたほうが良好であるが，次数 K に応じて決まるある間隔を超えると特性が急激に劣化する．そしてその直前に，マイクロホン間隔の最適値がある．

それでは，フィルタ次数 K と，最適なマイクロホン間隔 D の関係は，どうなっているか．Fig. 6.4をはじめとして複数の計算を行なった結果，両者は比例関係にあることがわかった．すなわち，マイクロホンアレイを2倍の大きさに広げれば，フィルタ次数も2倍必要になる．

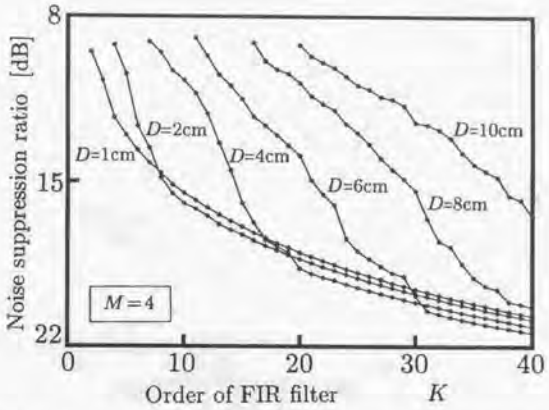


Fig. 6.4 マイクロホン数固定でフィルタ次数とマイクロホン間隔を変化させた場合

これは，以下の2とおりの解釈で直観的に説明できる．

第一は時間領域での解釈である．マイクロホンアレイを2倍の大きさに広げれば，マイクロホン間の遅延は2倍に広がる．妨害音を相殺するためには，やはりFIRフィルタの遅延幅も2倍にしなければならない．

第二は周波数領域での解釈である．p.14の式(3.15)を使って説明する．これは，完全推定の条件であるが，傾向を見るには十分である．さて，マイクロホンアレイを2倍の大きさに広げるということは，式(3.15)の空間伝達関数を示す行列 $H(\omega)$ の各要素の相対関係が周波数軸方向に2倍密度で細かく変化するようになったこととして表現できる．なぜならば，時間差 T_1 は，周波数特性では $\exp(j\omega T_1)$ の乗算に相当するからである．そうすると，一般には式(3.15)で結ばれたフィルタ特性 $F(\omega)$ も周波数軸方向で2倍密度の細かな変化で対応しなければならない．FIRフィルタで，細かな周波数特性を作ることは，次数をそれに反比例して長くしなければならない．すなわち，2倍の長さのフィルタが必要である．

ここでのまとめは，最適なマイクロホン間隔 D は，フィルタ次数 K に比例するということである．そして，それは理論的にも明白なことである．

6.7.2 マイクロホン数とマイクロホン間隔の関係

マイクロホン4，6，8個を，それぞれ正4面体，正8面体，正6面体の頂点に配置し，FIRフィルタの次数を8とした時の特性を調べた．**Fig. 6.5**は，中心からマイクロホンまでの距離を変数としてそれを表示したものである．

マイクロホン数を増加させる場合には，同時にマイクロホン群の大きさを増加させなければその効果がほとんどないことがわかる．

なお，この計算は独立な音源数がマイクロホンの個数より小さい条件での結果である．**3.1.3**で述べたように，音源数がマイクロホン数に一致し逆転すると状況が一変する．これは，数値計算でも確認している．

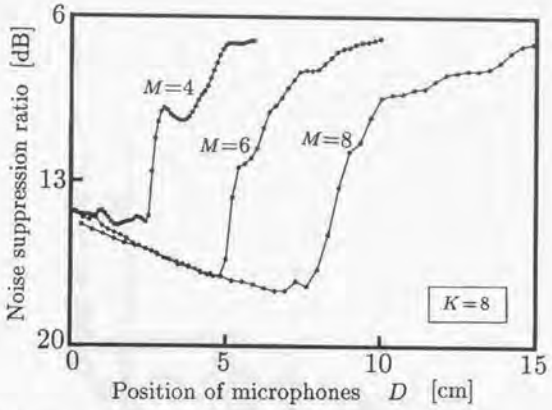


Fig. 6.5 フィルタ次数固定でマイクロホン数とマイクロホン間隔を変化させた場合

ここでのまとめは，最適なマイクロホン間隔 D は，マイクロホン数の増加に伴って増加するということである．

6.7.3 マイクロホン数とフィルタ次数の関係

マイクロホン数 M も，フィルタ次数 K も，ハードウエアや計算の手間が許すならば多いにこしたことはない．

しかし，実際には何らかの制限がる．例えば，FIR フィルタのタップ数 $N = MK$ は，FIR フィルタの計算量，LMS アルゴリズムの計算量，直接法の場合の行列 R や $\boldsymbol{p}$ の大きさ，などを規定する．すなわちシステムの仕事量を決める量である．

N に制限があると仮定して，マイクロホン数 M と，フィルタ次数 K のトレードオフを調べたのが，**Fig. 6.6** である．これは，横軸に N をとり，様々な K と M についての特性をプロットしたものである．

これらはすべて同一曲線上にのっている．すなわち，学習完了時の特性は，N のみで決まることがわかる．例えば，マイクロホン数を半分にしたならば，フィルタの次数を 2 倍にすれば同じ特性が得られる．

一般に，マイクロホン数を増やすことは，フィルタ次数を増やすことに比較してコストが高い．センサや AD 変換器などを増設しなければならないからである．したがってマイクロホン数の不足分（もちろん独立な音源数がマイクロホン数より少ないと場合）をフィルタ次数の増量で補えれば有難いことである．

しかし，音源数を超えてさらにマイクロホンを増やす意味がないわけではない．それを **Fig. 6.7** に示す．

これは，N がほぼ一定となる M と K の 5 通りの組み合わせについて，最適目標 $\psi_S(t)$ による LMS アルゴリズムでの学習曲線を示したものである．上のグラフの修正係数 λ を 10 倍した場合の学習曲線が，下のグラフである．これより，マイクロホン数を増やすことは，学

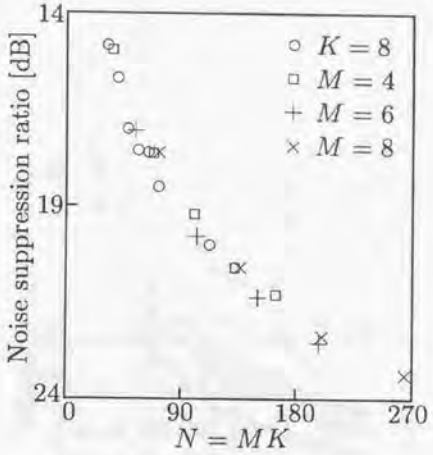


Fig. 6.6 タップ数と推定誤差特性の関係

習速度の向上をもたらすことがわかる．すなわち，最終的な特性は $N = MK$ となるどのような組み合わせ M,K でも一定であるが，LMS アルゴリズムの学習速度は，M が大きいほうが良好なようである．

6.7.4 マイクロホン配置の形

複数のマイクロホンをどのような形に配置するのがよいかという問題は，変化の自由度が大きく厳密に解くのは困難である．

そこで，解明の手がかりをつかむために以下の計算を行なった．**Fig. 6.8** 左に示すように，3 個のマイクロホン M_1,M_2M_3 を固定し，のこり 1 個のマイクロホン M_4 を平面 P 上で移動させ，誤差特性を計算した．フィルタの次数 $n = 16$，音源配置の平均数 12 である．

この誤差特性を P 上で表示した結果を **Fig. 6.8** 右に示す．色の黒い所ほど特性が悪いことを示している．M_4 の P 上での最適位置は白く表わされている部分である．

この結果から，マイクロホンは，同一平面上にのらないように 3 次元的な広がりをもって配置したほうが特性が良好であることが推察される．

しかし，マイクロホン配置の問題，特にその配置形の問題は，十分な検討がなされたとはとても言い難い．これは，今後の課題である．

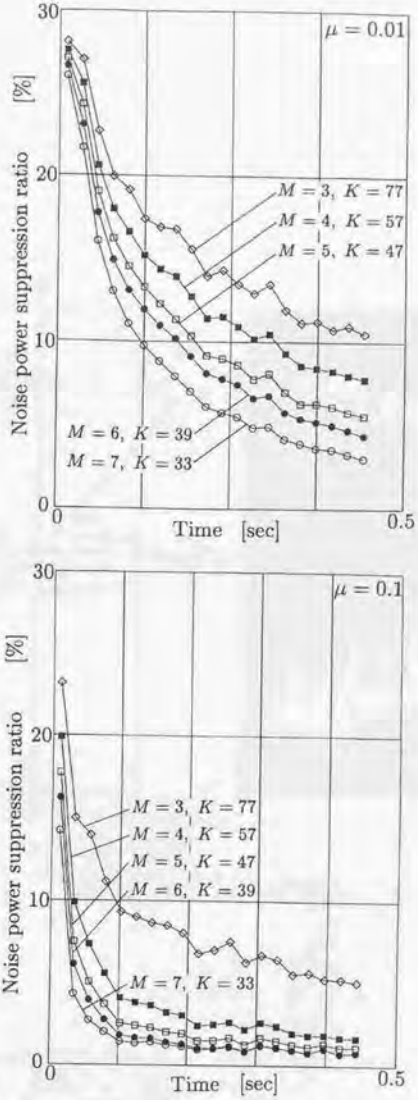


Fig. 6.7 LMS アルゴリズムによる学習曲線で見た M と K のトレードオフ

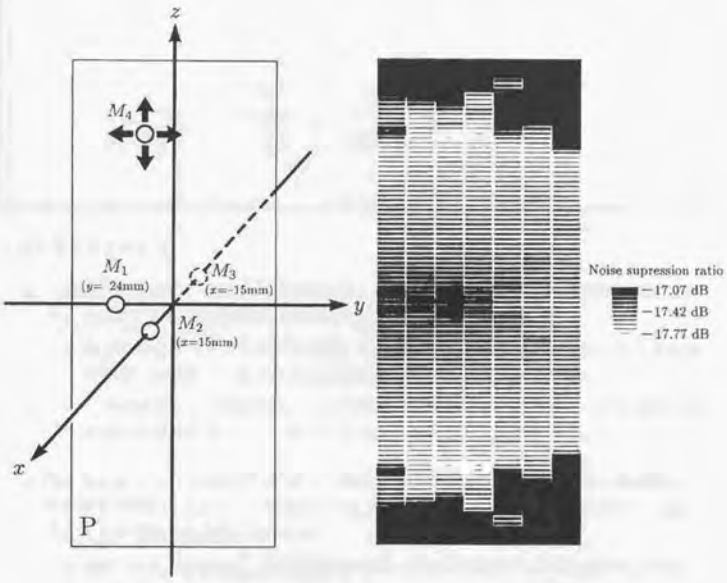


Fig. 6.8 マイクロホン配置の評価

第 7 章

まとめ

第 I 部をまとめよう.

- 「未知の受信信号」と,「未知の伝達関数」という状況で, 目的音と妨害音の混在したマイクロホン信号から自律的に信号抽出を行なう手法について論じた.
  - ◇ 従来の手法は,「トレーニング信号」を与えるか, または「伝達関数に対する先験的知識」の少なくともどちらかを必要とするものがほとんどだった.
  - ◇ 「受信信号」と「伝達関数」という単純な物理量でなく, あいまいで多様性のある情報を柔軟に受けとめて動作する適応フィルタ法の構築をめざした.

- Cue Signal とセンサ受信信号の積を内部目標にして線型フィルタを平均 2 乗誤差最小の規範で学習することで, 一定の条件のもとでは, 最適なフィルタ係数へ向かって適応することが可能であることを示した.
  - ◇ FIR フィルタを用いて, 定常妨害音から単一目的音を抽出する場合の条件を 定理 5.1 に示した.
  - ◇ FIR フィルタを用いて, 多数の目的音を抽出する場合の条件を 定理 5.2 に示した.
  - ◇ 一般の線型フィルタを用いて, 定常妨害音から単一目的音を抽出する場合の条件を 定理 5.3 に示した.

  これらをまとめて, ここで本質的な条件を整理してみれば,
  - ○ 目的音がエンベロープと搬送波に分離可能であること.
  - ○ 目的音と妨害音が独立であること.
  - ○ Cue Signal に対する条件は, $K[\alpha, \bullet]$ が目的音に対してのみ非 0 であるということ.

  となる. なお, 定理 5.3 では, 複雑な条件が多数でてきたが, それらは「エンベロープ類と搬送波類の厳密な区別」,「定常性の厳密な定義」,「非現実的フィルタの排除」をしたにすぎない.

- Cue Signal が妨害音の強度エンベロープと相関を持ってしまった場合（$K[\alpha, \alpha_j] \neq 0$）でも，妨害音の強度エンベロープを推定し，それと Cue Signal を直交化することで，妨害音を抑制することができる．

- Cue Signal 法は，目的音の強度エンベロープに相関のある時間関数（時系列）さえ得られれば適応可能である．つまり Cue Signal には，多種多様で広範囲の信号が利用可能である．この柔軟性のため，目的音と妨害音の区別のさまざまな規範の採用が可能となる．最終的にはセンシングシステムの自律性を高めることができる．

- Cue Signal 法に対する各種適応アルゴリズムの適合性は以下のとおりである．ここでは実時間動作をさせることを考えて比較した．

［LMS アルゴリズム］	理論的考察より多少問題があることが明らかである．ただ，シミュレーションと実験の結果よりある程度十分な適応が可能であることがわかったので，計算量の制約からやむを得ない場合には使ってもよいだろう．
［学習同定法］	理論的に重大な問題がある．使うべきではない．
［RLS アルゴリズム］	実時間動作を考えると若干負荷が重い．
［直接法］	Cue Signal 法に最も適している．実時間動作に対してもブロック化と仮 Cue Signal という 2 つの工夫をすることで対応できる．

- Cue Signal 法は受信信号のみから適応するタイプである．このため，システムノイズやセンサデバイスのバラツキなどの問題に自動的に対応する．その半面，残響の排除などはできない．

- 平均 2 乗誤差最小の規範は，本システム出力をどう使うかに依存してきまり，必ずしも必然的なものではない．

- 目的音は必ずしも分離可能である必要はなく，分離可能信号の和として記述できればよい．ただし，Cue Signal はそのすべての要素に対して適応の条件を満たしている必要がある．

- フィルタ次数は，マイクロホンの配置とサンプリング周期より決定すればよい．また，マイクロホン配置問題については，最適性を論じるまでには至らなかった．これは，今後の課題である．

- ここで述べた理論は，音響センサのみでなく，マルチセンサ情報または単一センサ情報から線型フィルタで信号抽出するすべての場合に応用可能である．

第 II 部

視聴覚情報のセンサフュージョン

第 8 章

センサフュージョンと視聴覚融合

第 II 部では Cue Signal 法を視聴覚のセンサフュージョンに応用する方法について論じる．この章は，その導入部である．**8.1** 節では，センサフュージョンとは何であるかを簡単に述べる．次に **8.2** 節では，視覚情報と聴覚情報のセンサフュージョンについて述べる．

8.1 センサフュージョン

センサフュージョン[42)–45)]は，複数のセンサ情報を融合・統合するための技術である．複数センサの情報処理を統一的に論じることで，単一のセンサの知能化を考えるだけでは到達できなかった測定量の計測や，より高い柔軟性・信頼性・適応性・自律性を持ったセンシングをめざす．

センサフュージョン技術には，1) 同一の物理量を測定する同一種類のセンサのみを複数用いる技術，2) 同一の物理量を測定するための異なる種類のセンサを用いる技術，さらには，3) 異なる属性（たとえば視覚情報と触覚情報など）を測定するセンサをも用いる技術，などが含まれる．

ここ数年，このような複数のセンサ情報の処理手法に関する研究が盛んである．その理由としては，

- センサの高精度化，高信頼性化への要求が高まって，それはもはや単一のセンサデバイスでは実現出来ないものになってきた．
- センサデバイスが，高性能化の目的で，測定レンジや対象ごとに特化した．より多くの条件下で計測を行なうために，それらを組み合わせる必要がでてきた．
- 自律型システムの開発でセンサの知能化に対する要求が高まった．すなわち単純な物理量の計測でなく，より抽象的・意味的な情報収集が必要となった．ほとんどが，複数のセンサデータを組み合わせて初めて得られるものである．
- 複数のセンサデータを実時間で演算処理できるハードウエアが手軽に利用可能になった．

などが考えられる．すなわち，複数センサ情報の統合の要求と，それを可能にする技術基盤

の2つが，研究が盛んになった根本にあると考えてよいだろう．

しかし，初期においては，複数センサデータの処理が計測工学のなかの一分野として認識されていなかった．研究者本人がセンサフュージョンであると自覚しないで研究していた場合も少なくない．むしろ，初期においては各々必要に応じてアドホックな研究が行なわれてきたと言ってもよいかもしれない．

しかし，上にあげた4つの理由から複数センサデータを処理する研究が急増し，その成果が蓄積され，自ずから一分野を成すようになってきた．

しかし，未だ若い研究分野であるので，工学としての体系化は完了していない．センサフュージョン技術の境界をどこに引くかということにも共通の認識は得られていない．また，現在は用語等にも完全な統一がとれていない状況である．そこで，現況を簡単に整理したあとで，本論文での用語の定義を行なうことにする．

R. C. Luo[46]は "multisensor integration"（統合）という用語と "multisensor fusion"（融合）という用語を使い分けている．彼の定義によれば，"multisensor integration" とは，「ある目的達成のために，複数センサからの情報を共動利用すること」であり，"multisensor fusion" とは，「integration 内の各段階での実際のデータの結合や融合．複数のセンサデータを結合・融合してひとつの表象にすること」である．言い替えれば，ひとつの integration という目的を達成するために，実際のデータの結合・融合—すなわち fusion—が行なわれる，という見方である．これは，**Fig. 8.1** のように示すことができる．この例では，ひとつの integration を達成するために，複数の fusion が組み合わされている．また，センサの選択，センサ情報の世界モデルによる表現，異なるセンサからの情報の変換，などに機能が integration の中に含めてある．

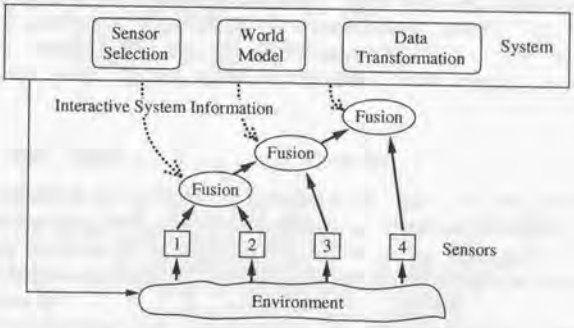


Fig. 8.1 R. C. Luo の Fusion と Integration

一方，石川は，カタカナの "センサフュージョン" という用語を「複数のセンサ情報の処理過程」と定義し，非常に広い意味で用いている．石川の分類を，**Table. 8.1** に示す．石川は，複数センサデータの処理技術は，基本的に Table. 8.1 に示す "複合"，"統合"，"融合"，"連合" の4つの形態が含まれるとし，これら全体の総称として，"センサフュージョン" という語を用いることを提案している．

本論文では「融合」「フュージョン」をともに最も広義の複数センサデータ処理技術の意味

Table 8.1 センサフュージョンの分類（石川[42][43]を僅かに改変）

分類	意味	各センサ情報(A,B)と処理の関係	処理の目的
複合 (multisensor)	複数個が合わさること	A, B → A + B **相補的，加法的処理**．相互関係は言及しないか，もともと独立．	単一機能性や局所性の回避，測定レンジの拡大など
統合 (integration)	支配が形成されること	A, B → f(A + B) **乗算的処理**，演算処理fに対する関係として規定．	精度・信頼性の向上，処理時間の短縮，故障診断など．
融合 (fusion)	緊密に一体となること	A, B → C **協調・競合的処理**．相互の関係からまとまった知覚表象を得る．	両眼融合（立体視），視触覚融合（物体・空間認識）など．
連合 (association)	関連が形成されること	A, B → (A と B の関係) **連想的処理**．相互の関係が抽出される．	予測，学習，記憶，モデル形成，異常の検出など．

で用いることにする．すなわち，本論文のタイトルにある融合とは，Luo の用語法に従えば fusion and integration であり，石川の用語法に従えばフュージョンである．

8.2 視覚情報と聴覚情報の融合

この節では，視覚情報と聴覚情報のセンサフュージョンについて考える．まず，**8.2.1**で視聴覚のセンサフュージョンの必要性を述べたあと，**8.2.2**で視覚情報と聴覚情報はセンシングという立場で考えたときに本質的にどのような差異があるのかを論じる．それをもとに，**8.2.3**では，視聴覚融合の難しさと，それを乗り切るために焦点となる4つのキーポイントについて述べる．最後に，**8.2.4**で，視聴覚融合の従来研究として音声認識技術について簡単にまとめる．

8.2.1 視覚・聴覚のセンサフュージョンの必要性

画像処理技術の発展やＣＣＤ撮像素子などの普及により，視覚センシングは，近年その高度化・知能化が急速に進められている．一方，聴覚センシングである音響計測技術もディジタル信号処理技術の発達によりその進歩は著しいといえる．しかし，両者をセンサフュージョン的意味で包括的に扱った知能化センサの開発は，まだほとんど手がつけられていないと言っても過言ではない．

センサの知能化を進めるとき，センサ研究者はしばしば生体をお手本にする．生体においては，視覚や聴覚など単独の器官での情報処理だけでなく，彼らが外界を認識し理解し行動するために，異種の感覚器官からの情報の融合・統合が行なわれている．

少し寄り道して，生体におけるセンサフュージョンの具体例をあげよう．まず，身近な例として，オーケストラによる演奏の鑑賞を考えてみる．特定の楽器に注目し，音の出る瞬間を視覚的に予測できれば，その楽器の音が聞き取り易くなる．これは，誰でも経験する身近な例である．

また，聴覚障害者は，言葉を理解するのに話者の唇の動き（読唇情報）を利用しているが，健聴者においてもこの読唇情報を音声認識の手がかりとしていることがよく知られている．

例えば，マガーク効果 (McGurk Effect)[47][48] と呼ばれる現象がある．これは，視聴覚融合の最も簡単でかつ顕著な証拠と言える．マガーク効果は，視覚系と聴覚系に矛盾した情報を与えた結果，両者が融合されてその中間的な事象が知覚される現象である．具体的に述べると，被験者に，「ga」を発音する唇を見せて，同時に「ba」という音声を聞かせる．すると，被験者はその中間の調音位置を持った「da」を知覚するのである．これを **Fig. 8.2** に示す．

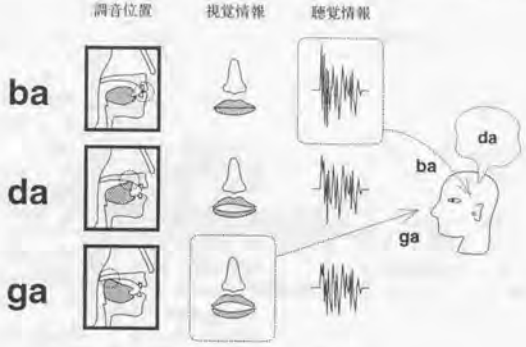


Fig. 8.2 マガーク効果

マガーク効果は，雑音などで音声の明瞭度が低い時に起こり，明瞭度が高い時には聴覚情報は視覚情報の影響を受けにくくなる．[49] つまり，我々の聴覚機構には，受信した聴覚情報の不十分な部分を，必要に応じて視覚情報によって補う能力があるのである．

視聴覚の相互干渉についての心理的研究はその他にも多数あり，文字呈示を行なうことによって音声知覚が影響されるという報告[50]や視覚情報によって音源の定位感に影響がでるという報告[51][52] などがある．

また，生理学的な研究からも視聴覚の統合は明らかにされつつある．例えば，Meredith と Stein はネコの上丘深層から，視覚刺激と聴覚刺激が同時に加えられたときのみ強く反応する細胞や，聴覚刺激が視覚刺激を抑制する細胞を見つけている．[53]

少し寄り道して，生体における視聴覚融合を見てきた．感覚器官から認識機構に至る処理系において視覚と聴覚は密接に結びついているといってよいだろう．生体をめざしてセンシングシステムの知能化を行なうならば，視聴覚融合技術は，必ず重要となってくる技術である．

それでは，工学的センシングに戻ろう．

さきほど述べたように，視覚センサおよび聴覚センサの知能化は近年大幅に成果をあげてきている．視覚あるいは聴覚の個々の器官の情報処理の部分については，かなり生体に近い機能を有する（場合によっては生体を超える）工学的センサが登場している．視聴覚のセンサフュージョンによってさらなるセンサの知能化を図る条件が今まさに揃っているといってよいだろう．これは，生体なみの柔軟性・適応性・自律性を持ったセンシングシステムに一歩近づくことでもある．視聴覚融合型の知能化センシングを研究しよう．

Table 8.2 視覚的情報と聴覚的情報の一般的傾向

		視覚的情報	聴覚的情報
情報内容	含まれる対象物情報	主に対象物表面に関する情報を含む（形状・表面状態）	対象物の内部から生ずる情報をも含む（内部・外部の振動）
	波動の位相情報	普通は強度のみ	重要
	対象ごとの分離	座標 (x,y) で切り分けられる	加算されて受信される
伝播の性質	信号の種類	動画像	音響信号
	情報の媒体	主に光	主に音
	波長	短い（1 μ m 以下）	長い（1 cm 以上）
	伝播の経路	直線的に伝播する	回折，反射などが多い
	他の物体による遮蔽	有り	小さい
	伝播の遅延	ほとんど無視できる	音速・残響の影響で無視できない
情報形態	生のセンサ出力	時間変化する2次元パターン	1～数点での時間関数
	形式的な記述	$f_i(x,y,t)$	$g_i(t)$
	パラメータ	センサ番号 i, 時間 t, 座標 (x,y)	センサ番号 i, 時間 t
	工学的に表現するときの情報形態（典型的な一例）	縦 192 ×横 256 ×両眼視 2 ×色 3 ×濃淡 8bit ×サンプリング周期 60Hz = 140Mbit/s	チャネル数 2 ×量子化レベル 16bit ×サンプリング周期 44.1kHz = 1.4Mbit/s
例	生体のセンサ	目，コウモリの聴覚	耳
	工学的センサ	ビデオカメラ，レンジファインダ イメージセンサ，サーモグラフィ マルチスペクトルスキャナー 超音波ホログラフィ	マイクロホン 振動センサ

8.2.2 視覚情報と聴覚情報の本質的な差異

まず，視覚情報と聴覚情報の定義をしておく．本論文では，視覚情報と聴覚情報を，Table. 8.2のように整理して考える．

ここでは，単に情報の伝達が光によって行なわれるのか音によって行なわれるのかで両者を区別するのでなく，その本質的な特性によって視覚的情報と聴覚的情報に分類してある．このように整理すると，視覚と聴覚がいかに対称的な性質を持っているかがよく理解できるであろう．

本節で述べる視覚情報とは，網膜上の2次元輝度分布に代表される情報である．工学的に表現すれば，座標 (x,y) および時間 t から輝度への関数 $f_i(x,y,t)$ $(i=1,2,\ldots,M)$（モノクロ単眼であれば $M=1$，カラーで両眼であれば $M=3\times 2$）で表わされるいわゆる動画像である．これは，対象となる3次元空間を2次元の空間に（近いものが遠いものを覆い隠すという様式で）射影した情報であって，$f_i(x,y,t)$ の定義域上の一点に対応する点が対象となる3次元（時）空間 (x,y,z,t) 上に一点存在することを示している．つまり，透明体を透して見るような特殊な場合を除いて $f_i(x,y,t)$ の定義域の一点は，対象空間のあるひとつの対象物のある一点を指している．これより，複数の対象物を (x,y,t) によって切り分けることが可能となる．

一方，聴覚情報は $g_i(t)$ $(i=1,2,\ldots,N)$（左右2つの耳で聞くのであれば $N=2$）で表わされる N個の時間関数である．$g_i(t)$ は対象となる空間にある複数の対象物の発する音を

すべて混合したものとして受信される．よって，$g_i(t)$ から対象物ごとの成分を分離するのは容易ではない．

このように考えていくと，超音波によるイメージングはここではむしろ視覚的情報として分類した方が良いことがわかる．つまり，コウモリの聴覚（処理部も含める）や，ロボットの超音波映像は音響信号を媒体として用いてはいるが，むしろ視覚的情報と分類したほうが見通しがよくなる．ちなみに，メンフクロウの中脳の下丘外側核の細胞は，（あたかも網膜に2次元画像がマッピングされるがごとく）空間上の特定位置からの音のみに反応することが調べられている．[54] これも，視覚的と言ってよいだろう．

8.2.3 視聴覚融合の4つのキーポイント

それでは，Table. 8.2をもとに，視聴覚融合を工学的に実現するとき，何が問題になってくるかを考えてみよう．視聴覚融合が実現しにくいのは，以下の2つに起因する．

第一は，両者が本質的に異なる情報形態をとっていることにある．（Table. 8.2の「情報形態」の欄参照）．これは，単にセンサデバイスが異なる形態（帯域，次元数）で出力するという問題ではなく，もともとは視聴覚情報の本質的な違いに起因する問題である．

第二は，両者がほとんどオーバーラップしない異なる情報を伝達してくるというところにある．（Table. 8.2の「情報内容」の欄参照）．これは，対象物から発信される情報がもともと異なるということだけでなく，対象からセンサまでの経路で異なる性質の媒体によって伝達されてくる（Table. 8.2の「伝達の性質」の欄参照）ことにも原因がある．いずれにしても，フュージョンが達成されるためには視覚情報・聴覚情報の両者に共通する情報が必要で，それが少ないのが問題なのである．

この2つの問題は，センサフュージョンを行なうにあたっては，具体的にそれぞれ，

【i】　視覚情報・聴覚情報を「どの座標上で」結合するのか

【ii】　視覚情報・聴覚情報を「何によって」結合するのか

という課題となってあらわれる．視聴覚融合のキーポイントとして，まずこの2つが指摘できる．

さて，共通の情報が少なく，異なる伝達経路，異なる信号フォーマットという視聴覚融合センシングの状況を絵で表すと **Fig. 8.3** のようになる．

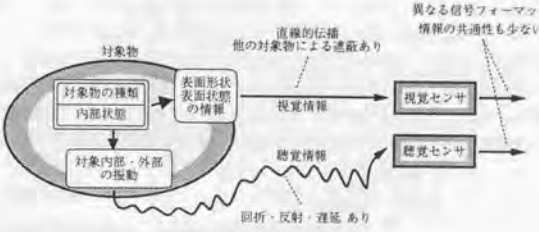


Fig. 8.3 視覚情報と聴覚情報を結合するもの

これを見ればわかるように，視覚情報・聴覚情報の両者の結合をたどるには，対象物そのものまで遡らなければならない．そして，視覚情報と聴覚情報の関係はそれぞれの対象物が

内的に持っているものであって，簡単で包括的な法則では記述できない．つまり，内部状態と視覚情報，あるいは内部状態と聴覚情報の関係は単純ではない．例えば，顔を撮した動画像と音声の関係，または，バイオリンの画像と音の関係などはそれぞれ別個のものとして存在する．これに完全に対処するためには，対象物を見分け，対象物毎の視聴覚関係を記述したデータベースを検索して両者を関係づけなければならなであろう．

そこまで徹底しないにしても，視聴覚融合を行なうには，この関係をセンシングシステムの中に（陽な形であれ陰な形であれ）持っている必要がある．融合の根拠が必要なのである．これが第3のポイントである．

【iii】　視覚情報・聴覚情報を「何を根拠に」結合するのか

ちなみに，視聴覚融合と対比するために両眼融合を考えてみよう．両眼融合の場合には，必然的に 1) 位置座標上で，2) 対象表面の反射率分布によって2つの視覚センサ情報を融合することができる．これを，**Fig. 8.4** に示す．つまり，対象物表面の形状や反射率まで遡れば十分なのである．対象物が何であるか見分けたり，その内部状態等を推定する必要はほとんどない．これと比較すれば，ここまで述べた3つのポイントの重要性と視聴覚融合の難しさを理解してもらえると思う．

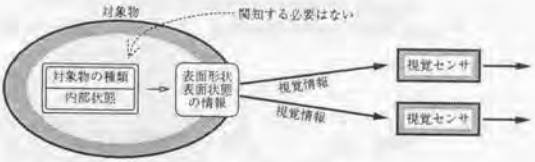


Fig. 8.4　両眼融合の場合に情報を結合するもの

さらに，実際に異種情報をどのような具体的手法でセンサの知能化に生かすかが問題である．すなわち異種情報による学習にむいた適応アルゴリズム等を考えなければならない．これが第4のポイントになってくる．

【iv】　視覚情報・聴覚情報を「どうやって」結合するのか

以上あげた，4つのポイントをまとめれば，「どこで（where），何で（what），その根拠は（why），どうやって（how）融合するかという問題」と言えるのである．

最後に，視聴覚融合の手法を大別してみる．視聴覚融合には，**Fig. 8.5** に示すように，視覚情報と聴覚情報の両者を対等に扱って，より高い意味レベルでの融合を行なうものと，片方を従属的に扱って信号レベルでの融合を行なうものの2つの手法に大別して考えることができる．本論文の第 II 部で述べる Cue Signal 法による視聴覚融合 Fig. 8.5 (a) に相当する．

8.2.4　音声認識

視聴覚融合の節の最後に，具体的な研究例を示そう．

先に述べたように視聴覚融合の知能化センシングは現在はまだほとんど研究されていない．しかし，唯一の例外として音声認識がある．これは，聴覚情報による音響信号分析と視覚情報による工学的読唇を融合する認識手法である．

たとえば，動画像から，唇の形に関する 2,3 のパラメータを抽出し音声の解析結果に併用するという方法がある．[55]

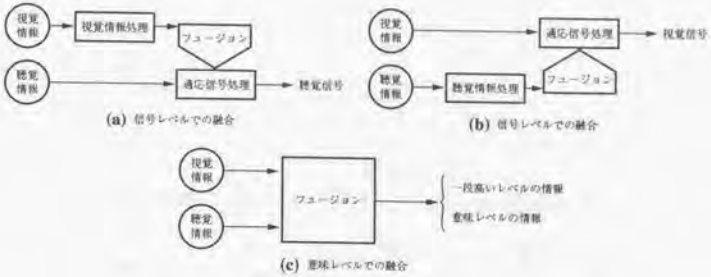


Fig. 8.5 信号レベルでの視聴覚融合と意味レベルでの視聴覚融合

視覚情報と聴覚情報というように，視聴覚両者の関係を事前に明確に記述することが困難な場合は，ニューラルネットの手法が有効な場合がある．**Fig. 8.6** に示すものは，ニューラルネットを使って視聴覚融合を行い，母音の識別を試みたものである．[56][57]

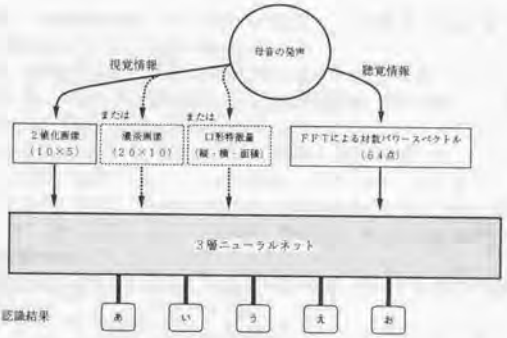


Fig. 8.6 ニューラルネットを用いた視聴覚融合による音声認識

この例では，聴覚情報として，FFT による対数パワースペクトル（データ数 64）を用いている．また，視覚情報として，話者の口周辺の濃淡画像（データ数 20×10）または口周辺の 2 値化画像（データ数 10×5）または 2 値化画像の幾何学的特徴量（データ数 3）を用いる．両者を 3 層のニューラルネット（中間層ユニット数 8～12，出力層ユニット数=認識母音数=5）で認識した結果，音声のみの場合の不特定話者認識率が，80%であったものが，音声と 2 値化画像で認識した場合は 92%まで上昇したことが報告されている．

この例においては，視覚情報と聴覚情報は 5 つの母音という記号において融合されている．また，融合の問題構造の難しさを，ニューラルネットという手法を採用することでうまく対処したものと見ることができよう．

Fig. 8.7 多系列の隠れマルコフ過程を用いた視聴覚融合による音声認識

視聴覚融合に，多系列の隠れマルコフ過程を利用して，音声認識（単語の識別）を行なう手法[58)59)]も提案されている．これを **Fig. 8.7** に示す．

聴覚情報は，音声信号を LPC ケプストラムでクラスタリングすることでシンボル化した時系列 $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)$ に変換される．一方，視覚情報は，唇の内周の縦と横の長さをクラスタリングすることでシンボル化した時系列 $\boldsymbol{x} = (z_1, z_2, \ldots, z_T)$ に変換される．

こうして得られた y_i と z_i を，隠れマルコフモデルにおける内部状態（どんな音素を発生しようかという意識と解釈できる）によって融合し，音声のみによる認識よりも高い認識率をめざすものである．音声に白色雑音を混入させた実験結果によれば，融合の効果は雑音の強度が大きい時ほど顕著であって，これは **8.2.1** で述べた生体におけるマガーク効果にも一致している点が興味深い．

なお，視聴覚融合による音声認識は，**8.2.3** で述べた 4 つのキーポイントで見ると，以下のように整理できる

「どの座標で」結合するのか　→　時間軸上で結びつける
「何によって」結合するのか　→　シンボル（文字）で結びつける
「何を根拠に」結合するのか　→　音声と口形の関係を根拠に
「どうやって」結合するのか　→　ニューラルネット，隠れマルコフ過程で結合する

なお，視聴覚融合による音声認識は，Fig. 8.5 の分類で言えば，(c) に相当する．

第 9 章

Cue Signal 法による視聴覚融合

この章では，Cue Signal 法を視聴覚のセンサフュージョンに応用する．まず，**9.1** 節で，その概要を述べる．次に，**9.2** 節では，Cue Signal を視覚情報（ビデオカメラで撮った動画像）から生成するためには具体的にどのような処理をしなければならないかを論じる．**9.3** 節には，基礎的な実験結果を記載する．

9.1 手法の概要

第I部で述べたように，Cue Signal 法では，生のセンサ信号以外で適応に必要な信号は Cue Signal のみであり，それに課せられた条件は平均値で規定されるという柔軟なものであった．

それゆえ，Cue Signal 法を応用して視聴覚融合型の知能化センシングシステムを構成することは容易に実現できる．ビデオカメラなどで得た視覚情報から，音源の強度変化（強度エンベロープ）を大雑把に推定し Cue Signal を生成すればよいからである．

Cue Signal 法による視聴覚融合型センシングシステムの具体的なブロック図を **Fig. 9.1** に示す．また，**Fig. 9.2** には，Cue Signal 法による知能化センシングの概念を，生体の場合と比較して示す．この 2 つの図を用いて，視聴覚融合の 4 つのポイントが，Cue Signal 法の場合にはどのように解決されるのかを述べよう．

生体（人間）の場合は，視覚情報と聴覚情報の関係（例えば，口の形と音声の関係，あるいは，楽器の操作様式と楽音の関係）を後天的に獲得しており，それをもとに目から入った情報と，耳から入った情報を対応づけている．時間軸上での融合である．さらに，音源定位の聴覚機能を用いて，空間座標上での融合も行なわれていると考えられる．

これに対して，Cue Signal 法では，時間軸上で両者を融合する．これは，最も単純で最も普遍的な選択である．普遍的とは，いかに異なる種類の情報間でも，時間軸上での対応づけは通常必ず成立するという意味である．

現実世界では，時間軸はすべての事象に共通の座標軸である．ただし，光速や音速が有限であることに起因する情報伝播の遅延は考慮したほうがよい．これらの遅延が無視できる程

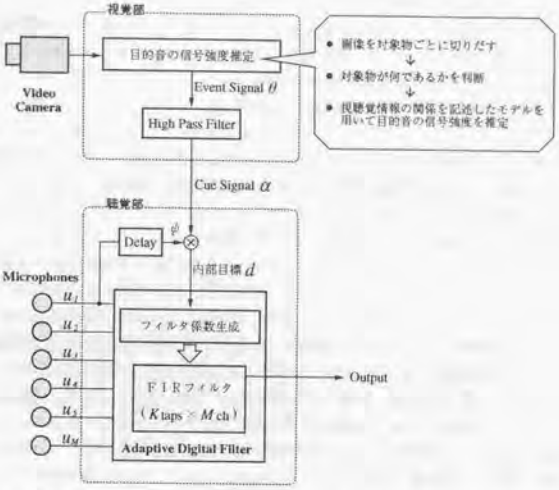


Fig. 9.1 Cue Signal 法による視聴覚融合型の知能化センシングシステムのブロック図

度の時間分解能で見たときには（Cue Signal を帯域制限したことがまさにそれにあたる），時間軸は単なる共通の座標軸であるだけでなく，ひとつの事象を単一の時間点で対応させることができるからである．

つぎに，Cue Signal 法による視聴覚融合は，「何によって」視覚情報と聴覚情報を結合することになるのかを明確にしておこう．両者を結合する「変数」として，どの様な量を選んだことになるのかである．

その「変数」は，言うなれば「事象の生起の度合」を表す量である．ここでいう「事象」とは，言葉を発する，楽器を鳴らす，ものが衝突した，など，視聴覚両方の情報変化をもたらす根源である．また，「事象の生起の度合」は，聴覚システム側から見れば，目的音強度であり，視覚システムで言えば，音を発している対象物の動画像が持っている情報である．そして，「事象の生起の度合」の推定値がまさに Event Signal に相当するのである．例えばある人

Fig. 9.2 視聴覚融合型の知能化センシング

間が音声を発したという事象の度合は，聴覚的には音声の音強度で表わされ，それは口の動きという視覚情報から推定可能であるので，両者を結合することが可能となる．

さて，視覚情報からの音強度推定には種々の方法が考えられる．しかし，**8.2.3** の p.82で述べたように，（陽な形であれ陰な形であれ）視覚から音強度を推定するための対象物モデルを導入する必要がある．このモデルの選択問題が，視聴覚融合の３番目のポイントである「なにを根拠に」結合するのかという問題に他ならない．

最も簡単なモデルは，視覚変化（形状変化・動き）の大きさが音強度に相関を持つという暗黙のモデルである．具体的な画像処理として，時間微分画像の強度を空間積分することで Cue Signal を作ることができる．また，ある音源が音を発するとき特定の方向への移動・振動を伴うものであれば，その方向に垂直な縞をもつ空間フィルタをかけて空間積分することで，この特定の音源の強度を推定できる．さらに，音声の場合は，口の縦方向の大きさとその時間微分値の加重和で音声強度の推定が可能である．

最後の例の基礎実験を行なっているので，参考までその結果[32]を紹介しよう．**Fig. 9.3** は，３人の被験者が通常の速度で朗読した場合の，口の縦方向の大きさの時間微分値 $v(t)$ と音声強度 $p(t)$ の相互相関を示したものである．唇速度の大きさを 0.2 秒遅延させてやれば音声強度が推定できることがわかる．なお，この実験では，単なる微分値を用いる代わりに，ビデオレートにして前後５画面分の口の大きさを使った線型フィルタ（FIR フィルタ）出力を用いることによって，音声強度の推定がさらに改善されることがわかっている．なお，５次の FIR フィルタを用いたということは，口の大きさ，その速度，加速度，３次微分量，４次微分量 を最適に加味したということである．

まとめると，視覚情報・聴覚情報の２つの異種情報を，事象の発生（音が発せられた）によって，時間軸上の１点で対応させるのが Cue Signal 法による視聴覚融合である．これは，**8.2.3** で述べた「視聴覚融合の４つのキーポイント」で整理して書けば以下のようになる．

「どの座標で」結合するのか　→　時間軸上で結びつける
「何によって」結合するのか　→　事象生起の度合で結びつける
「何を根拠に」結合するのか　→　視聴覚情報の関係を記述したモデルで結びつける
「どうやって」結合するのか　→　Cue Signal 法で線型フィルタを学習させる

9.2　視覚情報処理の具体的内容

Fig. 9.1 のブロック図のなかに書き込んだように，視覚部の役割は，Cue Signal の交流化以外には次の３つがある．

画像を対象物ごとに切り出す． ビデオカメラで得た動画像には多数の対象物の画像が含まれている．しかし，Cue Signal 法を使うためには，目的音のみの音強度変化をとらえる必要がある．このためには，まず得られた動画像のなかから目的音源のみを含む動画像を切り出さなければならない．さらに，本システムの大前提は自律的なセンシングであるので，この切り出しも自動的に行なわれるものでなければならない．また，**5.2.6**（p.41）で述べたように，妨害音源をも見てその音強度を推定し Cue Signal の直交化をはかる場合には，妨害音源画像も自律的に切り出す必要がある．

対象物が何であるかを判断する． この次の処理で，視聴覚の関係を記述したモデルを用いる．対象物に応じたモデルを取り出して利用するためには，対象物が何であるかを判断しなければならない．

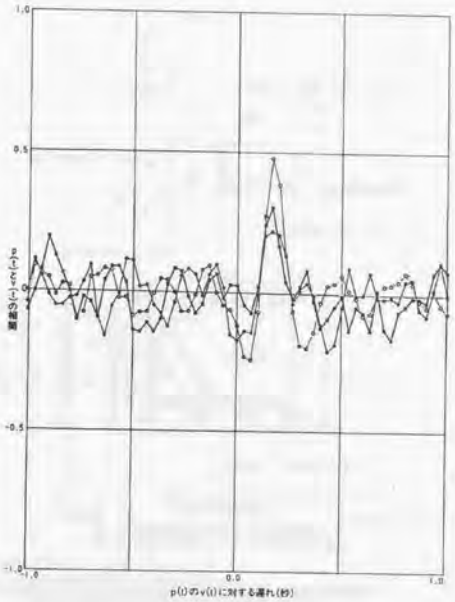


Fig. 9.3 口の縦方向の速度と音声強度の相互相関

モデルを用いて音強度を推定する. 切り出された動画像に，いろいろな種類の対象物それぞれに対してその視聴覚情報の関係を記述したモデルをあてはめて，目的音の音強度を推定する．実質的に視覚情報と聴覚情報の橋渡しをする部分である．

これらを，全部まともに行なうとなると，どれも大変な仕事である．しかし，いつもすべてまともに行なわなければならないというわけではない．例えば，実際には **9.1** 節で述べたような暗黙のモデルを使うなどの方法がある．(形状変化や動きなどの視覚変化の大きさが，音強度に相関を持つなどどいうのが，暗黙のモデルである)．また，対象物の判定にしても，すべての対象物を単一のモデルで記述すれば不要となる．

このように，上にあげた3つの役割は，むしろ最大限必要な仕事といえるものである．実際の状況に応じて省略すればよい．

9.3 実験

9.3.1 簡単な実験

これは，第I部の **5.4** 節（p.51～）で述べた実験を少し変更して，最も簡単な視聴覚融合をデモンストレーション的に実験したものである．

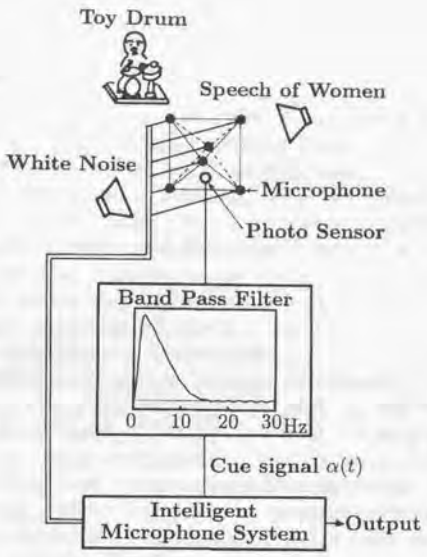


Fig. 9.4 視覚刺激を伴う音源の音を抽出する適応型音響センシングシステム

9.3.1.1 実験方法

実験方法も **5.4** 節のときと同じで，各音源（目的音，妨害音）の信号を別個にひとつづつ鳴らして採取し，各マイクロホンの出力信号は，ワークステーション内でそれら別個の信号の和をとることによって作成する非実時間の実験である．

実験は，無響室にさざまな音源を配置し，聴覚センサとして 6 本のマイクロホン，視覚的センサとしてフォトダイオードを設置して行なった．マイクロホンは一辺 155mm の正 8 面体の頂点に配置した．フォトダイオードはそれらマイクロホンの中央付近に配置し，受光方向を外界のある 1 方向に固定した．

Fig. 9.4 に実験の理論的なブロック図を示す．用いた音源は 3 つある．スピーカからの白色雑音，スピーカからの音声，玩具のドラムである．このドラムには豆電球がついており電源が入っているときに点灯するようになっている．そこで，この電源をリモコンで入・切することで，視聴覚情報を最も単純に結合した対象物とすることができる．

すなわち，この実験は，定常妨害音（白色雑音）と非定常目的音（音声）のなかから，視覚的刺激を伴う音源の音（ドラムの音）を抽出する実験である．

Fig. 9.4 で，Cue Signal 生成について説明する．ここで与えられた音源識別の規範は，「目的音は音の発生と同時に光を出す」という簡単なものである．Cue Signal は，フォトダイオード出力をノイズカットのためのバンドパスフィルタで濾過することによって生成できる．ここで言うノイズとは，室内照明などによる商用周波数の成分が代表的なものである．フィル

タの周波数特性は，Fig. 9.4の中央のワクの中に示してある．また，Cue Signal 生成以外の部分は **5.4** 節のときと同じであるのでここでは省略する．

9.3.1.2 実験結果

実験結果を **Fig. 9.5** に示す．これは，LMS アルゴリズムによる学習結果である．なお，ここでの評価量（SNR）は，信号と推定誤差の比で表わしている．

Fig. 9.5 の (a) が，マイクロホン信号 $u_1(t)$，(b) が Cue Signal $\alpha(t)$ である．実際のドラム音の波形は (d) に示されている．(d) の強度を (b) が推定していなければならないが，これは，Fig. 5.12（p.55）のときほどは推定できていないことがわかる．異種情報である視覚情報から Cue Signal を作成する場合は，聴覚情報同士の場合に比較して目的音強度変化の推定は難しいのでこれは当然であろう．(c) がシステム出力である．それでも，ある程度はドラムの音のみを抽出できていることがわかる．なお，FIR フィルタ係数の初期値はすべて 0 である．

Fig. 9.5 (e) に信号・誤差比で見た学習曲線を示す．太線が内部目標による（すなわち Cue Signal 法による）学習曲線を表わし，細線が最適目標による（すなわち 理想的なトレーニング信号による）学習曲線をあらわしている．それぞれの 3 本の学習曲線は LMS アルゴリズムにおける修正係数 μ を変化させて実験したものである．（四角，丸，菱形の順に値が約 2 倍になっている）．ちなみに，Cue Signal 法の場合は，中間の値である丸印が最も好ましい結果を与えている．(c) は，その値での出力波形である．また，最も上にある破線は，最適目標を用い，LMS アルゴリズムではなく直接法により式(3.25) の方程式を 0.22 s 間隔で解いていったものである．これは，未来のデータを使わないという制限内での最適解を示すものである．

以上の結果より，マイクロホン 1 個の場合に比較して 約 8 dB の信号・誤差比の向上が見られ，それは最適解に比較すると 約 8 dB の劣化に相当することがわかる．すなわち，Fig. 5.12 に比較すれば劣るものの，ある程度目的音を抽出できたと言えよう．

9.3.2 視覚的な目的音規範の実験

次の実験は視覚的な規範で目的音を定義しようというものである．具体的には，音源が音を発生するときの変形・移動の方向によって目的音であるかどうかを見分けるというものである

9.3.2.1 実験方法

9.3.1 の実験では視覚情報の処理には特別な処理系は必要ではなかった．しかし，視覚的な規範での目的音定義くらいの仕事になってくると，視覚情報用になんらかの処理系が必要となってくる．次の章では，視覚情報の実時間処理用の専用プロセッサの製作について述べるが，ここでは，パーソナルコンピュータと内蔵フレームメモリを用いて行なった簡単な実験について述べよう．

Fig. 9.6 に実験システムを示す．用いた音源は 3 つである．

第 1 の音源は小さな鈴を 10 個束にした楽器である．これにソレノイドをつけて，遠隔操作で演奏できるように細工した．この音源の視覚的特徴は，音が鳴るときに鈴が左右に振動するということである．

第 2 の音源はカスタネットである．これにも上方にソレノイドをつけて，遠隔操作で演奏できるように細工した．この音源の視覚的特徴は，音が鳴るときに音源上部が上下に振動するということである．

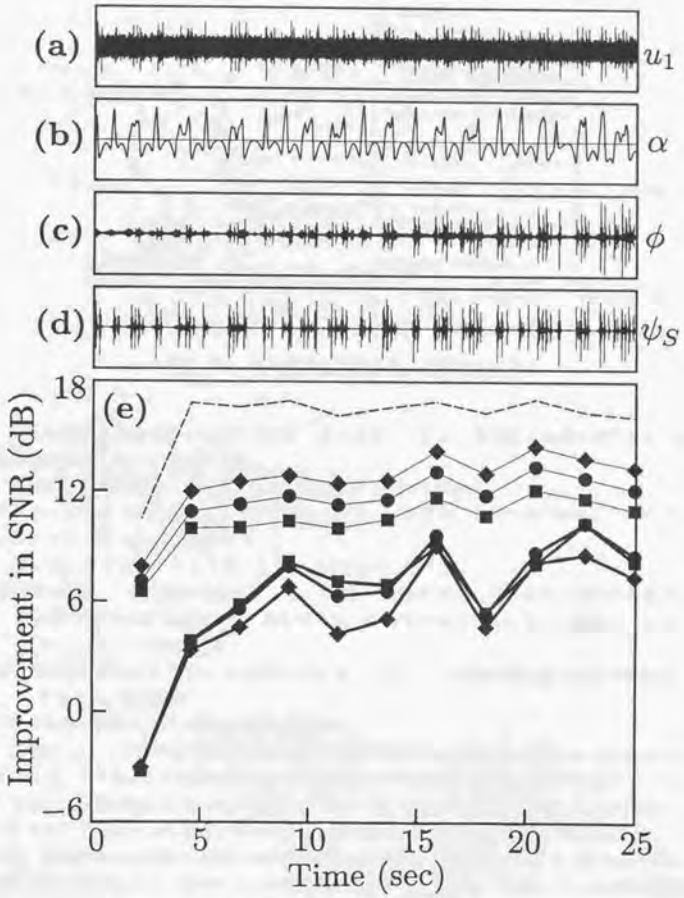


Fig. 9.5　視覚刺激を伴う音源の音の抽出実験の結果

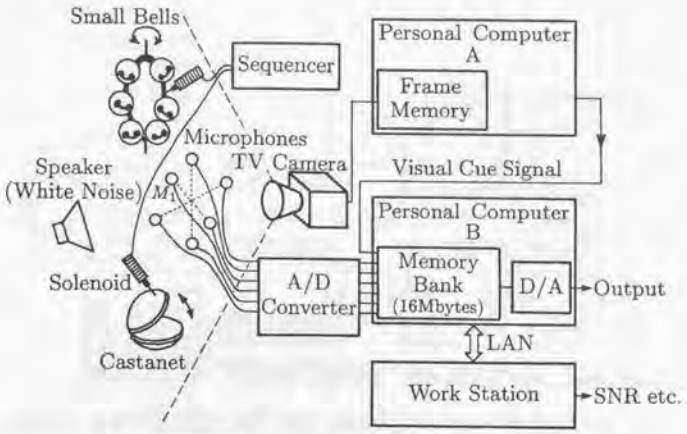


Fig. 9.6 視覚的な目的音規範の実験システム

第3の音源は白色雑音を発生するスピーカである．これは，視覚的な特徴を持たない（視覚的変化を有しない）音源である．

これら音源の映像を，広角レンズをつけたビデオカメラで映し，パーソナルコンピュータのフレームメモリに転送して，パーソナルコンピュータ上の一部機械語で書かれたプログラムによって Cue Signal を生成する．

Cue Signal 生成のアルゴリズムを **Fig. 9.7** に示す．これは，

目的音の規範 → 視覚的に定義する．ここでは，音を発生するときの音源の揺れの方向で目的音か妨害音かを識別する．具体的には，鈴は音が鳴るときに左右に振動し，カスタネットは上下に振動する．

視聴覚関係を記述するモデル → 暗黙のモデル．すなわち，視覚的変化の大きさが音強度を表すものと仮定する．

という仕様を満たすように構成したものである．

Fig. 9.7は，上半分が左右振動を伴う音源の音を抽出するための Cue Signal の生成アルゴリズムで，下半分が上下振動を伴う音源の音を抽出するための Cue Signal の生成アルゴリズムである．信号の流れにそって順に説明していこう．ビデオカメラで採取された映像は，ソフトウエアで実現された空間フィルタで左右方向あるいは上下方向の動きが検出される．これは，縦縞あるいは横縞の空間フィルタを画像に乗算し，空間積分することで行なっている．なお，図にあるように，空間フィルタが全画面にかけられていないのは，パソコンの演算速度による制約である．

空間フィルタ法によって得られた出力は，ハイパスフィルタにかけられ，強度エンベロープの推定値として適当な波形まで平滑化してやる．平滑化された信号は 4 s を1ブロックとして，交流化（$\overline{\alpha(t)} = 0$）および他の音源の推定強度との直交化〈p.42の式(5.98)〉が行なわれる．

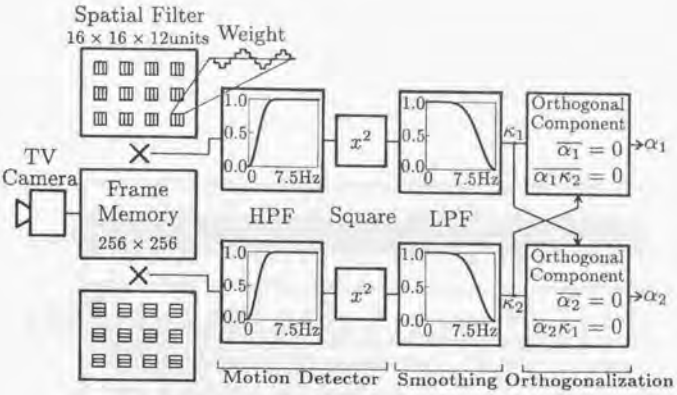


Fig. 9.7　音源の視覚変化（変形・移動）方向別に Cue Signal を生成するアルゴリズム

Fig. 9.6 にもどって説明する．パソコン A が視覚情報処理用である．カメラからのビデオ信号は，1/15s の周期でフレームメモリに取り込まれる．パソコン A は，ソフトウエア（積和演算）で空間フィルタを実現し，1/15s 間隔で出力する．パソコン B はデータ転送用である．6 チャネルの無指向性マイクロホン出力は，サンプリングレート 44.1kHz で A/D 変換され，パソコン A で出力された視覚情報とともに 16M バイトの RAM にいったん格納され，ワークステーションに転送される．

適応化アルゴリズムと FIR フィルタは，ワークステーション上で実現している．こうして作られたセンシングシステムの出力は，再びパソコン B に転送され音響信号としてモニタできるようにした．

FIR フィルタの次数は 32，その適応は 1 s 周期で $t = 0$ からその時間までの受信情報をもとした直接法である．フィルタの係数は，1 s おきに更新される．

9.3.2.2　実験結果

Fig. 9.8 に，左右振動を伴う音源の音を抽出するための Cue Signal（Fig. 9.7 の $\alpha_1(t)$）を用いた場合の実験結果を示し，**Fig. 9.9** に，上下振動を伴う音源の音を抽出するための Cue Signal（Fig. 9.7 の $\alpha_2(t)$）を用いた場合の実験結果を示す．

Fig. 9.8 (a) が目的音と妨害音を含んだマイクロホンの出力（$u_1(t)$），Fig. 9.8 (b) が Cue Signal（$\alpha_1(t)$）である．システムの出力を，Fig. 9.8 (c) に示す．数秒の学習で鈴の波形（Fig. 9.8 (d)）をほぼ抽出できていることがわかる．また，SN 比（信号と推定誤差の比）による学習曲線を Fig. 9.8 (e) に示す．最終的には 1 本のマイクロホン出力（Fig. 9.8 (a)）に比較して，約 18 dB の妨害音抑制が可能となった．

Fig. 9.9 についても同様に，数秒の学習でカスタネットの波形をほぼ抽出することができ，最終的な SN 比は 約 14 dB 向上した．

以上のように，全く同一のセンサ情報を用いて，視覚的な手がかりだけを頼りに，2 種類の音信号を別々に抽出することができた．

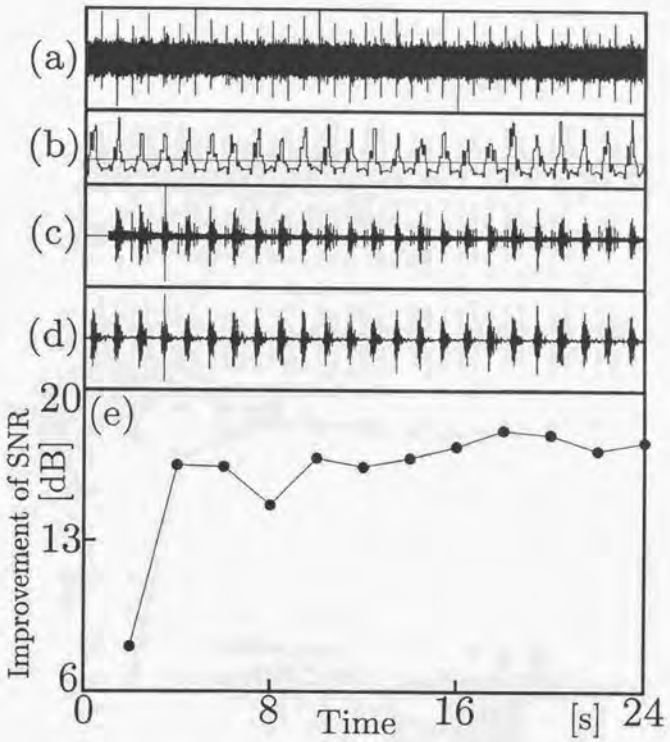


Fig. 9.8 左右振動を伴う音源の音を抽出する実験の結果

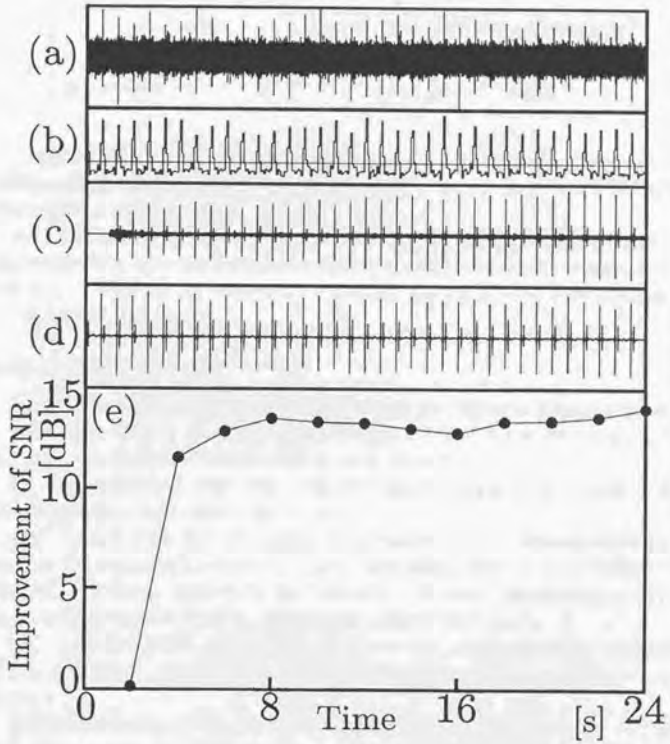


Fig. 9.9 上下振動を伴う音源の音を抽出する実験の結果

9.3.2.3 Cue Signal を直交化したことの効果

妨害音源をも見て適応化したことの効果を見るために，妨害音の音強度推定値で直交化した場合としない場合についての $T = 22$ s~24 s における SN 比の差を **Table. 9.1** に示す．目的音源を鈴とし，妨害音数2の場合と1の場合の両方で比較した．

Table 9.1 Cue Signal を直交化したことの効果

用いた音源	直交化した場合としない場合の SNR の差
鈴，カスタネット，白色雑音	+ 0.29 dB　(3.4%)
鈴，カスタネット	+ 0.41 dB　(4.9%)

2音源の音の強度エンベロープ間の相関が弱いため，（2つの楽器の強度エンベロープの相関係数は実測で 約 0.05 であった）その差は顕著ではないものの，数パーセントの改善が見られていることがわかる．

また，白色雑音を止めた場合に，その効果が大きくなっているがこれは **5.2.6** の結果（相対的に大音量を出している妨害音源はその音源をも見て抑制すべきである）に対応している．ちなみに，マイクロホン M_1 でのエネルギーレベルは，鈴に対してカスタネットは 0.8 dB 高く，白色雑音は 12.4 dB 高い．

9.3.2.4 視聴覚情報の同時性

ここでは，視覚情報と聴覚情報の時間軸方向での融合という観点から実験結果を検討する．

視覚情報から作成する Cue Signal は，目的音強度エンベロープの推定値である．ところで，この2つの異種情報が時間的にずれを持つことはないだろうか．

たとえば，音速は光速に比較して著しく遅いので 30m 先の目的音を対象にした場合，視覚情報は聴覚情報にくらべて約 0.1s 進んでしまう．

また，打楽器などのように，叩いた直後から音が出始めるために，視覚的な手がかりに聴覚的な信号が遅れる音源もあるだろう．これは，視聴覚関係を記述したモデルで対処すべき問題である．すなわち，暗黙のモデルを使う場合でも，一般的には，視覚情報から作った Cue Signal をそのまま使うのではなく，時間軸方向にも調節すべきであろう．

さて，この実験の場合の，目的音源の強度エンベロープと，視覚的 Cue Signal の相互相関を **Fig. 9.10** に示す．実線が鈴を目的音にした場合で破線がカスタネットを目的音にした場合である．

本実験で適応が成功したのは，両楽器の聴覚情報と視覚情報の時間差が小さく（0.1 秒以内），強度エンベロープと視覚的 Cue Signal が正の相関を持ち得たからであると言える．すなわち，暗黙のモデルが通用したわけである．

もし，この時間差にも自動的に追従して融合するよう工夫できれば，暗黙のモデルの応用範囲を広げることができるだろう．

9.3.2.5 視覚情報による不良センサデバイスの自律的切り離し

多数のセンサデバイスを持つシステムでは，たとえ一部のセンサが故障してもそれに自動的に対処できることが好ましい．

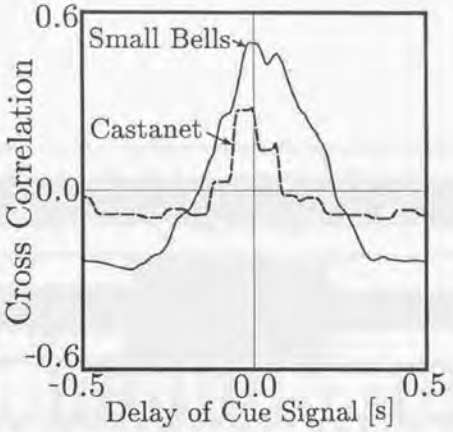


Fig. 9.10 強度エンベロープと視覚的 Cue Signal の相互相関

本システムでは，$\psi(t)$ を作るためのマイクロホンである M_1 以外のすべてのマイクロホンの故障に対して，視覚情報からそれに対処し，実質的にその信号を遮断することができる．

Fig. 9.11 は，Fig. 9.8 のときと全く同じ実験条件で，マイクロホン M_4 が故障した場合を実験したものである．具体的な実験方法は，$t = 7.5$ s ～ 15.5 s の間，マイクロホン M_4 の出力を強制的に白色雑音で置換した．

Fig. 9.11 (a),(b) がそれぞれマイクロホン M_1, M_4 の出力，Fig. 9.11 (c) が視覚的に左右方向に揺れ動く音源を抽出する手がかり量 $\alpha_1(t)$，Fig. 9.11 (d) が本システムの出力，Fig. 9.11 (e) が目的音の波形である．なお，Fig. 9.11 (b) で黒い長方形に見える部分が $u_4(t)$ を白色雑音（実際には乱数）で置き換えた部分である．

Fig. 9.11 (f) に，SN 比改善量（2 s 平均）を示す．縦軸は 7 dB ～ 16 dB の範囲である．M_4 故障の直後に低下するが，すぐに回復している．

また，Fig. 9.11 (g) には，M_4 に関するフィルタ係数 f_i の 2 乗和の 全フィルタ係数の 2 乗和に対する比率をあらわす．縦軸は 0 % ～ 17 % の範囲である．M_4 故障後に，M_4 に関する重みはすみやかに 0 になり，M_4 が復帰するとその重みも回復していることがわかる．すなわち，視覚情報で適応した結果，故障中には M_4 の出力を自動的に切り離し，M_4 が正常に戻ると，切り離された経路を自動的に復活することができている．

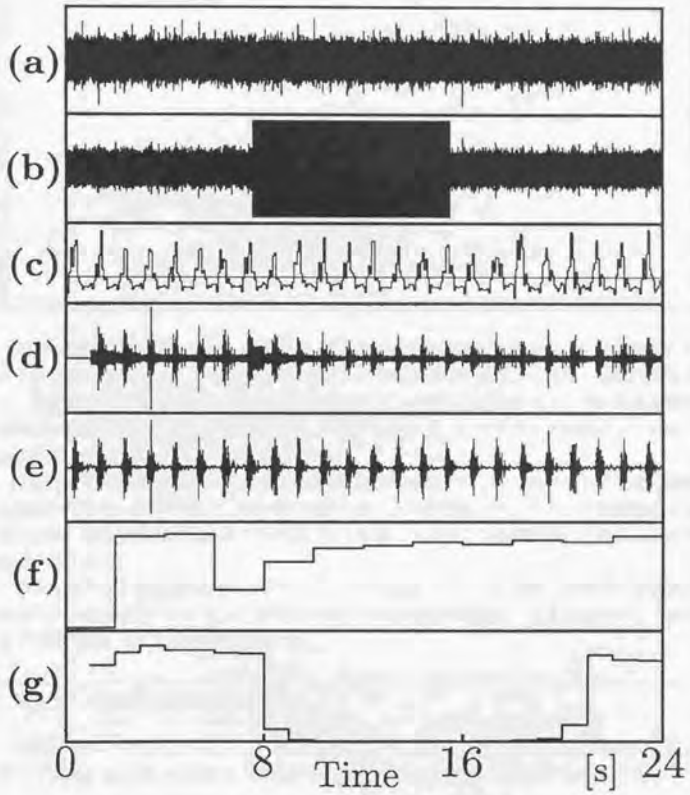


Fig. 9.11 不良センサデバイスに対する異種情報による自律的対処

第 10 章

リアルタイム処理系の試作

前章の実験を見ればわかるように，思いどおりの視聴覚融合型知能化センシングをリアルタイムで行なうためには，やはり専用の処理システムが必要である．つまり，実際の応用では，暗黙の視聴覚情報関係モデルやLMSアルゴリズムを使える場合があり，その場合処理能力の低い既存の処理系でも十分であるが，研究のためには，ハードウエアの制約なしに様々な実験をしたい．

そこで，専用の実験用リアルタイム処理系を製作することにした．このシステムは，現在も拡張中であり，製作が完了したわけではないが，この章では，本システムの設計方針，工夫した点，機能の概略，現時点での性能，それを使ってどのような実験ができるか，などを述べようと思う．

ここで紹介する視聴覚融合型知能化センシング用実験システムは，視覚用システムと聴覚用システムの２つのサブシステムより構成される．**10.1** 節では視覚用システムについて，**10.2** 節では聴覚用システムについて説明する．

10.1 視覚情報用サブシステム

視覚情報用サブシステムをどのような発想のもとに設計したかを述べるためには，まずこのサブシステムを用いてどのような実験を計画しているかを述べる必要があるだろう．

10.1.1 Cue Signal 生成アルゴリズムの具体例

たとえば，初歩的な画像処理技術を組み合わせて，自律的な Cue Signal 生成器としてそれらしく動かすためには，最低限でも **Fig. 10.1** くらいの処理をする必要がある．

ちなみに，Fig. 10.1 の動作を補足説明しておくと以下のようになる．

1. ［時間微分・絶対値］ 視覚的に変化した（対象物が変形した，動いた）部分のみに感度をもつ画像を得るための処理．

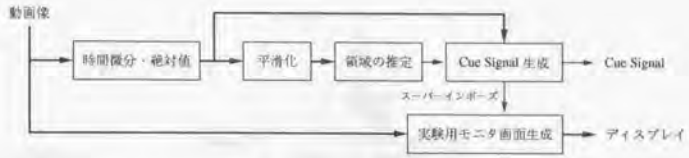


Fig. 10.1 簡単な Cue Signal 生成アルゴリズムの一例

2. ［平滑化］ 時間微分・絶対値処理の出力画像は対象物が変形したり動いたりした瞬間しか感度がないので，それを1次おくれ系で平滑化して視覚的変化を有する対象物を浮かび上がらせる．

3. ［領域の推定］ このようにして得られた画像を見て，対象物ごとの領域を判定する．例えば，対象物の大きさを32ピクセル四方程度と決めておき，その大きさの矩形領域を動かして画像上で輝度の高いところを探す．複数の矩形領域をお互いに反発しあうようなポテンシャルを与えて動かせば，複数の対象物領域を判定することができるであろう．ただし，そのなかから，どれが目的音源であるかは，別の方法で判断しなければならない．

4. ［Cue Signal 生成］ 最後に，3で得られた（領域の）座標位置と，1で得られた視覚的変化画像を用いて Cue Signal を生成する．視覚的変化画像のうち，得られた座標位置にある32ピクセル四方を切りとって積分し（平均を0にするための）ハイパスフィルタに通せば，Cue Signal らしきものが得られる．

5. ［実験用モニタ画面作成］ 1で得られた視覚的変化画像に Cue Signal の波形をオシロスコープのようにスーパーインポーズしたり，その他必要な数値や画像を随時モニタできるようにしておく．例えば，聴覚部の FIR フィルタの係数 f_n をバーグラフでリアルタイム表示するなどである．

もう少し高度な領域分割のアルゴリズムは考えられないだろうか．

我々は，視覚情報と聴覚情報を結合させるとき，事象の同時性を根幹の原理とした．画像は，ピクセルの集合である．そして，今やろうとしている領域分割を，「対象物ごとに領域を分割すること」と見るのではなく，「事象ごとにピクセルを結合すること」と考えてはどうだろうか．こうすれば，視聴覚融合と全く同じ原理で領域分割を行なうことができる．すなわち，事象の同時性である．具体的には，視覚的変化に同期性のあるピクセルを結合していき，時間的に相関のない視覚的変化をするピクセルは別の領域としていけば，事象の同時性を規範とした領域分割が可能となる．

Fig. 10.2 に，この原理にそって考えた Cue Signal 生成アルゴリズムを示す．

入力から順に説明していこう．まず，生の画像は3フレーム前の画像とピクセルごとに減算され，絶対値をとられる．これで，得られる画像は，変形・移動をした対象物に対応するピクセルが正の大きな値をもつ画像である．この画像と3つのマスク画像（Mask A, Mask B, Mask C）とでピクセルごとの乗算をし，それぞれ結果を空間積分する．3つのマスクには，対象物 A（事象 A），対象物 B（事象 B），対象物 C（事象 C）の現在での推定領域が（マスク値の正の領域として）保持されている．このため，3つの空間積分値のうち，現在視

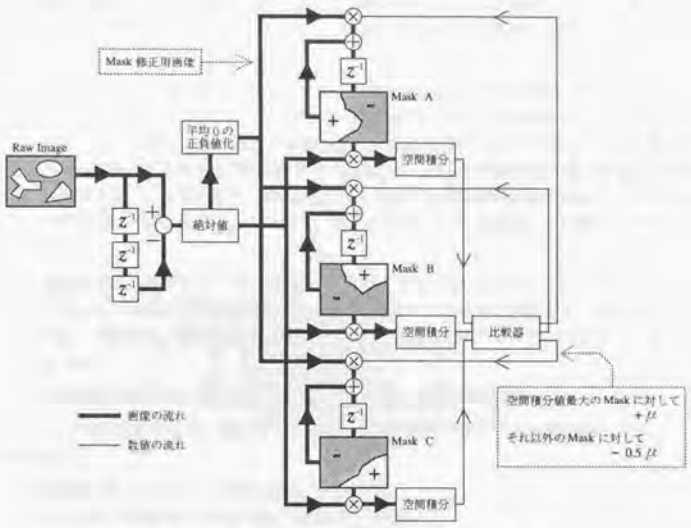


Fig. 10.2 対象物ごとの領域分割を自動的に行なう方法

覚的変化をしている対象物に対応するマスクと乗算した結果の空間積分値が最大となる．そこで，この3つの空間積分値を比較器で比較し，最大になったマスクに対しては，その領域を現在の「時間微分・絶対値画像」で補強し（加算する），そうでないマスクに対しては減算してやれば，自律的に領域の変化にマスク値が追従するようになるだろう．なお，マスク値の初期値は，平均0のランダム値でも入れておけばよい．もちろん，3つのマスクからどれが目的音源に対応するのかを決めるには，別の方法で判断しなければならない．

Fig. 10.2 に示した，Cue Signal 生成アルゴリズムの特徴は，ピクセルの集合に位相を入れずに扱っているということである．すなわち，あるピクセルと別のピクセルが隣あっているか否かという知識を全く用いていない．このため，ある対象物が別の対象物の後ろに隠れていて視覚的には2つに分断されて見えるときでも，全く問題なく領域の分割を行なうことができる．また，2つの対象物（例えば2台のロボット）が同期して動作している場合には，この2つは同一の事象を発生する対象物であるので，同一の領域に自動的に組み込まれるであろう．これは，Cue Signal 生成を考えたとき都合が良い．これらは，「対象物ごとに領域を分割すること」と「事象ごとにピクセルを結合すること」との本質的な違いから派生した特徴と言える．

10.1.2 視覚情報用サブシステムのアーキテクチャー

10.1.1 では，2つの処理アルゴリズム例をあげた．その他，視覚サブシステムでの処理として必要になってきそうな操作も考え合わせて，次のような処理が中心となると考えられる．

- 複数の異なる処理を流れ作業で行なう必要がある．これには，画像を1枚づつ流しながら，別々の処理プログラムをロードした複数のプロセッサで並列に処理するのがよいだろう．
- 1単位の処理（例えば空間微分）には，マイクロプロセッサやDSP（ディジタルシグナルプロセッサ）にして，1ピクセルあたり数命令～十数命令の演算が必要である．これに対して，現在のLSI技術で普通に得られる演算速度では，NTSCビデオ信号がもつ情報量をリアルタイムで処理するとすれば，1ピクセルあたりせいぜい数命令の演算しかできない．したがって，処理内容によっては，1単位の処理を複数のプロセッサで分担しなければならない．これには，画面を1/2や1/3に分割して処理するのがよいだろう．
- 領域分割のアルゴリズムや，目的音源の規範，Cue Signal生成のアルゴリズム，などの違いや，画面の分割処理の具合によって画像データの流れは多種多様に変化する．さらに，できれば，外界の変化（画像の変化）に応じて画像データながれを動的に変化させたい．
- ビデオ信号のフレーム間時間（1/60秒）ごとに処理を繰り返す．

これらを考えあわせると，次のような仕様の動画像処理プロセッサを製作すればよいことがわかる．

- 複数のプロセッサによる並列処理．(MIMD型)
- 各プロセッサはローカルに複数の画像メモリを持つ．
- それぞれの画像メモリはDMAで同時に複数画面の転送が可能であること．
- ブロードキャスト的DMAや，1対1のDMAなどがその都度柔軟に行なえること．
- 1単位の処理を画面分割して複数のプロセッサで分担処理できること．
- プロセッサの同期機能があること（1/60秒ごとなのでソフトウエア的実現で十分）

この仕様で，設計・製作した視覚情報処理サブシステムの全体図を**Fig. 10.3**下部に示す．

システムは9個のプロセッシングエレメント（PE）とそれらを管理するマスタープロセッサ，画像入力部，画像出力部，から構成される．PEは，動画像に対する様々な処理を行なうユニットであり，両隣のPEと画像をDMA転送するための8bit幅のローカルバス（DMA Link）で2重に結合されている．中央のマスタープロセッサは，9個のPEを管理するプロセッサである．具体的には，PEの相互のタイミングをとったり，画像以外のスカラーデータをPEと通信したり，PEに対してプログラムをロードしたりする．

PE内部を**Fig. 10.3**上部に示す．各PEは，スレーブプロセッサ（Motrola社のDSP56001），3面の画像メモリ（8bit幅のMemory A, B, C），画像メモリをDMA Linkに接ぎかえるためのトライステート出力Dフリップフロップ（D-FF），画像メモリをスレーブプロセッサに接ぎかえるための双方向バストランシーバ（  ），で構成される．なお，DMA Link用のポートは3組あり，そのうち2組は両隣のPEに接続され，残り1組は画像入力ユニットや画像出力ユニットなどに接続するための外部入出力ポートである．また，用いたDSPはデータバスが24bitなので，8bit幅のメモリデータバスを24bit中のどこに接続するのかということを決めなくてはならない．そこで，本システムでは，DSPと画像メモリの間に双方向バストランシーバを挿入し，画像メモリを8bitメモリ3面として確保することも，24bitのメモリ1面として確保することも可能なように設計した．これらの制御は，実行時にソフトウエアによるコマンド切り替えやメモリアドレスによる自動切り替えで任意に変更できる．

Fig. 10.3に示したアーキテクチャーにより，ブロードキャスト的DMAや，1対1のDMA

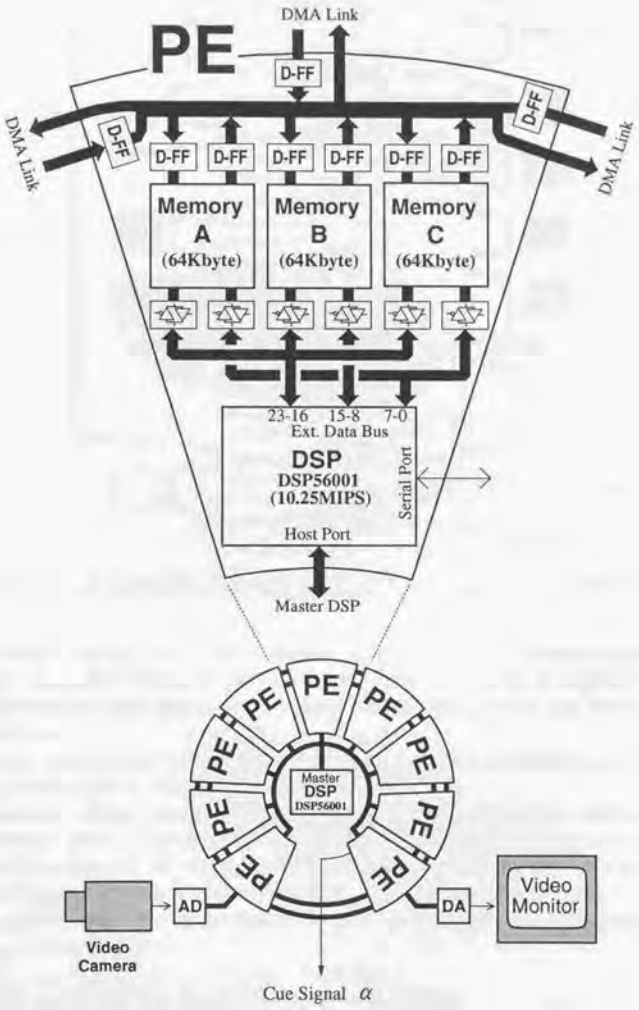


Fig. 10.3 リアルタイムで視覚情報を処理するために製作したサブシステム

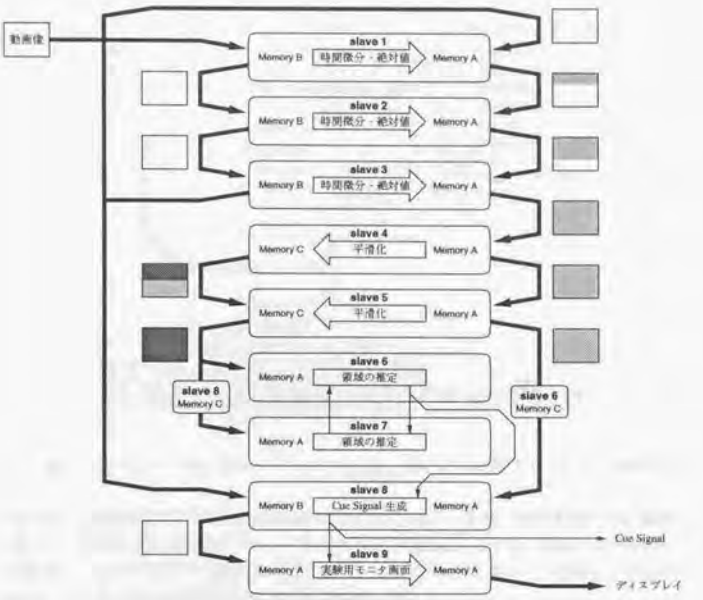


Fig. 10.4 一例としてあげた Cue Signal 生成アルゴリズム（Fig. 10.1）の具体的な実現

などが柔軟に行なえる．また，1対1の DMA は，DMA Link のリング上で交差しなければ複数個（最大5種類）の転送が同時に行なえる．ビデオ信号のフレーム間時間（1/60 秒）は，11 画面分の DMA 時間に相当するので，最大では 55 種類の DMA をフレーム間で行なうことができる．

ここで，Fig. 10.1 で述べたアルゴリズムが本システム上でどのように実現されるかを，一例として示しておこう．**Fig. 10.4** に，それを示す．

「時間微分・絶対値」処理を画面を3分割して3台のプロセッサに分担させ，「平滑化」処理は画面を2分割して2台のプロセッサで分担，「領域の推定」処理は2台のプロセッサで2つの対象物を追跡する．同一の作業を2台や3台のプロセッサで分担するには，各プロセッサが自分の担当の領域のみ新しいデータを上書きしていくことで容易に実現できる．

Fig. 10.4 を見れば，様々な実験に対してハードウエアの変更をせずに柔軟に対応できそうであることがわかるだろう．

10.1.3 視覚情報用サブシステムのソフトウエア環境

すべて自作のシステムであるから，プログラムの開発環境やデバックの環境などは自前で整備しなければならない．ここでは，ソフトウエア環境として工夫した点について述べる．

さて，実際に Fig. 10.4 をいきなりコーディングしようとすれば，かなり面倒な作業とな

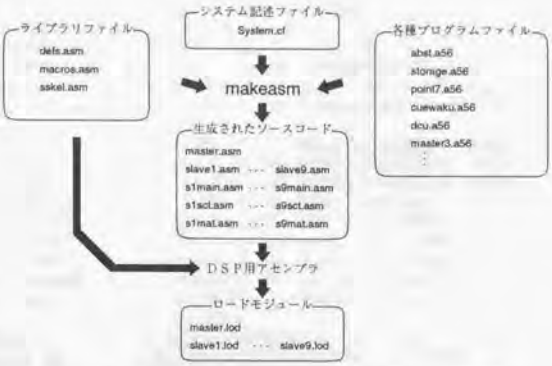


Fig. 10.5 プログラム makeasm によるソフトウエア開発環境

ることは明らかである．Fig. 10.4 のプログラムを書こうとしてまず困るのは以下の四点であろう．

第一に，画像処理のプログラムは高速性が要求されるので，各種の画像処理演算を DSP のアセンブリ言語を用いてプログラミングしなければならない．これは，非常に効率が悪い．

第二に，マルチプロセッサ相互の同期や通信をユーザーレベルですべて管理しなければならない．これもアセンブリ言語で書かなければならない．

第三に，DMA Link による画像の転送を行なうために Fig. 10.3 のトライステート出力 D フリップフロップ（D-FF）の制御もすべてユーザレベルで管理しなければならない．

第四に，画像メモリをスレーブプロセッサに接ぎかえるための双方向バストランシーバ（  ）などもユーザーレベルで管理しなければならない．

これらの困難を緩和するために，**Fig. 10.5** に示す手法をとった．これを説明しよう．

まず，時間微分・空間微分・平滑化・Cue Signal 生成など基本的なプログラムを蓄えておく．これが図の「各種プログラム」である．そして，実験を行なうにあたってユーザーは，各種プログラムのなかからどれを使用するかという情報と，画像データをどのように流すかという指令のみを記述する．これが「システム記述ファイル」（図の system.cf）である．こうしておいて，先に述べたような厄介な制御（マルチプロセッサ相互の同期や通信の制御，D フリップフロップ（D-FF）の制御，バストランシーバ（）の制御）のアセンブラプログラムは，システム記述ファイル（system.cf）から自動的に生成する．

これらの仕事を行なうのが，図では「makeasm」と書いてあるプログラムである．makeasm は，perl 言語にして 1000 行ほどのプログラムである．もちろん，アセンブラにかけられる形のソースコードを自動生成するためには，スレーブプロセッサ用ソースコードのテンプレート（図の sskel.asm）ファイルなども利用する必要がある．

システム記述ファイル（system.cf）の具体例を **Fig. 10.6** に示す．これは，Fig. 10.4 を実行させるためのものである．Cue Signal 転送など，スカラーデータ転送の部分などを若干省いてあるが，ほぼこんなものである．なお，各行のセミコロン（；）から後ろはコメントである．

```
                        ;   system.cf  for prj13
define  XSIZE  set 256   ; 画面は横256ピクセルとする
define  YSIZE  set 192   ; 画面は縦256ピクセルとする
define  SBIT   set 1     ; 各ピクセルの値は2の補数形式とする
processors 9             ; 現在装備されているスレーブプロセッサは9個
master ../a56/master3    ; マスタープロセッサにはmaster3をロードする

slave 1 abst             ; プロセッサ1は，画面の1/3を時間微分・絶対値
slave 2 abst             ; プロセッサ2は，画面の1/3を時間微分・絶対値
slave 3 abst             ; プロセッサ3は，画面の1/3を時間微分・絶対値
slave 4 storage          ; プロセッサ4は，画面の1/2を平滑化
slave 5 storage          ; プロセッサ5は，画面の1/2を平滑化
slave 6 point7           ; プロセッサ6は，対象物1の領域を判定
slave 7 point7           ; プロセッサ7は，対象物2の領域を判定
slave 8 cuewaku2         ; プロセッサ8は，Cue Signalを生成
slave 9 dcu5             ; プロセッサ9は，実験用モニタ画面を生成

videoin 1                ; ビデオカメラはプロセッサ1に接続
videoout 1               ; モニタディスプレイはプロセッサ1に接続
coefin 9                 ; 音響部からFIRフィルタの係数を，モニタ画面に
                         ; スーパーインポーズするためプロセッサ9に入力
dma 0     (6c)-8a                                ; DMA×1
dma 7     5a-(6c)     (8c)-7a                    ; DMA×2,  7を起動
dma 0     4a-5a                                  ; DMA×1
dma 6     3a-4a       5c-6a,(8c) 9a-videoout     ; DMA×3,  6を起動
dma 4,5,9 2a-3a  8b-9a 4c-5c                     ; DMA×3,  4，5，9を起動
dma 8     1a-2a  3b-8b                           ; DMA×2,  8を起動
dma 0     3b-1a                                  ; DMA×1
dma 3     2b-3b                                  ; DMA×1,  3を起動
dma 2     1b-2b                                  ; DMA×1,  2を起動
dma 1     videoin-1b                             ; DMA×1,  1を起動
```

Fig. 10.6 システム記述ファイル（system.cf）の具体例

8行目から16行目には，9つのプロセッサにどんな「各種プログラム」を割り当てるかが記述してある．ここで，同じ名前の「各種プログラム」を2つ以上書くと，領域分割の分担作業と見なされる．例えば，時間微分・絶対値処理プログラムであるabstは3台のプロセッサに割り当てられているので画面の1/3を処理するように自動的に修正が加えられる．

22行目から31行目がDMA Link経由の画像転送の記述部である．例えば，25行目の記述は以下の意味である．「PE 3のMemory AからPE 4のMemory Aに画像をDMA転送．同時に，PE 5のMemory CからPE 6のMemory Aに画像をDMA転送．同時に，PE 5のMemory CからPE 8のMemory Cに画像をDMA転送．同時に，PE 9のMemory Aから画像出力ポートに画像をDMA転送．これらが終了したら，PE 6の処理プログラムをスタートさせる」

「時間微分・絶対値」処理のプログラム（abst.a56）や「平滑化」のプログラム（storage.a56）などがストックしてあれば，ユーザーはこのFig.10.6のみを記述すればよい．あとは，makeasmに通せば最終的なアセンブリ言語コードが生成されるのでそれをアセンブラにかければすぐに実験が始められる．

次にデバック環境について簡単に触れておこう．

本システムのソフトウエアを開発するにあたって，動作の確認やプログラムのデバックは

困難な作業となる．その理由として次の2つをあげることができる．

- センサデータの実時間処理であること．現実世界の情報で動作するため，再現性のある動作をさせることが難しい．すなわち，閉じた世界でのデバックがやりにくい．
- 並列処理であること．システムが予想外の動作をしたとき，どのプロセッサに原因があるのかを突き止めにくい．

そこで，本システムでは，実際のセンサデータを処理中であっても，マスタープロセッサを介して各PEのローカルなメモリを自由に参照・変更できるようにした．これによって，デバックが容易になっただけでなく，実験のパラメータなども動作中に自由に変えられるようになり，実験の効率を向上させることができた．

10.2 聴覚情報用サブシステム

第I部で結論したように，Cue Signal 法には，ブロック化した直接法が最も適している．そこで，この処理をリアリタイムで実行する専用のシステムを自作した．

10.2.1 ブロック化した直接法の具体的な処理内容

ブロック化した直接法で Cue Signal 法 を実現すると，具体的には **Fig. 10.7** に示す演算を実行しなければならない．

これは，次の3つの仕事に大別できる．

[**仕事 1**] 相関行列 R と，相関ベクトル $\boldsymbol{p}$ を計算する．
[**仕事 2**] 連立1次方程式 $\boldsymbol{p} = R\boldsymbol{f}$ を解く．
[**仕事 3**] FIR フィルタ演算 $\phi(t) = \boldsymbol{f}\,\boldsymbol{y}(t)$ で目的音を抽出する．

このうち，仕事1と仕事3が音響信号のサンプリング周期で繰り返し演算しなければならない処理で，仕事2が数秒の周期で行なう処理である．そして，この3つの仕事は比較的独立性が高い．仕事1から仕事2へは，R と $\boldsymbol{p}$ を数秒に1回転送するだけであるし，仕事2から仕事3へは，$\boldsymbol{f}$ を数秒に1回転送すればよいからである．

そこで，リアルタイム化にあたり，この3つの仕事をパイプライン処理で実行させることを考えた．それを **Fig. 10.8** に示す．このようにすることの利点は，FIR フィルタによる信号抽出がほとんど遅延無しに行なえることである．逆に欠点は，適応に用いた時間ブロックと，信号抽出に用いる時間ブロックが異なる時間ブロックになってしまうことである．このことの影響は，このあと（**10.2.4**）の実験で確かめることとする．

10.2.2 聴覚情報用サブシステムのアーキテクチャー

視覚システムのときと同じように，聴覚システムに求められる処理の特徴を列挙してみよう．

- サンプリング周期が，視覚システムに比較して非常に短い．視覚システムは 166ms で，聴覚システムは 22.7μs であるので，3桁近い違いがある．
- しかし，サンプリング周期の間にやらなければならない処理量は少ない．例えば，相関行列 R や 相関ベクトル $\boldsymbol{p}$ の計算，FIR フィルタの計算などは，すべて積和演算にすぎない．
- サンプリング周期の間に通信しなければならない情報量も少ない．各プロセッサとも，せいぜい数ワードである．これは，視覚システムが画像という数万ワードのデータ転送

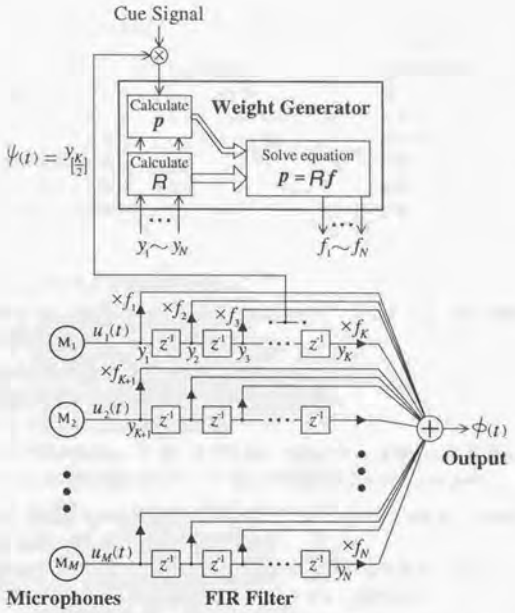


Fig. 10.7 ブロック化した直接法の具体的な処理内容

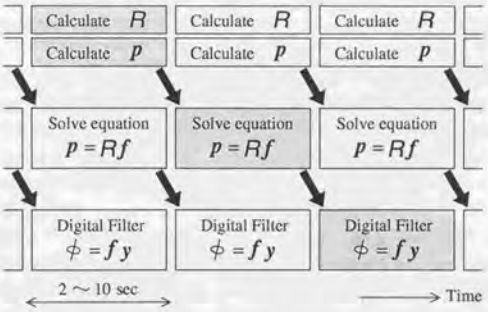


Fig. 10.8 ブロック化直接法のパイプライン処理

Table 10.1 視覚情報処理システムと聴覚情報処理システムに求められる特徴

	視覚情報処理システム	聴覚情報処理システム
処理の繰り返し周期	長い（1/60 秒）	短い（1/44100 秒）
1回の通信量	大きい（256 × 192 ワード）	小さい（数ワード）
データの分解能	粗くてよい（8 bit）	高い分解能が必要（16 bit）
ローカルメモリの DMA 転送	必要	不要
共有バスによるデータ転送	不要	必要
ハードウエア的同期機構	不要	必要

を必要としていたのと好対照である．

これらの事項と Fig. 10.7 Fig. 10.8 などを考えあわせて，次のような仕様の音響信号処理プロセッサを製作すればよいことがわかる．

- 複数のプロセッサによる並列処理．(MIMD 型)
- 各プロセッサはローカルに小規模のメモリを持つ．
- ローカルメモリの DMA 転送は不要．
- プロセッサ間の通信としては，小さなワード長のデータを高速に転送できること．
- サンプリング間隔が短いのでハードウエア的な同期メカニズムが必要．

このように，視覚システムと聴覚システムでは，設計に対して求められる特徴が正反対である．これを **Table. 10.1** に比較して示しておく．

さて，以上の仕様に基づいて，聴覚情報処理サブシステムを設計・製作した．全体図を **Fig. 10.9** 上部に示す．なお，これを見ればわかるように，現時点ではシステムの処理能力の制約のため，仕事 2 をワークステーションで代行させている．

システムは，30 個の DSP ユニット，AD 変換器，DA 変換器，ホストコンピュータ（ワークステーション）などから成る．

プロセッサ間の通信にはコモンバス方式を採用した．30 個程度のプロセッサの通信には，むしろ小回りが効くし，柔軟である．ソフトウエア的にもシンプルになる．なお，このコモンバスにはハードウエア的にプロセッサの同期をとる機能を付加してある．プロセッサ間の通信もハードウエア的にタイミングをとるので，オーバーヘッドの少ない小回りの効く通信が可能である．

視覚システムのマスタープロセッサに相当するものはない．サンプリング間隔が短いため，ソフトウエア的に同期や通信の指令を出していたのでは間に合わなくなるからである．

DSP ユニット内部を **Fig. 10.9** 下部に示す．各ユニットは，プロセッサ（DSP56001），小規模なローカルメモリ（RAM, ROM），DSP データバスを共有バスに接続するための双方向バストランシーバとその制御回路，で構成される．このように，各ユニットは素直でシンプルな設計になっている．

ここで，視覚システムから聴覚システムへの Cue Signal の通信について述べる．Cue Signal は非同期シリアル通信によって伝達している．動作サイクルの異なる 2 つのシステム間で，このような簡単な通信手段が成立するのは，Cue Signal の帯域がせまく情報量が非常に少ないためである．すなわち，Cue Signal 法による視聴覚融合は，視覚システムと聴覚システムを密結合する必要がないのが特徴である．これは，Table. 10.1 で示した事情を考え合わせれば，両システムを視聴覚それぞれの特性に合わせて最適に設計できることを意味しており，Cue

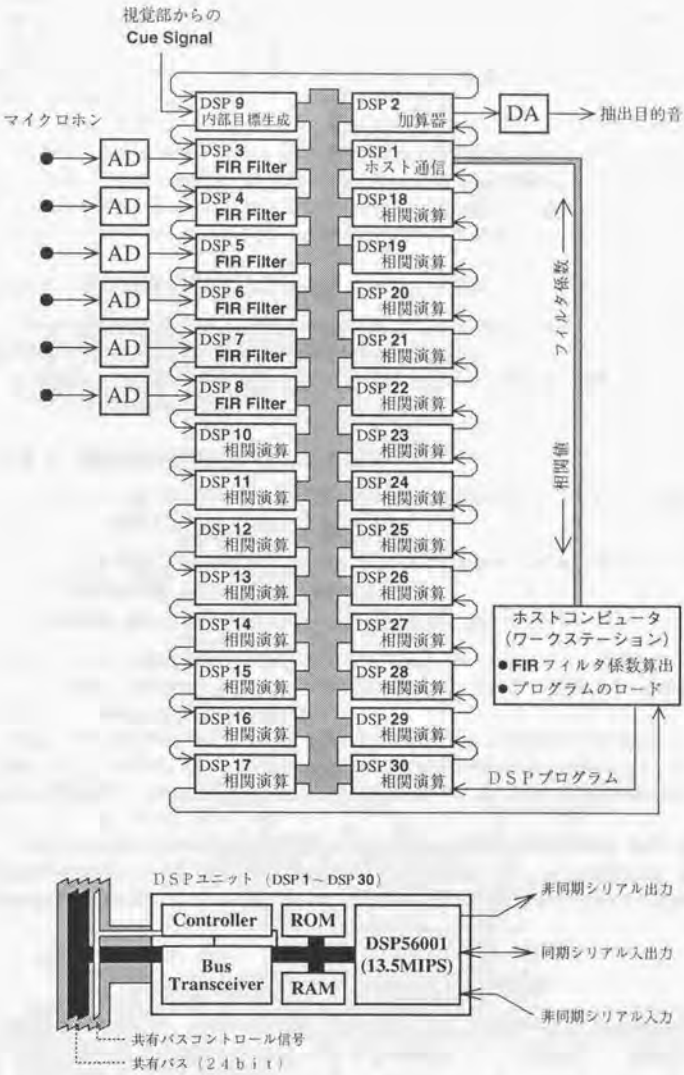


Fig. 10.9 リアルタイムで聴覚情報を処理するために製作したサブシステム

Signal 法の長所であると言えるだろう．

最後に，Fig. 10.7 と Fig. 10.8 で述べたアルゴリズムが本システム上でどのように実現されるかを述べておこう．Fig. 10.9 の上部の DSP ユニットの駒の中に，30 台のユニットでの処理の分担が書き込んである．相関行列 R と相関ベクトル $\boldsymbol{p}$ の計算に 21 台のユニットを割り当て，FIR フィルタ演算に 6 台のユニットを割り当てている．

なお，相関行列 R の計算は以下のように簡略化できる．すなわち，R はマイクロホン 6 本，フィルタ次数 32 の場合，192×192 の行列である．要素の数は 36864 もある．しかし，タップ信号 y_1 と y_2 の積和は，観測時間が十分長ければ y_2 と y_3 の積和とほとんど等しい．このように見ると，36864 の要素のうちで，ほんとうに独立な要素は，僅か 1152 しかない．R と $\boldsymbol{p}$ の計算が 21 台ですんだのは，この性質を用いたためである．

10.2.3 聴覚情報用サブシステムのソフトウエア環境

プロセッサ数が増えてくると，各プロセッサにプログラムを割り当てたり，共通の定義やプロセッサ番号に依存する定義などの管理がやはり必要である．

ここでは，視覚システム用に開発したツールである makeasm （Fig. 10.5 参照）を改造して，これらの管理にあてている．

10.2.4 聴覚情報用サブシステムの基礎実験

ここまで，述べてきたパイプライン的に実現したブロック化直接法のアルゴリズムが使えるためには，以下の 2 点を確認しておく必要がある．

- R と $\boldsymbol{p}$ を算出した時間ブロックと，FIR フィルタで目的音抽出を行なう時間ブロックが，異なる時間ブロックでも問題はないのか．
- 数秒程度の短いブロック長でも，音声などの非定常音に適応できるのか．

そこで，これらの問題を調べるため，製作した Fig. 10.9 のシステムを用いて基礎実験を行なった．実験は，目的音として（テープレコーダに録音した）音声を，妨害音として白色雑音を用いた．**Table. 10.2** に，実験結果を示す．

実験 1 は 20 秒のブロック長に対して，適応と信号抽出をを同一の時間ブロックで行なった実験である．これに対して実験 2 では適応と信号抽出を異なる時間ブロックで行なった．SN 比の改善量でみて，特性に劣化は見られない．よって，パイプライン化した影響は致命的ではないといえる．

実験 3 では時間ブロック長を 5 秒に短縮した．特性は僅かに劣化しているものの，まだ信号抽出が可能であることがわかる．ブロック長は，特性で見れば長いにこしたことはないが，環境の変化への追従速度などを考え合わせれば，数秒程度の長さが適当であろうと考えられる．

Table 10.2 聴覚情報用サブシステムの基礎実験結果

実験番号	ブロック長	適応と信号抽出	処理前の SN 比	処理後の SN 比	SN 比の改善
実験 1	20 s	同一の時間ブロック	−15.6 dB	5.4 dB	21.0 dB
実験 2	20 s	異なる時間ブロック	−15.5 dB	5.8 dB	21.3 dB
実験 3	5 s	異なる時間ブロック	−15.0 dB	2.2 dB	17.2 dB

第 11 章

考察

ここでは，Cue Signal 法による視聴覚融合に関連した 3 つのテーマについて論じる．**11.1** 節では，視聴覚融合モデルを構成する立場から，Cue Signal 法による視聴覚融合が一体何をやっているのかを改めて考え直してみる．**11.2** 節では，視覚センサで音源位置を計測してマイクロホンアレイの指向特性をむける方法に比較したとき，Cue Signal 法の長所は何であるのかを論じる．**11.3** 節では，今後の発展問題として Event Signal を考え直してみる．すなわち，Event Signal を単なる音強度の推定「値」から，その分布や，信頼度などに関する知識を含んだ情報への拡張を試みる．たとえば，安藤は速度ベクトル分布計測において，測定信頼度の自己評価量を導入した．[60] また，微分両眼視による立体再現では，求められた凹凸を表す画像以外に，測定の評価量を表す画像を生成し，複数の計測の累積合成に活用する手法を提案している．[61] ここで述べる Event Signal の拡張は，そのような研究に触発されて試みたものである．

11.1 視聴覚融合のモデル

11.1.1 事象生起による融合

まず，視聴覚融合のために，対象物における視聴覚情報の結合をどのようにモデル化するかについての述べる．**Fig. 11.1** にさまざまな様式を示す．**8.2** 節で述べたように，視聴覚情報は通常は対象物の内部状態によって結合されている（図中 (a) の「一般的モデル」）．内部状態とは，たとえば‘あ’という文字を発音するという意思などである．しかし，これでは一般的すぎて取り扱いにくい．そこで，モデルの単純化を行なう．

(b) の「事象の分解」は，起こり得る事象を列挙し，内部状態をその線型和で表現しようとするものである．たとえば母音の発音という事象を，‘あ’,‘い’,‘う’,‘え’,‘お’という 5 つの事象で分解して表わす．このモデルであれば，‘あ’と‘え’の中間の発音などを表現することもできる．

また，(c) の「事象生起の離散化」は，それらの事象の生起を 2 値化したものである．その下 (d) の「事象生起の離散化（排他的）」は，各事象のうちどれかひとつしか起こらない場合のモデルである．これは例えば，明確に母音を発音する場合に相当するだろう．

(e) の「単一事象生起」は，有る事象が生起した度合のみを表現する．すなわち，話者がどのくらい強く発声したか，などをモデル化して表わす．その下 (f) の「単一事象生起（事象：

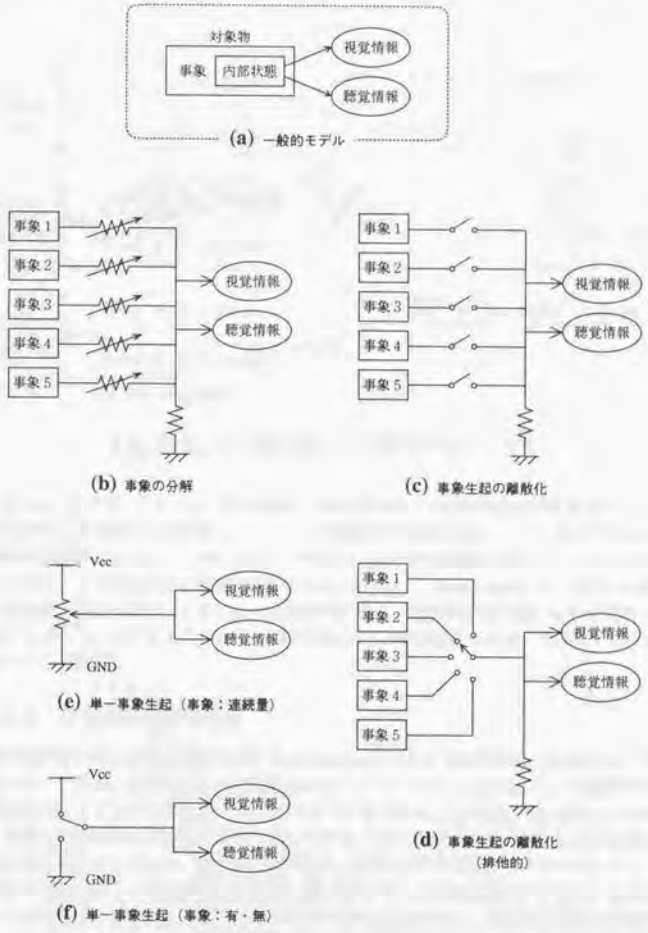


Fig. 11.1　視聴覚融合のための対象物モデル

有・無)」は，それをさらに2値化したものであるが，ここまで単純化するのは行き過ぎであろう．

「単一事象生起」モデルが，Event Signal による視聴覚融合を説明するものである．もちろん，**6.6**節で述べた考え方は，まさにこの図 (b) の「事象の分解」に示した多事象モデルに相当するものである．

次に，センシングシステムまで含めてこのモデルをあてはめるとどうなるか．これを，**Fig. 11.2** に示す．

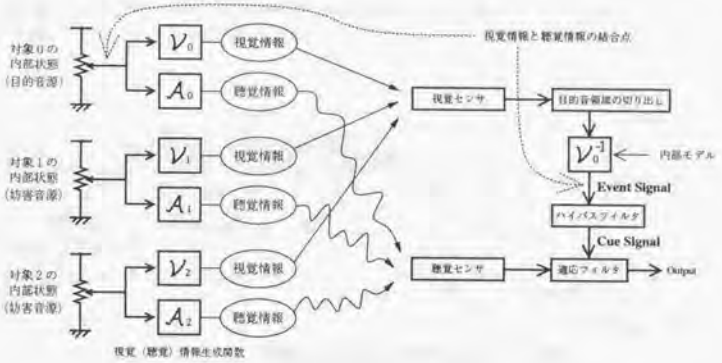


Fig. 11.2 Cue Signal 法による視聴覚融合のモデル図

Fig. 11.1 (e) で示したように，各対象物の内部状態は単一の時間関数に単純化されている．内部状態量は視覚情報生成関数 $\mathcal{V}_j$ によって視覚情報に変換される．一方，聴覚情報は，聴覚情報生成関数 $\mathcal{A}_j$ によって生成される．両者とも，一対多の関数である．システムの視覚部の役割は，この内部状態を視覚的に推定することである．Event Signal は，対象の内部状態の推定量に他ならない．すなわち，視覚部の役割は，視覚情報生成関数 $\mathcal{V}_j$ の逆変換 $\mathcal{V}_j^{-1}$ であると言うことができよう．なお，この変換には各々の対象物に応じた内部モデルが必要となることが多い．

11.1.2　空間座標上での融合

視聴覚融合に関しては，事象生起量 (Event Signal) による方法以外に，もうひとつの可能性がある．それは，聴覚センサを立体聴覚のセンシングシステムに仕立てて，音響信号からも対象物のおおよその位置を得てしまおうという方法である．この考え方を，**Fig. 11.3** に示す．異なる位置に設置された複数の聴覚センサは，対象空間上の各点に対応したそれぞれ固有の伝達特性を有している．これを利用すれば，視覚センサほど直接的でないにしても，音源位置を推定することが可能である．この方法であれば，3次元空間上の「位置」で融合することができる．すなわち，対象の内部状態まで遡る必要がない．単純物理量だけの世界で融合を片づけることができるのである．これは大きな利点であろう．しかし，このとき聴覚はすでに視覚的になっていると言えるかもしれない．ただ，先に述べたように，生体のセン

シングにおいてこちらの方法も利用されていることは，生理学的研究からも確実である．

Cue Signal 法の特徴は，結合の単純性にある．すなわち，事象生起の度合を表すただひとつの時系列で視聴覚を結合している．この単純性ゆえ，空間座標を仲介した結合など他の手法による結合を併用することも容易であると言える．

本論文では Cue Signal 法による視聴覚融合を述べたが，それは Cue Signal による方法以外の視聴覚融合を否定するものではない．むしろ将来は Cue Signal 法だけでなく，様々な方法を組み合わせるべきであると主張したい．

しかし，その前に Cue Signal 法単独の融合を深く検討する必要がある．Cue Signal という単純量でどこまで効果があがるのかを見極める必要がある．本論文で，Cue Signal 法に限って検討したのはこのような理由からである．

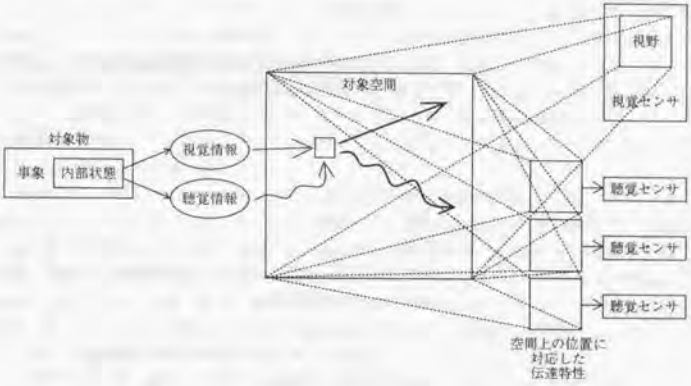


Fig. 11.3 対象空間の位置による視聴覚融合

11.2 FIR フィルタの全係数を可変にする意味

例えば，空間座標を用いた視聴覚融合手法では，視覚センサを用いて音源の位置を測定し，その方向にマイクロホンの指向特性を合わせるべく，各マイクロホンに適切な遅延をかけて和をとるという方法が考えられる．このとき適応の自由度は M である．

これに対して，Cue Signal 法による視聴覚融合では，FIR フィルタの全係数を最適化することができる．このときの自由度は $N = MK$ である．このように FIR フィルタの全係数を可変にすることの意味（すなわち，各マイクロホンに対して任意の周波数特性を作ることの意味）は，理論的には **3.1.3** での議論などからも明らかであるが，前者と比較して実際にどの程度効果があるのかは，まだわからない．

そこで，様々な自由度をもった線型フィルタに対して，目的音推定誤差を調べてみよう．

評価の方法は音源の位置とスペクトルを与えて，数値計算で推定の平均2乗誤差を求めることとする．なお，フィルタ係数収束後の特性を調べるため，適応目標 $d(t)$ としては 最適目標 $\psi_S(t)$ を与える．具体的には以下の方法で評価する．

まず，評価基準である．平均2乗誤差を最小にするフィルタ係数を $\boldsymbol{f}$ としたとき，平均2乗誤差 $\langle e(t)^2 \rangle$ は，以下のように目的音劣化に起因する成分 $\langle e_S(t)^2 \rangle$ と妨害音混入に起因する成分 $\langle e_N(t)^2 \rangle$ に分解できる．

$$\langle e(t)^2 \rangle = \boldsymbol{f}^T R \boldsymbol{f} - 2\boldsymbol{f}^T \boldsymbol{p} + \langle d(t)^2 \rangle \tag{11.1}$$

$$= \underbrace{\boldsymbol{f}^T R_S \boldsymbol{f} - 2\boldsymbol{f}^T \boldsymbol{p}_S + \langle d(t)^2 \rangle}_{\text{目的音劣化に起因する成分}} + \underbrace{\boldsymbol{f}^T R_N \boldsymbol{f} - 2\boldsymbol{f}^T \boldsymbol{p}_N}_{\text{妨害音混入に起因する成分}} \tag{11.2}$$

ここで，$R_S, \boldsymbol{p}_S$ は，目的音に対する相関行列，相関ベクトルを表すものとし，$R_N, \boldsymbol{p}_N$ は，妨害音全体に対する相関行列，相関ベクトルを表すものとする．

平均2乗誤差最小の規範は，$\langle e_S(t)^2 \rangle$ と $\langle e_N(t)^2 \rangle$ の和を最小にするので，目的音強度が大きければ前者を重視した適応が行なわれ，逆に妨害音強度が大きければ後者を重視した適応が行なわれる．そこで，両者をグラフの横軸と縦軸にとって，目的音強度を変化させて両誤差をプロットしていけば2乗誤差のトレードオフを表す曲線が得られる．実際の平均2乗誤差は，その曲線に目的音と妨害音の強度差に応じた傾きの接線を引けば，その接点によって表わされる．たとえば，目的音強度と妨害音（総）強度が等しいとき，誤差曲線と傾き -1 の直線との接点が平均2乗誤差を表わす．いずれにしても，原点に近い曲線ほど推定の平均2乗誤差が小さいことになる．そこで，様々な自由度をもった線型フィルタに対する推定誤差をこの誤差曲線で評価することとする．

サンプリング周波数は，$F = 44.1$kHz とし，妨害音は，0～22.05kHz の白色のスペクトルを持つとした．目的音は，0～11.025kHz に一様スペクトルをもつもの（帯域制限信号），低域にエネルギーの集中しているもの（強度 $\propto 1+\cos(2\times\pi\times$周波数$/F)$），白色，の3種類を用いた．また，空間上には反射壁がないものと仮定し，音源は，目的音源1個 ($x=30$cm, $y=40$cm, $z=80$cm)，互いに独立で同一強度の妨害音源3個 ($x=-100$cm, $y=-30$cm, $z=-20$cm)，($x=90$cm, $y=-50$cm, $z=-50$cm)，($x=-80$cm, $y=-70$cm, $z=80$cm)，を配置し，6本のマイクロホンは座標軸上に原点から 11cm 離して配置した．

評価の対象とした線型フィルタは全部で6種類である．これを，**Fig. 11.4** に示す．

Fig. 11.4 の (a) は，1本のマイクロホンのみを用い，単にそのゲインを調節するものである．妨害音と目的音を弁別する能力は全くないが，一般にゲインを下げることで平均2乗誤差は減少する．例えば，目的音と妨害音の強度が等しいとき，ゲインを 1/2 にすれば，平均2乗誤差は半分になる．このように目的音抽出能力が全くないにもかかわらず起こる形式的な誤差の減少を参考として示すために (a) を計算した．

(b) は，1本のマイクロホンを用いて目的音と妨害音のスペクトルの差異から信号抽出を行なおうとするものである．

(c) から (f) は，6本のマイクロホンを用いた場合である．(c) が遅延を調節してマイクロホンアレイの指向特性を制御するものである．視覚センサで目的音源方向を測定しフィルタを適応させる方法（空間座標上での融合）では，この (c) のフィルタを用いることとなる．

(d) は，(b) と (c) を組み合わせたものである．遅延を調節して目的音源方向による信号抽出を行なったあと，FIR フィルタでスペクトルの差異を利用した信号抽出を行なう．

(e) は，(c) が各マイクロホン出力の単純和であったものを，重みをつけた和としたものである．視覚センサで目的音方向のみでなく目的音や妨害音の位置がわかっているとき，目的音に近く妨害音に遠いマイクロホンには重い重みをつけて，目的音に遠く妨害音に近いマイクロホンには軽い重みをつければ，(c) に比較して良好な信号抽出が期待できる．

(f) が，一般の FIR フィルタである．Cue Signal 法では，このすべての係数を適応させることができるのが特徴である．

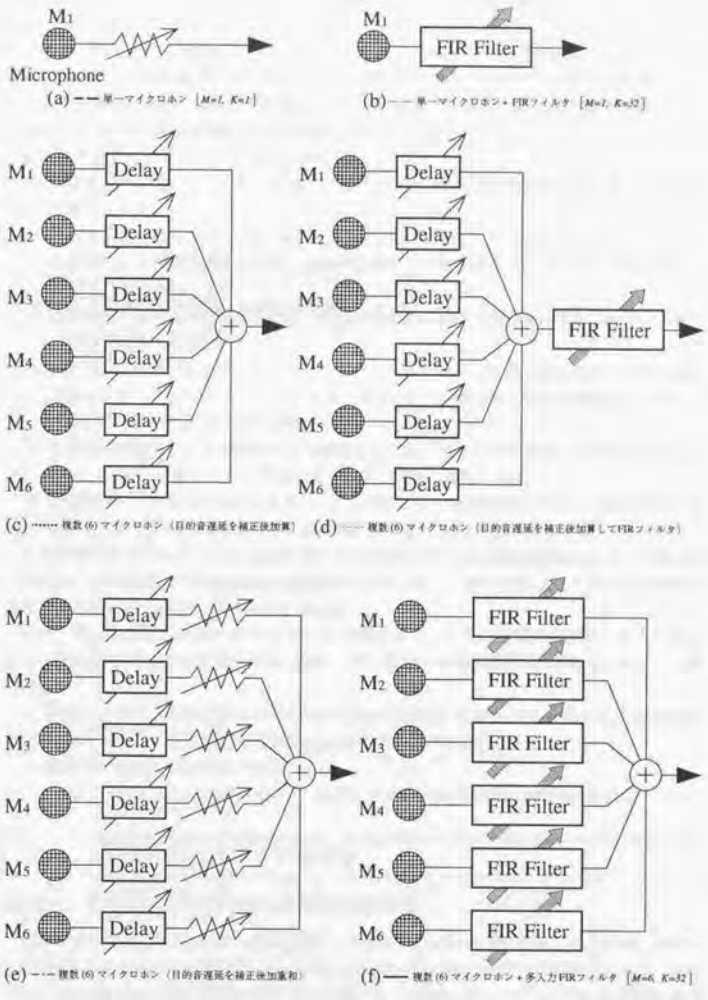


Fig. 11.4 平均２乗誤差を評価した線型フィルタ

これら，6種類の線型フィルタに対する誤差曲線を，目的音のスペクトル別に **Fig. 11.5**，**Fig. 11.6**，**Fig. 11.7** に示す．Fig. 11.5 が 0-11.025kHz で一様なスペクトルを持つ帯域制限された目的音，Fig. 11.6 が 低域にエネルギーの集中している目的音（強度 $\propto 1+\cos(2\times\pi\times$周波数$/F)$），Fig. 11.7 が 0-22.05kHz の白色のスペクトルを持つ目的音の場合である．

以上3つの結果は，音源数 J が マイクロホン数 M より少ない場合である．しかし，**3.1.3** で述べたように，音源数がマイクロホン数より多いと (f) のフィルタの効果は一変するので，このような場合も調べておく必要がある．そこで，Fig. 11.5 の設定のまま音源数のみを 4 から 10 に増加させた誤差曲線を **Fig. 11.8** に示す．

これらのグラフより以下のことがわかる．

- 推定誤差で評価して，一般の FIR フィルタ (f) が，如何なる目的音強度に対しても抜群に優れている．
- 目的音と妨害音のスペクトルの差異が Fig. 11.5 や Fig. 11.6 の程度であれば，スペクトルの差異による信号抽出よりも，音源の方向による音源抽出のほうが推定誤差を減少させる効果が大きい．
- 複数のマイクロホンを使うことで，平均2乗誤差に占める目的音の劣化に起因する成分の割合が小さくなる．
- Fig. 11.7からわかるように，たとえスペクトルの差異による信号抽出が行なえない場合であっても，一般の FIR フィルタを用いる (f) は，音源方向の差異を遅延調節に利用した (e) に比較して推定誤差が格段に小さい．
- 音源方向の差異から遅延調節による抽出をした後スペクトルの差異をも利用する (d) よりも，一般の FIR フィルタを用いる (f) が，格段に優れている．
- 音源数がマイクロホン数よりも多い（$J > M$）という条件のもとでも，一般の FIR フィルタ (f) が（優位性は多少縮まるものの）依然良好な特性を示している．

これらの結果を見れば，Cue Signal 法による時間軸上での視聴覚融合型適応は，音源方向を利用した空間座標上での視聴覚融合型適応に比較して，一般の FIR フィルタの全係数を最適化できることが大きな長所であると言える．

ただ，以上の評価は理想的な Cue Signal が得られて，十分な適応時間が与えられた場合の Cue Signal 法と，音源位置が正確に計測できた場合の空間座標上融合の比較である．実際には，

- 視覚センサで，どの程度 **5.2** 節で述べた条件を満たした Cue Signal が生成できるのか．
- 視覚センサで，どの程度音源位置を正確に測定できるのか．
- 適応の時間は，どの程度とれるのか．

なども含めて評価しなければならない．このような総合的評価は今後の課題である．

11.3 Event Signal のメタ信号化

11.3.1 計測目的に応じた出力形態の必要性

センシングは，必ずなんらかの目的があって行なっているはずである．たとえば，センシング結果をもとに制御を行なうとか，次の意思決定の材料にするなどの動機があって，初めてセンシングを行なう意味があると言えるだろう．センシングシステムは単なる数値を出力するための物ではない．計測のための計測ではありえない．

とすれば，センシング結果が何に使われるかを知らなければ最も適したセンシングシステムを設計することはできないのではないか．

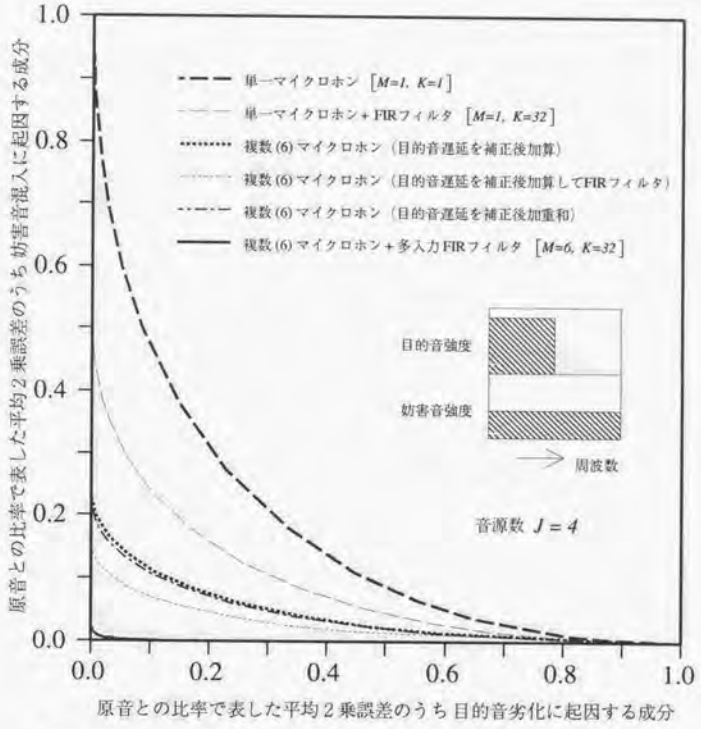


Fig. 11.5 平均2乗誤差の軌跡（目的音：帯域制限信号，妨害音：白色×3）

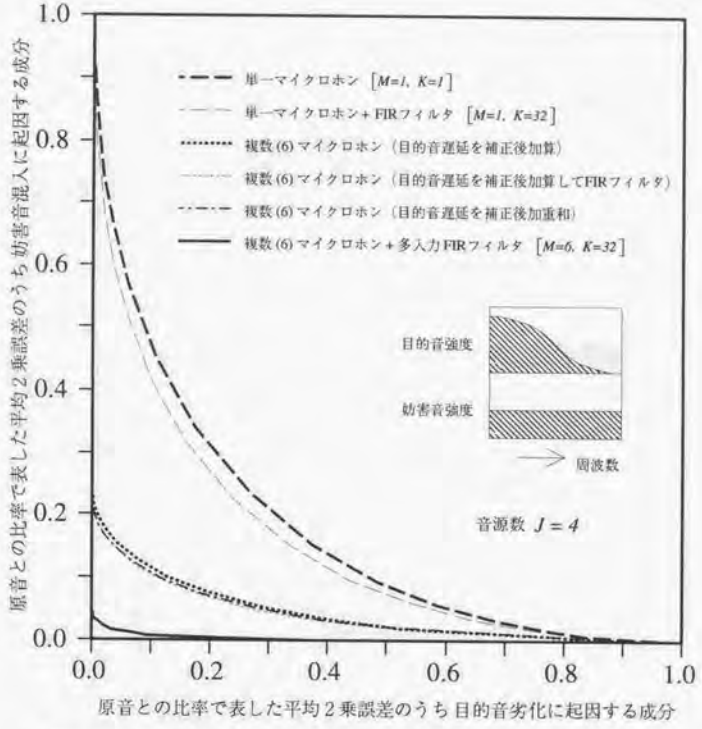


Fig. 11.6 平均2乗誤差の軌跡（目的音：低域強調信号，妨害音：白色×3）

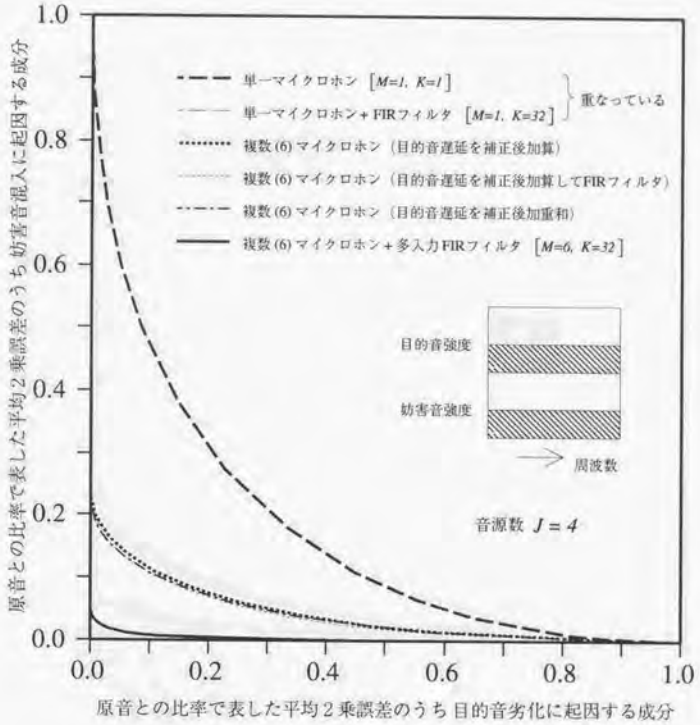


Fig. 11.7 平均2乗誤差の軌跡（目的音：白色，妨害音：白色×3）

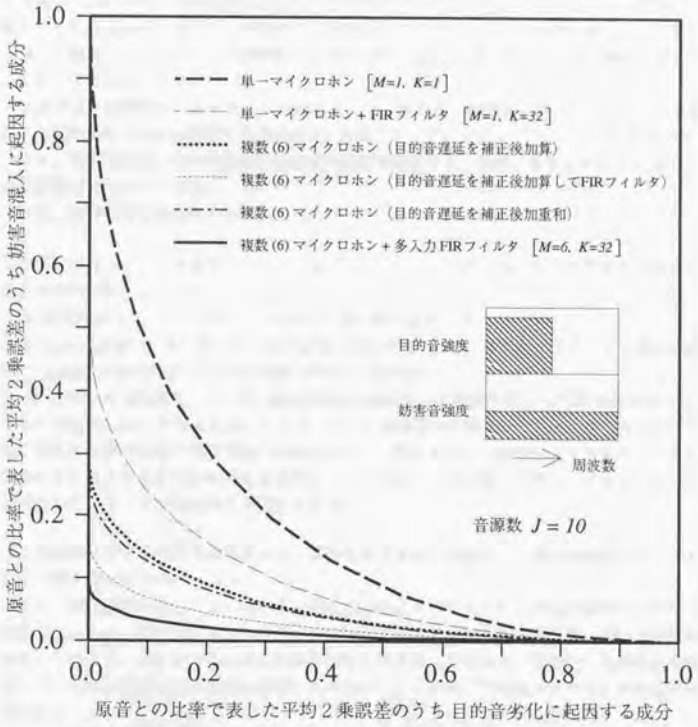


Fig. 11.8 平均2乗誤差の軌跡（目的音：帯域制限信号，妨害音：白色×9）

例えば，この第II部で述べている視聴覚融合システムにおける視覚部の役割は，Cue Signal を出力することである．そしてそのために，Event Signal を計測した．Event Signal は対象物の音強度を計測したものであるが，計測の目的は音強度の点推定値を得ることではなく，あくまでも Cue Signal の生成である．

本論文では，ここまで Event Signal の計測システムを，単純な信号 $\theta(t)$ を出力するもの（各時点 t で，単なるスカラー量である $\theta(t)$ を出力するもの）として考えてきた．しかし，$\theta(t)$ 計測の目的は，$\theta(t)$ の出力そのものにあるのではなく，Cue Signal $\alpha(t)$ の生成にあるはずである．とすれば，単純信号 $\theta(t)$ のみを出力するのは不十分なのではないだろうか．すなわち，Cue Signal 生成という最終的な目的を考えたとき，$\theta(t)$ まで情報を落としてしまうのは，不適当ではないのか．視覚情報には $\theta(t)$ だけでは言い尽くせない有用な情報が含まれているはずである．

本節では，単純信号である Event Signal $\theta(t)$ を，信号群，時間をパラメータとした分布関数，命題論的信号などに拡張する可能性と，拡張したことによるメリットについて考察する．

なお，本論文では，これらの拡張した広い意味での信号を，単純な信号と対比する意味で，**メタ信号**とよぶことにする．

まず，次の簡単な例題を考えてみよう．

【例題】ガスメータ　家庭用ガスメータを考える．今，1ヶ月ごとのガス使用量を計測することの目的が次の2つであるとしよう．

- 使用量に応じて各家庭に1ヶ月分のガス料金を請求する．
- あるガス管にから分岐する100世帯分のガス使用量と，ガス会社がそのガス管に供給したガス量を比較してガスの漏れがないか調べる．

さて，あるガス流量計で1ヶ月間 流量計測をした結果，累積使用量 y の事後確率分布 $f(y)$ として **Fig. 11.9** (a) が得られたとしよう．(たとえば流量計内部でカウントされたあるパルス数に対応する真の流量分布が Fig. 11.9 (a) だったと思えばよい．確率的にミスカウントが起こるのでこのような非対称な誤差分布が生じるとする）このとき，ガスメータは1ヶ月の使用量として，どのような数値を出力すべきか？

この問題に対する解答は複雑である．結論から言えば，流量として単なる数値を出力するだけでは十分ではない．

まず，料金請求を考えよう．例えば，**Fig. 11.9** (b) に示すように，料金の段階に比較して誤差 $y_{\max} - y_{\min}$ が小さかったと仮定する．真の使用量に対応する料金はせいぜい2種類である．このとき，最もよい料金の算出法は明らかである．すなわち，料金枠に真の料金が含まれている確率が高い方の料金を採用すればよい．ここでは，「分布のメジアン」が本質的な量となる．

次に，逆に **Fig. 11.9** (c) に示すように，料金の段階に比較して誤差 $y_{\max} - y_{\min}$ がずっと大きかったと仮定してみよう．このとき，適切な料金の算出方法は評価方法の違いで複数に分かれうる．第一は，請求料金を真の料金の期待値に一致させようとする方針である．このとき，「分布の期待値」が本質的な量にほぼ近い．第二は，間違って請求される額の最大値を最小にしようとする方針である．このとき，「分布の上限と下限の中心 $(y_{\max} + y_{\min})/2$」が本質的な量となろう．第三は，カウントミスによって生じる不確定分は消費者に請求しないとする方針である．このときは，「分布の下限 $y_{\min}$」が本質的な量である．もちろんその他にも，2乗誤差最小など無数の方法が考えられる．しかし，少なくとも (b) のとき最も合理的と思われた「分布のメジアン」が，(c) の場合は何の意味も持たなくなっていることは明らかであ

ろう．

最後にガス管での漏れを調べる場合を考えよう（**Fig. 11.9**(d)）．各家庭のガスメータの誤差に相関がないとすれば，１００世帯ぶんの総使用量の推定値は，各家庭の分布の期待値を合計すれば得られる．このとき，本質的な量は，「分布の期待値」である．また，別の方針もある．ガスの漏れが確実に存在することを主張するために，分布の上限の和をとりたい場合である．このとき，「分布の上限 $y_{\max}$」が本質的な量となる．

この例でわかることは，「センシング結果がどのように使われるかを決定する以前には，測定結果を単なるスカラー量（単純信号）まで落としてしまってはいけない」ということである．特に，事後確率分布の形や大きさが変動するときには，平均やメジアンなどを出力するだけでは依然不十分である．上の例において，各家庭で計測する必要があるのは推定使用量ではない．どんな目的で使われるかが決定するまでは事後確率分布そのものをとっておく必要がある．

すなわち，最終的な計測は，料金計であり，ガス漏れ検出計であるわけで，各仮定のガスメータというのは，それら最終的意思決定へ至る中間的な情報であると言える．このような場合，単純信号（スカラー量）でなくて，メタ信号（複合量）を出力しなければならない．

これは，Cue Signal の中間的な情報である Event Signal の状況とよく似ている．視覚情報からは，スカラー量としての Event Signal を生成する以前に様々な知識が得られている．例えば，対象物が隠れてしまって Event Signal の推定が信頼のおけないものになったということがわかっているかもしれない．または，Event Signal の値の確率分布が得られている場合もあるだろう．これらの情報を全て用いて Cue Signal を生成すべきではないのか．すなわち，Event Signal も メタ信号に拡張すべきではないのか，そしてどのようにして Cue Signal を生成したらよいのか，というのがこの節で試みる議論の主題である．

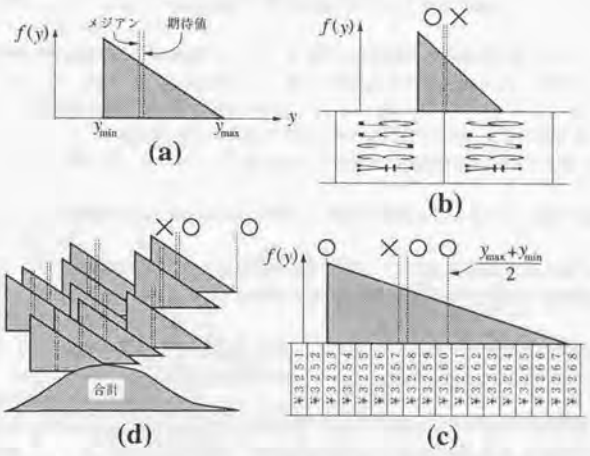


Fig. 11.9　計測目的に応じた出力形態の必要性（例題:ガスメータ）

Table 11.1 Meta Event Signal の具体型

記号	メタ信号の型	メタ信号の具体的表現
(a)	通常の信号	$\theta(t)$
(b)	事後確率分布	$f(t,\theta)$
(c)	分布のパラメータ	例えば，平均値 $\mu(t)$ と分散 $\sigma(t)$ の組
(d)	命題の真偽信号	$\mathrm{True}_p(t)$ $(p=1,2,\ldots,P)$
(e)	測定可能信号の付加	$f(t,\theta)$ と 測定可能信号 $\mathrm{Av}(t)$ の組
(f)	命題の受容信号	$\mathrm{Prop}_p(t)$ $(p=1,2,\ldots,P)$
(g)	確率的加重和	$f_e(t,\theta)$ が 確率 $p_e(t)$ $(e=1,2,\ldots,E)$
(h)	時間的連結	$f(\theta_1,\theta_2,\ldots,\theta_H)$
(i)	要素ごとの時間的連結	$f_e(t,\theta)$ と $f(e_1,e_2,\ldots,e_H)$
(j)	要素ごとの完全連結	$f_e(t,\theta)$ が 確率 $f(e)$ $(e=1,2,\ldots,E)$
(k)	複合型（1）	同一の量を観測した複数の $[\theta]$
(l)	複合型（2）	異なる量を観測した複数の $[\theta]$

11.3.2 Meta Event Signal の具体型

それでは，Event Signal のメタ信号である **Meta Event Signal** について考えてみよう．なお，本論文では，Meta Event Signal を単純な信号である Event Signal $\theta(t)$ と区別するために，$[\theta]$ と表記することにする．

視覚情報や視覚情報，内部知識などすべてを駆使して対象物の音強度を推定することを考える．このとき，Meta Event Signal として考えうる形態をを **Table. 11.1** に列挙してみた．また，各々の形態を **Fig. 11.10** に図示する．

なお，以下では，Event Signal の場合を念頭において記述するが，ここで論じることは，Event Signal に限らず，センシングの出力のメタ信号化にほぼ共通に応用できると考えている．

まず，(a) は通常の Event Signal $\theta(t)$ である．時間の関数としてスカラー量が対応する．単純な信号である．

(b) は，スカラー量のかわりに，それぞれの時点 t について，θ を確率変数とみた分布 $f(t,\theta)$ を出力するものである．観察したという行為の後の推定確率分布であるので，事後確率分布と言ってもよいだろう．

(c) は，(b) の分布の型を事前に仮定しておき，分布のパラメータのみを出力する方法である．たとえば，平均値 $\mu(t)=\mathrm{E}[\theta]$ と分散 $\sigma(t)=\mathrm{E}\left[(\theta-\mathrm{E}[\theta])^2\right]$ の組をもって $[\theta]$ とするなどは典型的な例だろう．

(d) に示したのは命題的出力ともいうべきものである．命題とは，例えば，「真の値 θ が定数 θ_c に対して，$\theta<\theta_c$」などである．このような命題 P 個の真偽を示す P 個の 2 値量を出力する．この出力を，ここでは True_p $(p=1,2,\ldots,P)$ と書くこととする．センシングによってはこのようは命題的出力を取り出せるものは珍しくない．なお，この (d) は上の (a)〜

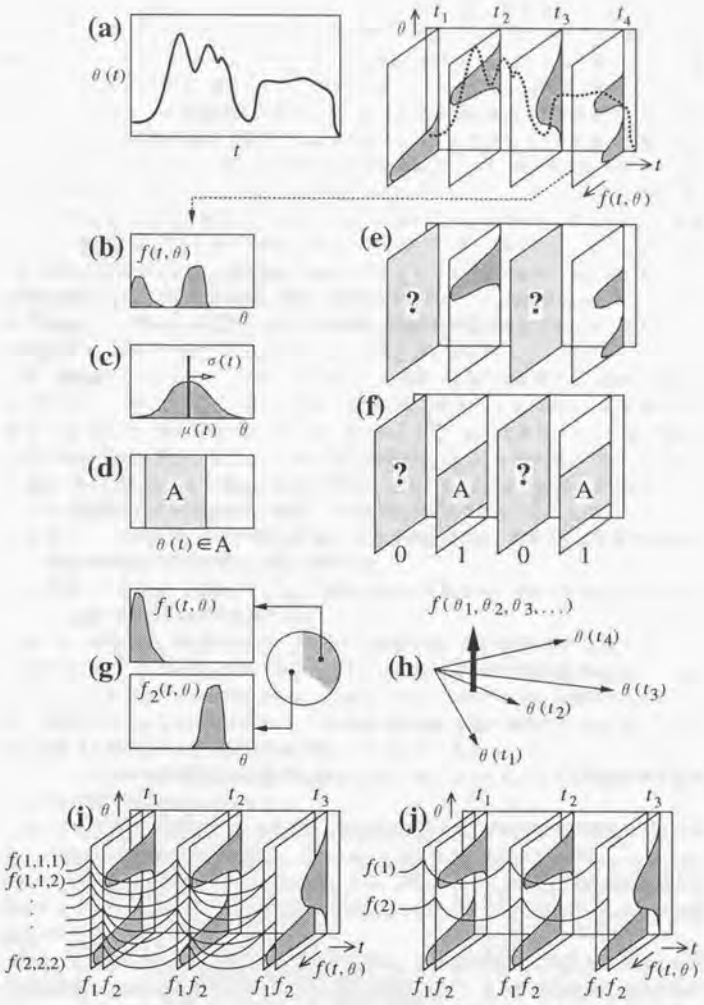


Fig. 11.10 Meta Event Signal の具体的な形態

(c) と組み合わせることもできる.

以上4つが基本型である. これらに対する拡張は, 以下の方法が考えられる. ここでは, 便宜上, (b) に対する拡張として記述しよう. (ただし, (f) は (d) の拡張である).

(e) に示した第一の拡張は, 測定結果として,「測定不可能」という値を許すことである. 具体例で言えば, 対象物が物陰に隠れてしまったり, 話者が後ろを向いてしまった場合には, 視覚センサからの $\theta(t)$ 推定は,「測定不可能」になる. センシング出力にこの状態を許そうというのがこの拡張である. (e) の形式的な表現としては, $f(t,\theta)$ の他に時点 t での測定が可能であったことを示す2値関数 $\mathrm{Av}(t)$ を併記すればよいだろう. (つまり, (b) と (d) の組み合せと見ることもできる). もちろん, 測定不可能とは, 測定結果としての一様分布 $f(t,\theta) = \text{constant}$ とは全く意味が違うし, 対処の仕方も当然異なるものである. これについては, 後で述べる.

(f) の拡張は, (d) を一般化して (e) の考え方を導入したものである. すなわち, 各命題 Prop_p に対し,「真」「偽」を 1,0 で出力するのではなく,「真」「真とも偽とも言えない」を 1,0 で出力するものである. ここでは,「測定不可能」は, すべての命題の真偽が不明という出力で表現される. また, これが, (d) の一般化であるのは, (d) の命題 True_p を, (f) 上では「True_p」と「True_p の否定」の2つの命題に分割すれば表現可能であるからである. 具体的に True_p と Av の組が Prop_p であるという見方をしてもよい.

第二の拡張は, (g) のように, 測定結果をこれら複数の表現の加重和として表わそうとするものである. 例えば,「現在の測定結果は, 確率 70%で分布 $f_1(t,\theta)$ であり, 確率 20%で分布 $f_2(t,\theta)$ であり, 確率 10%で測定不可能であった」などの出力である. すなわち, 要素としての Meta Event Signal $[\theta_e]$ $(e = 1,2,\ldots E)$ を確率 p_e $(\sum p_e = 1)$ で重みづけして, 新しい Meta Event Signal $[\theta]$ を構成する方法である. 念のため, (g) の具体例を示しておく.

- いま観察した対象の種類が, 対象1か対象2かが不明である.
- センシングの結果, この対象物が対象1である確率は 70%, 対象2である確率は 30%, 未知の対象物である確率は 10%と推定した.
- 対象1であれば, θに関して $f_1[\theta]$ の分布であると推定され, 対象2であれば, θに関して $f_2[\theta]$ の分布であると推定される.

第三は, 時間方向への拡張である. まず最も一般的には, ある時間区間の H 個のサンプリングに対して, $t_1,t_2,\ldots,t_H$ における θ の値 $\theta_1,\theta_2,\ldots,\theta_H$ の複合密度関数 $f(\theta_1,\theta_2,\ldots,\theta_H)$ を与えることである. これが (h) である. これは, 考えうる全ての $\theta(t)$ の波形それぞれに1対1の確率を与えることに他ならない. この方法の欠点は, 実際の測定から $f(\theta_1,\theta_2,\ldots,\theta_H)$ という莫大な情報を得るのは現実的ではないということだろう.

そこで, なんらかの便宜的な表現が必要になってくる. もちろん, それは現実世界の表現として妥当なものでなければならない.

(i) に示したのは, 各時点で (g) の各要素 e のどの分布をとるかの確率を他の時点でどの分布をとったかの関係で表わすものである. これには, $t_1,t_2,\ldots,t_H$ における e の値 $e_1,e_2,\ldots,e_H$ の複合密度関数 $f(e_1,e_2,\ldots,e_H)$ を与えればよい. これによって, 時間的な連結の妥当性を表現することができる. なお, この遷移確率が直前の状態だけによる場合は, マルコフ過程的モデルと言えよう.

また, その最も特殊な場合が, (j) である. これは, 各マルコフ系列に相互乗入れがない場合に相当する. 具体的には, 系列 $f_1(t,\theta), f_2(t,\theta), f_3(t,\theta),\ldots$ を確率的な重みで重ね合わせたものとして表現できる. このモデルは, さきほど (g) の説明で箇条書であげた具体例に対する最も自然な拡張になっている. それは, いま見ている対象物が対象1か対象2であるかが不明である場合でも, もしある時点で対象1であると仮定すれば, 次の時点で同じ場所にあるものは対象1であると断言してよいからである.

最後の第四の拡張である (k) と (l) は，**Fig. 11.11** のように，複数の Meta Event Signal を束ねたものである．(k) が同一量に対する複数の観測結果を束ねたものであるのに対し，(l) は複数の別個の量を測定し束ねたものである．

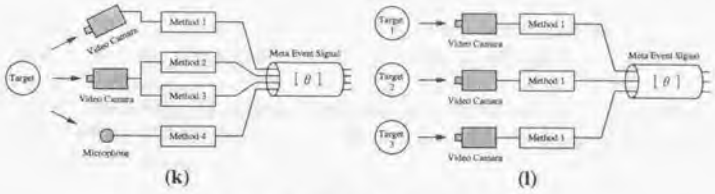


Fig. 11.11 複合型の Meta Event Signal

(k) は，異なるセンサで得られたり，異なる画像処理アルゴリズムで得られた複数の観測結果から構成される Meta Event Signal である．(g) と (k) の違いは，(g) が各要素 e のどれかひとつが有効な情報であるのに対し，(k) は束ねられた要素のすべてが有効な情報であるという点である．

(l) は，例えば音源ごとの強度エンベロープの推定値を束ねたものである．**5.2.6** で述べたように，妨害音源をも見たほうがより良い適応が可能である．

以上，Table. 11.1 にそって述べてきた．これらすべての表現形態をまとめて Meta Event Signal $[\theta]$ とよぶ．

Meta Event Signal の具体型を整理したので，次は Meta Event Signal の型別に Cue Signal への変換方法を論じよう．**11.3.3** では，分布型の Meta Event Signal に対する議論を行ない，**11.3.4** では，命題型の Meta Event Signal に対する議論を行なう

11.3.3 分布型 Meta Event Signal から Cue Signal への変換

ここでは，分布型の Meta Event Signal から，Cue Signal を生成する方法を考えることで，Meta Event Signal として必要十分な情報とは何かを考える．なお，分布型 Meta Event Signal とは，Table. 11.1 の (b),(c) の基本型とそれに対する (g)～(l) の拡張をさすものとする．

11.3.3.1 準備

第 I 部の **5.2.4**（p.33）で述べたように，Cue Signal α を使ったときの音源 S_j の抽出度合（正確には，適応時のゲインの目標値）は，$K[\alpha, a_j]$（定義は式(5.31)）によって決定される．ここでは，ある十分長い時間ブロックでの I 個のサンプリングデータで話を進めることとしよう．また，時間のインデックスを i で表わすことにする．まず，式(5.31) を離散時間表現に書き直す．

$$K[\alpha, a_j] = \frac{\sum_{i=1}^{I} \alpha_i {a_j}_i^2}{\sum_{i=1}^{I} {a_j}_i^2} \tag{11.3}$$

上式が，目的音（すなわち $j = 0$）についてはできるだけ大きくなり，妨害音（すなわち $j \neq 0$）についてはできるだけ小さくなるような Cue Signal α_i $(i = 1, 2, \ldots I)$ を与えられた $[\theta]$ から求めるのが，ここでの問題である．

さて，式(11.3)の分母は音源固有の a_{ji} のみの関数で，計測には依存しない．つまり，分母は，どんな Cue Signal を使おうとも一定である．そこで，以下の検討では，分子

$$K_j = \sum_i \alpha_i {a_{ji}}^2 \tag{11.4}$$

のみを考える．今，a_{ji} を固定して考えれば，これは，音源 S_j に対する適応後のシステム出力振幅に比例する量（正確には目標値）である．たとえば，K_0 が目的音出力の大きさ（の目標値）の尺度となる．

ここで，方針を明確にする．目的は，妨害音のなかから，目的音を抽出することである．それゆえ，1目的音1妨害音であれば，次式の値 D_1 を最大にする Cue Signal α_i を選択するのが妥当であると考えられる．

$$D_1 \stackrel{\text{def}}{=} \frac{K_0}{K_1} \tag{11.5}$$

または，妨害音に対しての K_j を固定して，K_0 を最大化してもよい．この評価関数を D_0 で表わす．

$$D_0 \stackrel{\text{def}}{=} K_0\big|_{K_1,K_2,\ldots} \tag{11.6}$$

D_1 も D_0 も，${a_{ji}}^2$ と α_i のみの関数である．もちろん，ここでは ${a_{ji}}^2$ の真の値は未知である．${a_{ji}}^2$ についての情報は，$[\theta]_i$ から知るしかない．つまり，式(11.5)や式(11.6)をすぐ解いて α_i を得ることはできないが，$[\theta]$ で表現された式(11.5)や式(11.6)の期待値を最大にする α_i を採用すればよいであろう．（なお，期待値をとるという必然性は，実はない．ここに既に任意性が入っている）．

それでは，以上の準備のもとに，種々の分布型 Meta Event Signal $[\theta]$ に対する Cue Signal の生成法を考えてみる．**11.3.3.2** が，Table. 11.1 の (b),(c),(g)～(j) の場合で，**11.3.3.3** が，それに (k) の拡張を加えた場合，**11.3.3.3** が，それに (l) の拡張を加えた場合である．

11.3.3.2　目的音の単一観測

単一の Meta Event Signal $[\theta]$ から Cue Signal α を求める問題である．なお，ここでは，$[\theta]$ には，妨害音のエンベロープ a_{ji} に関する情報が全く含まれていない場合を考える．

$$\begin{aligned}
\mathrm{E}[D_0] &= \mathrm{E}\left[\sum_i \alpha_i {a_{0i}}^2\Big|_{K_1,K_2,\ldots}\right] && (11.7)\\
&= \sum_i \alpha_i \mathrm{E}\left[{a_{0i}}^2\right]\Big|_{K_1,K_2,\ldots} && (11.8)\\
&= \sum_i \alpha_i \mathrm{E}\left[[\theta]_i\right]\Big|_{K_1,K_2,\ldots} && (11.9)\\
&= \sum_i \alpha_i \int \theta f(i,\theta) d\theta \Big|_{K_1,K_2,\ldots} && (11.10)
\end{aligned}$$

ここで，式(11.7)から式(11.8)の変形では，$[\theta]$ が妨害音についての情報を含まないこと，すなわち，確率変数 ${a_{0i}}^2$ と 確率変数 ${a_{ji}}^2$ $(j=1,2,\ldots)$（これは拘束条件 $K_1, K_2, \ldots$ を介して α_i を動的に制限している）が独立であることを用いた．

この項目での，結論はこうである．すなわち，単一の Meta Event Signal から Cue Signal を生成するのであれば，式(11.8)が示すように，単に ${a_{0i}}^2$ の期待値を求めればよい．そして，この確率変数 ${a_{0i}}^2$ の挙動は $[\theta]$ が教えるので，結局は，$[\theta]$ の期待値をとればそれが全てである．例えば，Table. 11.1 の (b) の場合は，各時点での分布の平均値が必要な全ての情

報を持っている．(c) の場合は，得られているパラメータに平均値（期待値）があれば，それを採用して，あとは捨てればよい．なければ，手持ちの情報から平均値を推定する．また，(b) をさらに一般化した (g)～(j) の場合でさえ，各時点での分布の平均値だけで十分である．これは，式(11.10) という最小化問題を解くためには，$f(i,\theta)$ すべてを伝達する必要はなく，$\int \theta f(i,\theta) d\theta$ のみを伝達すればよいからである．

以下に典型的な場合を具体的に述べる．

【1】妨害音のエンベロープ a_{ji} について，全く知識がない場合．

唯一使える先験的知識は，${a_{ji}}^2$ が非負の数であることである．すなわち，${a_{ji}}^2$ をフーリエ変換したときに周波数 0 の成分は必ず存在する．しかし，他の周波数成分に関しては位相が未知であるので，K_j を求めるために α_i と内積をとって期待値をとれば 0 になってしまう．そういうわけで，具体的には，以下のような，拘束条件のもと D_0 を最大化する α_i を求めればよい．

$$\frac{1}{I}\sum_i \alpha_i = 0 \tag{11.11}$$

$$\frac{1}{I}\sum_i {\alpha_i}^2 = 1 \tag{11.12}$$

ここで，式(11.12) は係数を正規化するためのものである．この問題は，ラグランジェの未定定数法を利用して，

$$D_L = \mathrm{E}\left[\sum \alpha_i {a_{0i}}^2 + \lambda(I - \sum {\alpha_i}^2) + \mu(\sum \alpha_i)\right] \tag{11.13}$$

$$= \sum \alpha_i \mathrm{E}\left[{a_{0i}}^2\right] + \lambda(I - \sum {\alpha_i}^2) + \mu(\sum \alpha_i) \tag{11.14}$$

から，求まる．計算の結果，Cue Signal は，

$$\alpha_i = \frac{1}{\sqrt{\frac{\sum (\mathrm{E}[{a_{0i}}^2] - \mathrm{E}[a_{0i}^2])^2}{I}}} (\mathrm{E}\left[{a_{0i}}^2\right] - \overline{\mathrm{E}\left[{a_{0i}}^2\right]}) \tag{11.15}$$

となる．定数倍を省略して書けば，

$$\alpha_i = \mathrm{E}\left[{a_{0i}}^2\right] - \overline{\mathrm{E}\left[{a_{0i}}^2\right]} \tag{11.16}$$

である．これが，定理 5.1の条件を満たす具体型である．

どのような $[\theta]$ であろうとも，期待値をつないでバイアス分のみ全体をシフトすれば最良の Cue Signal が得られることが示された．

【2】妨害音のエンベロープ a_{ji} について，完全に知識がある場合．

もし，妨害音のエンベロープが一定であれば，式(11.11) と式(11.12) を拘束条件にすれば良いので，【1】と全く同じである．そうでない場合は，別に計算しなければならない．ここでは，妨害音が 1 個の場合を考える．妨害音のエンベロープ波形は，a_{1i} であるから，拘束条件は，式(11.11) の代わりに，

$$\frac{1}{I}\sum_i \alpha_i a_{1i} = 0 \tag{11.17}$$

となる．これを同じく未定定数法で解けば，係数の定数倍を省略して，

$$\alpha_i = \mathrm{E}\left[{a_{0i}}^2\right] - \frac{\overline{a_{1i}\,\mathrm{E}\left[{a_{0i}}^2\right]}}{\overline{{a_{1i}}^2}} a_{1i} \tag{11.18}$$

である．これは，**5.2.6** で述べた Cue Signal の直交化に他ならない．いずれにしても，$\mathrm{E}\left[{a_{0i}}^2\right]$ だけしか必要ないというのが，この **11.3.3.2** の「目的音の単一観測」の場合の答である．

11.3.3.3 目的音の複数観測

次に，異なるセンサで得られたり，異なる画像処理アルゴリズムで得られた複数の観測結果から構成される Meta Event Signal の場合を考えてみよう．これは，Table. 11.1 の分類で言えば (k) に場合に相当する．

式(11.7) から式(11.9) までの変形はこの場合も有効である．結局，式(11.9) の $\mathrm{E}\left[{a_{0i}}^2\right]$ を複数の Meta Event Signal から生成すればよい．

このとき，はじめて Meta Event Signal として扱う意味がでてくる．例えば，Table. 11.1 の (b) の場合で言えば，各分布 $f(i,\theta)$ の期待値をとって平均するのではなく，分布どうしを融合してから期待値をとらなければならない．

2つの $f(i,\theta)$ が独立な測定であるときには，分布どうしの融合にはベイズ的手法が有効である．すなわち，片方の $f(i,\theta)$ をベイズ法でいうところの事前分布とみて，もう片方の $f(i,\theta)$ を尤度関数とみれば，事後分布は，両者の積によって表わされる．

11.3.3.4 目的音と妨害音の観測

今度は，目的音に関する Meta Event Signal $[\theta]_{0i}$ だけでなく，妨害音に関する Meta Event Signal $[\theta]_{ji}$ が得られている場合を考える．すなわち，Table. 11.1 の (l) の場合である．

ここでは，妨害音が1個の場合を考える（$j=1$）．以下では，$[\theta]_{0i}$ と $[\theta]_{1i}$ が独立な測定である場合と相関のある測定の場合に分けて考える

$[\theta]_{0i}$ と $[\theta]_{1i}$ が独立な測定である場合　D_1 で評価する．期待値をとる分数の分母と分子が独立な確率変数であると見ることができるので，

$$\mathrm{E}[D_1] = \mathrm{E}\left[\frac{\sum_i \alpha_i {a_{0i}}^2}{\sum_i \alpha_i {a_{1i}}^2}\right] \tag{11.19}$$

$$= \mathrm{E}\left[\frac{1}{\sum_i \alpha_i {a_{1i}}^2}\right] \cdot \sum_i \alpha_i \mathrm{E}\left[{a_{0i}}^2\right] \tag{11.20}$$

これは，既に簡単ではない．目的音に関する $[\theta]_{0i}$ については期待値がすべてを表わしているが，妨害音に関する $[\theta]_{1i}$ に対しては，妨害音1個の場合ですら期待値だけでは不十分であることがわかる．

$[\theta]_{0i}$ と $[\theta]_{1i}$ が独立な測定でない場合　同じく，D_1 で評価する．しかし，今度は以下のようにしか変形できない．

$$\mathrm{E}[D_1] = \mathrm{E}\left[\frac{\sum_i \alpha_i {a_{0i}}^2}{\sum_i \alpha_i {a_{1i}}^2}\right] \tag{11.21}$$

$$= \iint\cdots\int \frac{\sum_i \alpha_i \theta_{0i}}{\sum_i \alpha_i \theta_{1i}} f(\theta_{01},\ldots,\theta_{0I},\theta_{11},\ldots,\theta_{1I})\, d\theta_{01} d\theta_{02}\ldots d\theta_{1I} \tag{11.22}$$

すなわち，目的音，妨害音ともに期待値をとるだけでは不十分で，もしすべての時間とすべての音源に関する複合密度関数が得らているのならば，式(11.22) から直接 Cue Signal を算出するのが最良である．Meta Event Signal を単純 Event Signal へ簡略化することは許されない．

11.3.3.5 分布型の場合のまとめ

- 単一観測の場合 → 各時点で分布の期待値をとれば必要十分
- 複数の独立な観測の場合 → 分布の積をとってから期待値をとる（ベイズ的融合）
- 妨害音も観測した場合 → 分布そのものが意味をもつ

もちろん，この結果は式(11.5)や式(11.6)という評価（目的音と妨害音の強度比の期待値）をしたことに依存している．もし，目的音と妨害音の強度比の分散まで問題にするのであれば，このまとめのかぎりではない．

11.3.4 命題型 Meta Event Signal から Cue Signal への変換

命題型 Meta Event Signal とは，Table. 11.1 の (d) の基本型とそれに対する拡張である．

11.3.3 で議論した分布型メタ信号の場合は，2つのメタ信号を融合するのに有効な手法はベイズ的方法であった．これに対して，命題型メタ信号の場合は，ベイズ理論に代わって Shafer–Dempster の理論が有効である．ここでは，Shafer–Dempster の理論をそのまま紹介することはせず，その趣旨を活かしつつ，本論文用に再構築した理論を述べる．

11.3.4.1 Shafer–Dempster 理論

ベイズ的融合に対する別の融合として Shafer–Dempster の理論（evidential logic）が知られている．これは，元々，Dempster が数学理論として発表した多値写像における上限確率と下限確率の論文[62]を，Shafer が発展させて "A Mathematical Theory of Evidence" という本[63]にまとめたものである．この理論は，Shafer も自身も言っているように，ベイズ理論を特殊な場合として含むので，ベイズ理論の一般化と見ることも可能である．

Shafer–Dempster の手法を導入するには，各命題を集合論的に扱わなければならない．ここでは，最も直接的に問題を集合論的に書き換えることとする．最も手っとり早い方法は，θ のとりうる値の集合 Θ を定義することである．このようにすれば，各命題を Θ の部分集合として表わすことができる．例えば，「真の値 θ が定数 θ_c に対して，$\theta < \theta_c$」という命題は，

$$A = \{\theta \mid \theta < \theta_c,\ \theta \in \Theta\} \tag{11.23}$$

という集合 A で表現できる．また，測定不能で全く何の情報も得られなかったことは，集合 Θ で表現できる．

Shafer は，まず**基本確率分配**（basic probability assignment） m を定義した．m は基本命題をその基本命題が言える確率に写像するものである．すなわち，m は Θ の部分集合から，$[0,1]$ への写像である．また，m のすべての和は 1 とする．

次に Shafer は，**Belief 関数**（Belief function）を以下のように定義した．

$$\mathrm{Bel}(A) \stackrel{\mathrm{def}}{=} \sum_{B \subset A} \mathrm{m}(B) \tag{11.24}$$

これは，「命題 A の言える確率は，その命題の十分条件になっている基本命題の確率の和である」とするもので，Belief という意味からも合理的であろう．

Shafer が，**Dempster の融合則**（Dempster's rule of combination）と言っているのは，2つの Belief 関数を，ひとつの Belif 関数に結合する方法である．これは，両者の基本確率分配 $\mathrm{m}_1, \mathrm{m}_2$ から，新しい基本確率分配 m を以下のように計算することで行なわれる．

$$\mathrm{m}(A) = \frac{\sum_{\{i,j \mid A_i \cap B_j = A\}} \mathrm{m}_1(A_i)\,\mathrm{m}_2(B_j)}{1 - \sum_{\{i,j \mid A_i \cap B_j = \emptyset\}} \mathrm{m}_1(A_i)\,\mathrm{m}_2(B_j)} \tag{11.25}$$

ここで，上式の分母は $\sum_{A\subset\Theta} \mathrm{m}(A) = 1$ とするための正規化係数と考えてよい．

さて，belief 関数の定義や，この融合則からもわかるように本質的なのは Belif 関数ではなくて基本確率分配 m のほうである．そこで，本論文では基本確率分配によって記述する．

基本確率分配を用いると，Table. 11.1 や Fig. 11.10 に示した種々の Meta Event Signal のうちの 3 つの型である (d),(f),(g) は，すべて自然な形で記述できる．それを，**Fig. 11.12** (d),(f),(g) に示す．

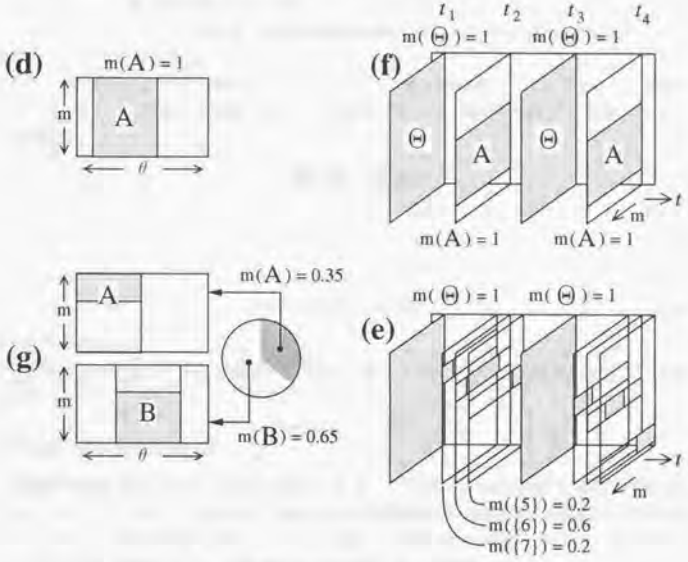


Fig. 11.12 基本確率分配 m による Meta Event Signal の表現

Dempster の融合則は，2 つの独立な観察による命題（直交した命題）を融合するときに用いることができる．

11.3.4.2 分布型 Meta Event Signal の命題型への変換

実際問題として，ここまで述べた $\theta < \theta_c$ のような命題型 Meta Event Signal のみから Cue Signal を生成するのは困難である．なぜならば，このような命題型情報は定量性に乏しいからである．それゆえ，現実には，分布型と命題型を併記した 複合型 Meta Event Signal を扱うこととなるだろう．

このとき，分布型と命題型の両方を統一して扱うことができなければならない．そこで，分布型 Meta Event Signal を命題型 Meta Event Signal に変換する方法を論じよう．

まず，簡単のため θ が離散値をとると仮定する．実際のディジタル信号処理では離散値を扱うのであるから，これで問題ないであろうし，離散化の刻みを一様に小さくすれば連続量

も近似できるであろう．

分布型から命題型への変換は，

$$\mathrm{m}(\{\theta\}) = f(\theta) \qquad \theta \in \Theta \tag{11.26}$$

$$\mathrm{m}(\text{それ以外}) = 0 \tag{11.27}$$

によって，分布型の表現である度数関数 $f(\theta)$ を，命題型の表現である基本確率分配 m に変換するのが妥当であろう．度数関数の定義より $\sum_{\theta\in\Theta} f(\theta) = 1$ であるので $\sum_{\theta\in\Theta} \mathrm{m}(\{\theta\}) = 1$ となり基本確率分配の条件はもちろん満たされている．

このようにしたことで，分布型，命題型の確率的加重和による Meta Event Signal を m で表現することができるようになる．

たとえば，$\Theta = \{1,2,3\}$ の場合でそれぞれの 3 つの値の確率が 2 : 3 : 5 であることが測定されたが，この測定が測定不可能であった（出鱈目であった）確率が 50%である場合の m による表現は，

$$\mathrm{m}(\{1\}) = 0.1 \tag{11.28}$$

$$\mathrm{m}(\{2\}) = 0.15 \tag{11.29}$$

$$\mathrm{m}(\{3\}) = 0.25 \tag{11.30}$$

$$\mathrm{m}(\{1,2,3\}) = 0.5 \tag{11.31}$$

$$\mathrm{m}(\text{それ以外}) = 0 \tag{11.32}$$

などとすればよい．

これで，Table. 11.1 の (e) も記述することができるようになった．それを，**Fig. 11.12** (e) に示す．

11.3.4.3 目的音の複数観測

複数の命題型 Meta Event Signal を融合して単一の Meta Event Signal に変換するにはどうしたらよいだろうか．これには，Dempster の融合則をそのまま用いれば良さそうである．

ところが，これはうまくいかない．これを **Fig. 11.13** に示す．図示したのは，式(11.28)〜式(11.32) という観察結果 1（50%の測定不可能性を含んだ観察）と，

$$\mathrm{m}(\{1\}) = 0.1 \tag{11.33}$$

$$\mathrm{m}(\{2\}) = 0.35 \tag{11.34}$$

$$\mathrm{m}(\{3\}) = 0.55 \tag{11.35}$$

$$\mathrm{m}(\text{それ以外}) = 0 \tag{11.36}$$

という観察結果 2（測定不可能性を含まない観察）を，Dempster の融合則で結合したものである．マス目の面積が式(11.25) の分子で総和演算する個々の確率の大きさを示している．観察 1 での測定不可能性が 50%であったにもかかわらず，それとは独立な観察 2 を組み合わせただけで観察 1 での測定不可能性が 71%にまで上昇してしまっていることがわかる．しかも，θ の離散化の刻みを細かくしていけば，この確率は，限りなく 100%に近づいてしまう．

そこで，本論文では，Dempster の融合則（式(11.25)）を次のように修正して用いることとを提案する．

$$\mathrm{m}(A) = \frac{\sum_{\{i,j|A_i\cap B_j=A\}} \frac{\mathfrak{n}(A)}{\mathfrak{n}(A_i)\,\mathfrak{n}(B_j)}\,\mathrm{m}_1(A_i)\,\mathrm{m}_2(B_j)}{\sum_{\{i,j\}} \frac{\mathfrak{n}(A_i\cap B_j)}{\mathfrak{n}(A_i)\,\mathfrak{n}(B_j)}\,\mathrm{m}_1(A_i)\,\mathrm{m}_2(B_j)} \tag{11.37}$$

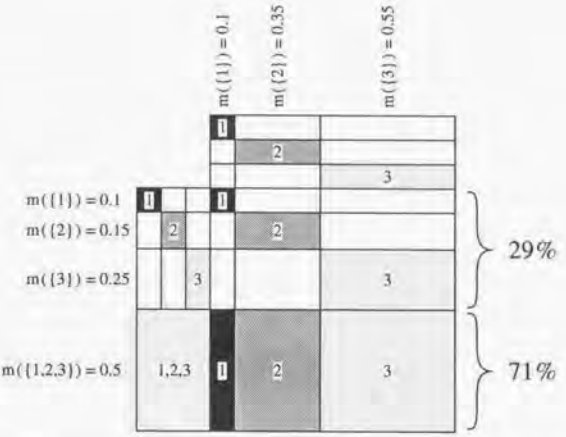


Fig. 11.13 Dempster の融合則を分布型と命題型の混合情報に適用したときの問題点

ここで，$\mathrm{n}(A)$ は集合 A の要素の個数を表わすものとする．連続系では，$\int_{\theta\in A} d\theta$ の意味である．また，上式の分母は $\sum_{A\subset\Theta} \mathrm{m}(A) = 1$ とするための正規化係数で，深い意味はない．

さて，行なわれた修正は，分布型や分布型の確率的加重和のみを考えている場合には，式 (11.25) と同じである．違うのは，複数の要素に対する m が絡んだ（すなわち典型的な命題型の）場合に，分布型の場合と対等になるように補正係数 $\mathrm{n}(A)/\mathrm{n}(A_i)\,\mathrm{n}(B_j)$ をかけたことである．

あるいは，次の説明のほうが素直かもしれない．命題型の基本確率分配 $m(\{1,2,\ldots,K\}) = p \neq 0$ を，融合のときだけ $m(\{1\}) = m(\{2\}) = \cdots = m(\{K\}) = p/K$ と仮の一様分布にバラし（バラすので確率は要素の数 K で割り算する），融合後に各要素を再連結すると見るのである．こう見れば，補正係数 $\mathrm{n}(A)/\mathrm{n}(A_i)\,\mathrm{n}(B_j)$ の分母が，バラしたときの確率の変換に相当し，分子が再連結したときの確率の変換に相当することがわかるだろう．これは，式 (11.37) を，以下のように書き換えれば，より明白である．

$$\frac{\mathrm{m}(A)}{\mathrm{n}(A)} \propto \sum_{\{i,j|A_i\cap B_j=A\}} \frac{\mathrm{m}_1(A_i)}{\mathrm{n}(A_i)} \frac{\mathrm{m}_2(B_j)}{\mathrm{n}(B_j)} \tag{11.38}$$

または，さきほどの「計測不可能」の例で言えば，「計測不可能」を「定量性を有しないという印をつけた一様分布」と解釈して融合するのが式(11.37) の方針であると言ってもよい．

Fig. 11.13 の場合を，本論文で主張する方法で融合すれば **Fig. 11.14** のようになる．

11.3.4.4 Cue Signal への変換

こうして得られた最終的な命題型 Meta Event Signal を Cue Signal にどう変換したらよいか．

まず，簡単な方から考える．すなわち，多数の Meta Event Signal を融合した結果，最終

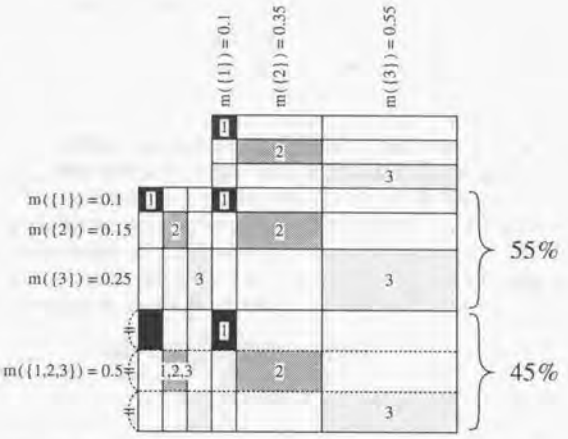


Fig. 11.14 本論文の方法で分布型と命題型の混合情報を融合した場合

的に「$\mathrm{m}(A) \neq 0$ ならば，A の要素は単一の θ」という Meta Event Signal になった場合である．これは，各 θ それぞれの確率が記述されているにすぎない．すなわち，命題型情報が分布型情報に還元されたことになる．この場合には，**11.3.3** で述べたそれぞれの方法で Cue Signal に変換すればよい．

問題なのは，「複数の θ から成るある A に対して $\mathrm{m}(A) \neq 0$」という場合である．このとき，この A の要素間での相対的確率は定量化されていない．定量化されていないので，期待値などの評価をすることは不可能である．最も簡単で確実な解決方法は，そのような時間区間は適応の対象からはずしてしまうことである．これについては，次の **11.3.5** で少し触れる．あるいは，便宜的に，そのような A 内の θ は等確率であるとしてしまう方法も考えられる．しかし，$\theta < \theta_0$ などの命題を考えてみるまでもなく，この便宜的方法が多くの問題を含んでいることは明かだろう．

なお，このような困った事態に陥るのは，ある時間区間ですべての観測結果が非分布型の Meta Event Signal になってしまった場合のみである．これは，むしろ特殊な状況と言えよう．

11.3.4.5　命題型の場合のまとめ

- 基本確率分配 m で表現し，処理する．
- 分布型も命題型に変換する．
- 融合には，式(11.37) を用いる．
- 最終的には分布型に還元できるのが普通なので，分布型（**11.3.3**）で示した方法で Cue Signal に変換すればよい．

11.3.5 Cue Signal の拡張

よく考えてみれば，**11.3.3.2**での結論（分布がわかっているときは，その期待値のみが必要な情報である）は，適切とは思えない．それは，強度推定の分散が大きくても小さくても同じ重さで扱うという点である．

こうなってしまった原因は Cue Signal を単純な信号とした点にあるのではないか．つまり，ここまでは，最適な Cue Signal を作るという目的で Event Signal のメタ信号化をはかってきたが，実は究極の目的は Cue Signal 作成ではなく線型フィルタの適応なのである．

そこで，Cue Signal も拡張すべきではないかと考えるのは当然であろう．

考えうる最も簡単な方法は，適応の ON/OFF 信号を用意することである．あるいは，少し発展させて適応の ON/OFF 信号を連続量にすることも考えられる．ここでは，そのような信号を **Enable Signal** $\kappa(t)$ と表記することにしよう．これは，適応の規範を決めた 2 乗平均誤差（式(3.23)）を以下のように変更することに相当する．

$$\langle \epsilon(t)^2 \rangle = \langle \kappa(t)^2 [d(t) - \phi(t)]^2 \rangle \tag{11.39}$$

$$= \left\langle \left[\kappa(t)\, d(t) - \sum_{n=1}^{N} f_n\, \kappa(t)\, y_n(t) \right]^2 \right\rangle \tag{11.40}$$

まず，各適応アルゴリズムごとにこのようなことが簡単に実現できることを確認しておく．一般的には，式(11.40) をみればわかるように，フィルタ係数更新に関してのみ内部目標 $d(t)$ とタップ信号 $y_n(t)$ を $\kappa(t)$ 倍すればよい．特に LMS アルゴリズムの場合は，修正係数 μ を $\kappa(t)^2$ 倍するだけですむ．また，直接法の場合は，$\boldsymbol{p}$ と R の計算のための積和時に各要素を $\kappa(t)^2$ 倍して足しこんでいけばよい．

次に，この Enable Signal を使うことで適応能力を向上させることができるのかどうかを考えよう．

簡単のため，2 値化 Enable Signal の場合で考える．適応を 2 値化 Enable Signal で ON/OFF することの長所と短所は以下の事項である．

長所 1　都合のよい時間区間のみを採用できるので，目的音と Cue Signal の相関を大きくすることができる．つまり，$K[\alpha, a_0]$ を大きくできる．

短所 1　時間平均の長さが短くなるので，妨害音と Cue Signal の相関を消すのに十分でなくなってくる．すなわち，$K[\alpha, a_1]$ が大きくなってしまう．

短所 2　適応のための統計量をとる時間が短くなってしまう．

Enable Signal 採用の是非はこれらの要因のトレードオフの問題と言える．このうち，長所 1 と短所 1 のトレードオフは Meta Event Signal より定量的に推定できる．しかし，短所 2 については，信号波形に依存してしまう．例えば，白色雑音の場合は 1 秒以下の短い時間区間でも十分学習できるが，音声などのように長い時間観測しなければ統計的性質のつかめないものに対しては少なくとも数秒の計測が必要であろう．

というわけで，このトレードオフの問題にを厳密に論じるためには，多くの実験を積み重ねる必要があり，現状では困難である．

そこで本論文では，厳密な検討は今後に譲ることとし，問題を単純化して考えておくことにする．

Fig. 11.15 に，検討用に単純化したモデルを示す．

これは，ある時間区間を H 個の小領域に分割して考えてみようというものである．まず，要素数 H の集合 B を $i \in B$ が，ちょうど各小領域の中央の時点になるように定義しておく．

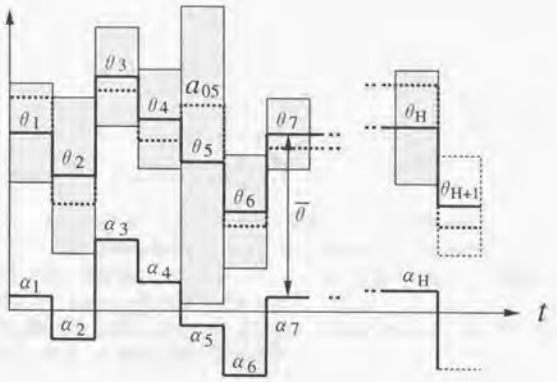


Fig. 11.15 Enable Signal の検討用モデル

方針は，各小領域で目的音強度 $a_{0i}{}^2$ と妨害音強度 $a_{1i}{}^2$ を確率変数として見て，さきほどの長所 1 と短所 1 のトレードオフを定量的に調べることとする．短所が現れるのは，妨害音が強度変動をするときであるので，ここで考えるのは非定常な妨害音ということになる．また，ここでは，次のことを仮定する．

- 小領域内での目的音強度 $a_{0i}{}^2$ は一定値であるとする．妨害音強度 $a_{1i}{}^2$ についても同じ．
- 小領域間での目的音強度 $a_{0i}{}^2$ の変動分は無相関であるとする．妨害音強度 $a_{1i}{}^2$ についても同じ．
- 短所 2 が問題にならない程度の十分長い時間区間が利用できるものとする．

まず，はじめに長所 1 を，具体的に見積もってみれば以下のようになる．

$$\mathrm{E}[K[\alpha, a_0]] = \mathrm{E}\left[\frac{\sum_{i\in B}\alpha_i {a_{0i}}^2}{\sum_{i\in B}{a_{0i}}^2}\right] \tag{11.41}$$

$$= \mathrm{E}\left[\frac{\sum_{i\in B}(\theta_i-\overline{\theta}){a_{0i}}^2}{\sum_{i\in B}{a_{0i}}^2}\right] \tag{11.42}$$

$$\simeq \frac{\sum_{i\in B}(\theta_i-\overline{\theta})^2}{\sum_{i\in B}\theta_i} \tag{11.43}$$

ここで，Cue Signal は，Meta Event Signal の期待値を用いている．また，$\overline{\theta}=\sum_{i\in B}\theta_i/H$ である．なお，式(11.42) から式(11.43) の変形が近似値になっているのは，分母と分子の期待値を別々にとったからである．

つぎに，短所 1 を見積もってみよう．$K[\alpha, a_1]$ が大きくなるといっても，期待値をとれば 0 になってしまうので，まず，分散を計算する．

$$\mathrm{E}\left[K[\alpha, a_1]^2\right] = \mathrm{E}\left[\frac{(\sum_{i\in B}\alpha_i {a_1}_i{}^2)^2}{(\sum_{i\in B}{a_1}_i{}^2)^2}\right] \tag{11.44}$$

$$= \mathrm{E}\left[\frac{(\sum_{i\in B}(\theta_i-\overline{\theta}){a_1}_i{}^2)^2}{(\sum_{i\in B}{a_1}_i{}^2)^2}\right] \tag{11.45}$$

$$\simeq \frac{\mathrm{E}\left[\left(\sum_{i\in B}(\theta_i-\overline{\theta})a_{1i}{}^2\right)^2\right]}{\mathrm{E}\left[\left(\sum_{i\in B}a_{1i}{}^2\right)^2\right]} \tag{11.46}$$

$$= \frac{\sum_{i\in B}(\theta_i-\overline{\theta})^2\mathrm{E}\left[\left(a_{1i}{}^2-\mathrm{E}\left[a_{1i}{}^2\right]\right)^2\right]}{H\mathrm{E}\left[a_{1i}{}^2\right]^2+H\mathrm{E}\left[\left(a_{1i}{}^2-\mathrm{E}\left[a_{1i}{}^2\right]\right)^2\right]} \tag{11.47}$$

$$= \frac{\mathrm{E}\left[\left(a_{1i}{}^2-\mathrm{E}\left[a_{1i}{}^2\right]\right)^2\right]}{\mathrm{E}\left[a_{1i}{}^2\right]^2+\mathrm{E}\left[\left(a_{1i}{}^2-\mathrm{E}\left[a_{1i}{}^2\right]\right)^2\right]}\cdot\frac{\sum_{i\in B}(\theta_i-\overline{\theta})^2}{H} \tag{11.48}$$

$$= C_a\frac{\sum_{i\in B}(\theta_i-\overline{\theta})^2}{H} \tag{11.49}$$

ここでも，式(11.45)から式(11.46)の変形で，分母と分子の期待値を別々にとるという近似が入っている．また，最後の変形では，式(11.48)の前半の因子（妨害音エンベロープの統計的性質のみに依存する定数）を C_a とおいた．これは，妨害音が完全に一定強度のときのみ0であるので，ここでは，正の定数と考えればよい．

それでは，SN比を見積もってみよう．式(11.43)を式(11.49)の平方根で割れば，長所1と短所1のトレードオフを見積もることができる．

$$\frac{\mathrm{E}\left[K[\alpha,a_0]\right]}{\sqrt{\mathrm{E}\left[K[\alpha,a_1]^2\right]}} \simeq \frac{1}{C_a}\frac{\sqrt{\sum(\theta_i-\overline{\theta})^2}}{\sum\theta_i}H \tag{11.50}$$

$$= \frac{1}{C_a}\frac{\sqrt{\overline{(\theta_i-\overline{\theta})^2}}}{\overline{\theta_i}}\sqrt{H} \tag{11.51}$$

$$= \frac{1}{C_a}\sqrt{\sum\left(\frac{\theta_i}{\overline{\theta}}-1\right)^2} \tag{11.52}$$

これは，次のことを意味している．

今，H 個の小領域があるわけだが，ここに $H+1$ 個めの小領域での情報 θ_{H+1} が得られたとする．これを，Enable Signal $\kappa=1$ として適応に利用するほうが良いのか，あるいは $\kappa=0$ として無視するほうが良いのか，というのが Enable Signal の有効性の問題である．式(11.52)がそれに答えてくれる．すなわち，新たな θ_{H+1} を参加させれば必ず式(11.52)は増加する．（$\theta_i=\overline{\theta}$ の一点でのみ増化は0）．すなわち，意外にも，常に Enable Signal $\kappa=1$ にしておくのが最良なのである．

もちろん，この結果は，目的音のみの Event Signal で，かつ，妨害音が強度変動するという限定された条件である．また，ここでは，Event Signal を目的音強度の期待値に固定して評価したが，Enable Signal の導入に伴って Event Signal を変動させて最適化をはかる余地が残されており，今後さらなる検討が必要だろう．

式(11.52)から言えるもうひとつの性質は，新たな小領域を追加することの効果である．$\overline{\theta}$ から遠い θ_i を持つ小領域ほど，目的音抽出に有効な時間区間であるといえる．逆に，$\overline{\theta}$ に近接する θ_i を持つ小領域は，適応にはほとんど役立たない時間区間である．

これは，言いかえれば「目的音強度が最大の時間区間や，目的音が無音の時間区間が適応に重要である」という当然のことを言っているにすぎない．しかし，「目的音強度が平均的である時間区間が，適応にはそれほど好都合ではないものの，適応を妨害する要因にはならない」ということは新たな発見であったと言ってもいいだろう．

第 12 章
まとめ

第 II 部をまとめる.

- 視覚情報と聴覚情報を融合した知能化センシングについて論じた.
  - ◇ 視覚情報と聴覚情報には，Table. 8.2 で示したような本質的な差がある.
  - ◇ このため，視覚情報と聴覚情報を融合したセンシングを考えるためには，次の 4 つがキーポイントとなる
    - 【i】 視覚情報・聴覚情報を「どの座標上で」結合するのか
    - 【ii】 視覚情報・聴覚情報を「何によって」結合するのか
    - 【iii】 視覚情報・聴覚情報を「何を根拠に」結合するのか
    - 【iv】 視覚情報・聴覚情報を「どうやって」結合するのか
  - ◇ 視覚情報と聴覚情報に対して，時間座標上での融合と空間座標上での融合の 2 つの手段が考えられる.
  - ◇ 融合にあたっては，視覚情報と聴覚情報の関係を記述した何らかのモデルが必要となることが多い

- Cue Signal 法を応用した，視聴覚融合型の知能化センシングについて論じた.
  - ◇ これは，上の 4 つのポイントで見れば，以下の考え方に基づいた視聴覚融合である

    「どの座標で」結合するのか → 時間軸上で結びつける
    「何によって」結合するのか → 事象生起の度合で結びつける
    「何を根拠に」結合するのか → 関係を記述したモデルで結びつける
    「どうやって」結合するのか → Cue Signal 法による適応フィルタ

- 視覚部の役割としては,「画像を対象物ごとに領域分割」し,「対象物が何であるかを判断」して,「モデルを用いて音強度を推定」しなければならない．これらをすべてまともに行なうのは大変である．しかし，状況に応じて適宜省略できる場合が多い．
- 「画像を対象物ごとに領域分割」する手法として,「対象物ごとに領域を分割する」のではなく「事象ごとにピクセルを結合する」という考え方に基づいたアルゴリズムを提案した．これは，事象の同時性を基本原理としたもので Cue Signal 法との整合性が良い．
- Cue Signal 法による視聴覚融合の実験の結果，視覚的な手がかりを目的音規範とした信号抽出が可能であることが示された．
- Cue Signal 法による線型フィルタの適応は，位置座標上での融合による適応（各マイクロホンの遅延を調節して和をとることでマイクロホンアレイの指向特性のピークを目的音源に向ける）に比較して，FIR フィルタの全係数を最適化できることが大きな長所である．
- 視聴覚融合型の知能化センシングを実験するためのリアルタイムシステムを試作した．
  - 視覚システムは，9 個の DSP による並列処理である．画像の転送を高速かつ柔軟に行なえるよう工夫した．
  - 聴覚システムは，30 個の DSP による並列処理である．音響信号のサンプリング間隔が短いので，プロセッサ間の通信は小量のデータをきめ細かく伝達できるよう設計した．
  - 視覚情報と聴覚情報それぞれに最適なものを設計しようとすると，両情報の本質的な特質の違いにより，正反対の特徴を持ったシステムとなった．これは，Table.10.1 に示すとおりである．
  - しかし，Cue Signal の帯域がせまく情報量が非常に少ないため，視覚システムと聴覚システムを容易に結合することができた．
  - ただし，今後の視聴覚融合研究を考えるとき，両システムが異質であることは障害になるかもしれない．
- Event Signal のメタ信号への拡張を試みた．
  - 計測結果を何のために使うのかを考えたとき，センシング出力のメタ信号化が必要になる．
  - Event Signal のメタ信号化としては，Table.11.1 や Fig.11.10 に示したように様々なものが考えられるが，大きく分布型と命題型にわけることができる．
  - 分布型の Meta Event Signal が得られた場合は，それが目的音強度に関するもののみである場合には，分布の期待値をとるのが最適である．しかし，妨害音に関する Meta Event Signal が得られた場合には，分布の期待値だけでは十分でない．
  - 命題型の Meta Event Signal は，基本確率分配で記述する．分布型も命題型に変換すれば，命題型の世界で統一的に扱うことができる．そして，複数の情報の融合には，本論文で提案した方法（式(11.37)）を用いる．最終的には分布型に還元できるのが普通なので，分布型（**11.3.3**）で示した方法で Cue Signal に変換すればよい．

- Cue Signal の拡張を試みた．具体的には，適応を ON/OFF するための Enable Signal を付加する効果について論じた．
  - 目的音のみの Event Signal で，かつ，妨害音が強度変動するという限定された条件では，Cue Signal が 0 の瞬間のみが Enable Signal が有効となるぎりぎりの境界であることがわかった．

- 以下の問題は，扱うことができなかった．今後の課題である．
  - 領域分割された画像のなかで，どれが目的音源に関する画像なのかを判定する方法．
  - 視覚情報から，様々な Meta Event Signal を生成する具体的手法．

謝辞

本研究をすすめるにあたり，時に具体的に，また時には大域的見地にたって終始指導して下さった東京大学工学部計数工学科 山崎弘郎教授に深く感謝いたします．本研究は，先生の計測工学への広く深い知見なしには成立しえないものでした．

筆者は，山崎研究室の大学院生であった2年間は，半導体レーザを用いた光学的表面粗さ計測を研究していました．また，助手になってから3年間は，本論文の第I部でまとめた音響センサの知能化のテーマに平行して，表面弾性波による触覚センサの研究なども手掛けておりました．山崎教授はそれらの研究に対しても幅広い視野にたって適切な指導をして下さいました．お教えいただいた知識は，筆者の計測工学の視野を広げるのに役だったばかりでなく，本論文の執筆にあたって間接的に活かされたのではないかと思っています．また，先生が著名な研究者であったため，数多くの他の研究者と知り会う機会にも恵まれ，貴重な体験ができたことも多々ありました．さらに，研究内容に関してだけでなく，研究者としての人生の歩み方などに関しても大先輩として有益なお話しを数多く聞かせていただきました．ここに改めて感謝いたします．

また，東京大学工学部総合試験所 安藤繁助教授にも，直接的あるいは間接的にご指導いただきました．また，第11章の冒頭にも書いたように，Meta Event Signal についての考察は，安藤先生の考え方がヒントになって考えついたものです．ここに厚く感謝の意を表します．安藤先生は，筆者が修士学生時代には（筆者にとっては）運悪く電気通信大学に移られてしまったのですが，，筆者が助手になった年に東大に戻られたので，山崎研・安藤研の合同輪講などで指導していただく機会に恵まれました．輪講での安藤先生のコメントは筆者にとって非常に有益でした．しかし，それ以上に参考になったのは，先生の研究への取り組み方でした．自分の興味を研究の駆動力として活かし，数々の具体的な成果をあげていくという先生の研究方法は非常に参考になりました．

東京大学工学部計数工学科 石川正俊助教授 が，製品科学研究所から本学に着任されたのは，筆者が助手になってしばらくしてのことでした．石川先生は，センサフュージョン研究で日本の第一人者でいらっしゃいますが，そのときが，筆者が偶然にも視聴覚融合型知能化センサの研究を始めたころだったこともあって，その後，数々の場面でご教示いただくことができました．石川先生が東大に戻られなかったら，本論文の第II部は存在すらしなかったかもしれません．また，石川先生が研究のとりまとめをされている科学技術庁のセンサフュージョンのプロジェクトのメンバーにも加えていただき，センサフュージョン関連の仕事をされている数多くの研究者と自由な雰囲気で討論を行なえたことは，筆者にとって非常に有益でした．それ以外にも，石川先生は「センサフュージョンと並列処理研究会」などを主催され，筆者にも参加する機会を与えて下さいました．これらを通して，先生のバイタリティーあふれる研究態度と，多くの研究者をまとめて先端的な仕事を進めていく姿をまのあたりにできたことは今後の筆者にとって貴重な体験になったと思います．

そのほか，今日の筆者があるのは，学部学生時代から指導してくださった計数工学科の諸先生方のおかげであります．研究を進めていく上で，学生時代のノートを参照したことも度々

ありました．また，助手になってからも直接ご教示いただいた他，先生方の執筆された文献等を通して勉強させていただきました．ここに，厚く御礼申し上げる次第です．

特に，本論文を完成させるにあたっては，山崎弘郎教授，北森俊行教授，森下巖教授，廣津千尋教授，藤村貞夫教授，安藤繁助教授，石川正俊助教授の７人の先生方に，お忙しいなか数々の具体的な助言と有益なご指導をいただきました．ここに深く謝意を表します．

また，計算機室の小野雄三氏には，無響室を作るにあったって，部屋の使用を快諾いただきました．ここに御礼申し上げます．

さらに，計算機室のワークステーションを使いやすいように管理してくれた諸氏に感謝いたします．例えば，本論文でのイラストの大部分は，計算機室のワークステーションにインストールされている idraw で描いたものです．

学外においても数多くの先生方のお世話になりました．放送教育開発センター 大橋力教授には，科学研究費重点領域研究でお世話になっただけでなく，その後も，先生の主催される音響環境関係の研究会に参加させていただき，貴重な話を数多く聞かせていただきました．

もちろん山崎研究室において，多くの先輩の方々のお世話になったことは言うまでもありません．

慶應義塾大学 本多敏助教授 並びに，山形大学 平中幸雄助教授 には，筆者の大学院生時代より，直接数多くの指導をしていただきました．特に平中助教授は，筆者が助手になってからも，助手の仕事内容について先輩として暖かく教えて下さいました．筆者の１年先輩である佐賀大学 寺本顕武助教授，下川雅嗣氏 および同期の原山昌巳氏にもお世話になっています．

電気通信大学の三橋渉講師には，筆者が修士の学生のときより輪講などでお世話になりました．また，なかなか論文の執筆を開始しない筆者を叱咤激励して下さいました．

諸先輩だけでなく，本研究をここまで進めてくることができたのは，山崎研究室で卒業研究をされた多くの学生諸氏の協力によるところが非常に大きいと言えます．ここに，厚く感謝いたします．

幸いにも，彼らはみな興味をもって研究してくれました．そして，非常に熱心に取り組んでくれました．そういうわけで，本論文には，彼らとの共同研究と言える部分がかなりあります．以下，年代を追って振り返ってみたいと思います．

1986 年度，筆者が助手になって初めて担当した卒業研究に配属されてきたのが，浦中洋氏と篠崎公則氏でした．当時進めていた研究[64]は，Cue Signal 法を考案する前でしたが，自律的な音響センサをめざしていたことは現在と同じでした．両氏には，Cue Signal 法の前身であった「一対比較法」[65][66]のシステムの製作や，現在でも活用している簡易無響室の製作に協力してもらいました．

その後，２年間は卒業研究として別の課題（岡村広紀氏，藤波甲一氏の遠隔音響環境の疑似再現に関する研究，[67][68] 大重貴彦氏，福村直博氏の表面弾性波によるセンシングシステムの研究）を研究してもらっていましたが，1989 年度から，再び本研究の一部を卒業研究としてとりあげることにしました．古賀弘樹氏と鎌苔秀氏には，視聴覚融合の基礎研究に協力してもらいました．例えば，**Fig. 9.3**（p.88）の結果は両氏の実験によるものです．また，佐藤知春氏と鈴木竜自氏には，マイクロホン配置やフィルタ次数の研究に協力いただきました．本論文では，**6.7** 節にその成果の一部をまとめてあります．

1990 年度からは，本論文の第 10 章で述べたリアルタイムシステムの製作を始めました．この年配属されてきたのが池田思朗氏と小林伸治氏です．両氏には，視覚情報用サブシステム（**Fig. 10.3**）の製作に協力いただきました．池田氏にはハードウエア部分を中心に研究に参加してもらいました．特に，システムの画像入力部，画像出力部は氏の設計・製作によるものです．また，小林氏にはハードウエアの製作以外にも，ソフトウエア環境の構築に大きな

貢献をしてもらいました．プログラム開発の中核をなす makeasm（**Fig. 10.5**）を完成させたのは小林氏であります．

1991 年度には，聴覚情報用サブシステム（**Fig. 10.9**）の製作を，加瀬光氏，堀上周吾氏とともに行ないました．加瀬氏には，プリント基板の設計・製作で協力いただいた他にも，パイプライン処理するブロック化直接法の検討をしてもらいました．また，堀上氏には，ハードウエア製作以外に，**Fig. 10.4** に示した視覚部のアルゴリズムの実験や小林氏製作の makeasm を聴覚用にアレンジする作業などもやってもらいました．

その他，リアルタイムシステムの製作にあたっては，当研究室の修士学生であった森隆氏に DSP ボードの製作を手伝ってもらったのをはじめ，4 年夏学期に計測工学実験第 2 で配属されてきた多くの学生に協力してもらうことができました．ここに感謝いたします．

最後になりましたが，山崎研究室での同僚である山口晃生氏には，いろいろな面で数えきれないほどお世話になり，ここに感謝の意を表します．日頃，研究に関することで議論してもらった他，助手としての事務的な仕事などでもお世話になりました．また，同一の締め切りに苦しめられながら，お互いに励ましあいながら仕事をしたことも度々ありました．その他，現在の山崎研究室の計算機環境は，ほとんど彼ひとりで構築したものであります．これがなければ，第 5 章の実験は出来なかったと思います．これらに対して，心から感謝したいと思います．

予想外に長くなってしまいました．ここまで書いてきて痛感するのは，素晴らしい先生方や尊敬すべき諸先輩，頼りになる同僚や頼もしい後輩たちなどに恵まれて，筆者は非常に幸運であったということです．このような環境にあったならば，もっと良い論文がまとめられたはずと思われる読者がいたとするならば，それはひとえに筆者の努力不足と力不足によるところが大きいのではないかと考えます．その点を少し反省して本論文を締めくくりたいと思います．

本研究に関係して，以下の援助を受けることができました．ここに謝意を表します．

昭和 62 年度	文部省科学研究費補助金	重点領域研究 (1)	62602006
昭和 63 年度	文部省科学研究費補助金	重点領域研究 (1)	63602042
平成 1 年度	文部省科学研究費補助金	重点領域研究 (1)	01602006
平成 2 年度	文部省科学研究費補助金	重点領域研究 (2)	02202214
平成 2 年度	文部省科学研究費補助金	奨励研究 (A)	02855107
平成 3 年度	文部省科学研究費補助金	奨励研究 (A)	03855092
平成 4 年度	文部省科学研究費補助金	奨励研究 (A)	04855086
平成 3 年度	科学技術庁科学技術振興調整費		
	「センサフュージョンの基盤的技術の開発に関する研究」		

参考文献

1) 有本卓：カルマンフィルタ, 17, 産業図書 (1977)

2) 谷萩隆嗣：ディジタル信号処理の理論 3, 150, コロナ社 (1986)

3) J. J. Shynk : Frequency-Domain and Multilate Adaptive Filtering, IEEE Signal Processing Magazine, **9**–1, 14/37 (1992)

4) J. B. Allen, D. A. Berkley, and J. Blauert : Multimicrophone Signal-Processing Technique to Remove Room Reverberation from Speech Signals, J. Acoust. Soc. Am., **62**–4, 912/915 (1977)

5) B. Widrow and M. Hoff, Jr., : Adaptive Switching Circuits, IRE WESCON Convention Record, pt. 4, 96/104 (1960)

6) G. Ungerboeck : Theory of the Speed of Convergence in Adaptive Equalizers for Digital Communication, IBM J. Res. Develop., **16**–6, 546/555 (1972)

7) Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter, Proc. IEEE, **64**–2, 1151/1161 (1976)

8) J. Nagumo and A. Noda : A Learning Method for System Identification, IEEE trans. Automatic Control, **AC-12**, 282/287 (1967)

9) 野田, 南雲：システムの学習同定法, 計測と制御, **7**–9, 597/605 (1968)

10) R. R. Bitmead and B. D. O. Anderson : Performance of Adaptive Estimation Algorithms in Dependent Random Environments, IEEE Trans. Automatic Control, **AC-25**, 788/794 (1980)

11) T. C. Hsia : Convergence Analysis of LMS and NLMS Adaptive Algorithms, Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, 667/670 (1983)

12) R. L. Plackett : Some Thorems in Least Squares Methods, Biometrica, **37**, 149 (1950)

13) B. Widrow *et al.*: Adaptive Noise Cancelling: Principles and Applications, Proc. IEEE, **63**–12, 1692/1716 (1975)

14) W. F. Gabriel : Adaptive Processing Array Systems, Proc. IEEE, **80**–1, 152/162 (1992)

15) J. L. Flanagan, J.D. Johnston, R. Zahn, and G. W. Elko : Computer-steered Microphone Arrays for Sound Transduction in Large Rooms, J.Acoust.Soc.Am., **78**–5, 1508/1518 (1985)

16) Y. Kaneda and J. Ohga : Adaptive Microphone Array System for Noise Reduction, IEEE Trans. Acoust., Speech, Signal Processing, **ASSP–34**–6, 1391/1400 (1986)

17) B. Widrow *et al.*: Adaptive Antena Systems, Proc. IEEE, **55**–12, 2143/2159 (1967)

18) 高橋弘太, 山崎弘郎 : 知能化マイクロホンシステム－線型フィルタの階層的適応手法－, 計測自動制御学会論文集, **25**–10, 1119/1125 (1989)

19) Kota Takahashi and Hiro Yamasaki : Self-Adapting Multiple Microphone System, Sensors and Actuaters, **A21-A23**, 610/614 (1990)

20) 高橋弘太, 山崎弘郎 : 視覚情報で適応化する知能化音響センサー視聴覚のセンサフュージョン－, 計測自動制御学会論文集, **27**–3, 276/282 (1991)

21) 高橋弘太: 視聴覚情報の統合, 日本ロボット学会誌, **8**–6, 766/771 (1990)

22) 高橋弘太, 山崎弘郎 : 視覚と聴覚のセンサフュージョンシステム, 計測と制御, **31**–9, 975/979 (1992)

23) Kota Takahashi and Hiro Yamasaki : Real-Time Sensor Fusion System for Multiple Microphones and Video Camera, Second International Symposium on Measurement and Control in Robotics, 250/256 (1992)

24) 高橋弘太: 視覚・聴覚融合システム, 山崎弘郎・石川正俊編 センサフュージョン－実世界の能動的理解と知的再構成－, コロナ社, 193/204 (1992)

25) 高橋弘太, 山崎弘郎 : 最小の知識で適応する学習型集音システム, 電気学会センサ技術研究会資料, **ST–89–6**, 41/50 (1989)

26) 高橋弘太, 山崎弘郎 : 階層構造を持った知能化音響センシングシステム, 電子情報通信学会技術研究報告, **EA-90**–22, 17/24 (1990)

27) 高橋弘太, 佐藤知春, 鈴木竜自, 山崎弘郎 : 知能化集音センサにおけるマイクロホン群の最適配置, 計測自動制御学会第 29 回学術講演会予稿集, 317/318 (1990)

28) 高橋弘太, 池田思朗, 小林伸治, 山崎弘郎 : 視覚情報で適応する聴覚センサのためのリアルタイムシステム, 計測自動制御学会第 30 回学術講演会予稿集, 369/370 (1991)

29) 高橋弘太, 山崎弘郎 : 視覚を用いて学習する聴覚システム, 第 34 回自動制御連合講演会前刷, 特セ-85/特セ-86 (1991)

30) 高橋弘太: 視聴覚融合による知能化センシング, ミューアルファ, **1**–8, 46/51 (1991)

31) 佐藤知春, 鈴木竜自 : 知能化音響センシングの最適設計のための基礎的研究, 東京大学工学部計数工学科卒業論文 (1990)

32) 古賀弘樹, 鐵恭秀 : 視覚情報による音響センサの知能化, 東京大学工学部計数工学科卒業論文 (1990)

33) 古賀弘樹, 鐵恭秀, 高橋弘太, 山崎弘郎 : 視覚センサを持った適応型集音システムの研究, 計測自動制御学会第 29 回学術講演会予稿集, 319/320 (1990)

34) 池田思朗, 小林伸治：視覚情報による音響センサの知能化, 東京大学工学部計数工学科卒業論文 (1991)

35) 加瀬光, 堀上周吾：視覚・聴覚をリアルタイムで融合する知能化センシングシステム東京大学工学部計数工学科卒業論文 (1992)

36) 加瀬光, 髙橋弘太, 山崎弘郎：視聴覚融合センシングシステムのための実時間音響プロセッサ, 第 31 回計測自動制御学会学術講演会 (1992)

37) 日野幹雄：スペクトル解析, 265, 朝倉図書 (1977)

38) A. V. Oppenheim and R. W. Schafer : Digital Signal Processing, Prentice-Hall (1975)

39) 電子通信学会編：新版 聴覚と音声, 331, コロナ社 (1980)

40) O. L. Frost, III : An Algorithm for Linearly Constrained Adaptive Array Processing, Proc. IEEE, **60**–8, 926/935 (1972)

41) 柳井晴夫, 竹内啓：射影行列・一般逆行列・特異値分解, 東京大学出版会 (1983)

42) 石川正俊：センサフュージョンシステム－感覚情報の統合メカニズム－, 日本ロボット学会誌, **6**–3, 251/255 (1988)

43) 石川正俊：センサフュージョンの現状と課題, 計測自動制御学会講演会予稿集, i/ix (1990)

44) 山崎弘郎・石川正俊編：センサフュージョン－実世界の能動的理解と知的再構成－, コロナ社 (1992)

45) J. M. Richardson and K. A. Marsh : Fusion of Multisensor Data, Int. J. Robot. Res., **7**–6, p.78 (1988)

46) R. C. Luo and M. G. Kay : Multisensor Integration and Fusion in Intelligent Systems, IEEE trans. Syst. Man Cybern., **SMC-19**–5, 901/931 (1989)

47) H. McGurk and J. McDonald : Hearing lips and seeing voices, Nature, **264**, 746/748 (1976)

48) J. McDonald and H. McGurk : Visual Influence on Speech Perception Process, Perception and Psychophysics, **24**, 253/257 (1978)

49) 積山薫, 東倉洋一：読唇情報が音声知覚に果す役割, テレビジョン学会技術報告, **13**–44, 31/36 (1989)

50) 近藤公久, 筧一彦：視覚情報の音声知覚に及ぼす影響, テレビジョン学会技術報告, **13**–33, 13/18 (1989)

51) 弓削とよ, 伊福部達：音像定位に及ぼす周辺視の影響, 日本音響学会聴覚研究会資料, H–83–1, 1/6 (1983)

52) 小宮山摂：視覚情報による音像定位の変化, テレビジョン学会技術報告, **13**–44, 25/30 (1989)

53) Meredith and Stein : Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration, J. Neurophysiol., **56**, 640 (1986)

54) E. I. Knudsen and M. Konishi : Science, **200**, 795/797 (1978)

55) 栗田知好, 本多清志, 垣田有紀 : 口唇画像情報を併用する音声の分析, 電子情報通信学会技術研究報告, SP88-94, p.41 (1988)

56) 田村進一, 呉簡彤, 河合秀夫, 黒須顕二 : 口形画像と音声特徴を併用する統合ニューロ音声認識, 電子情報通信学会技術研究報告, PRU89-19, 1/7 (1988)

57) 呉簡彤, 田村進一, 光本浩士, 河合秀夫, 黒須顕二, 岡崎耕三 : 音声・口形特徴量を併用するニューラルネットを用いた母音認識, 電子情報通信学会論文集 D-II, **J73**-8, 1309/1314 (1990)

58) 青野俊宏, 石川正俊 : 確率過程を用いたセンサフュージョン—多系列隠れマルコフモデルを用いた視聴覚融合—, 第 2 回自律分散システムシンポジウム予稿集, 115/118 (1991)

59) 青野俊宏, 石川正俊 : 確率過程を用いた視聴覚融合, 第 34 回自動制御連合講演会前刷, 特セ-87/特セ-90 (1991)

60) 安藤繁 : 画像の時空間微分算法を用いた速度ベクトル分布計測システム, 計測自動制御学会論文集, **22**-12, 1330/1336 (1986)

61) 安藤繁, 田部井俊幸 : 動的な 3 次元累積合成機構を有する微分両眼視法, 計測自動制御学会論文集, **24**-6, 628/634 (1988)

62) A. Dempster : Upper and Lower Probabilities Induced by Multivalued Mapping, Ann. Math. Statist., **38**, 325/339 (1967)

63) G. Shafer : A Mathematical Theory of Evidence, Princeton, NJ: Princeton Univ. Press (1976)

64) 浦中洋, 篠崎公則 : 主体的信号選択性をもつ集音システムの研究, 東京大学工学部計数工学科卒業論文 (1987)

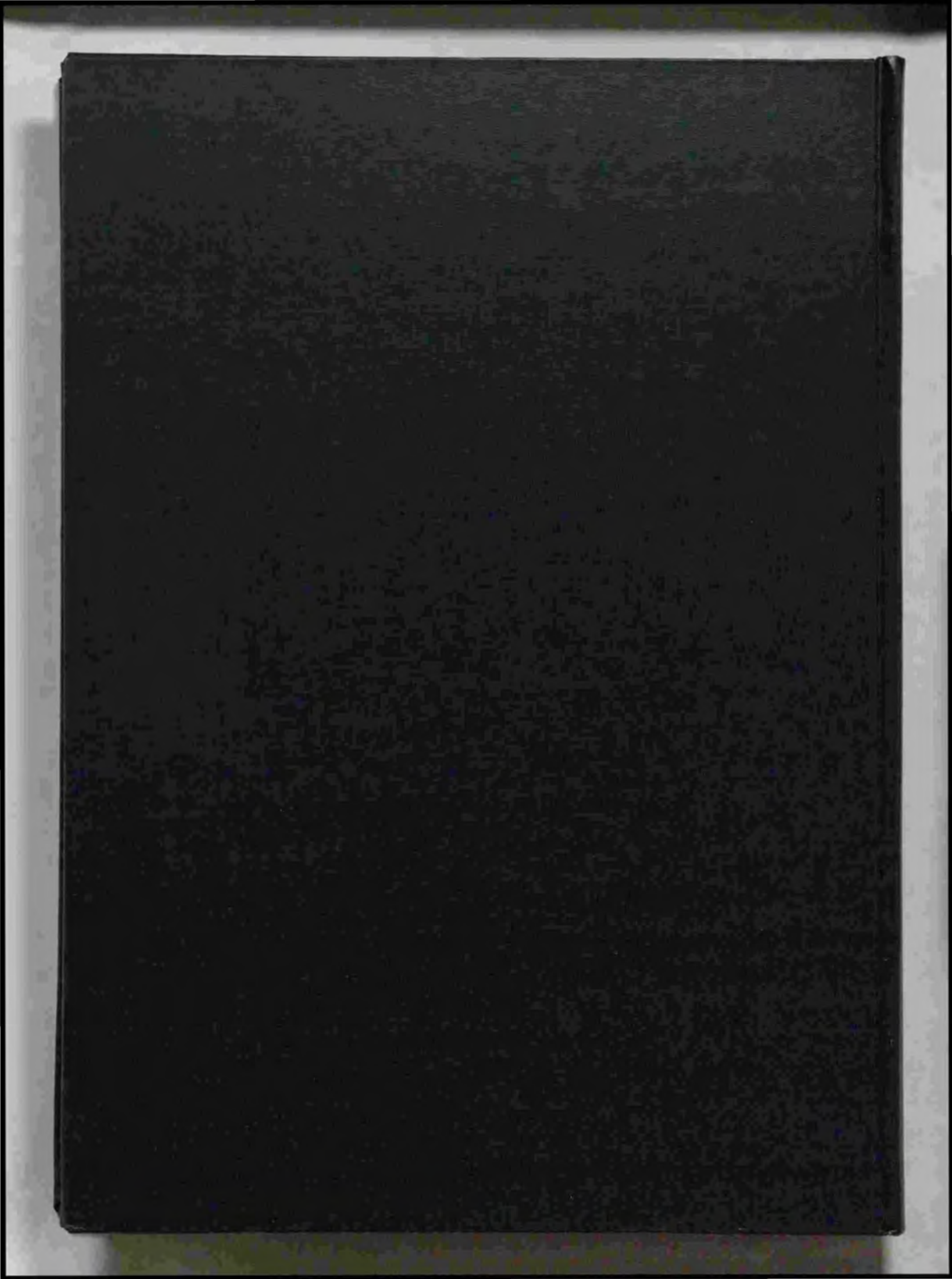
65) 高橋弘太: 聴覚システム, コンピュートロール, **21**, 53/58 (1988)

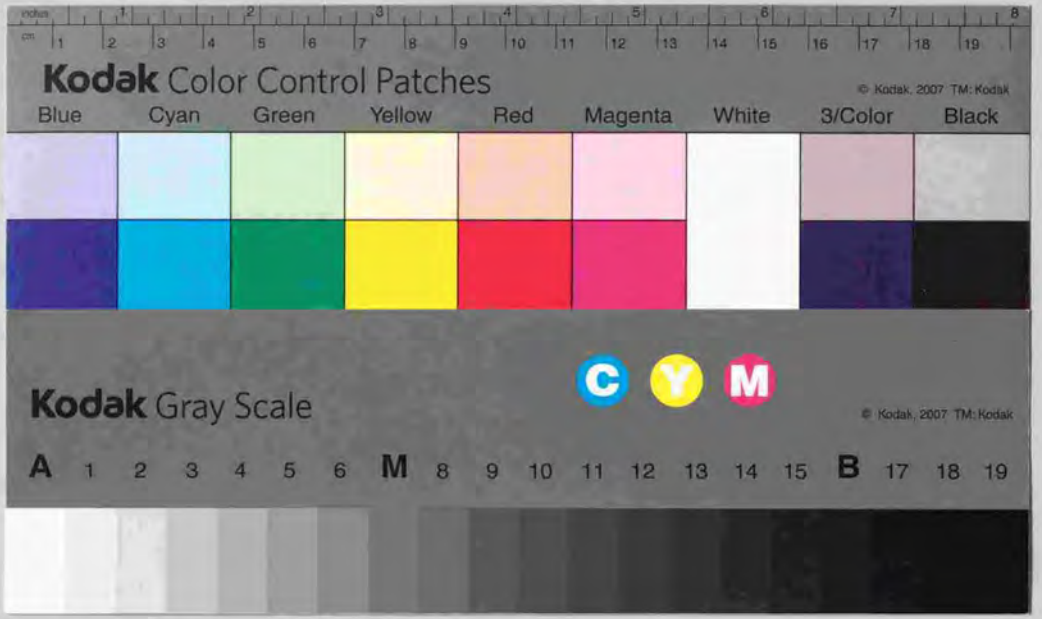
66) 高橋弘太, 浦中洋, 篠崎公則, 山崎弘郎 : 複数マイクロホンとディジタルフィルタによる適応型集音システム, 計測自動制御学会第 26 回学術講演会予稿集, 45/46 (1987)

67) 岡村広紀, 藤波甲一 : 遠隔音響環境の疑似再現に関する研究, 東京大学工学部計数工学科卒業論文 (1988)

68) 高橋弘太, 岡村広紀, 藤波甲一, 山崎弘郎 : 聴覚特性を利用した信号処理による音環境再現手法, 計測自動制御学会第 27 回学術講演会予稿集, 543/544 (1988)

索引




inches 1 2 3 4 5 6 7 8
cm 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Kodak Color Control Patches
© Kodak, 2007 TM: Kodak
Blue Cyan Green Yellow Red Magenta White 3/Color Black
Kodak Gray Scale
C Y M
© Kodak, 2007 TM: Kodak
A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19