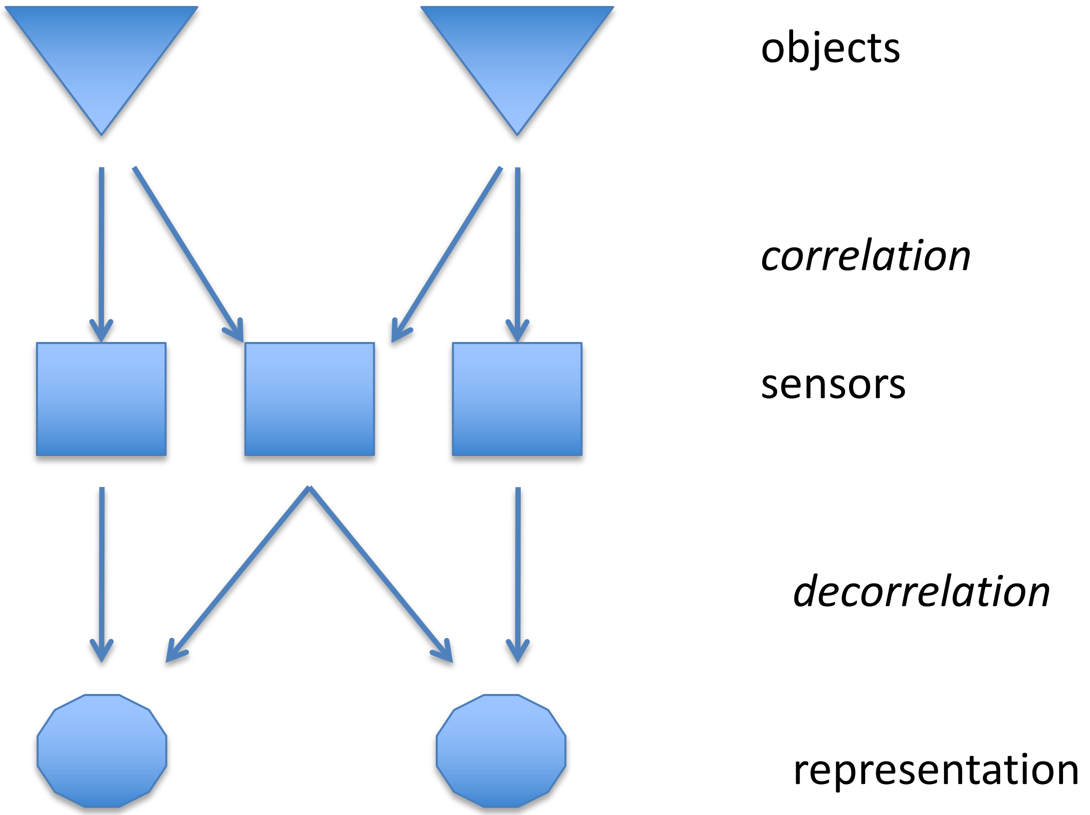
02458 Cognitive Modeling

Natural world statistics

A modest proposal

- The perceptual system is optimized to represent information relevant for survival
  - The dynamic range of neurons match the dynamic range of stimuli
  - Independent image components (e.g. objects) are coded in independent channel
- Testable hypothesis
  - If we know the statistics of natural images

Efficient representations



Joint (pixel) probabilities

- Images are parameterized as vectors of pixels
  - For each color channel
  - For now, we just look at luminance (b/w)
- How do we parameterize *P(I)*?
  - As a joint probability $P(I_1,I_2,\ldots,I_N)$
  - Pixels are not independent $P(I_1,I_2)\sim=P(I_1)(I_2)$
  - How do we describe dependence?

Covariance

- Variance: $E\left(\left(x-\mu_x\right)^2\right) \cong E\left(\left(x-\bar{x}\right)^2\right)$
- Covariance: $E\left(\left(x-\mu_y\right)\left(y-\mu_y\right)\right) \cong E\left(\left(x-\bar{x}\right)\left(y-\bar{y}\right)\right)$
- For independent variables:

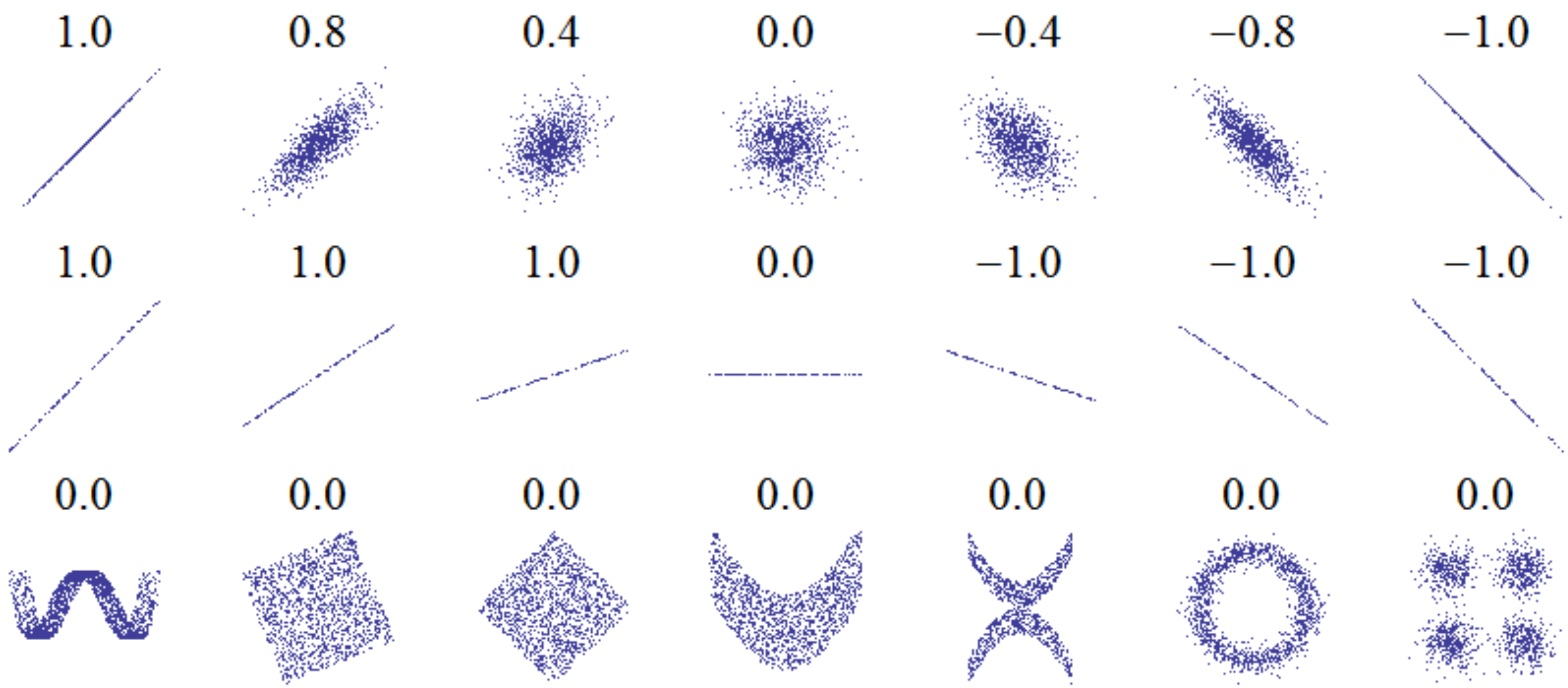
$$E\left(\left(x-\bar{x}\right)\left(y-\bar{y}\right)\right) = \sum_{i,j} P\left(x_i, y_j\right)\left(x_i-\bar{x}\right)\left(y_j-\bar{y}\right) =$$

$$\sum_{i,j} P\left(x_i\right)P\left(y_j\right)\left(x_i-\bar{x}\right)\left(y_j-\bar{y}\right) = \sum_i P\left(x_i\right)\left(x_i-\bar{x}\right)\sum_j P\left(y_j\right)\left(y_j-\bar{y}\right) =$$

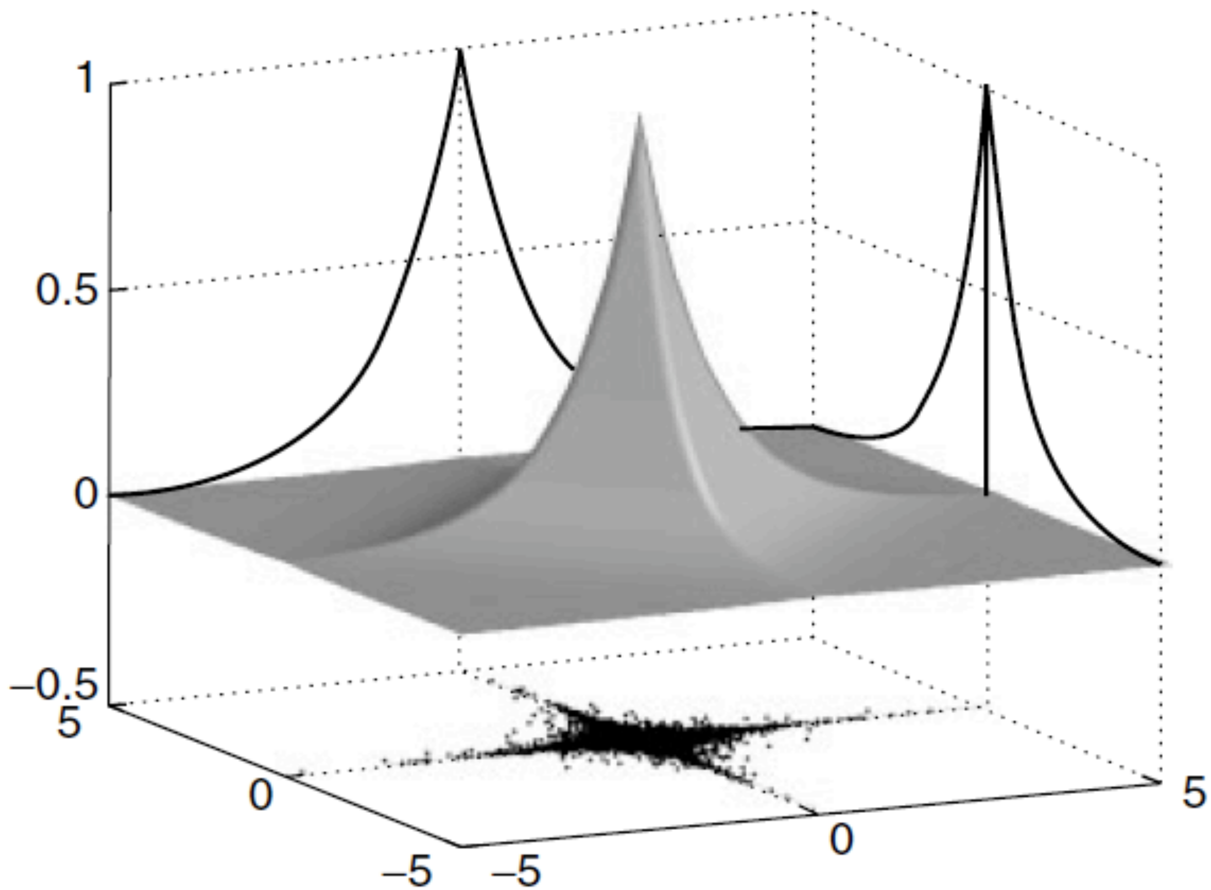
$$\left(\sum_i P\left(x_i\right)x_i - \bar{x}\sum_i P\left(x_i\right)\right)\left(\sum_i P\left(y_i\right)y_i - \bar{y}\sum_i P\left(y_i\right)\right) =$$

$$\left(\bar{x}-\bar{x}\right)\left(\bar{y}-\bar{y}\right) = 0$$

Covariance does not imply independence



Natural statistics are not Gaussian

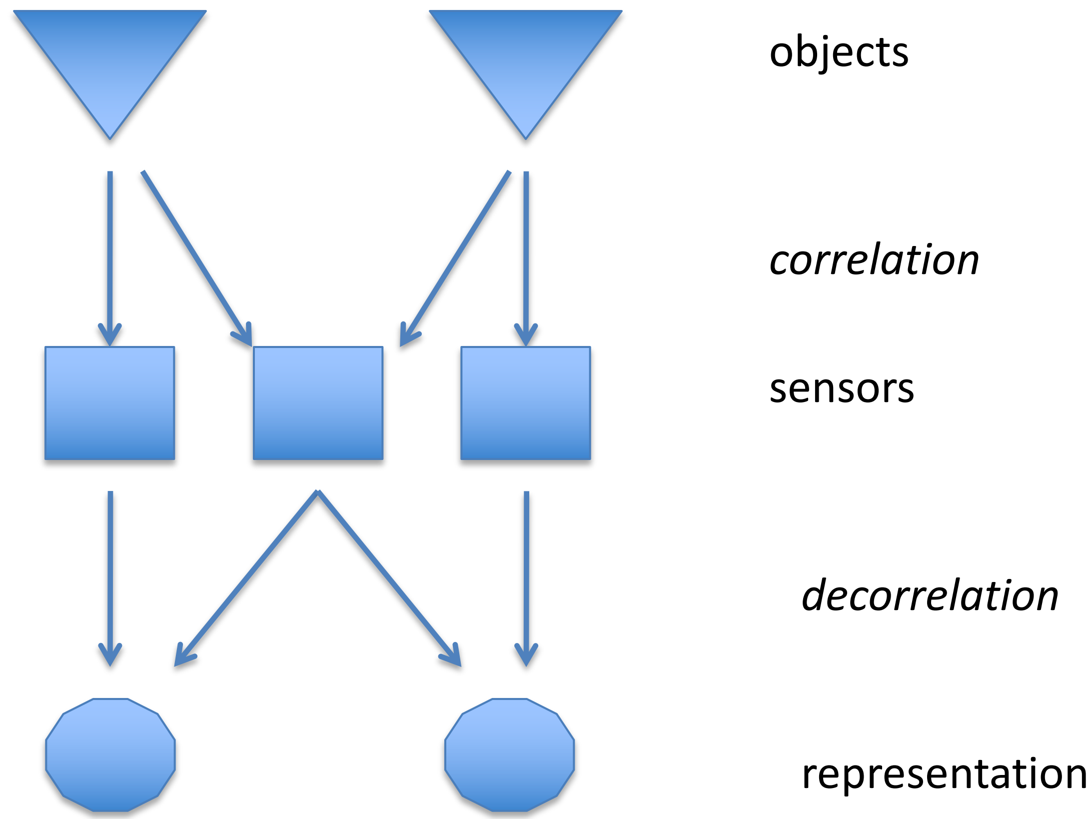


By Stone, JW

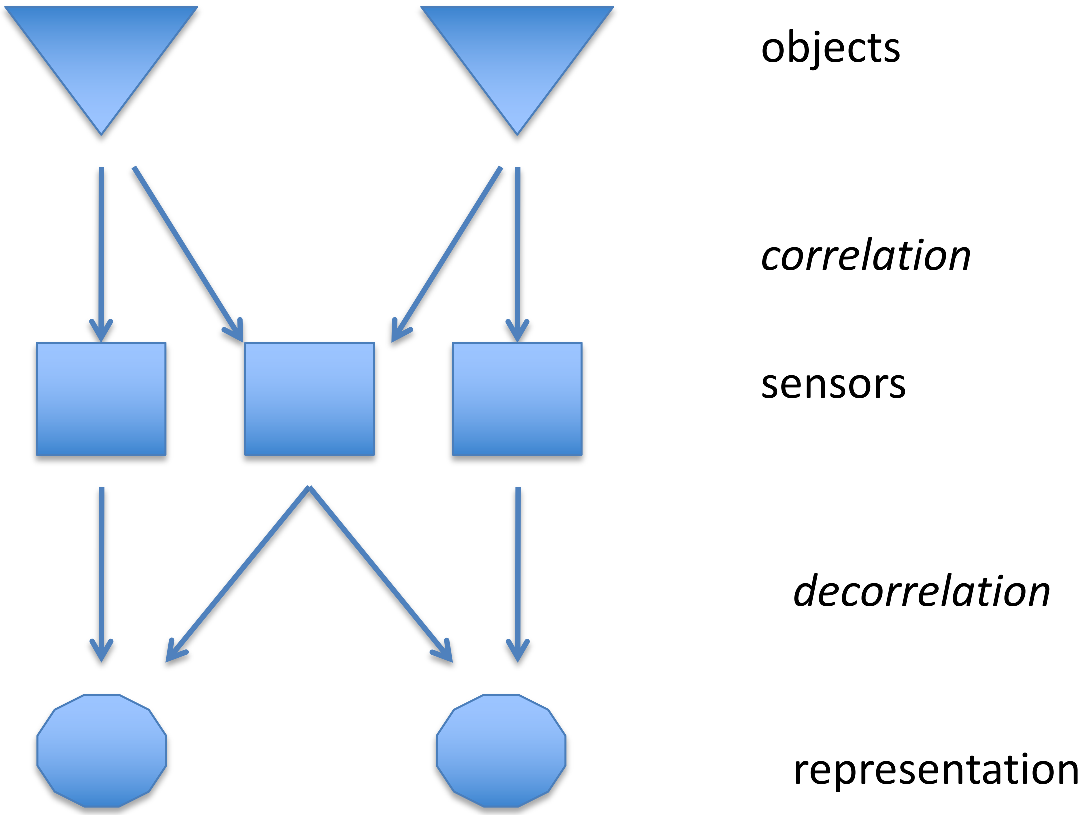
Natural statistics are sparse

- Sparseness means
  - Frequent, Low amplitude events
  - Infrequent, high amplitude events
- Sparseness is approximately
  - The fourth moment, kurtosis, $E((x-\mu)^4)$
- Sparseness implies structure
  - Gaussian distribution implies randomness

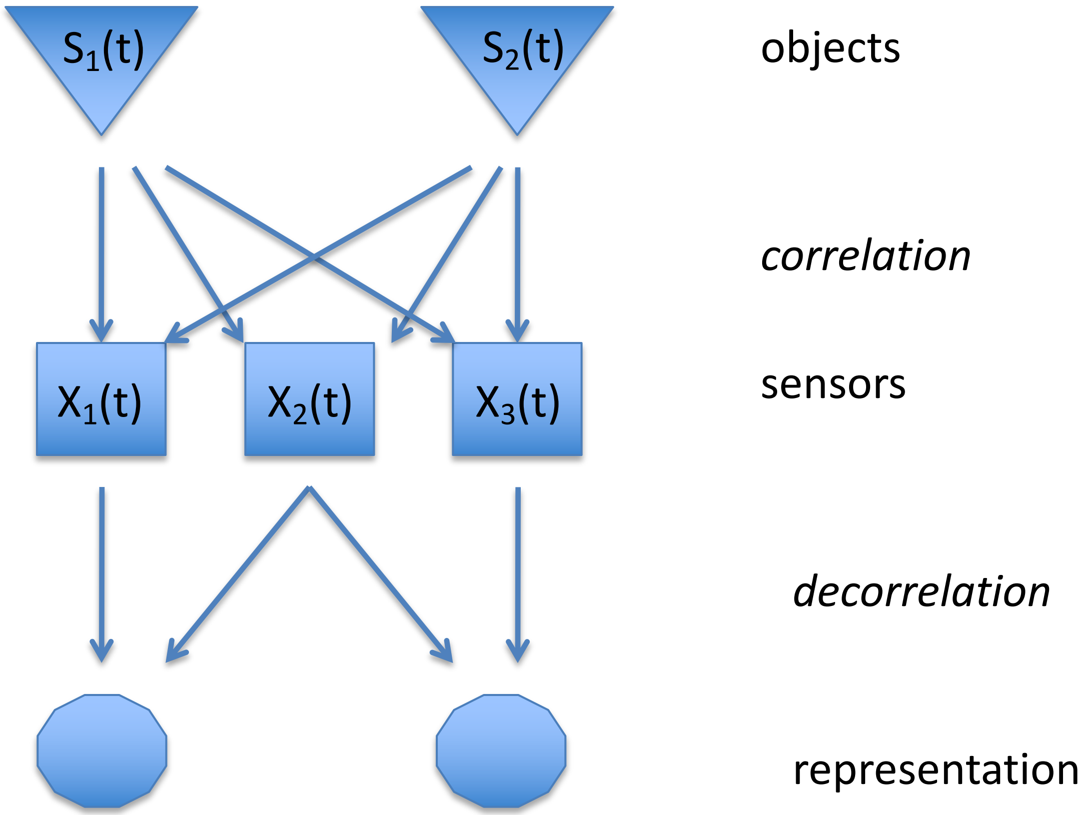
Efficient representations



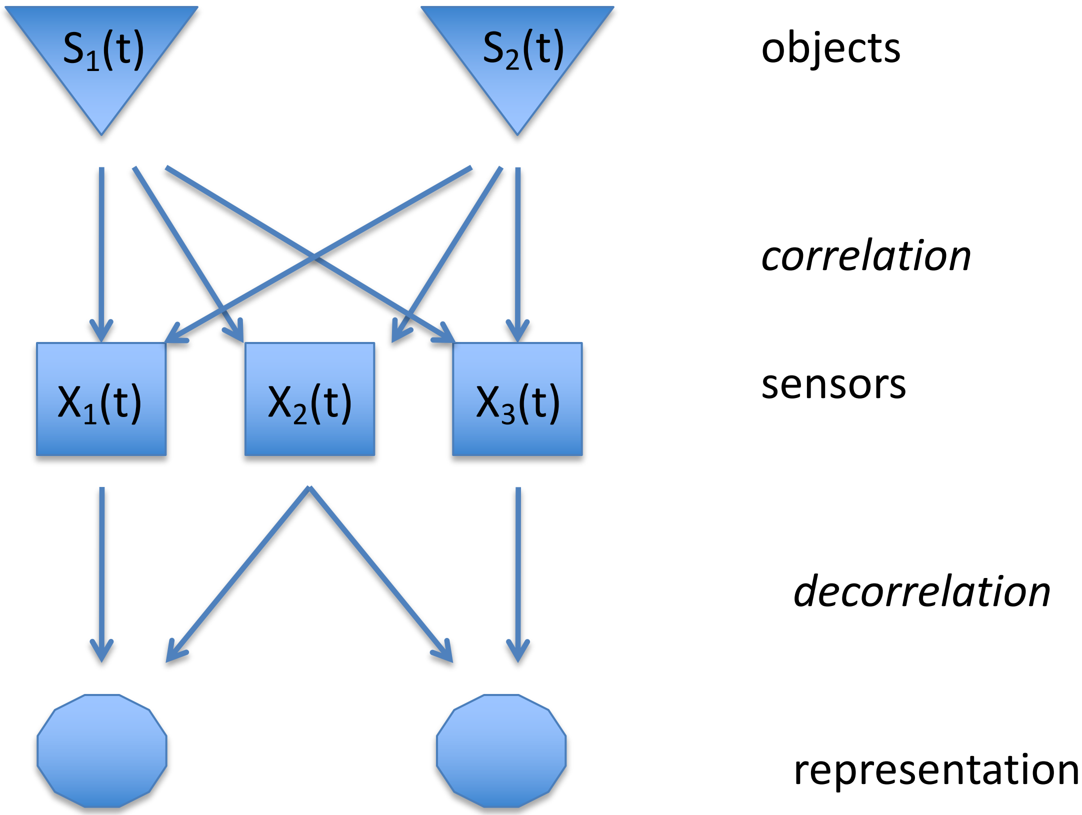
Efficient representations



Efficient representations



Efficient representations



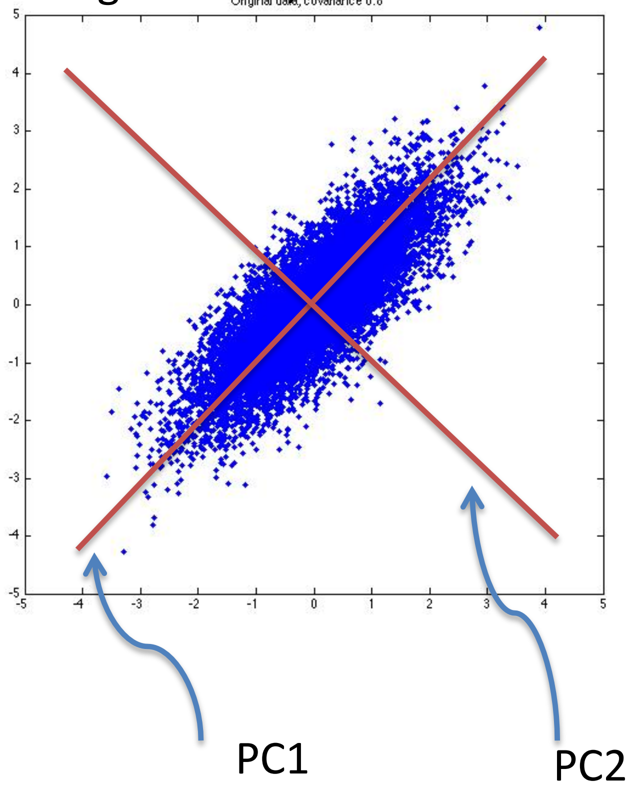
Efficient representations

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t)$$
$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t)$$
$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)$$
$$\Updownarrow$$
$$\overline{x}(t) = \overline{a}_1 s_1(t) + \overline{a}_2 s_2(t) + \overline{a}_3 s_3(t)$$
$$\Updownarrow$$
$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} \overline{a}_1 & \overline{a}_2 & \overline{a}_3 \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{bmatrix}$$
$$X = AS$$

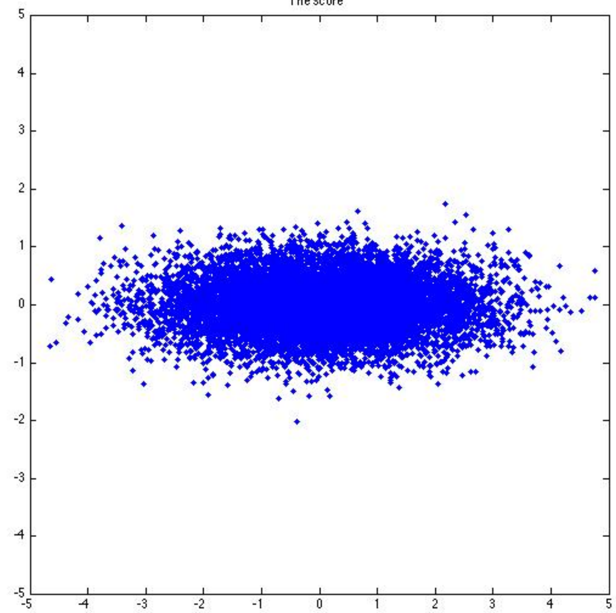
The problem:
Knowing the sensor activation x(t)
We need to find the stimulus/signal s(t)
AND
The weights / mixing matrix A

Principal component analysis (PCA)

Original data, covariance 0.8

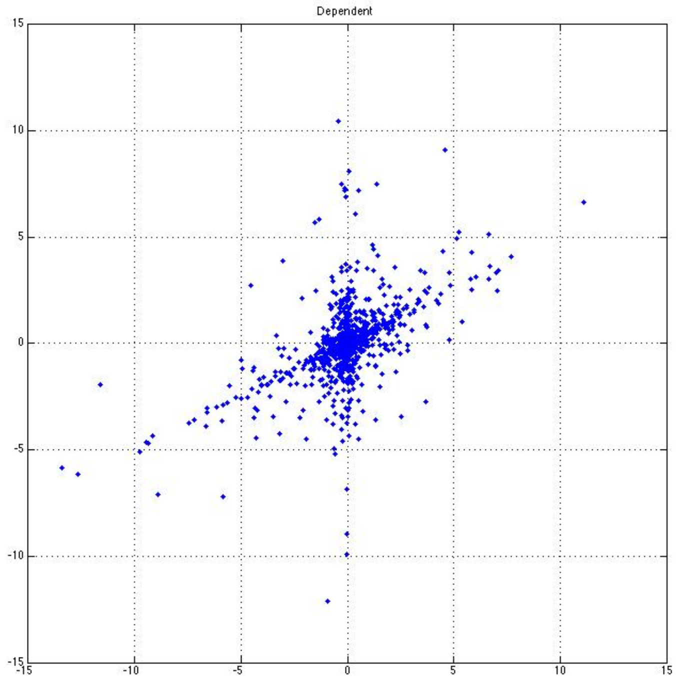


The score

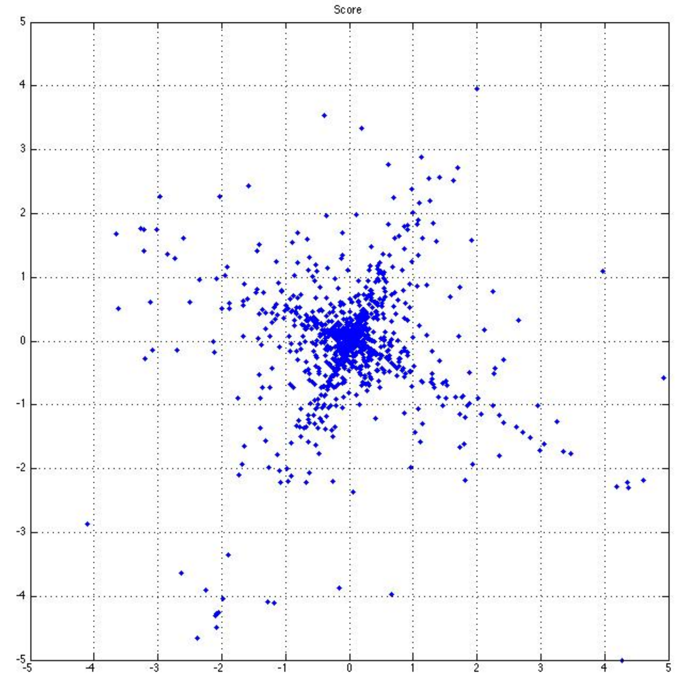


Notice greatest variance along 1^{st} dimension

Principal component analysis (PCA)



Non-Gaussian dependent joint distribution

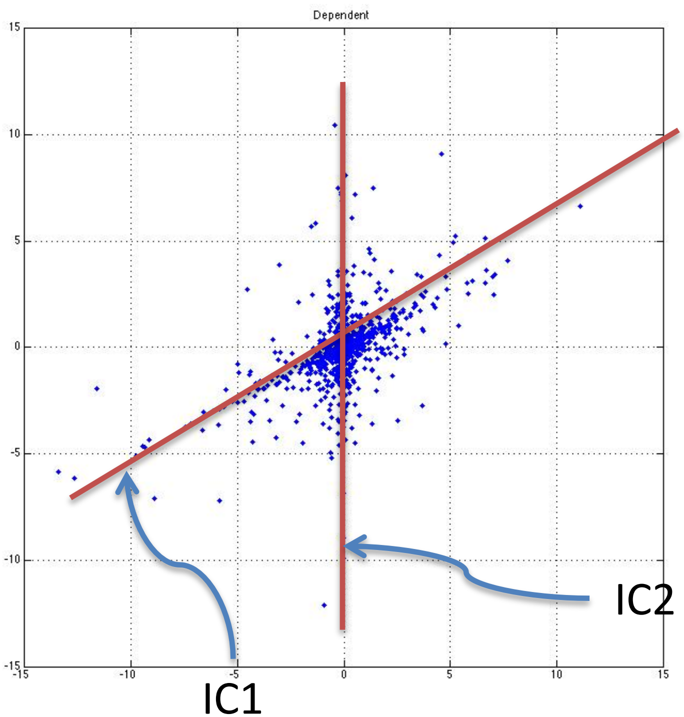


The PCA score is still dependent
“Rays” radiate from origo in directions not following the axis

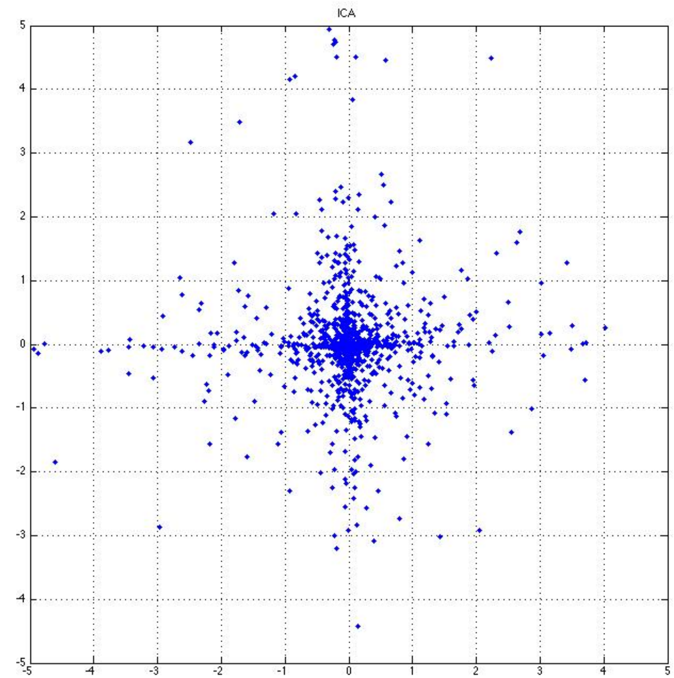
Need for something else

- PCA minimizes covariance / maximize variance
  - Assuming orthogonality
- We need to minimizes dependence / maximize sparseness
  - Without assuming orthogonality
- Independent Component Analysis
  - Maximizing kurtosis on non-orthogonal components

Independent component analysis (ICA)



Non-Gaussian
dependent joint
distribution

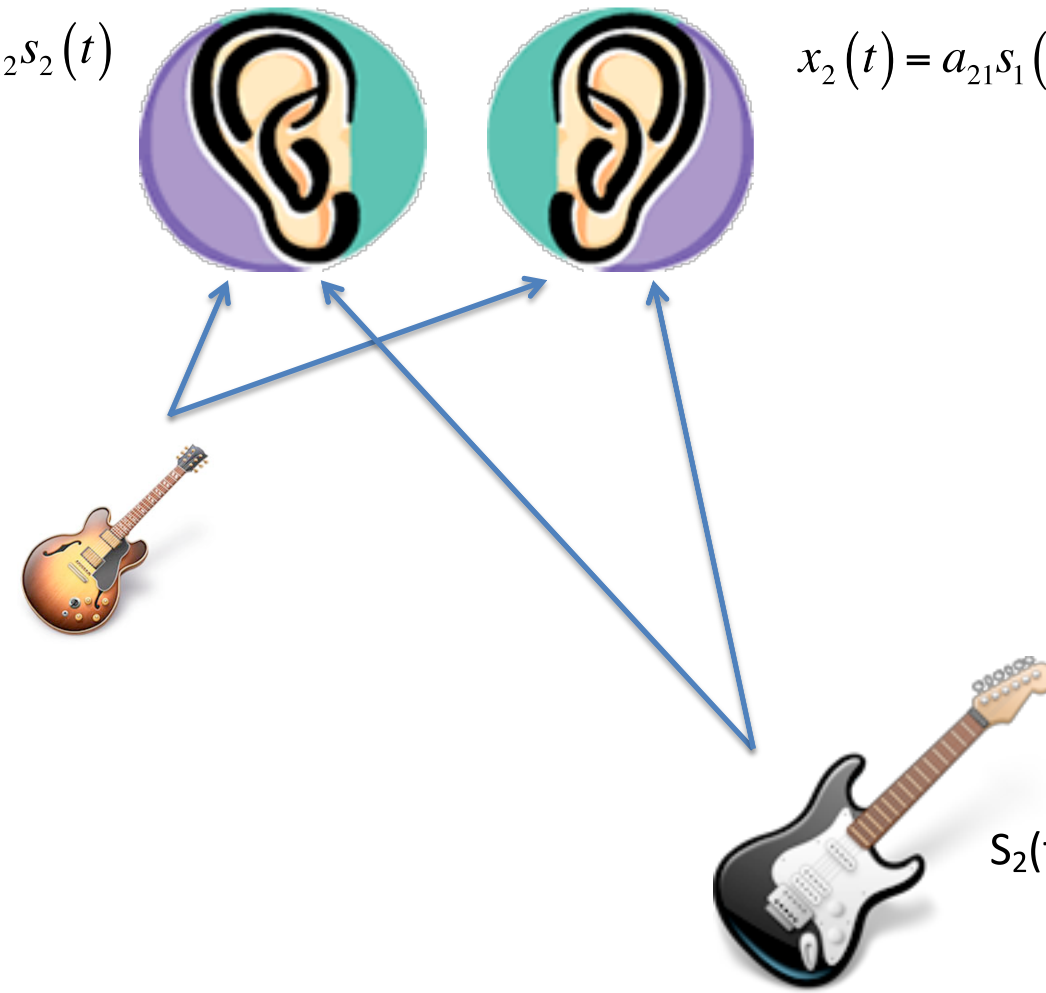


The ICA score is
independent
“Rays” radiate from origo in
directions following the axis

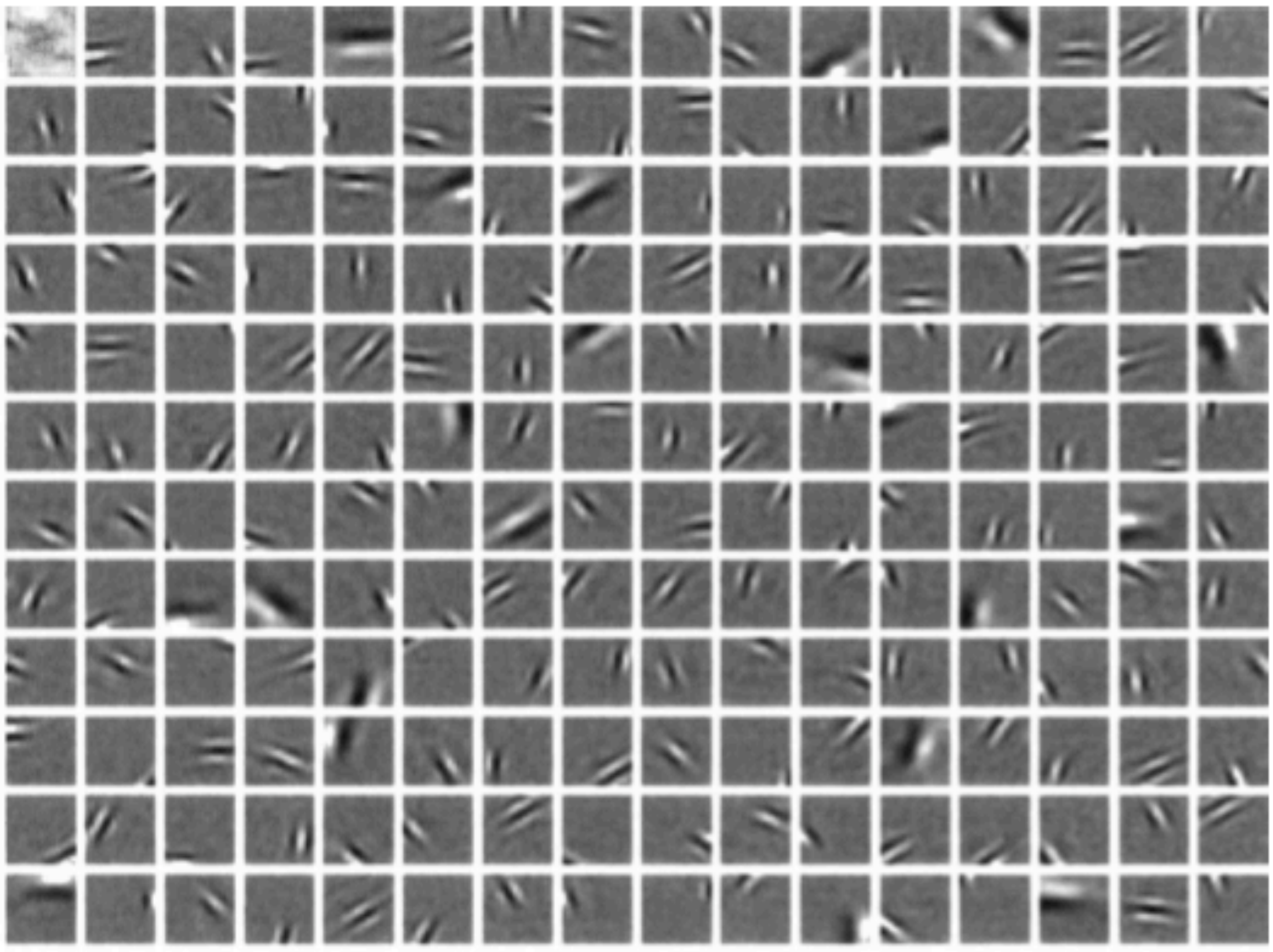
The cocktail party effect

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$$

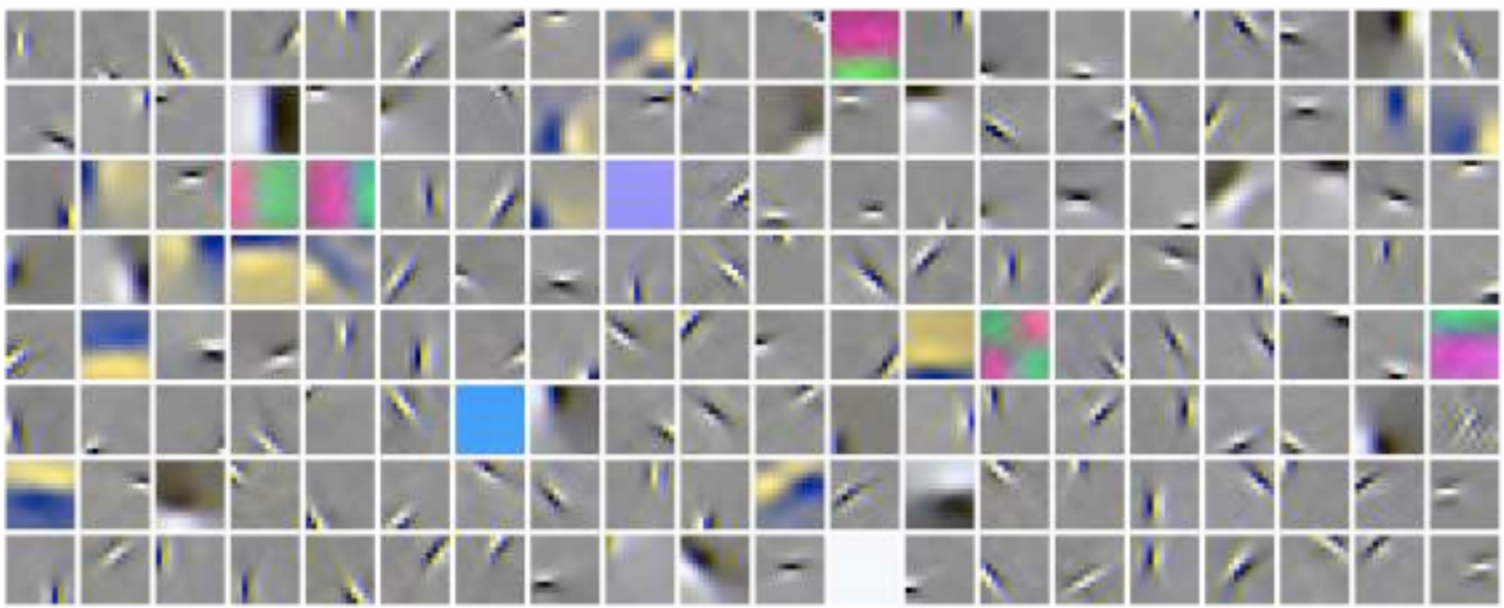


Sparse coding in human perception



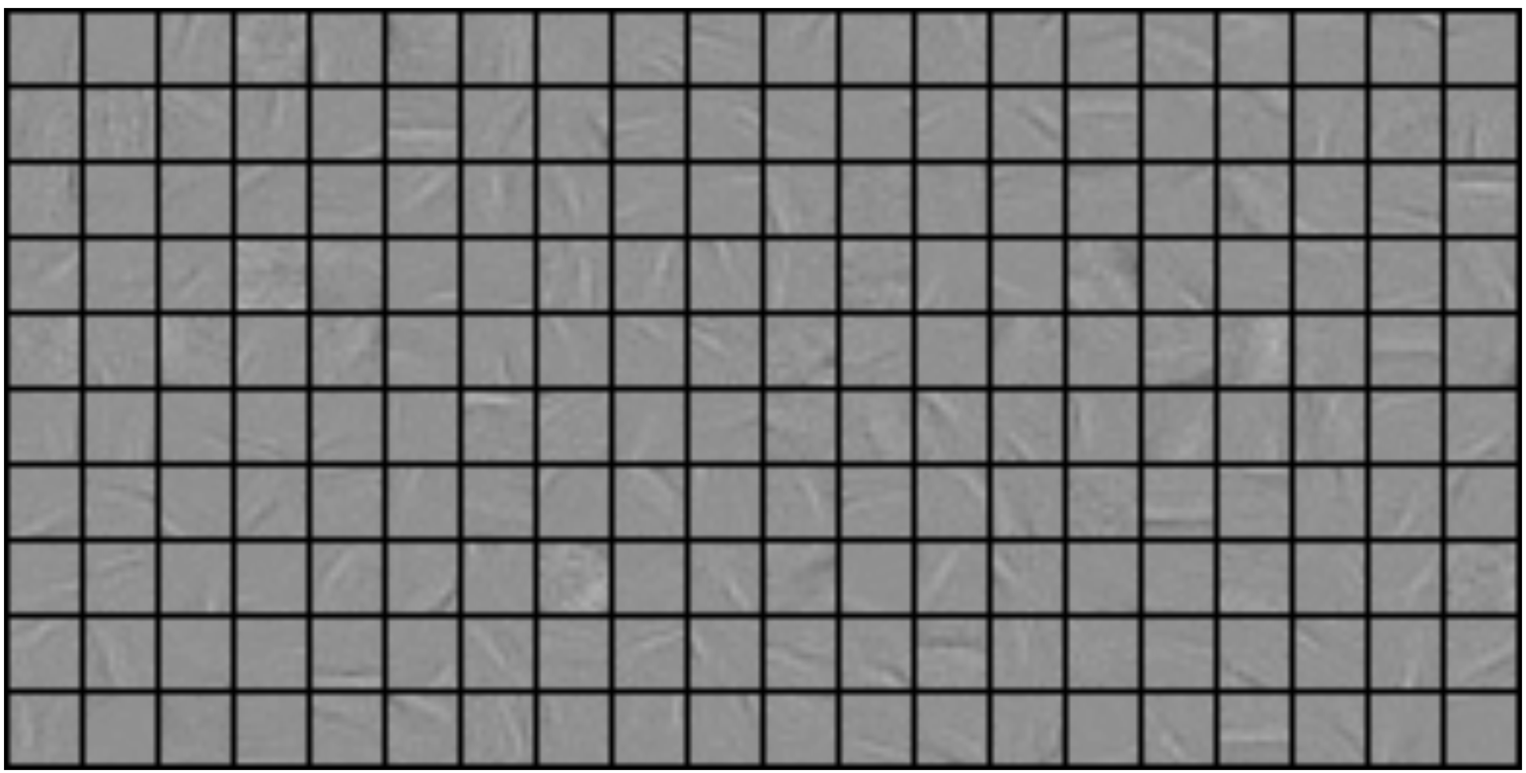
Sparse coding in human perception

- Hoyer and Hyvärinen, 2000
  - Extends to color (and stereo) images



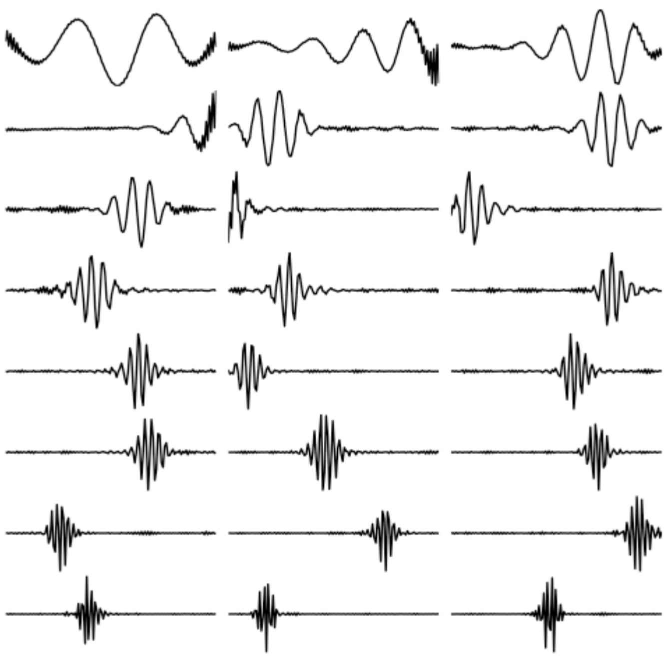
Sparse coding in human perception

- Olshausen, 2000
  - Extends time-varying images

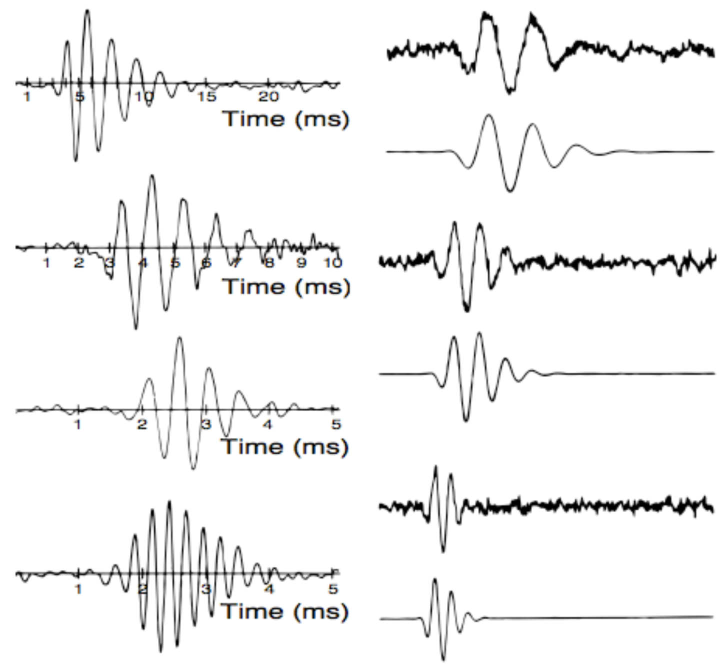


Sparse coding in human perception

- Lewicki, 2002
  - Proceeds to natural sounds



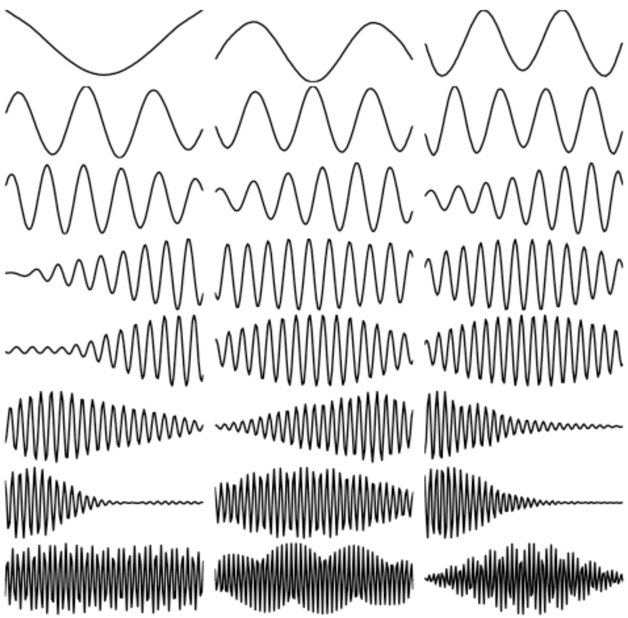
Independent components



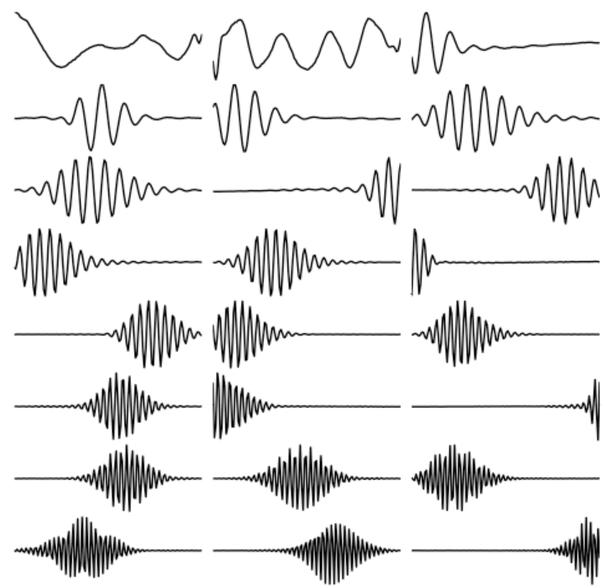
Cochlea filters

Sparse coding in human perception

- Including different auditory stimulus domains



Animal vocalizations



Human speech

Why stop with perception?

- What about higher order scene components?
  - Objects, identity, etc.

- What about concepts?

Text

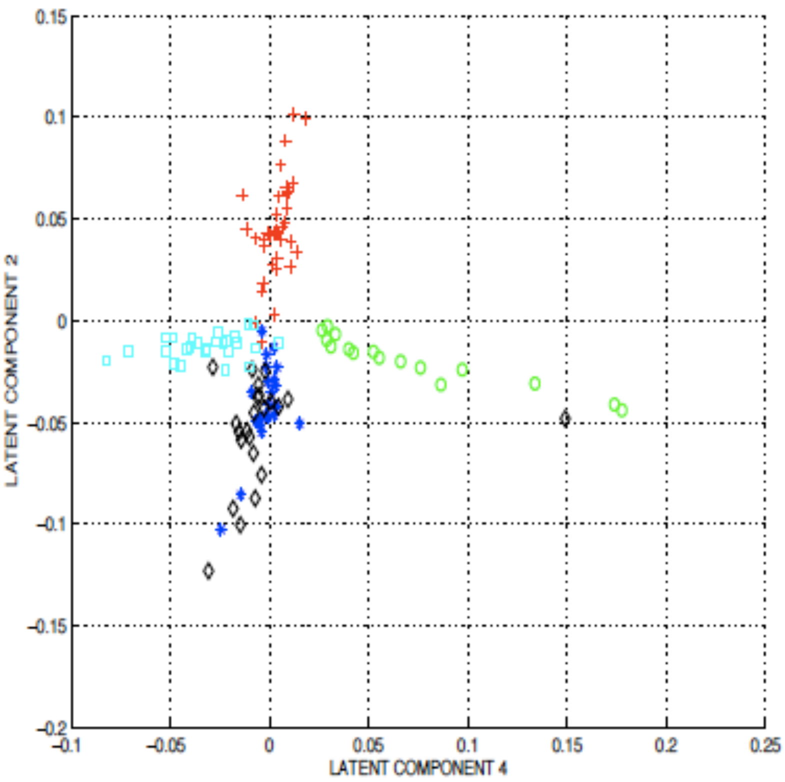
- Term/document matrix
  - Each document is a vector in term-space
    - But take away stop terms like and, if, or, a
- Latent Semantic Analysis
  - PCA on the term/document matrix
  - What are the principal axes?

Text

- Term/document matrix
  - Each document is a vector in term-space
    - But take away stop terms like and, if, or, a
- Latent Semantic Analysis
  - PCA on the term/document matrix
  - What are the principal axes?
  - Topics, themes or genres

Text

- Plotting the score reveals
  - Rays
  - Correspondence with human labels



Music

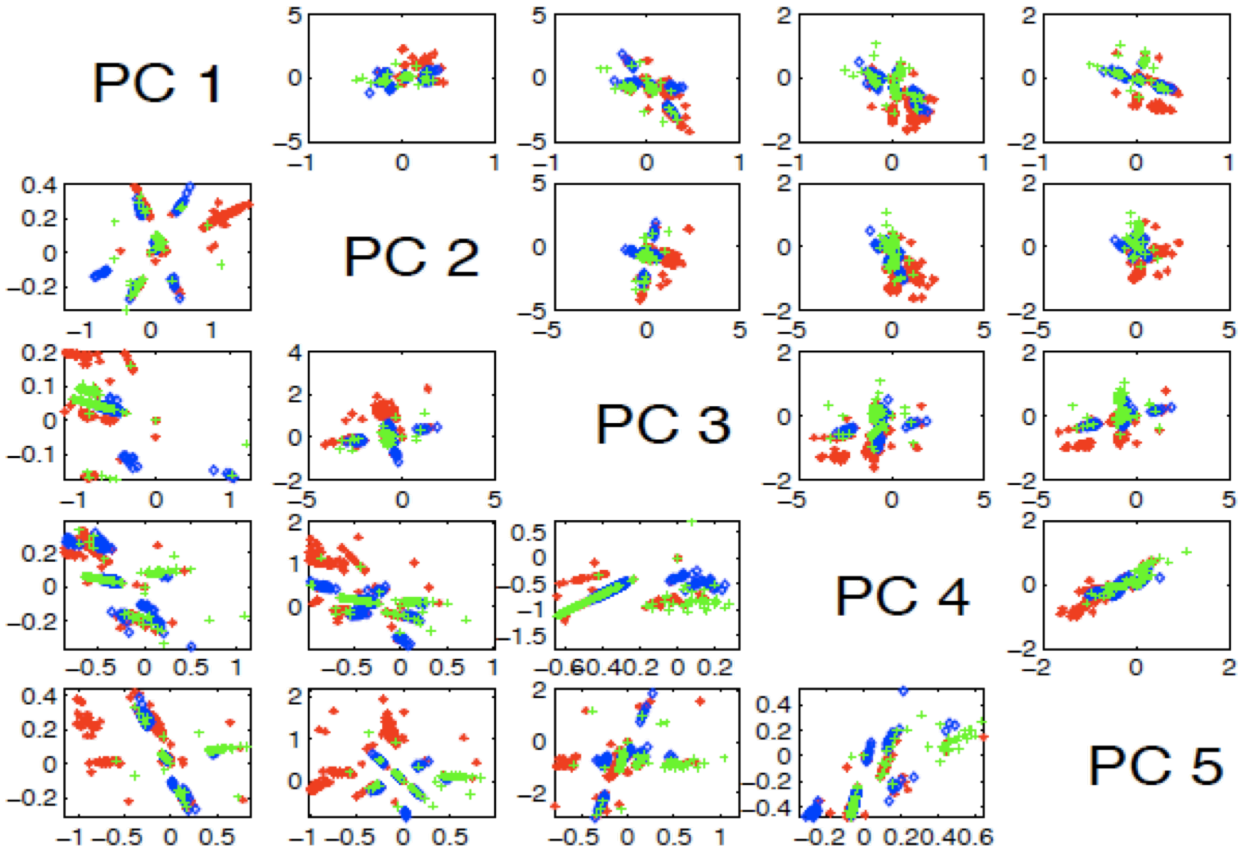
- Song/sound sample matrix
  - Sound sample is vectorized
    - Mel-frequency cepstral coefficients (MFCCs)
  - What are the principal components?

Music

- Song/sound sample matrix
  - Sound sample is vectorized
    - Mel-frequency cepstral coefficients (MFCCs)
  - What are the independent components?
    - Similar sounding music, genres

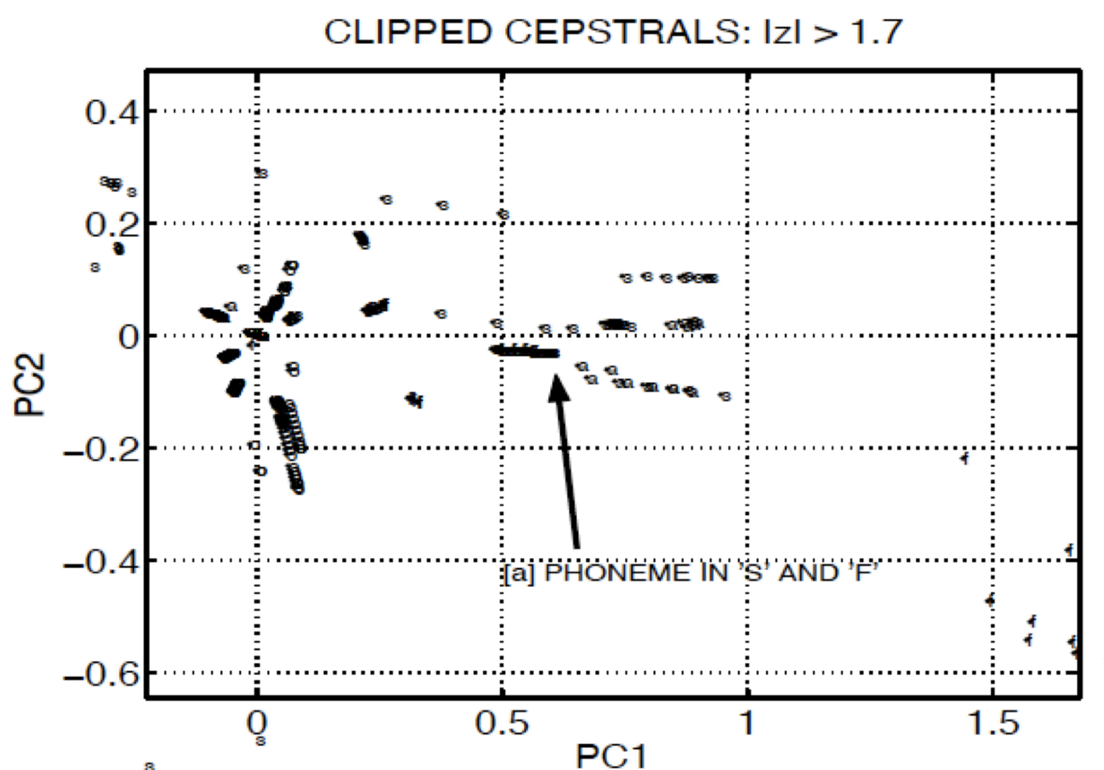
Music

- Comparing the rays with human classifications



Speech

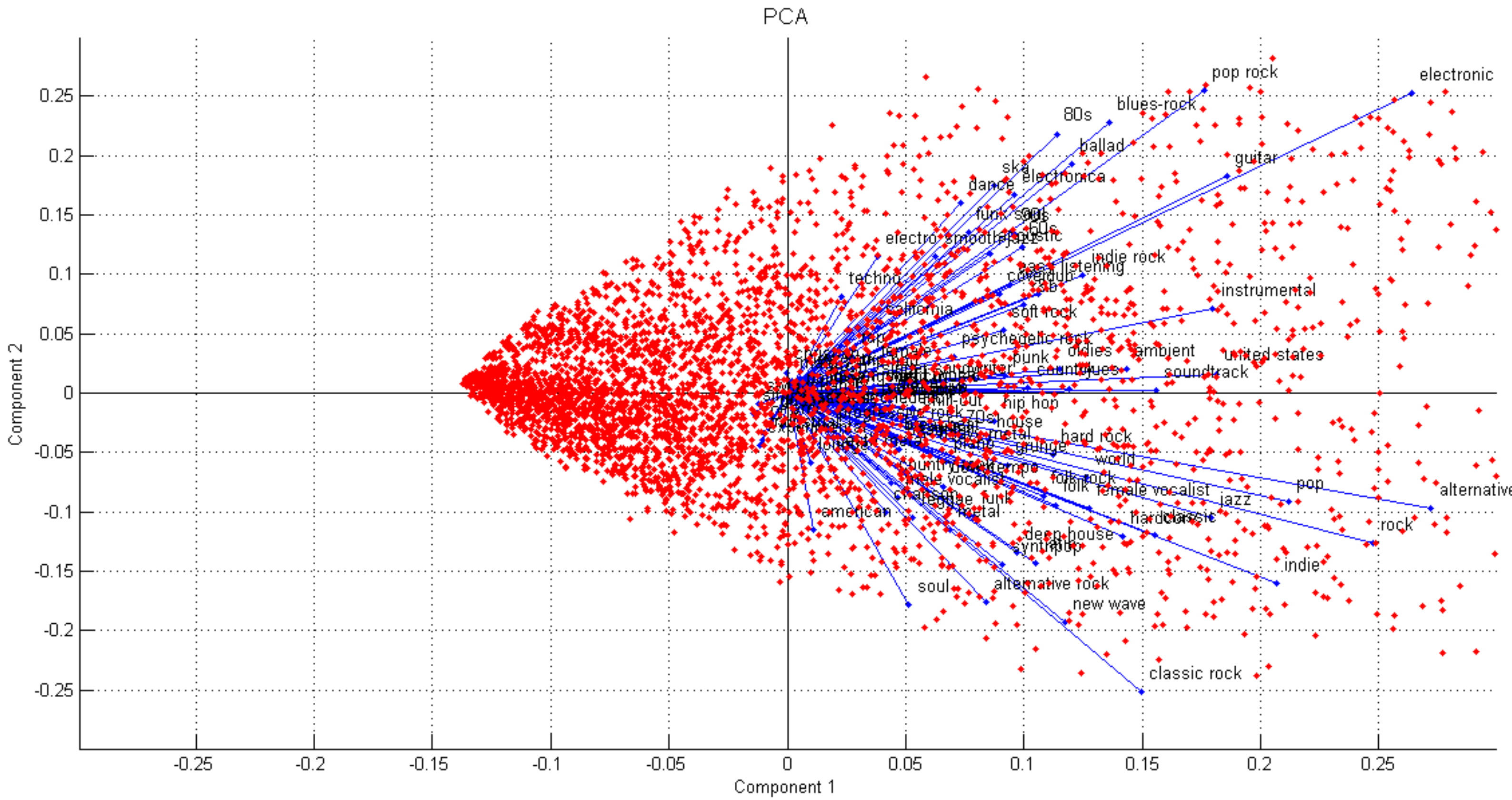
- Speech sample matrix
  - If sampled on the right time scale, speech samples should be equal to phonemes



Tag-based music analysis

- Echonest is a music data base
- Gets labels/tags on tunes from blogs (and other)
- Last FM is a web based music provider
  - Allows users to tag tunes
  - Share a large set of tag counts
- Document / term <-> tune/ tag term

PCA based analysis of tune-tag matrix



PCA based analysis of tune-tag matrix

Component 1

'electronic'	[0.2930]
'rock'	[0.2702]
'pop rock'	[0.2477]
'pop'	[0.2206]
'classic rock'	[0.1894]
'alternative rock'	[0.1814]

Component 2

'electronic'	[-0.2512]
'blues'	[0.2481]
'hip hop'	[-0.2172]
'techno'	[-0.2064]
'electro'	[-0.2059]
'house'	[-0.1808]

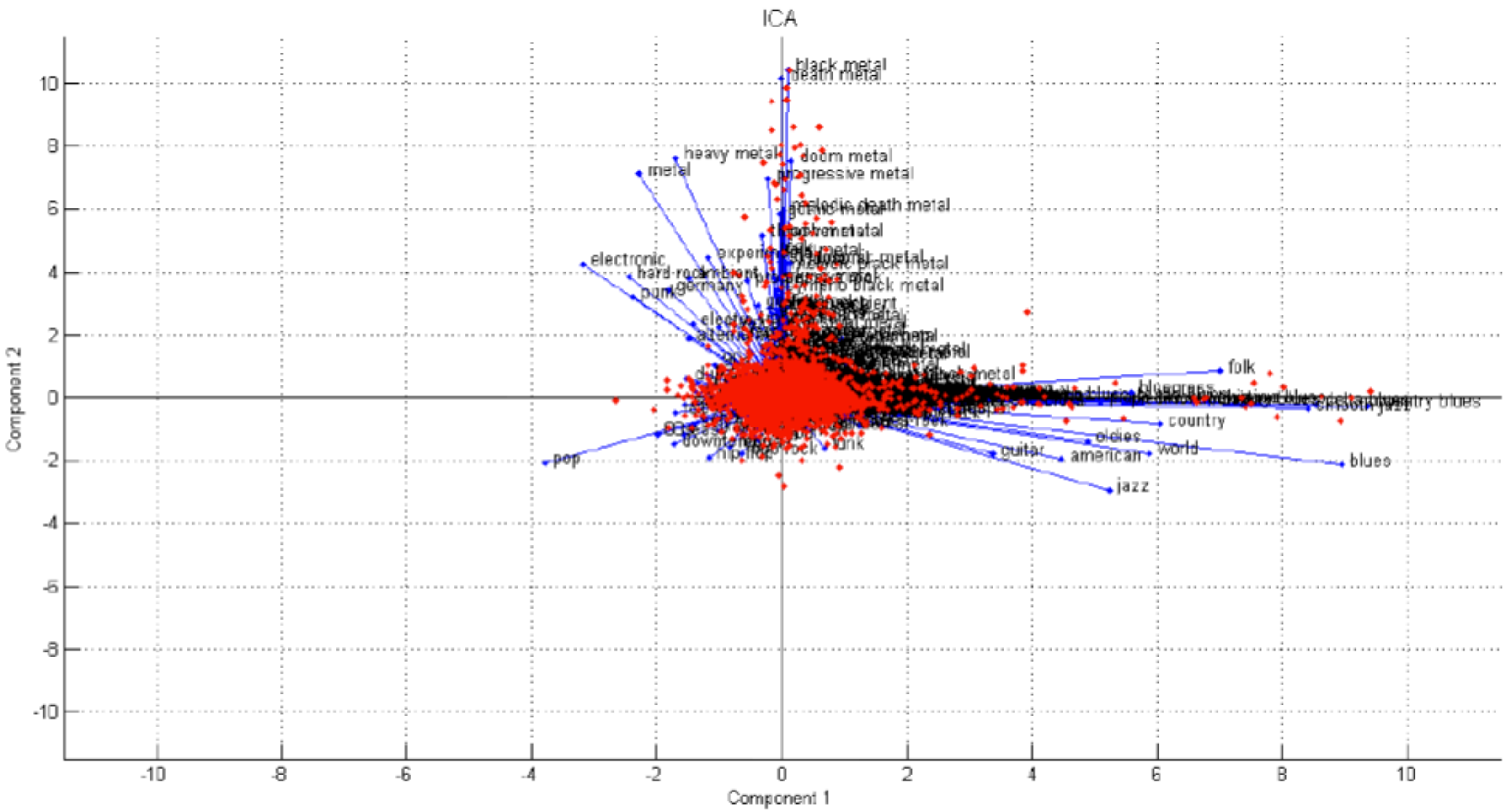
Component 3

'jazz'	[-0.2569]
'punk'	[0.2564]
'alternative'	[0.2412]
'metal'	[0.2332]
'rock'	[0.2326]
'alternative rock'	[0.2214]

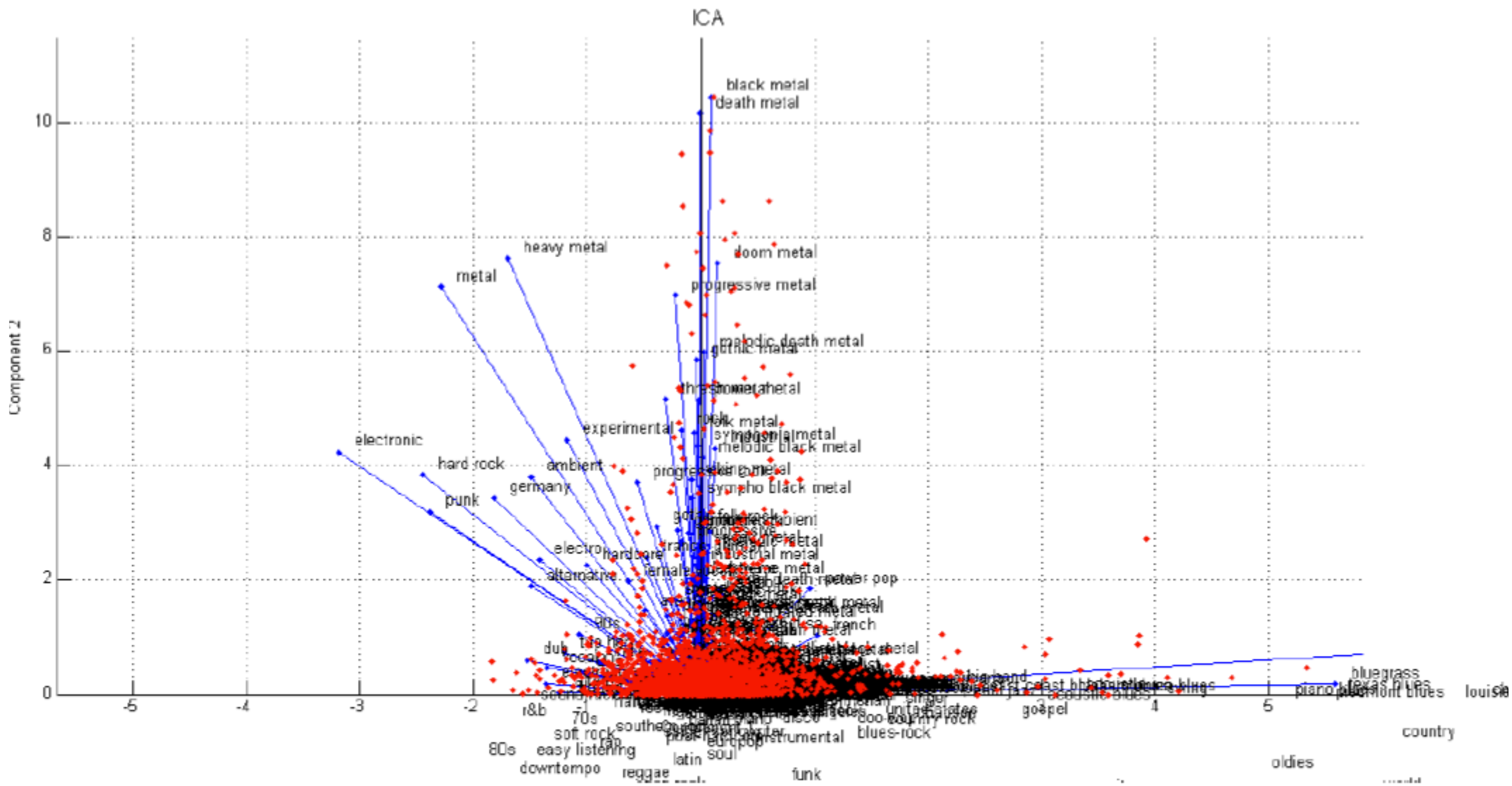
Component 4

'hip hop'	[-0.3920]
'rap'	[-0.2747]
'ambient'	[0.2547]
'folk'	[0.2391]
'reggae'	[-0.2021]
'soul'	[-0.1930]

ICA based analysis of tune-tag matrix



ICA based analysis of tune-tag matrix



ICA based analysis of tune-tag matrix

Component 1		*Component 3*		*Component 4*	
'progressive house'	[-13.0420]	'metalcore'	[9.7857]	'country blues'	[-11.2841]
'trance'	[-11.6010]	'hardcore'	[9.0779]	'delta blues'	[-10.3246]
'progressive trance'	[-10.5130]	'metal'	[8.6138]	'blues'	[-10.0484]
'techno'	[-10.3755]	'screamo'	[8.5570]	'chicago blues'	[-9.4272]
'tech house'	[-10.3539]	'emo'	[8.5221]	'smooth jazz'	[-8.0635]
'house'	[-9.8672]	'alternative metal'	[8.2246]	'louisiana blues'	[-7.5511]

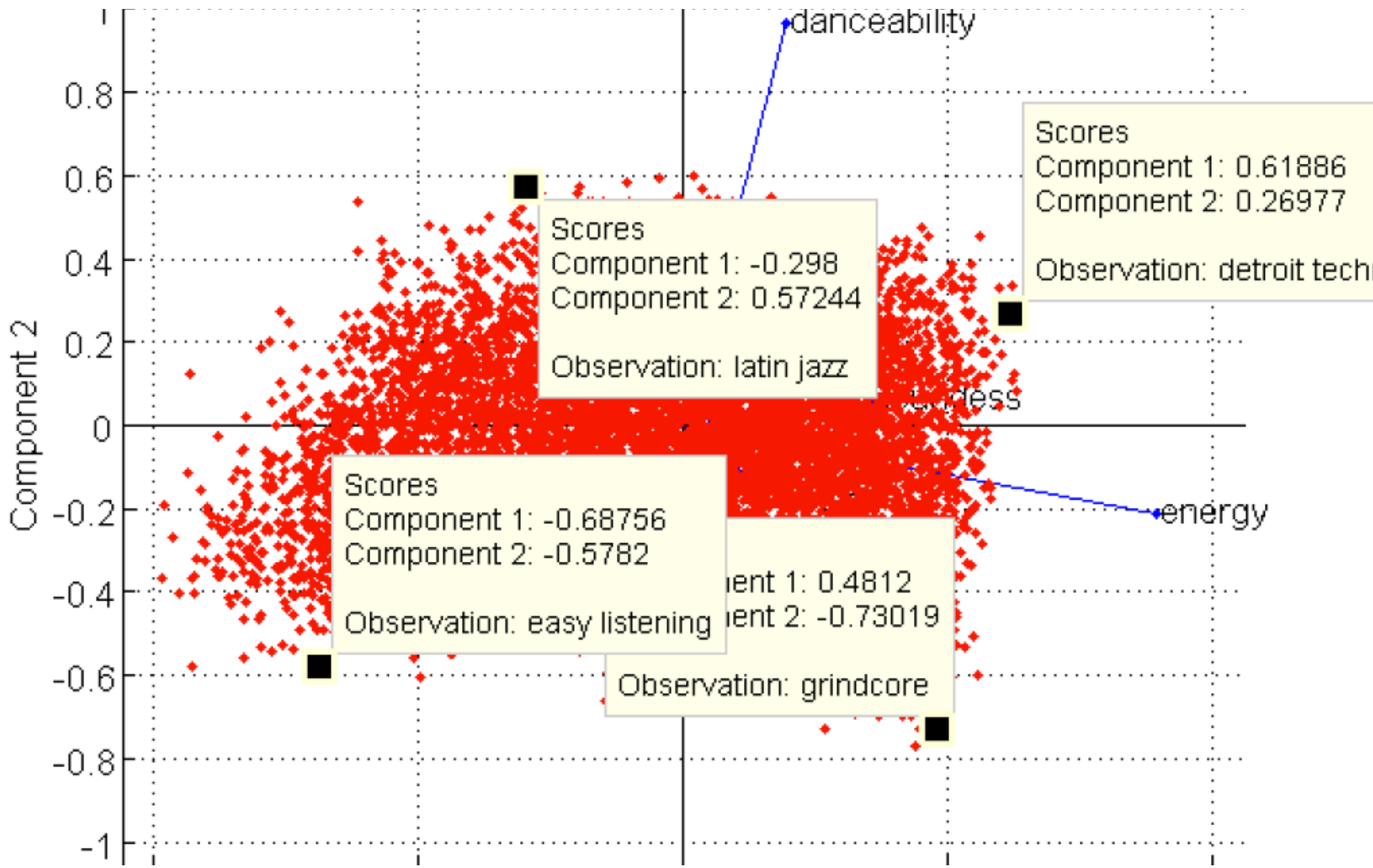
Component 5		*Component 9*		*Component 10*	
'country'	[-11.1097]	'black metal'	[-10.3062]	'downtempo'	[-7.1691]
'classic country'	[-10.3586]	'death metal'	[-10.0929]	'easy listening'	[-6.6198]
'oldies'	[-8.1613]	'doom metal'	[-7.9393]	'acid jazz'	[-6.5100]
'honky tonk'	[-7.6438]	'metal'	[-7.8662]	'trip hop'	[-6.4428]
'country rock'	[-7.2752]	'heavy metal'	[-7.8264]	'future jazz'	[-6.2716]
'world'	[-7.2509]	'progressive metal'	[-7.2647]	'disco'	[-6.1470]

Component 12		*Component 19*		*Component 33*	
'roots reggae'	[-10.0137]	'ccm'	[-14.2041]	'rap'	[-10.3298]
'dancehall'	[-9.9293]	'christian'	[-12.5425]	'gangster rap'	[-8.8597]
'dub'	[-8.8466]	'gospel'	[-10.0297]	'hip hop'	[-8.0857]
'reggae'	[-8.6909]	'contemporary christian'	[-7.4200]	'hardcore rap'	[-6.9934]
'lovers rock'	[-8.4640]	'christian rock'	[-6.4889]	'new york'	[-6.7396]
'ska'	[-6.8764]	'worship music'	[-4.3180]	'east coast rap'	[-6.6330]

Acoustical features of tunes

- Echonest also provides acoustical features
  - Energy (loudness variation)
  - Danceability (beat strength and consistency)
  - Tempo (BPM)
  - Loudness (overall)

PCA based analysis of acoustical features



Now to the exercises..