

マルチメディア工学 追加

相互情報量の利用、他

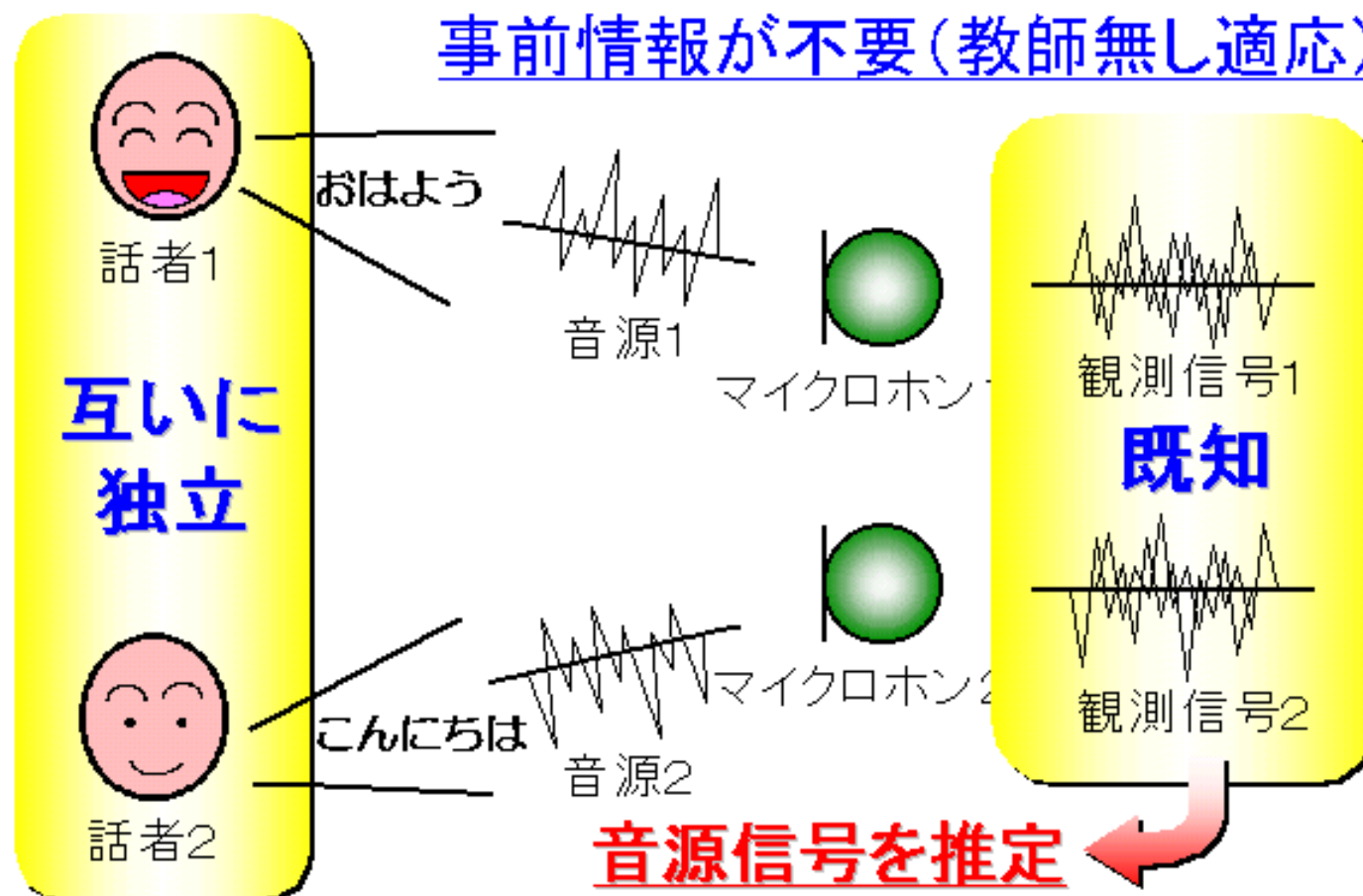
独立成分分析 (ICA; Independent Component Analysis)

位置合わせ手法 (Registration)

田村進一

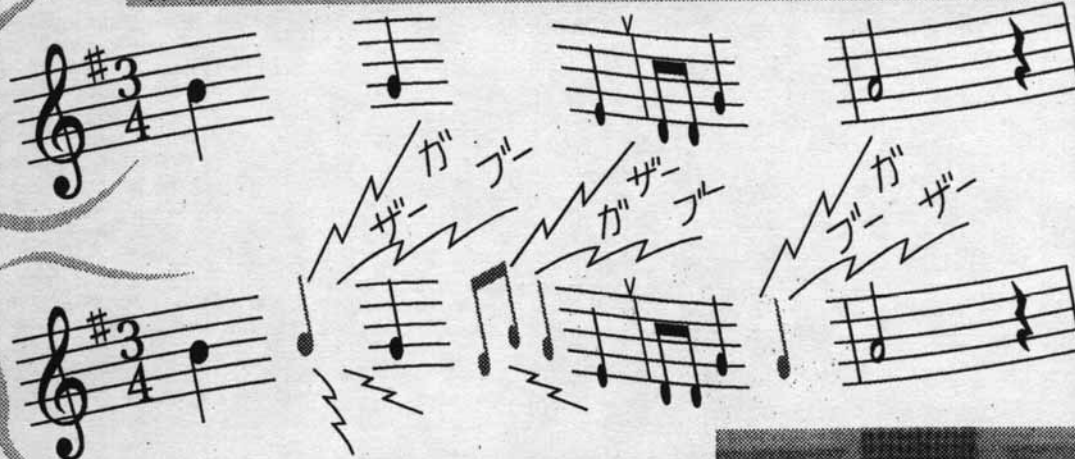
ICAに基づくブラインド音源分離

事前情報が不要(教師無し適応)



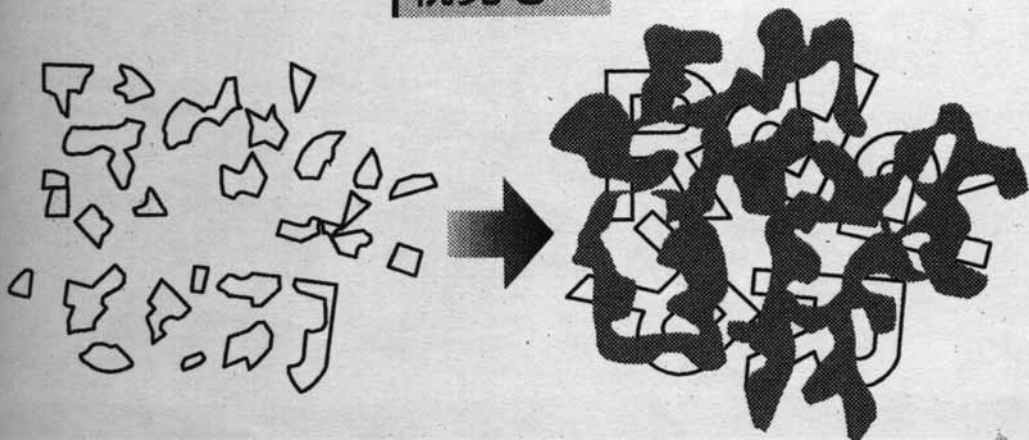
途切れていてもどうして分かる？

雑音が入らないと途切れて曲として聞こえにくい



雑音が入ると途切れていてもつながって聞こえる

視覚も…



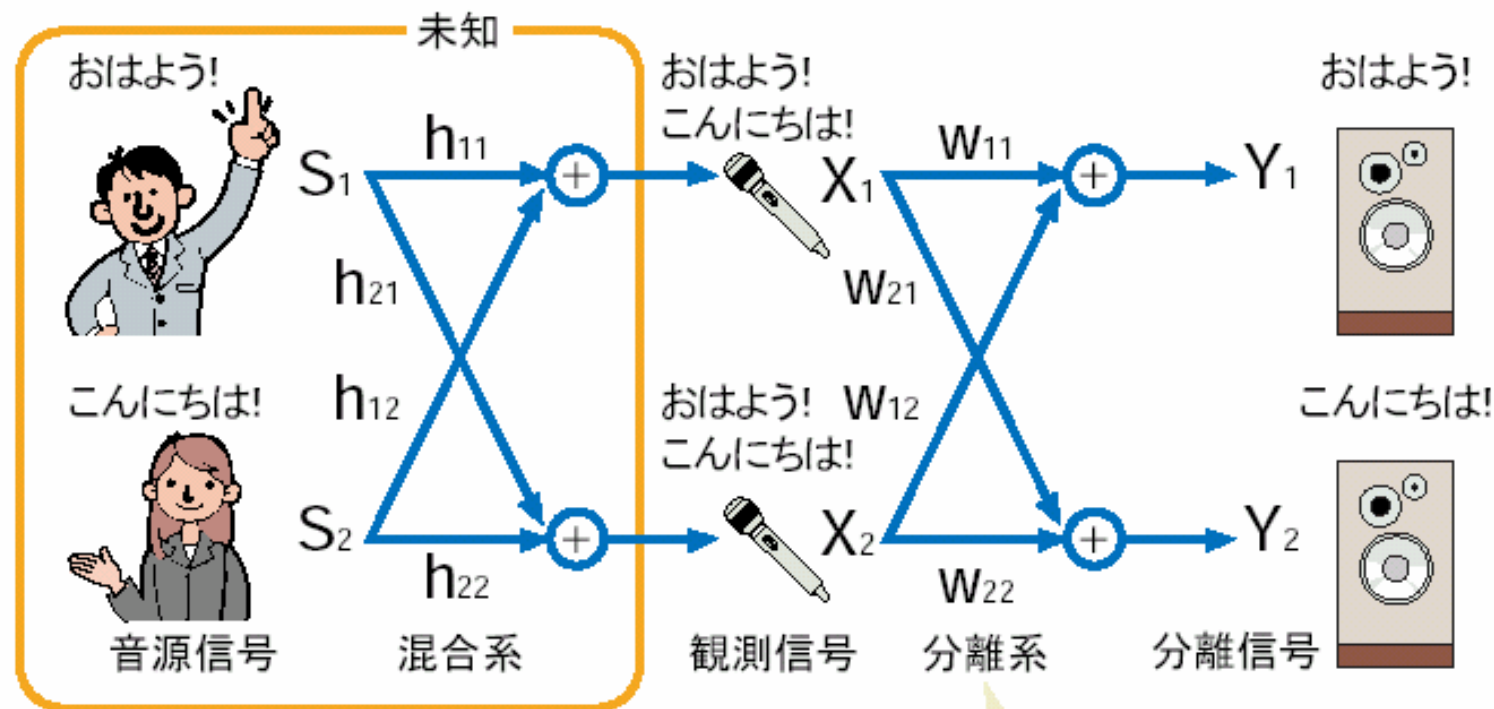
じゃまものがあると文字がよく見えてくる



ざわつくパーティー会場でも特定の人の声が聞き取れる

* ICA: Independent Component Analysis

●独立成分分析（ICA）に基づくブラインド音源分離の原理

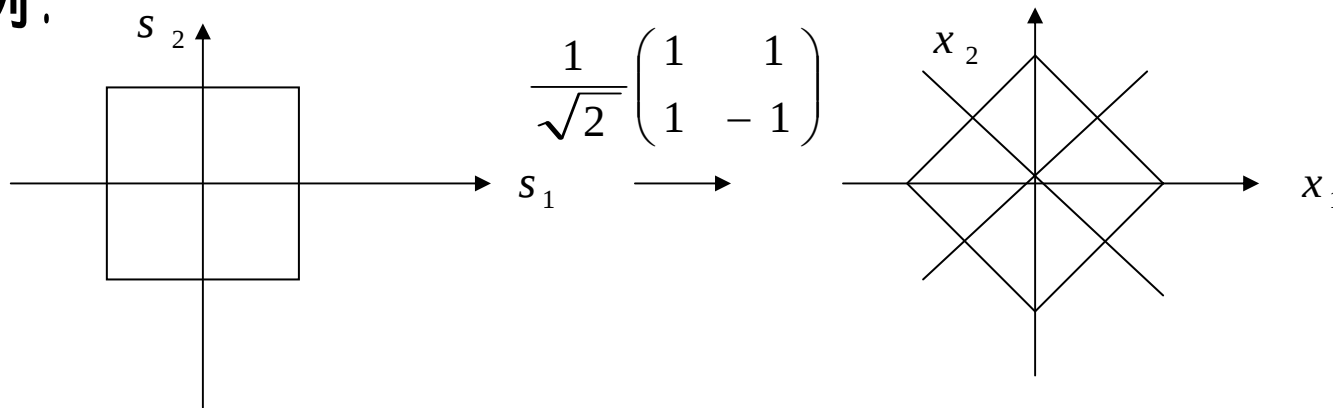


音源信号と混合系は未知であり、観測した信号だけから元の音を分離抽出する

相関と独立性

- 独立であることと相関がないということは異なる。
独立ならば相関はないが 相関がないからといって独立とはいえない。

例:



s_1, s_2 : 独立な $[-1, 1]$ 上の一様分布

x_1, x_2 : 無相関だが独立ではない

- 多変量正規分布の場合は、相関がないことと、独立であることは同値である。

独立性評価 (例)

$$p(\mathbf{y}) = \prod_{i=1}^n p(y_i)$$

Kullback-Leibler divergence (D)

相互情報量

$$D(\mathbf{W}) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{i=1}^n p(y_i)} d\mathbf{y}$$

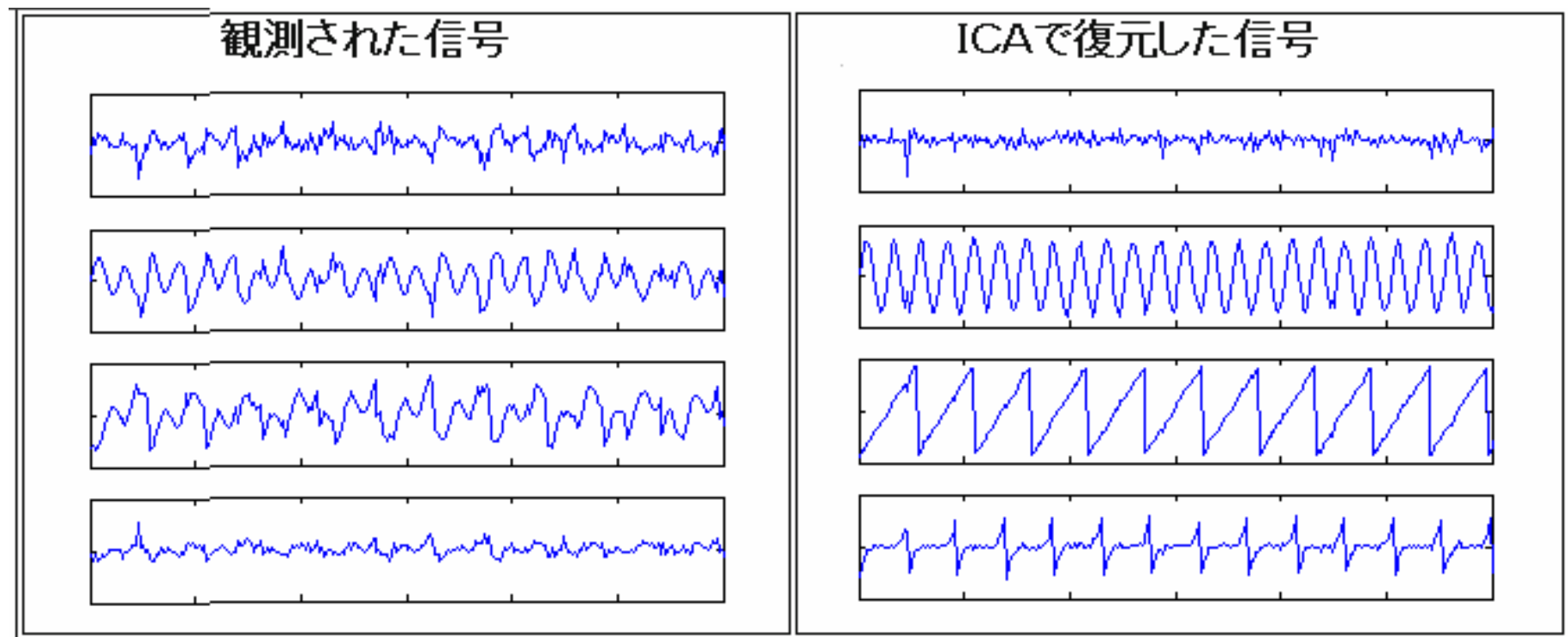
$$= \sum_{i=1}^n H(y_i) - H(\mathbf{y})$$

where

$$H(\mathbf{y}) = -\int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y}$$

$$H(y_i) = -\int p(y_i) \log p(y_i) dy_i$$

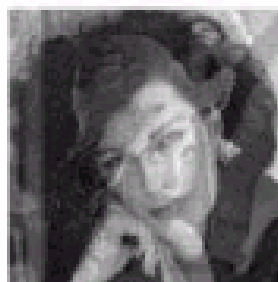
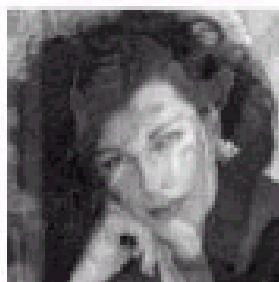
(例) y_2 を知ること
によって, y_1 につ
いて得られる
情報量



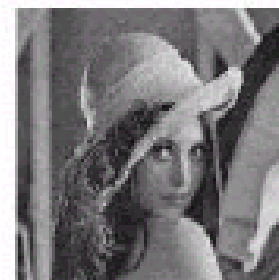
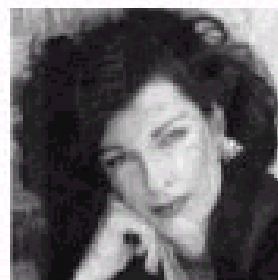
4 信号分離

分離した4信号の確率密度関数と、それらが独立と仮定したときの確率密度関数の差(相互情報量)が小さくなるように、分離の重みを調整していく

観測された画像

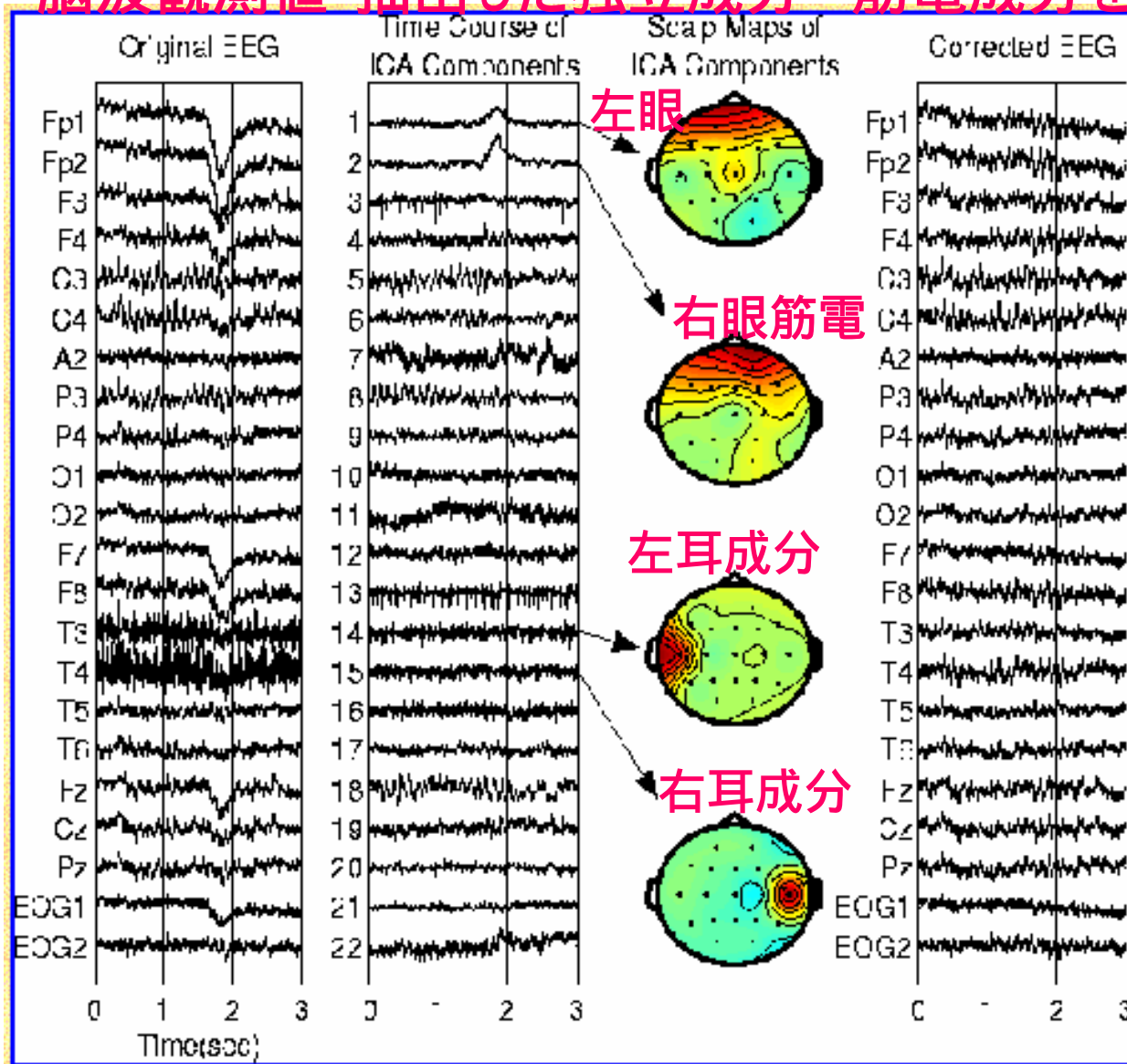


ICAで復元した画像



4 画像分離

脳波観測値 抽出した独立成分 筋電成分を引いた脳波

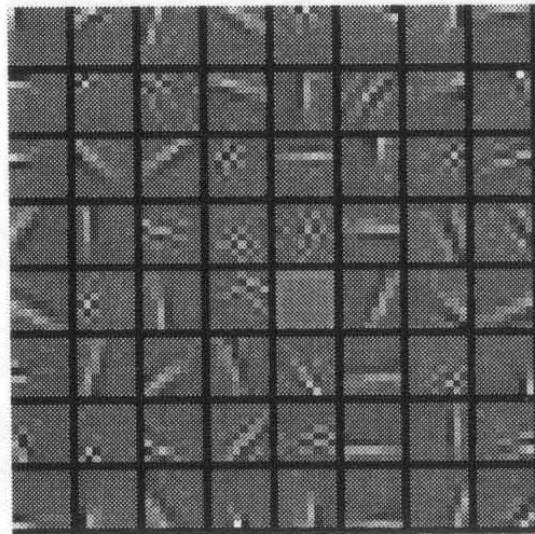


脳波からの眼球筋電信号等分離と補正

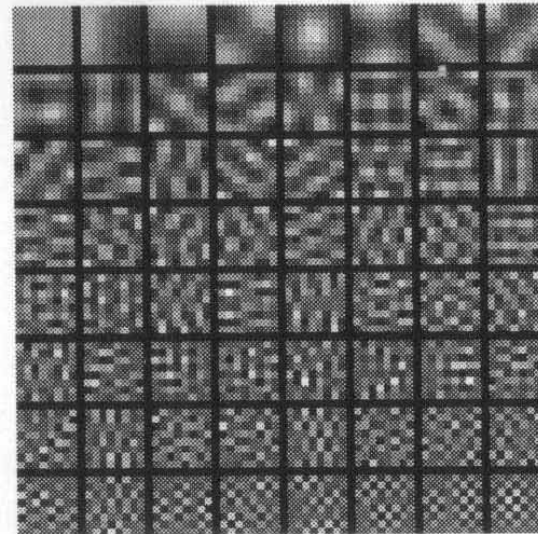
4枚の自然画像から、基底関数の抽出



Fig. 2 12,000 patch images (size:8x8) are randomly selected from above four natural images for learning of basis function.



(a) ICA basis.



(b) PCA basis.

Fig. 3 Basis functions of natural images.

- (a)抽出された独立成分は、サルなどの方向性受容野にそっくり
- (b)抽出された主成分はフーリエ展開基底様である

多数の画像からの同一人画像の抽出

70枚中に同一人の写真●は計6枚

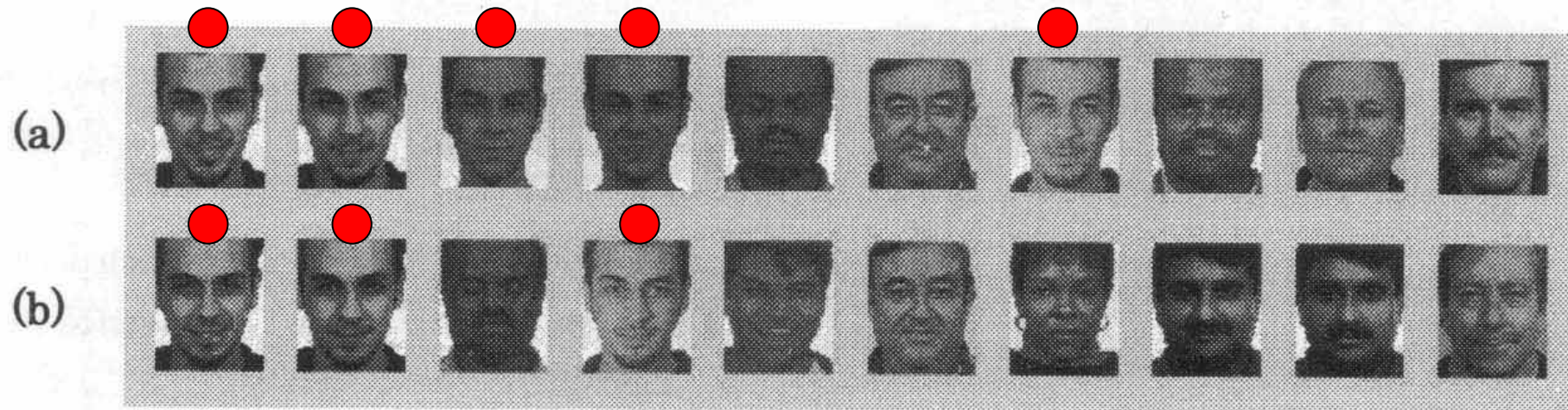


Fig. 5 Recognition results. (a) ICA; (b) PCA.

類似度の上位10枚の写真

独立成分(ICA)を基底にした展開成分で類似画像検索した方(a)が、主成分分析法(PCA)(b)よりも、同一人の画像が多く検索された

顔画像に対する基底関数

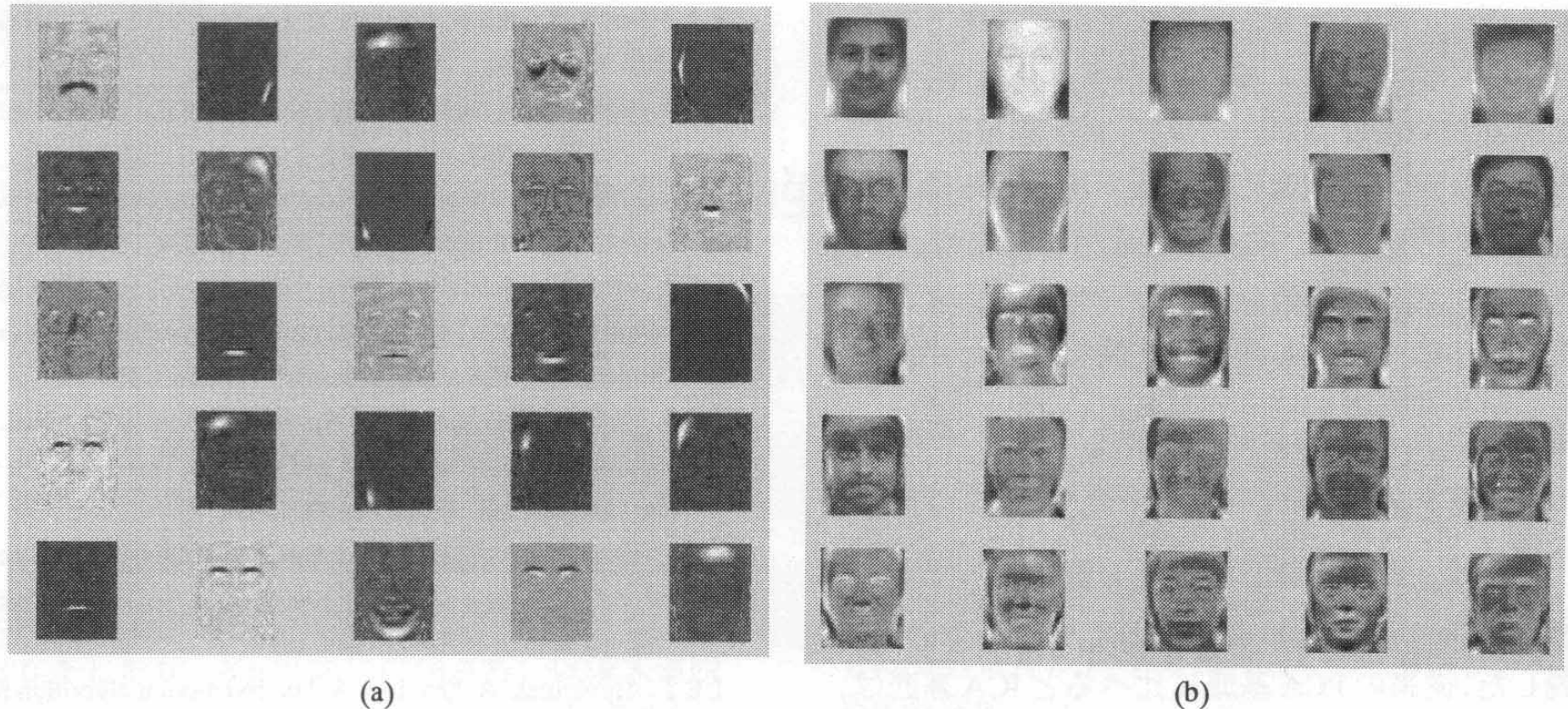
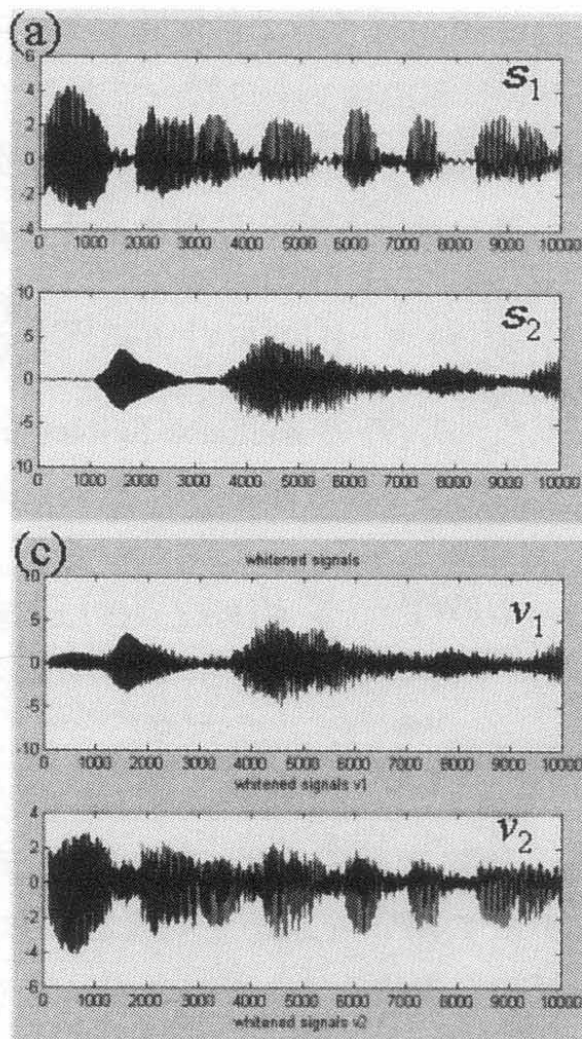


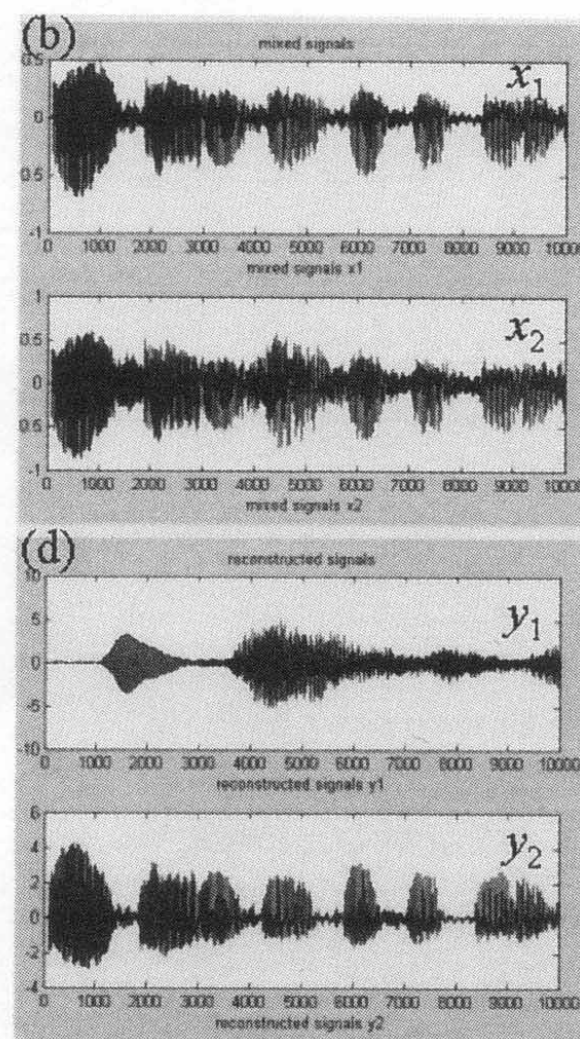
Fig. 4 Basis images for the factorial representation. (a) ICA; (b) PCA.

- (a) ICAでは、顔部品が基底となっている。個人識別や表情表現に有利
- (b) PCAでは、顔全体の変化成分が基底

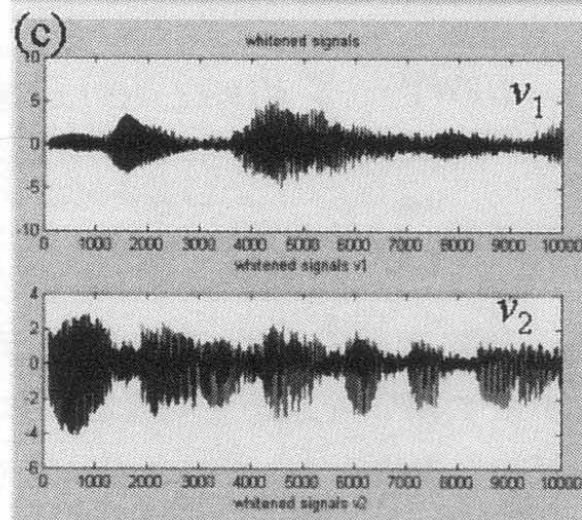
原音声



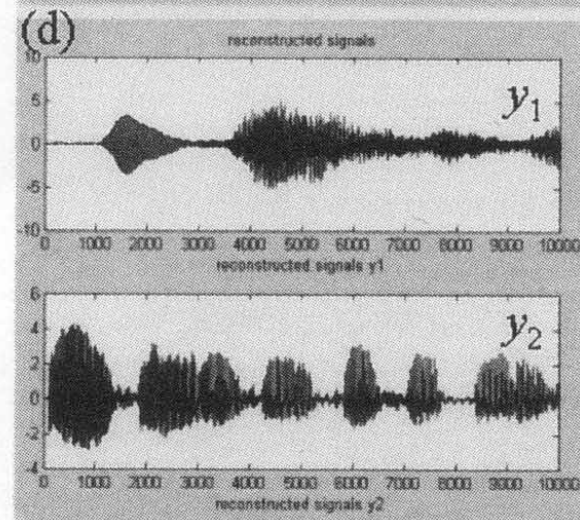
混合音



PCAによる分離



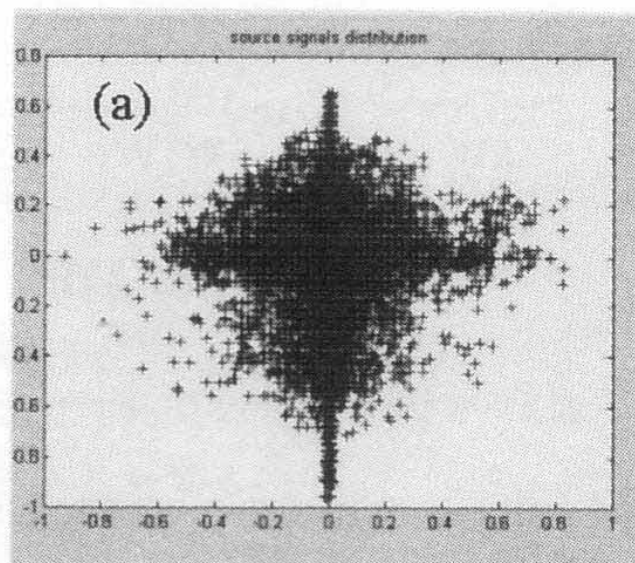
ICAによる分離
小信号部に注目



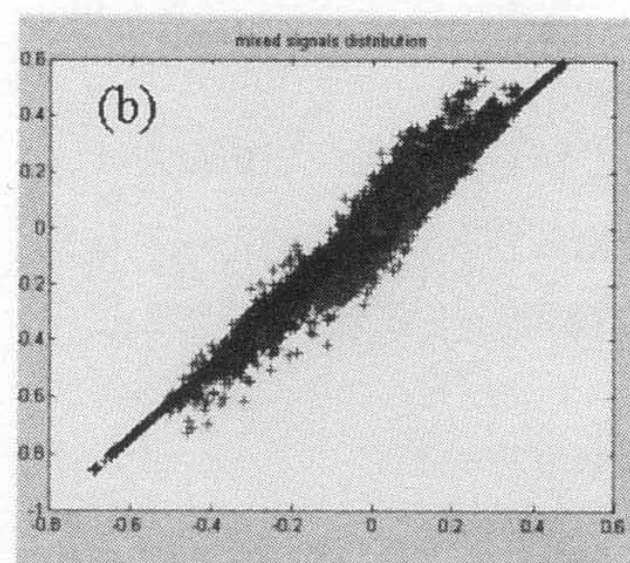
Demonstration of the separation of two speech signals from two mixtures. (a) source signals; (b) mixed signals; (c) separated signals using PCA (whitening); (d) separated signals using ICA.

2音声信号の分離

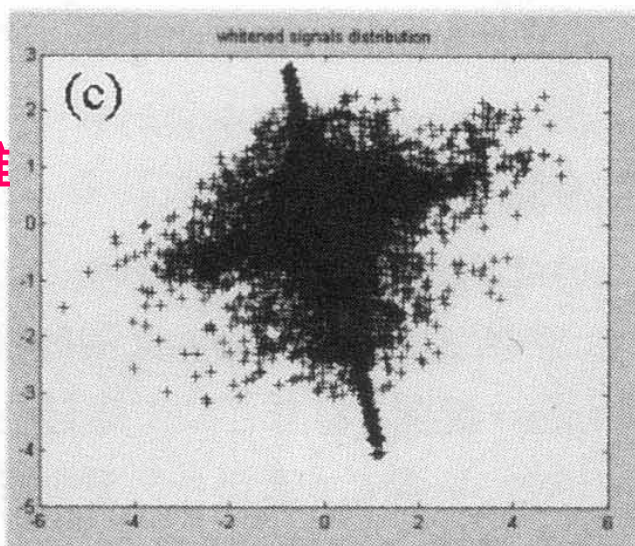
原音声



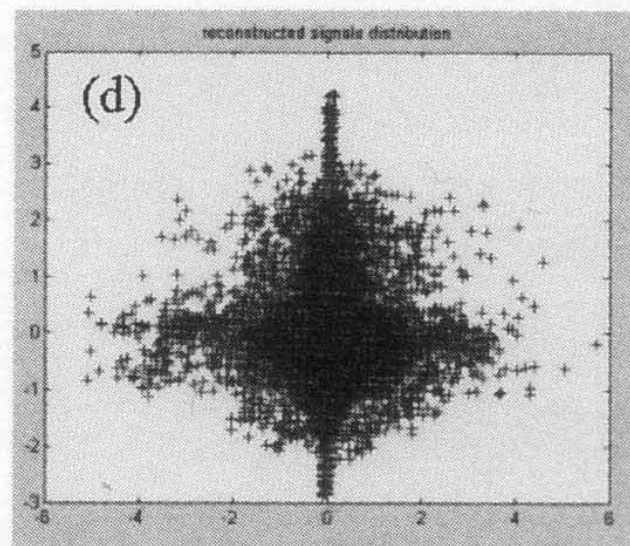
混合音



PCAに
よる分離



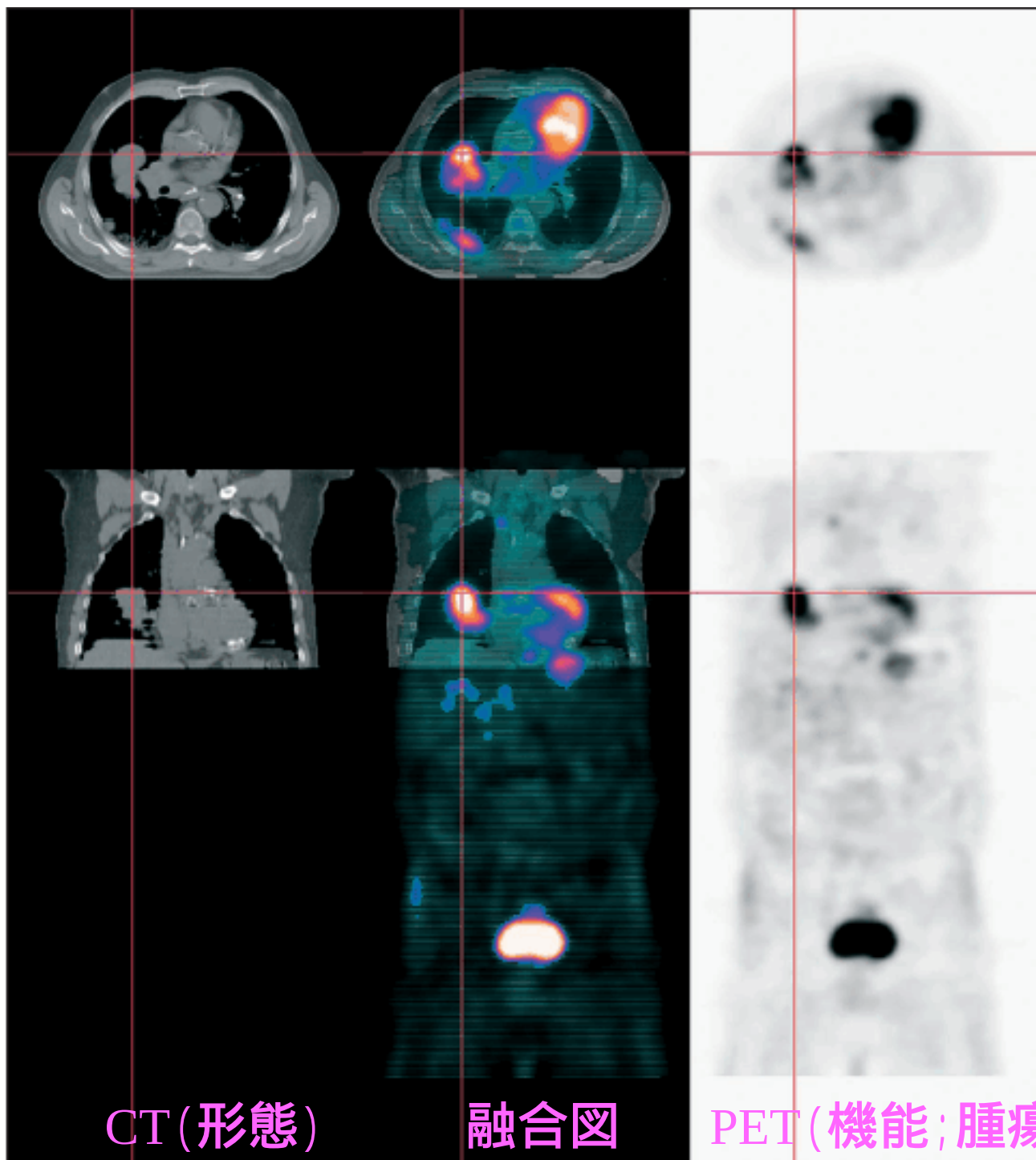
ICAに
よる分離



Scatter plots. (a) source signals; (b) mixed signals; (c) separated signals using PCA (whitening); (d) separated signals using ICA.

2音の振幅散布図

相互情報量による registration

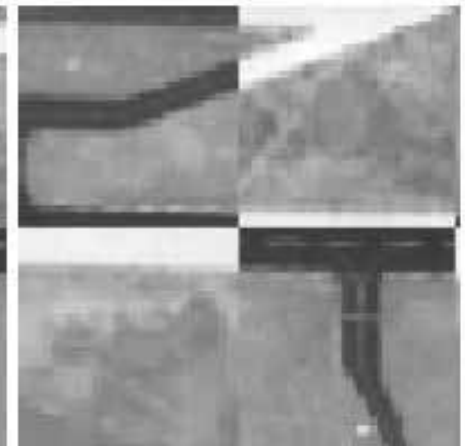
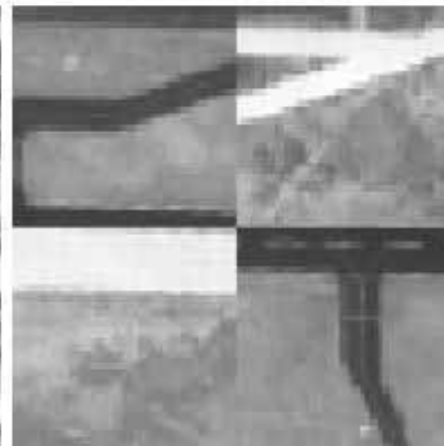


異種画像の
位置合わせ

近赤外線写真

可視光写真

位置合わせ前チェッカーボード表示



位置合わせ前 拡大図

後

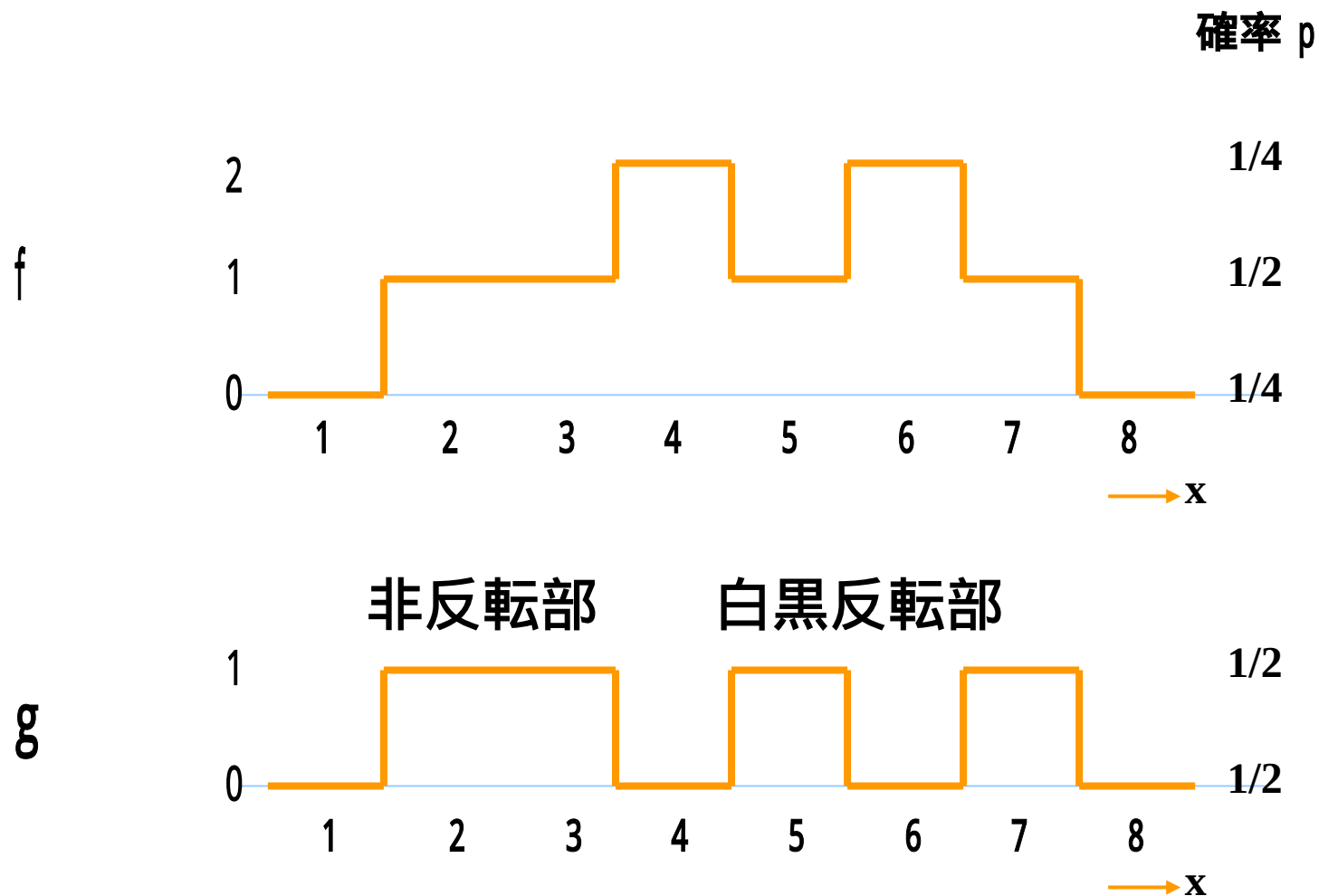
前(ずれている) 後(修正されている)

Figure 5. Non-parametric or 'fluid' registration of a NIR image (I_1) and a visible intensity image (I_2). The first row shows both images and an overlay of their initial positions. The second row shows detail views of the overlay of I_1 and $I_2 \circ \mathcal{F}$ before and after registration. The first 2 images of the second row show the alignment of a detail of the stadium. The 3th and 4th image show the alignment of a detail of the airstrip.

近赤外線写真と可視光写真との位置合わせ

滑走路が、赤外線では白く写り、可視光では黒く写っている(反転)ことに注意

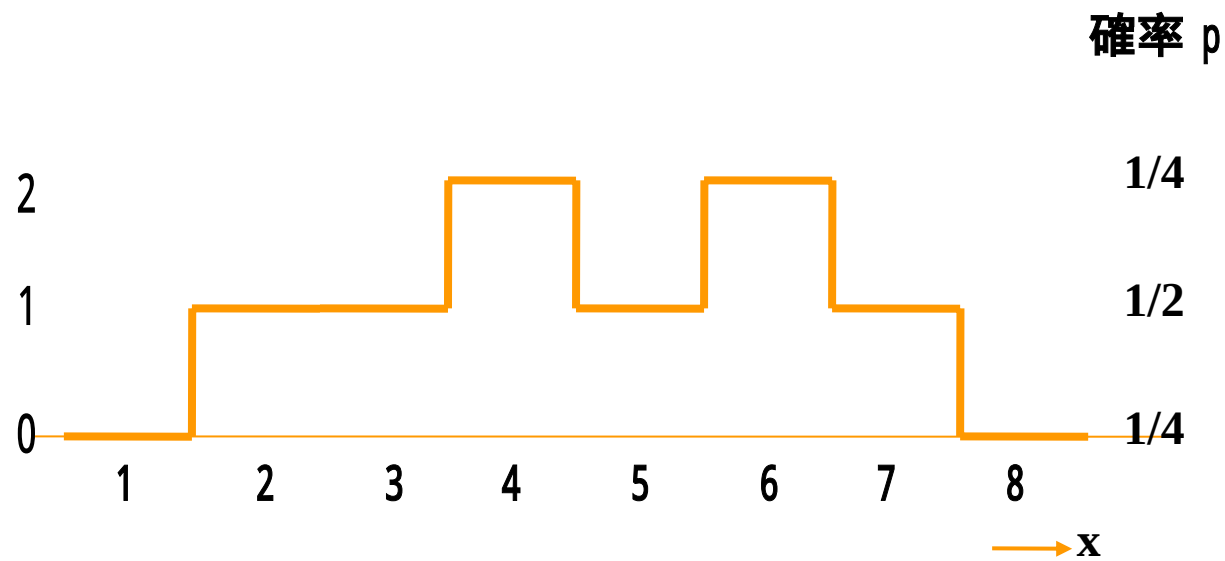
http://earth.esa.int/rtd/Events/ESA-EUSC_2004/Ar24_Strecha.pdf



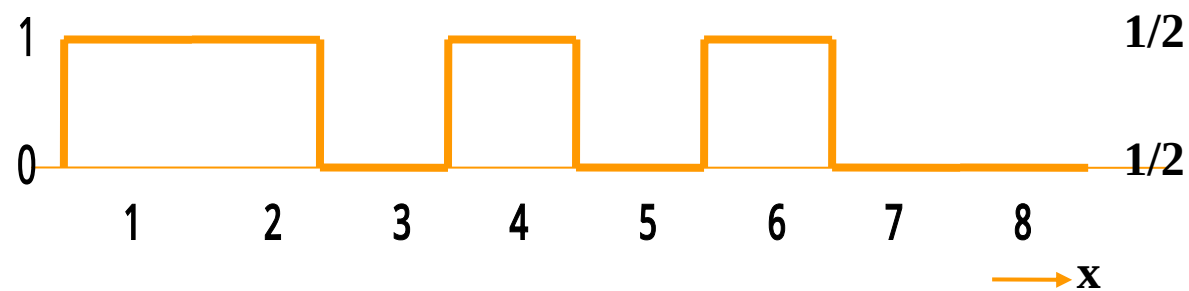
1次元画像

ポジ部とネガ部の混在した画像の位置合わせ
(この位置で正しく対応しているとする)

$f(x)$



$g(x+1)$



g をずらせば, f とよく一致する部分もある

正規化相関

$$\Phi(\lambda) = \frac{\sum \{f(x) - f_{av}\} \cdot \{g(x - \lambda) - g_{av}\}}{[\sum \{f(x) - f_{av}\}^2 \cdot \sum \{g(x - \lambda) - g_{av}\}^2]^{1/2}}$$

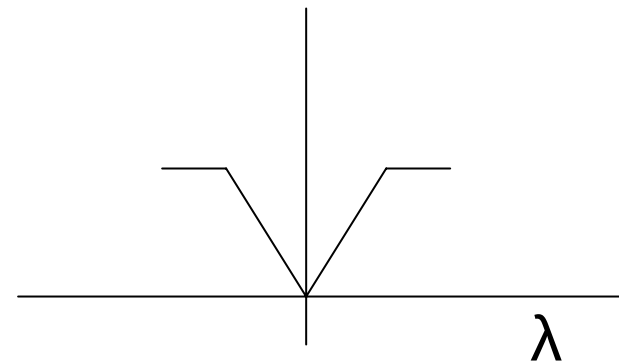
$$\Phi(0) = \frac{1}{2}[(-1)(-1) + 0(1) + 0(1) + 1(-1) + 0(1) + 1(-1) + 0(1) + (-1)(-1)] / []^{1/2} \\ = 0$$

$$\Phi(1) = \frac{1}{2}[(-1)(1) + 0(1) + 0(-1) + 1(1) + 0(-1) + 1(1) + 0(-1) + (-1)(-1)] / []^{1/2} \\ = 1 / 2\sqrt{2}$$

$$\Phi(-1) = 1 / 2\sqrt{2}$$

周辺よりも $\Phi(0)$ が小さい

→ 位置検出できず



ポジ部とネガ部が混在する場合, 正規化相関では位置合わせできない.
正規化相関は, 比例的関係のみカウント

相互情報量

$$\begin{aligned} I(A ; B) &= H(A) - H(A | B) \\ &= - \sum_{i=1}^r P(a_i) \log P(a_i) + \sum_{i=1}^r \sum_{j=1}^s P(b_j) P(a_i | b_j) \log P(a_i | b_j) \\ &= \sum_{i=1}^r \sum_{j=1}^s \{ -P(a_i, b_j) \log P(a_i) + P(b_j) P(a_i | b_j) \log P(a_i | b_j) \} \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \{ -\log P(a_i) + \log P(a_i | b_j) \} \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log \frac{P(a_i | b_j)}{P(a_i)} \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log \frac{P(a_i, b_j)}{P(a_i) P(b_j)}. \end{aligned} \tag{5.10}$$

相互情報量

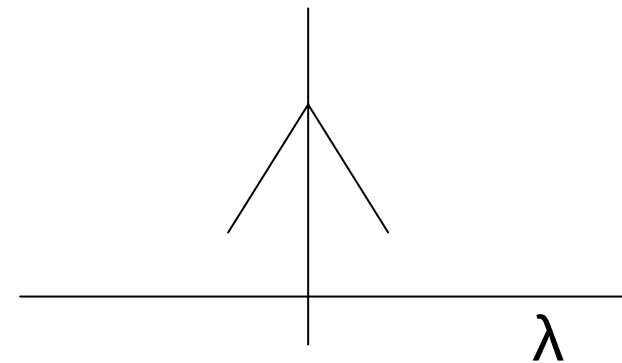
位置ズレなし ($\lambda = 0$)

$$I[f(x); g(x)] = \frac{1}{8} \left[2 \log \frac{\frac{1}{4}}{\frac{1}{4} \cdot \frac{1}{2}} + 4 \log \frac{\frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2}} + 2 \log \frac{\frac{1}{4}}{\frac{1}{4} \cdot \frac{1}{2}} \right]$$
$$= 1$$

$\lambda = 1, -1$

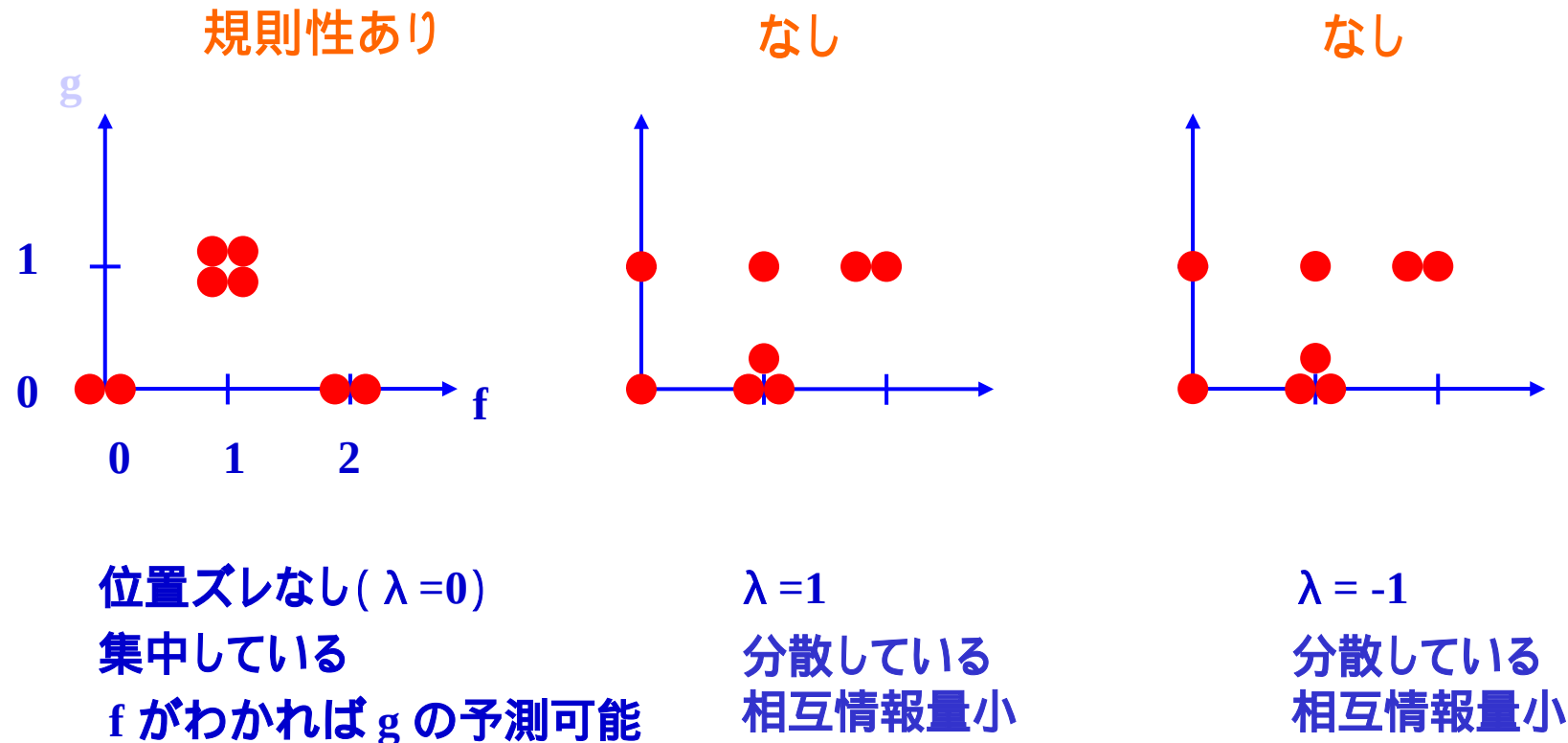
$$I[f(x); g(x - \lambda)] = \frac{1}{8} \left[\log \frac{1/8}{\frac{1}{4} \cdot \frac{1}{2}} + \log \frac{1/8}{\frac{1}{4} \cdot \frac{1}{2}} + 3 \log \frac{3/8}{\frac{1}{2} \cdot \frac{1}{2}} + \log \frac{1/8}{\frac{1}{2} \cdot \frac{1}{2}} + 2 \log \frac{1/4}{\frac{1}{4} \cdot \frac{1}{2}} \right]$$
$$= 0.344$$

正しく位置を決定できる



ポジ部とネガ部が混在しても, 相互情報量では位置合わせできる

同時生起確率 $p(f,g)$ (散布図)



f と g の相互情報量大

f と g が比例していないが、
一定の関係があるため、
相互情報量大 $\rightarrow$ 位置合わせ可能

位置がずれているときの
CTとMRの画素値散布図

位置ずれなしのときの散布図
(a)より集中している

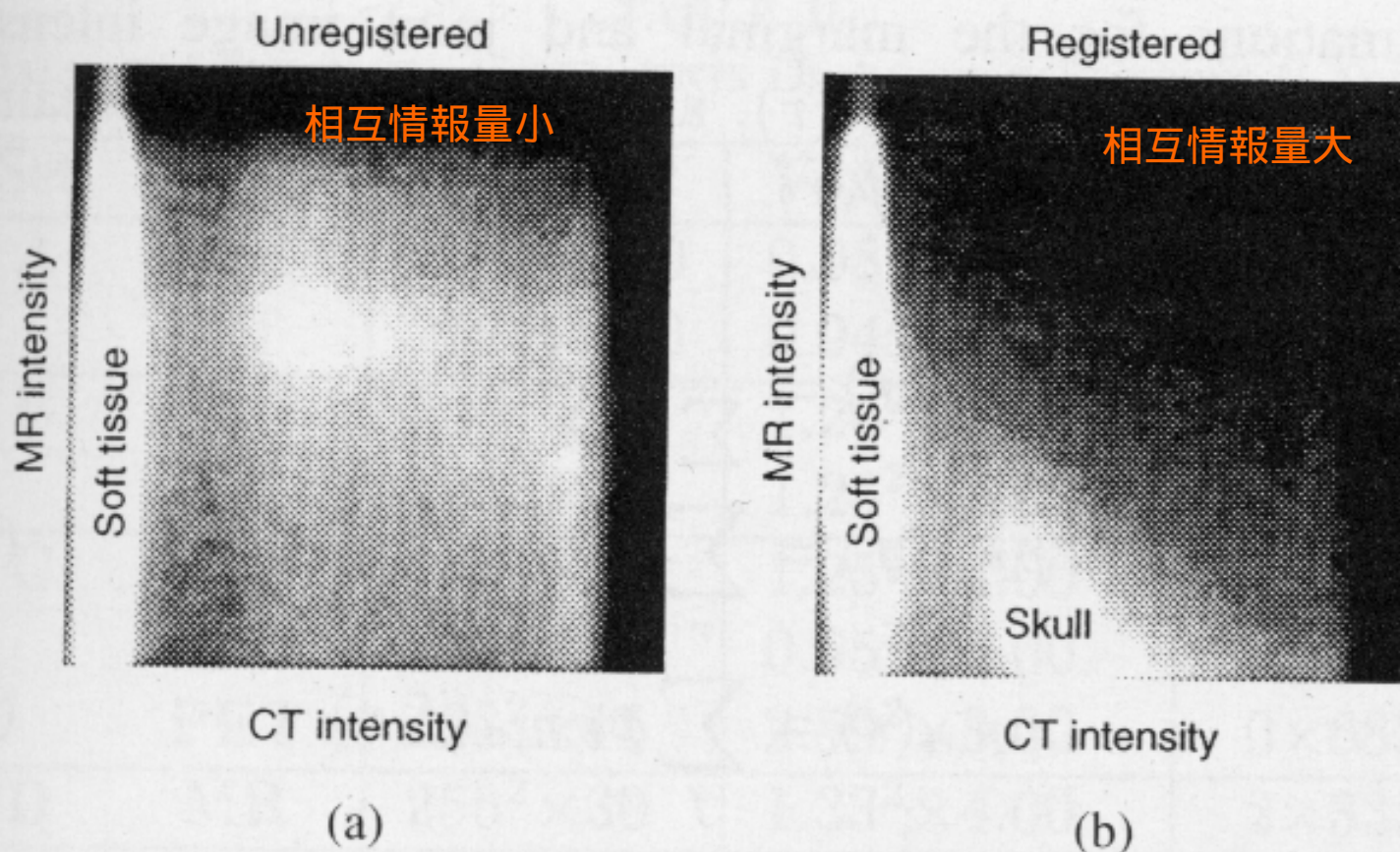
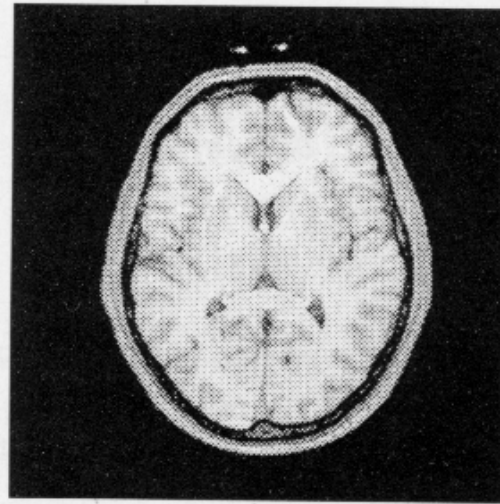


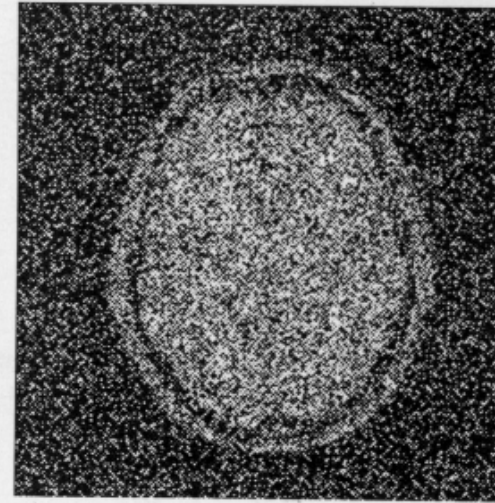
Fig. 1. Joint histogram of the overlapping volume of the CT and MR brain images of dataset A in Tables II and III: (a) Initial position: $I(CT, MR) = 0.46$, (b) registered position: $I(CT, MR) = 0.89$. Misregistration was about 20 mm and 10° (see the parameters in Table III).

相互情報量で位置ずれの判定可能

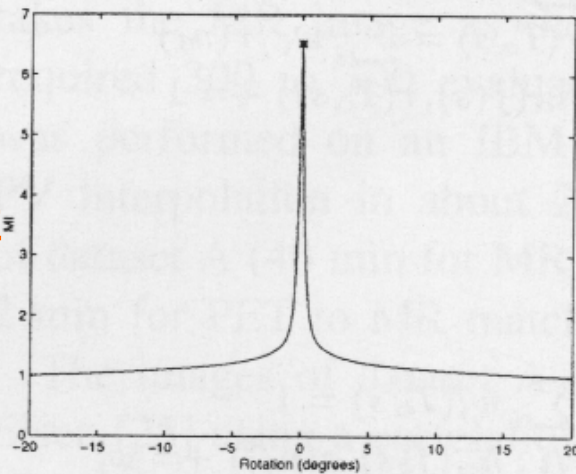
256 × 256
MR画像



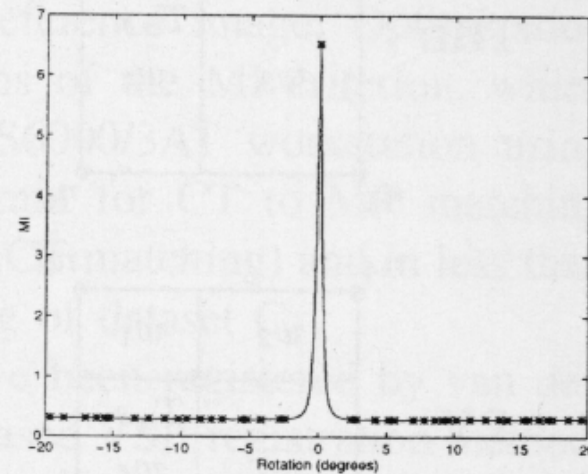
30,000対の画素
値を入れ替え



相互情報量

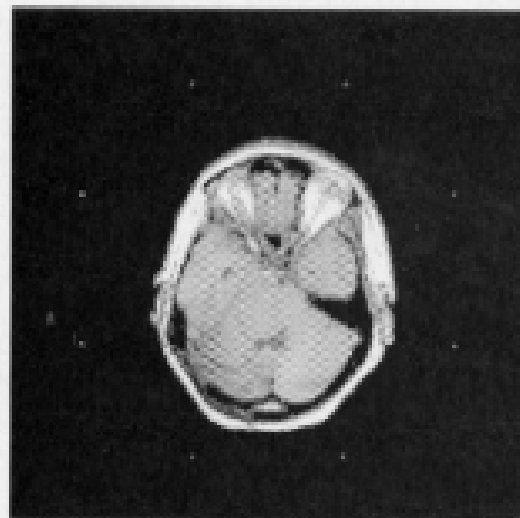


(a) 回転ずれ →

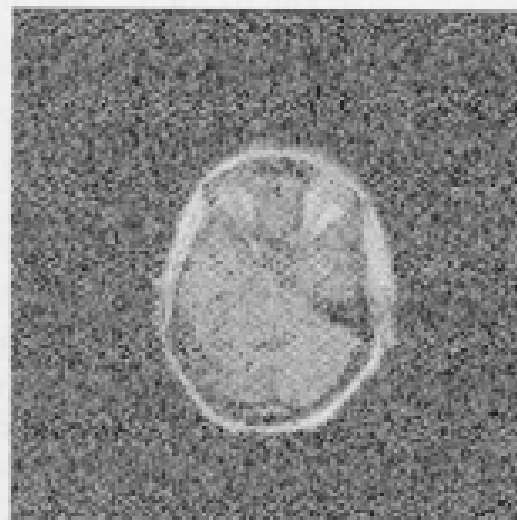


(b)

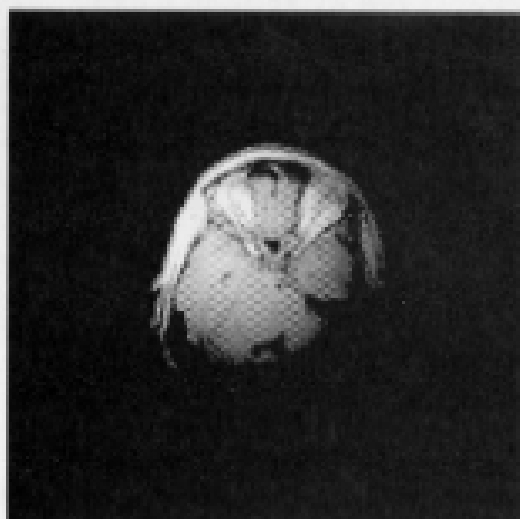
Fig. 2. Spatial correlation of image intensity values increases MI registration robustness. Top: (a) original 256 × 256 2-D MR image and (b) image of (a) shuffled by swapping 30 000 randomly selected pixel pairs. Both images have the same image content. Bottom: MI registration traces obtained using partial volume distribution (PV) interpolation for in-plane rotation of each image over itself. Local maxima are marked with “*”.



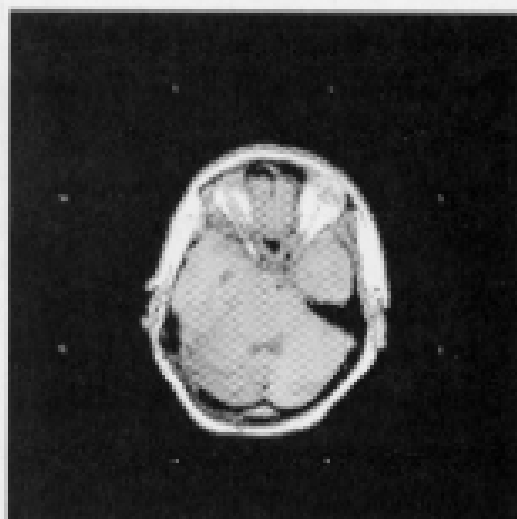
(a)



(b)



(c)



(d)

Fig. 8. (a) Slice 15 of the original MR image of dataset D, (b) zero mean noise added with variance of 500 grey-value units, (c) quadratic inhomogeneity ($k = 0.004$), and (d) geometric distortion ($k = 0.00075$).

各種の歪みに対しても相互情報量基準にしたがう位置合わせは強い

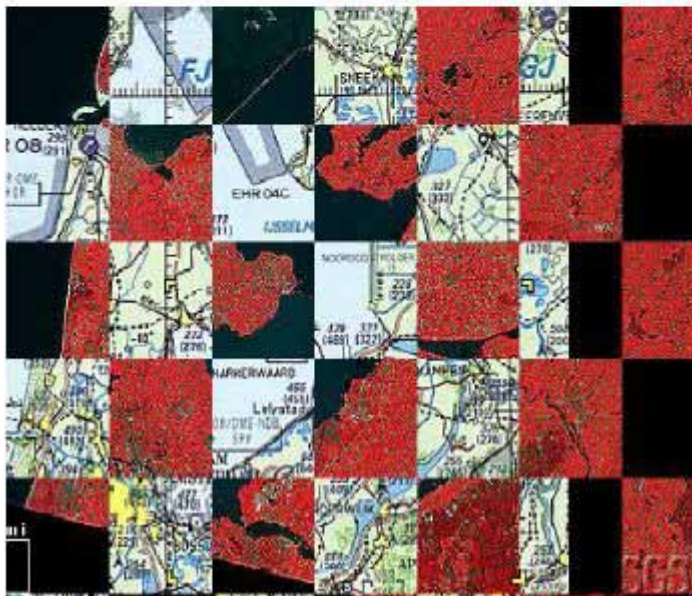
航空写真



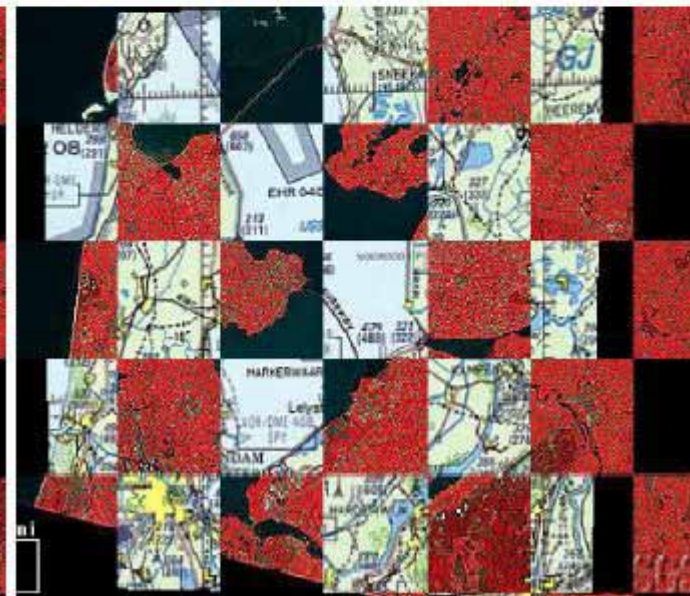
地図



位置合
わせ前

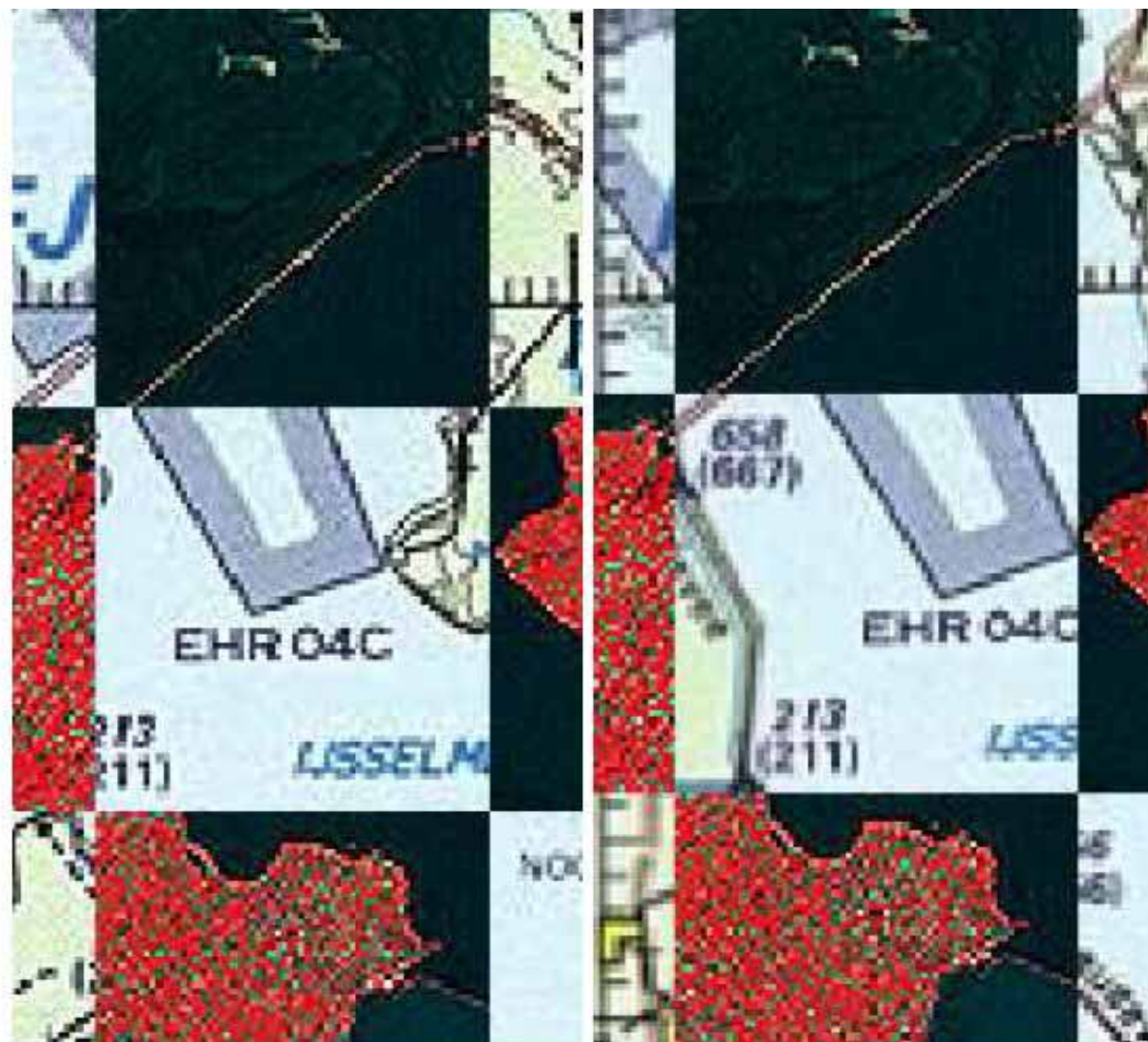


位置合
わせ後



相互情報量による航空写真と地図の位置合わせ

位置合
わせ前



位置合
わせ後

Figure 3. Homography registration of a map with color image using MI. top: original images; middle: before (left) and after registration (right); bottom: zoomed view before (left) and after (right) registration.

相互情報量による位置合わせ (拡大図)

http://earth.esa.int/rtd/Events/ESA-EUSC_2004/Ar24_Strecha.pdf

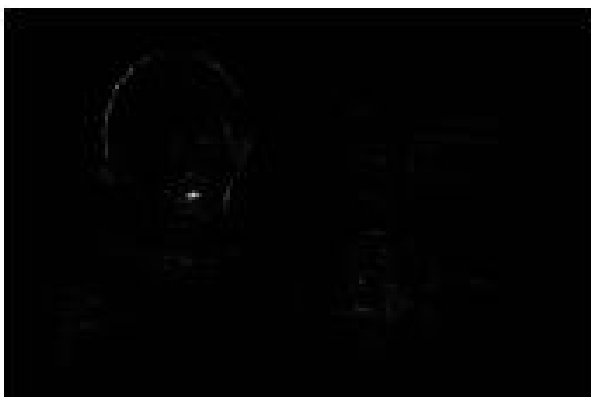
音- 画像間の相互情報量による音源探索



(a) 原画 . 左:発話 右:口パクパク



(b) 標準偏差画像



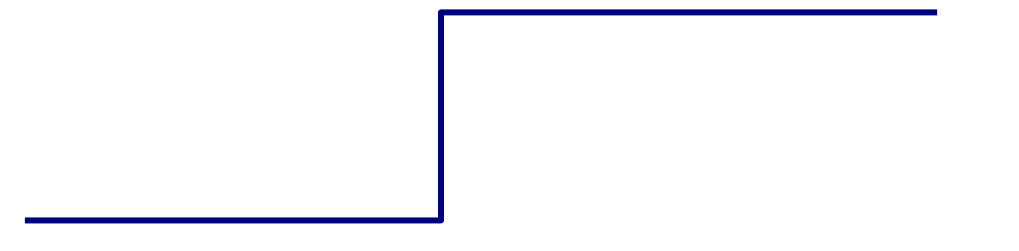
(c) 音声と画像間の相互情報量
口の位置を正しく捉えている



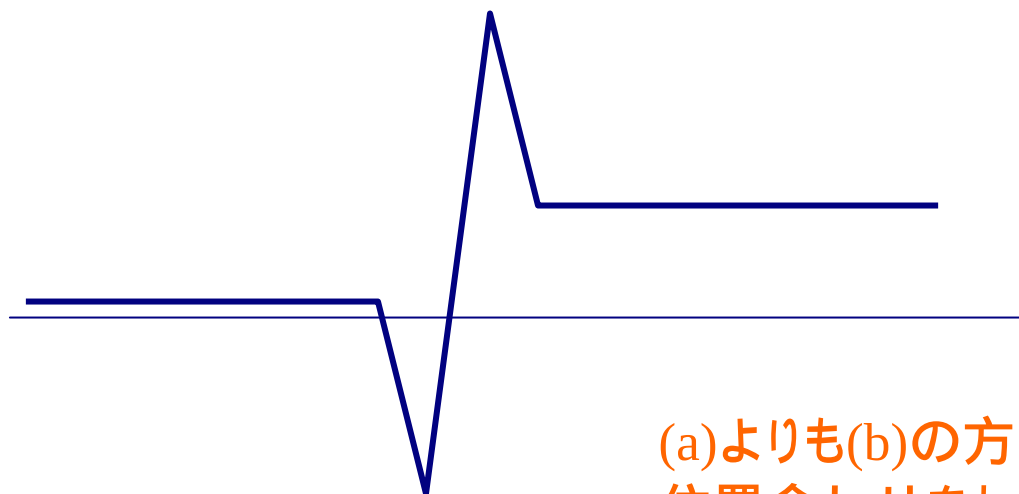
(d) 異なる音声との相互情報量
音源を特定できない

2次微分の効果

(a) 原画像



(b) 原画像 +
2次微分

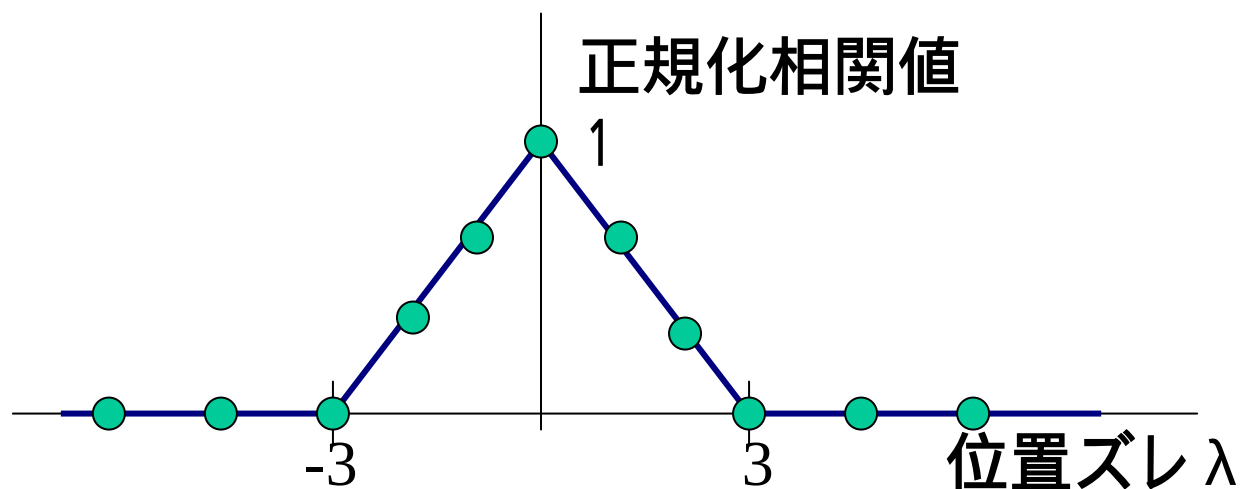


(a)よりも(b)の方が正規化相関による
位置合わせをし易いことを示す

原画像

正規化相関値の広がり大きい
位置合わせしにくい

$f(x)$	0	0	0	1	1	1
$2*[f(x)-f_{av}]$	-1	-1	-1	1	1	1
$2*[f(x+1)-f_{av}]$	-1	-1	1	1	1	1
$2*[f(x+2)-f_{av}]$	-1	1	1	1	1	1

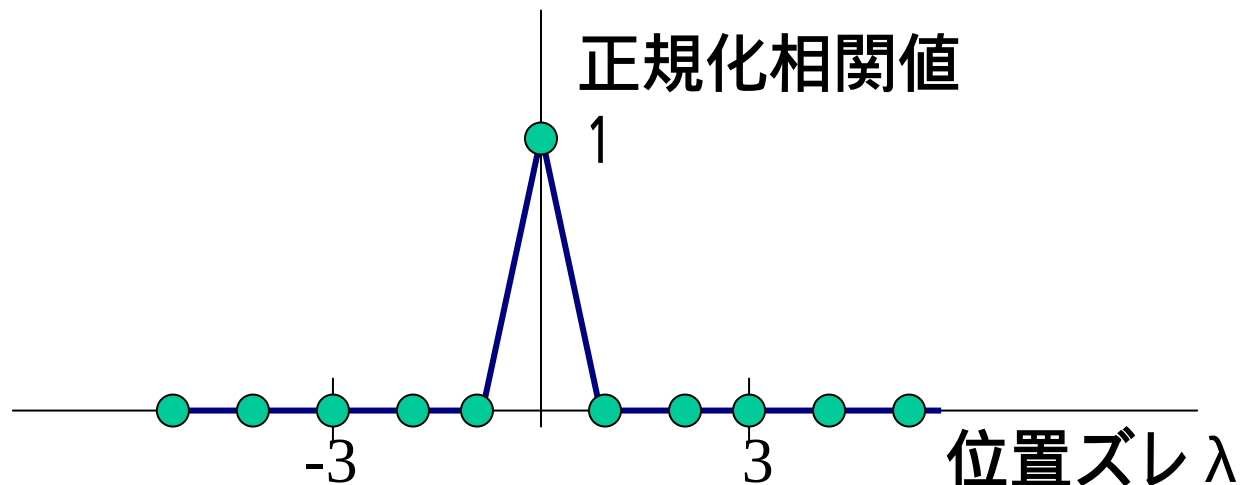


原画像 + 2次微分

正規化相関値の広がり小さく,
位置合わせしやすい

$f(x)$	0	0	-1	2	1	1
$2*[f(x)-f_{av}]$	-1	-1	-3	3	1	1
$2*[f(x+1)-f_{av}]$	-1	-3	3	1	1	1
$2*[f(x+2)-f_{av}]$	-3	3	1	1	1	1

例 $\sum 2*[f(x)-f_{av}]*2*[f(x+1)-f_{av}]=1+3-9+3+1+1=0$
 $f_{av}=0.5$



相互情報量の利用 まとめ

- 相互情報量とは,
一方 (f) から, もう一方 (g) がどれくらい予測可能であるかの尺度.
- それに対して, 相関は, f と g の比例的関係の尺度
(一部反転などがあると使用不可)
- 独立成分分析 (ICA)
データの独立性基準 (たとえば 分離された信号 f, g 間の相互情報量が最小になるように) により, 信号を分離
- 位置合わせ (registration)
異種画像間では相互情報量基準が効果的. 音と画像間でも可能

MP3 圧縮

F F T (fast Fourier transform)と同じ種類のM D C T (modified discrete cosine transformation

= 変形離散コサイン変換)を使って周波数領域を32分割し、最小可聴以下の信号を削除する。

マスキング効果で聴こえない音を削除する。

サブバンドの最大振幅が1になるように正規化し、データを Huffman符号化(出現頻度の高いデータを少ないビット数で、頻度の低いデータを長いビット数で表すことにより、データ量を減少させる方法=可変長符号)で削減する。

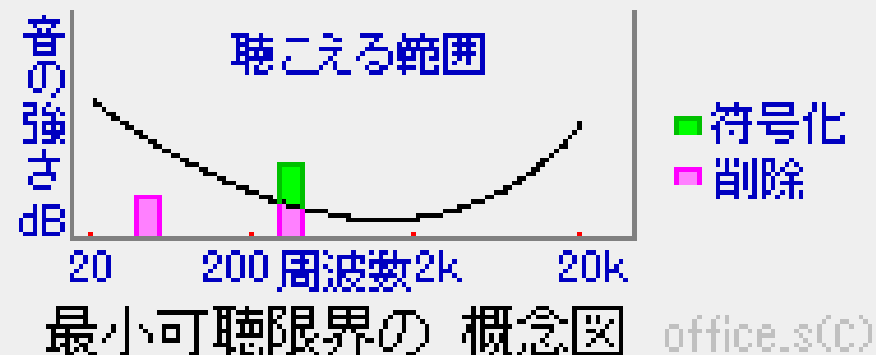
データの多いバンドに多量のビットを割り当てる。

ステレオ信号 L(Left) と R(right) は、 $L+R$ ・ $L-R$ に分けて符号化する
MS(Middle/Sides)ステレオ、等の手法で聴覚特性に合わせた効率のよい圧縮を行なうこともできます。

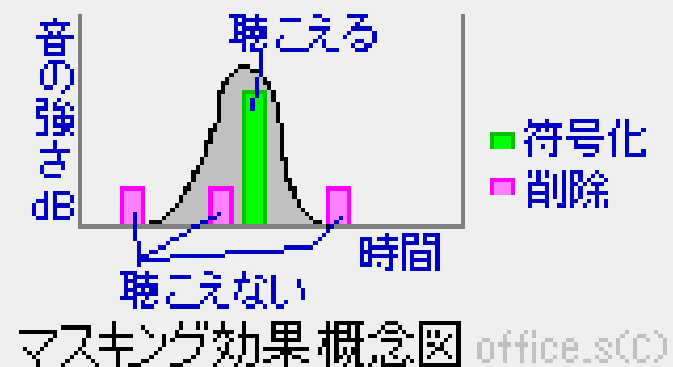
http://homepage1.nifty.com/office_s/w_knowledge.htm#knowledge-compres-mp3

MP3において不要と判断した情報とは:

最小可聴限界の外側の音は聴こえないので、削除します。



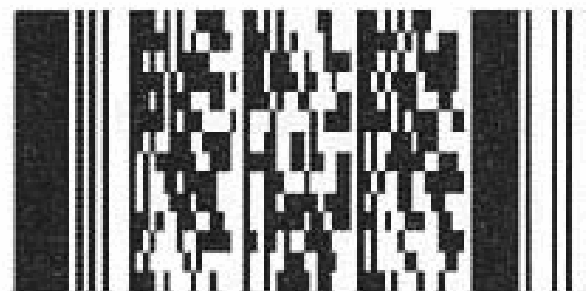
大きな音の直前直後にある周波数が同程度の小さな音や、大きな音の中に埋もれた周波数が同程度の小さな音などは、聴き取れません。これをマスキング効果と呼びます。そこで、マスキング効果によって聴き取れない音は削除します。





QRCode

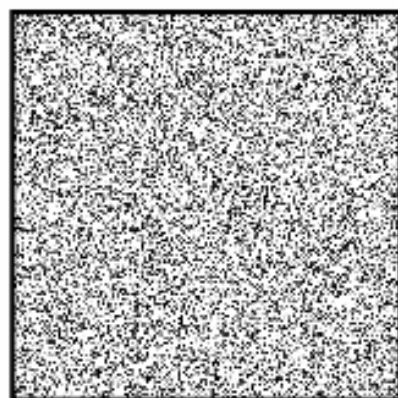
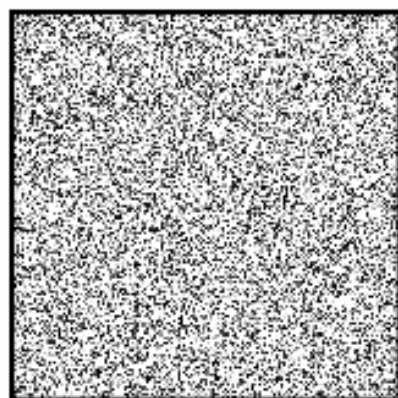
2次元バーコード



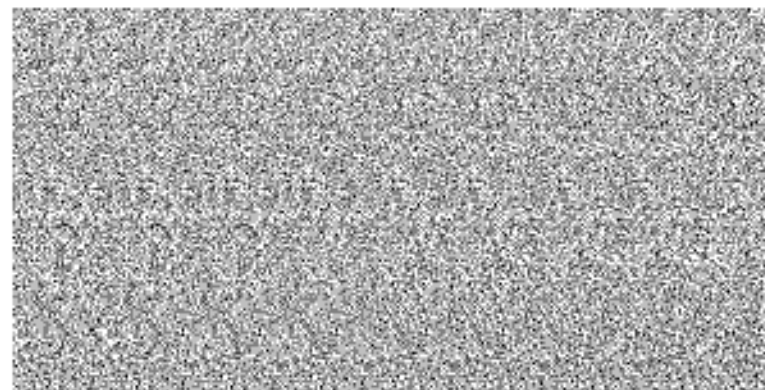
PDF417



JANコンポジットの例



左右眼RDS

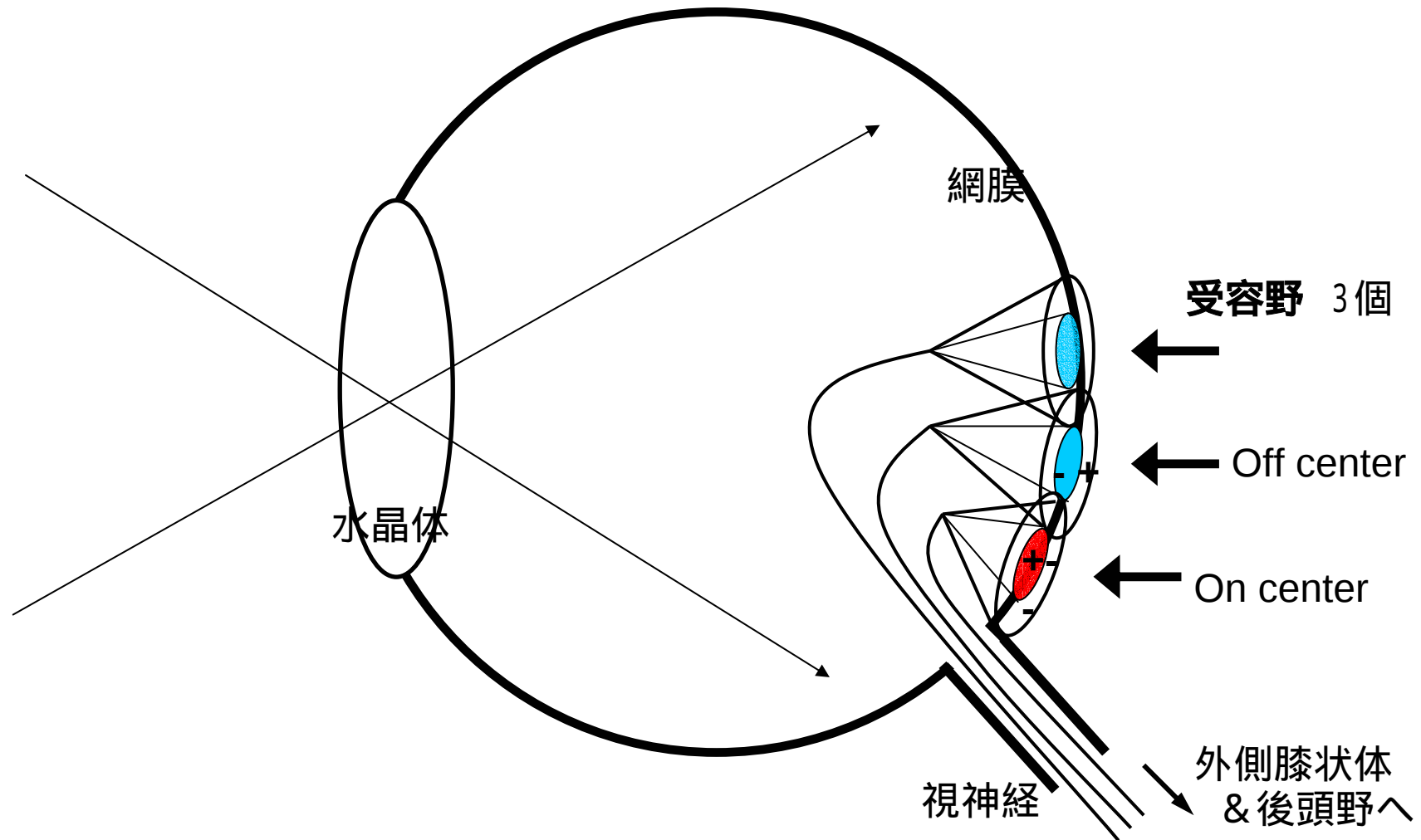


単一画像RDS

情報圧縮しにくい画像例

RDS=Random Dot Stereogram

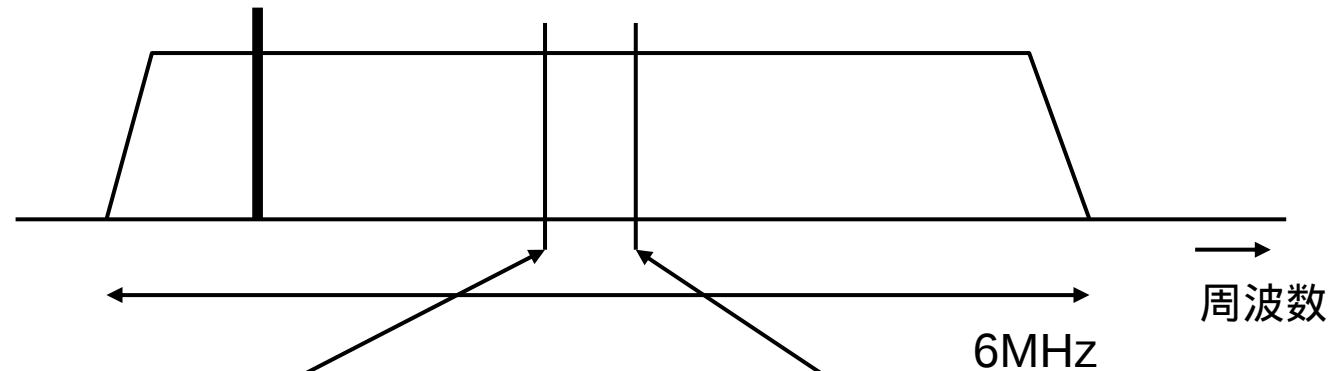
受容野



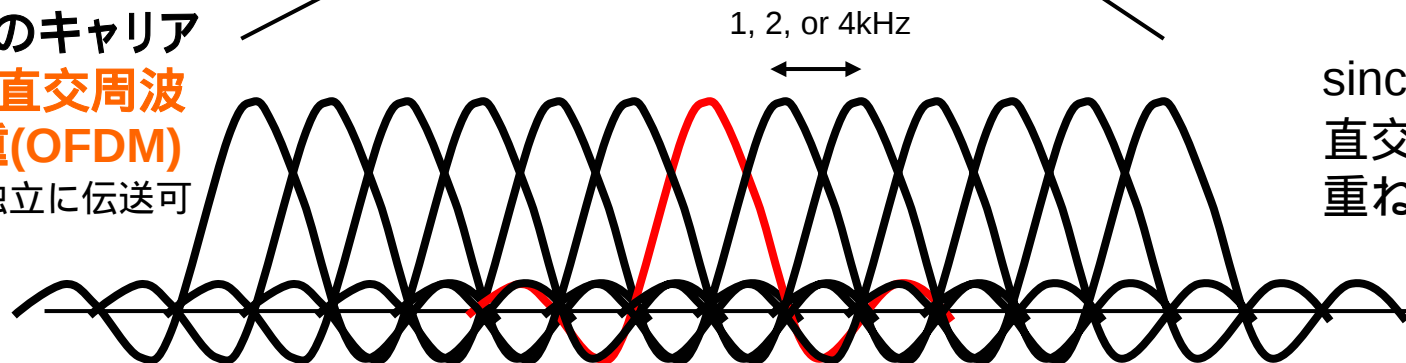
TV放送

地上波アナログ
TV放送

1個の
キャリア (搬送波)

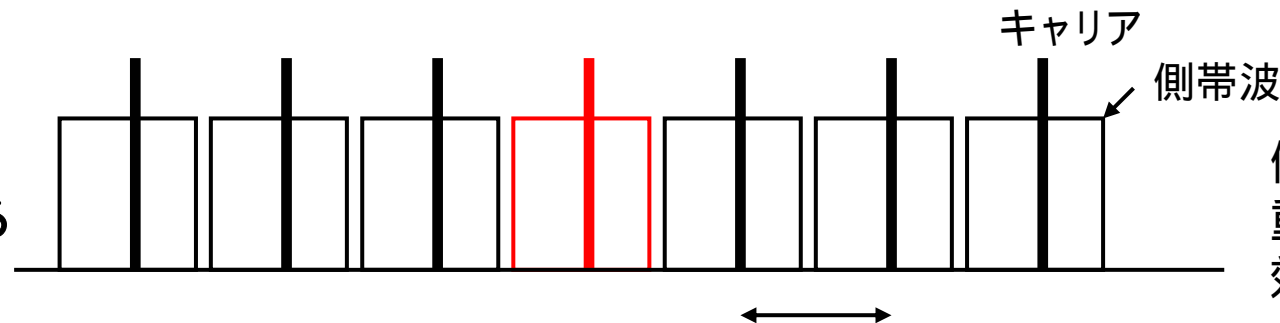


地上波デジタルTV
では, 多数のキャリア
を使用する**直交周波
数分割多重(OFDM)**
各キャリアは独立に伝送可



sinc関数は
直交しており,
重ねて密に伝送

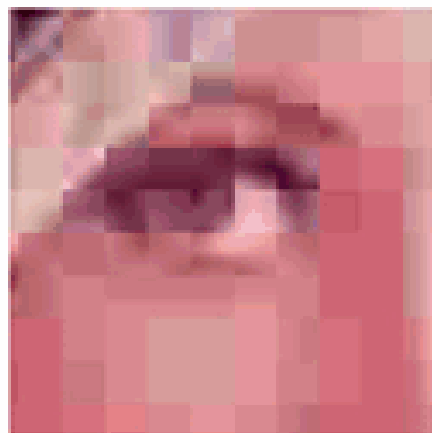
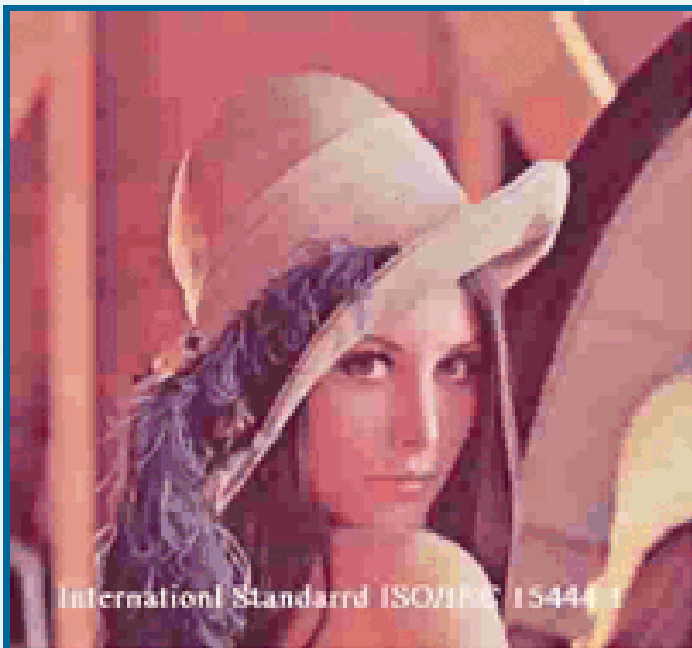
中波放送などは
完全に分離させる
**周波数分割多重
(FDM)**



側帯波が
重ならない.
効率低い

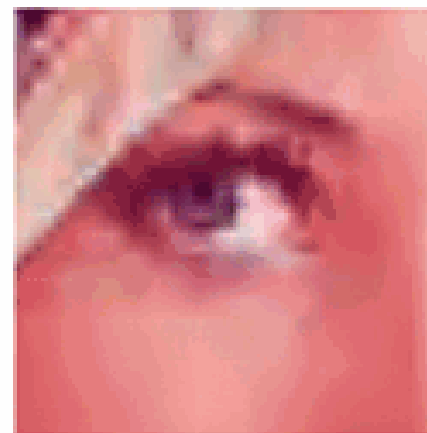
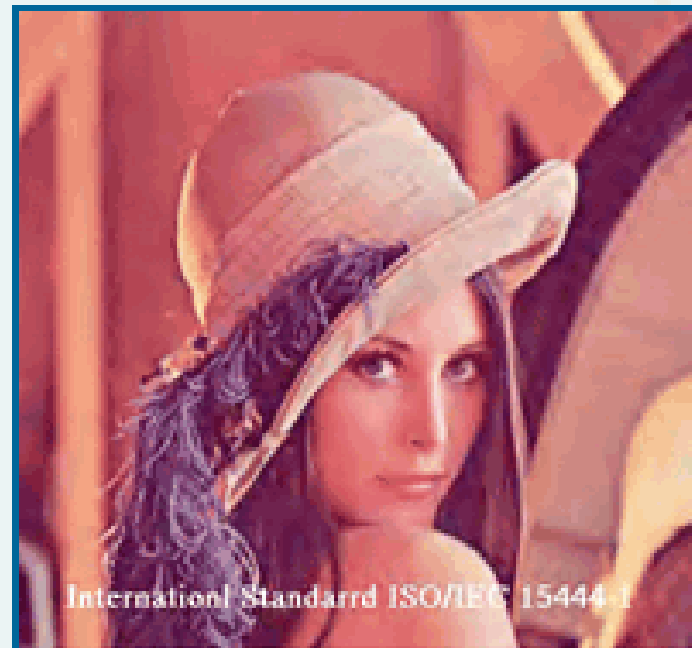
中波放送では 9 kHz 間隔

JPEG



ブロックノイズあり

JPEG2000



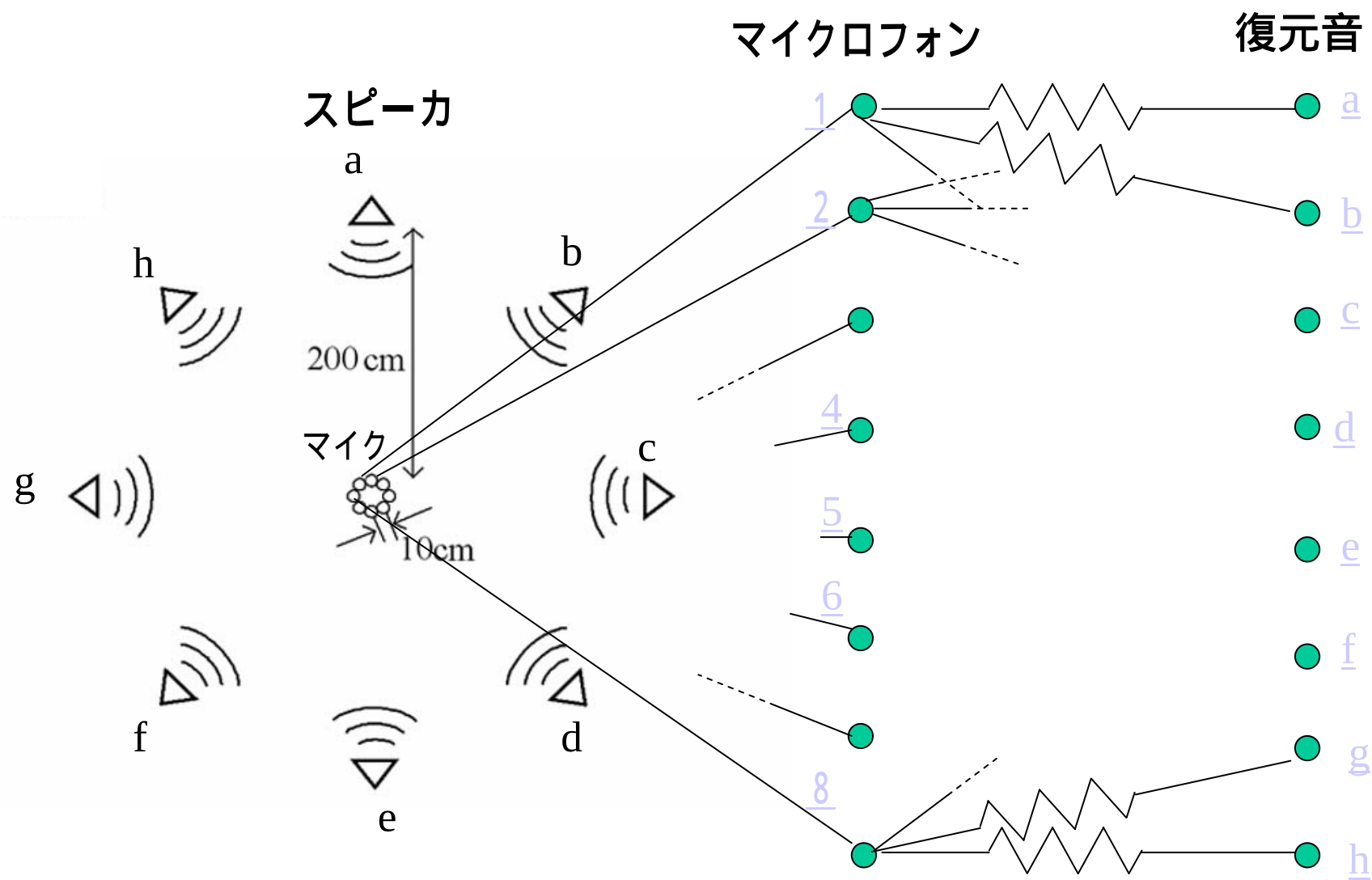
なし

JPEG2000 (wavelet 変換利用)



独立成分分析による音声分離(聖徳太子プロジェクト)

http://www.brain.kyutech.ac.jp/~matsuoka/S_project.htm



独立成分分析による音声分離 (聖徳太子プロジェクト)

医用画像処理ソフトウェアパッケージMEDAL による画像処理例

http://www.image.med.osaka-u.ac.jp/reza/medal_home.html