INDEPENDENT COMPONENT ANALYSIS
ICA

Un buon link da cui trarre informazioni nonchè uno dei codici, per l'implementazione della ICA in Matlab, che sta avendo più fortuna è:

http://www.cis.hut.fi/projects/ica/fastica/

“Independent component analysis (ICA) is a method for finding underlying factors or components from multivariate (multi-dimensional) statistical data. What distinguishes ICA from other methods is that it looks for components that are both statistically independent, and non Gaussian.”

A.Hyvarinen, A.Karhunen, E.Oja
‘Independent Component Analysis’

ICA

Blind Signal Separation (BSS) or Independent Component Analysis (ICA) is the identification & separation of mixtures of sources with little prior information.

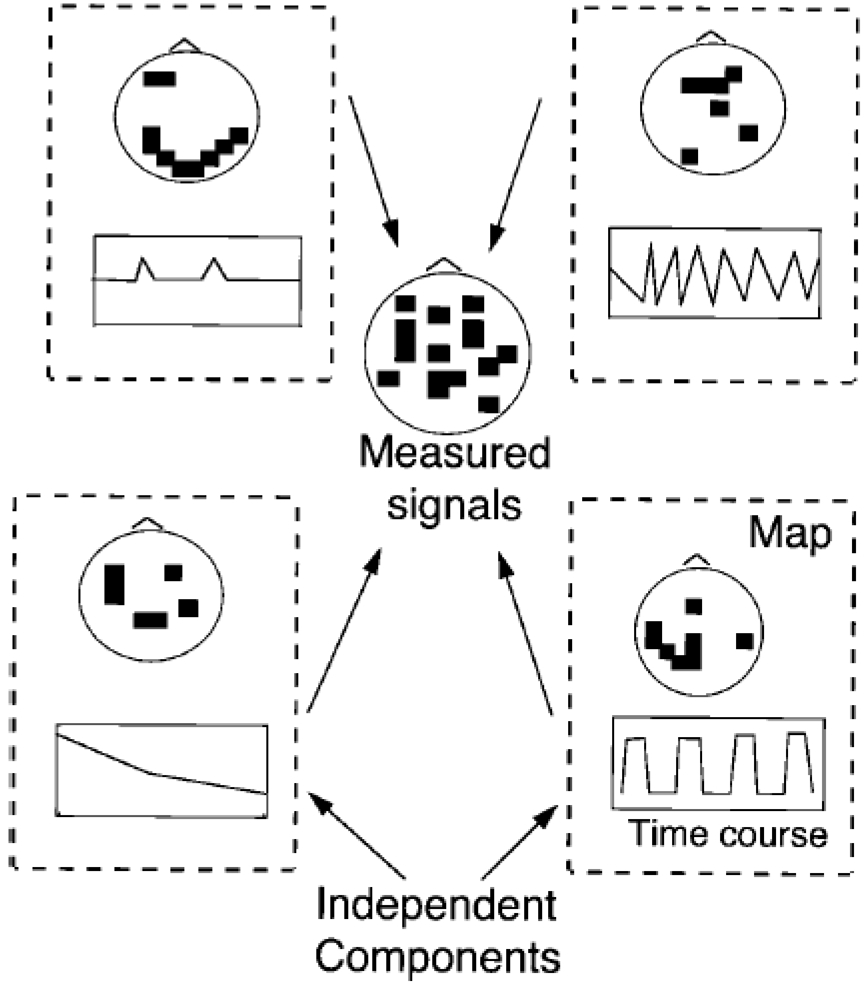
- Applications include:
  - Audio Processing
  - Medical data
  - Finance
  - Array processing (beamforming)
  - Coding

… and most applications where Factor Analysis and PCA is currently used.

- While PCA seeks directions that represents data best in a $\boldsymbol{\Sigma|x_0 - x|^2}$ sense, ICA seeks such directions that are most independent from each other.

  Often used on Time Series separation of Multiple Targets

Lo scopo dell'Independent Component Analysis (ICA) è quello di risalire alle sorgenti che hanno preso parte alla creazione del segnale, o dei segnali misurati.
Esempio:

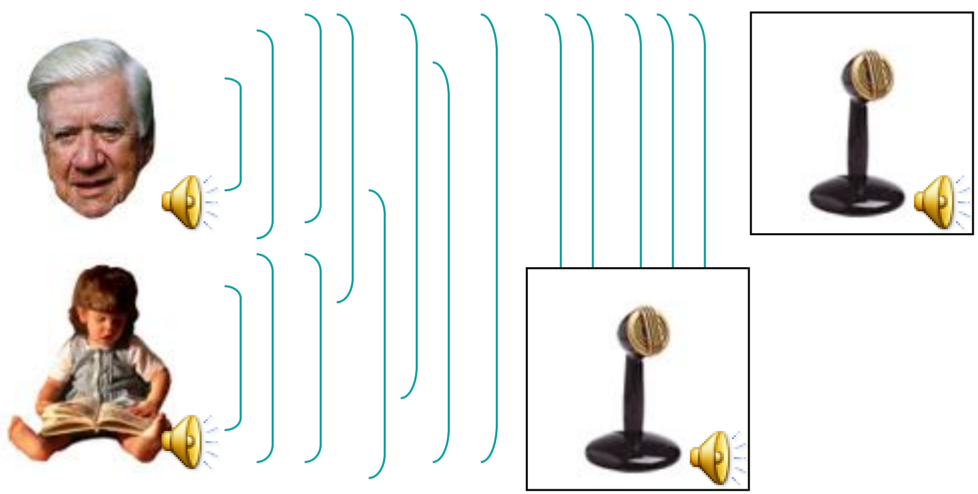


Il segnale osservato in un dato voxel è visto come la somma dei contributi di tutte le componenti indipendenti presenti nel data set (ad es. in questo caso 4)

McKeown et al., Human Brain Mapping, 6:160-188, 1998

Cocktail-Party Problem

- Simple scenario:
  - Two people speaking simultaneously in a room
  - Speeches are recorded by two microphones in separate locations



Cocktail-Party Problem

- Let $s_1(t)$, $s_2(t)$ be the speech signals emitted by the two speaker
- Recorded time signals, by the two microphones, are denoted by $x_1(t)$, $x_2(t)$
- The recorded time signals can be expressed as a linear equation:

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

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

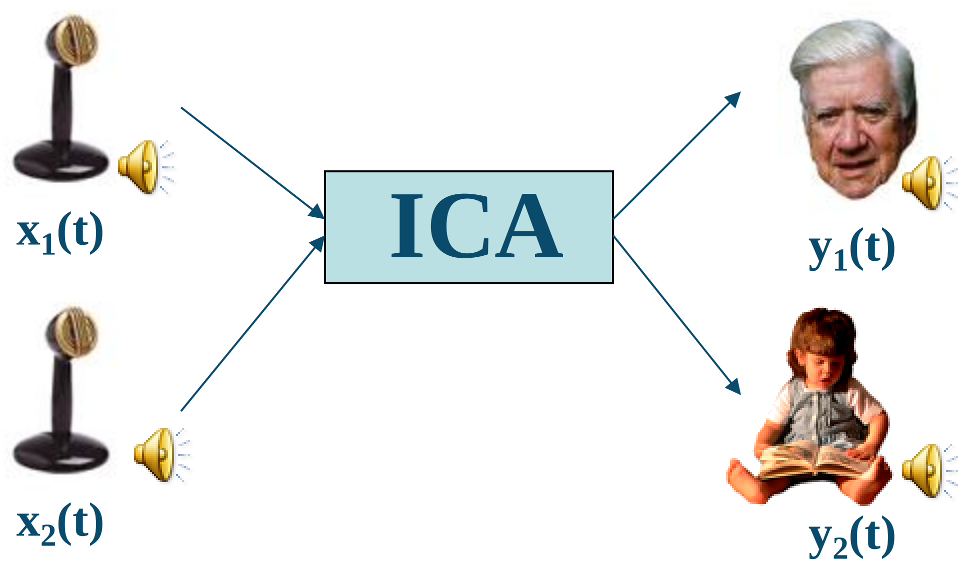
where parameters in matrix A depend on distances of the microphones to the speaker, along with other microphone properties

- Assume $s_1(t)$ and $s_2(t)$ are statistically independent

Cocktail-Party Problem

GOAL:

– Recover the unmixed speech signals, without knowing A or $s_i(t)$



FORMALIZZAZIONE

Supponiamo che ∃ M segnali di <u>media nulla</u> s_1 s_2 s_3 .... s_M ma siano osservabili solo N combinazioni lineari delle variabili s_i

Chiamiamo $\mathbf{x}_j$ la j-esima variabile osservabile:

$$\mathbf{x_j = a_{j1}s_1 + a_{j2}s_2 + \ldots + a_{jn}s_n \quad i=1,\ldots,M \quad j=1,\ldots,N}$$

In forma matriciale si può scrivere:

$$\mathbf{X} = \mathbf{A}\,\mathbf{S}$$

Se la matrice A non è nota il problema puo' essere risolto facendo alcune assunzioni sulle proprietà statistiche delle sorgenti s_i

N.B. Dovrebbe essere considerato anche un termine aggiuntivo n per il rumore

$$X = A\,S + n$$

Ipotesi per l' applicabilità di ICA:

- N >= M
- le sorgenti s_i NON hanno distribuzione gaussiana
- le sorgenti s_i sono statisticamente indipendenti
- la matrice A ha rango massimo
- per ora n=0 , ma il modello puo' essere esteso anche se di più difficile risoluzione

Ipotesi 1) poiché siamo interessati ai soli cambiamenti del segnale, i valori medi non portano alcuna informazione aggiuntiva quindi si sottrae il valor medio (ad x^{meas}_j)

$$X_{mean} \leftarrow X^{meas} - E[X^{meas}]$$

Il problema quindi diventa $X_{mean} = X = A\,S$.

Ipotesi 2) Whitening o sphering: serve ad ottenere dati scorrelati e con varianza pari a 1, ossia cerchiamo la matrice V tale che:

$$Z = V\,X \quad \text{dove } E[\,Z\,Z^T\,] = 1\ (=I)$$

$$\Downarrow$$

$$\text{Se:} \qquad C = E[X \cdot X^T] \qquad \text{allora} \qquad V = C^{-1/2}$$

C = matrice quadrata e simmetrica

Infatti:

$$E[Z \cdot Z^T] = I = E[V \cdot X \cdot X^T V^T] = C^{-1/2} \cdot C \cdot C^{-1/2}$$

Lo sbiancamento è un passaggio molto utile, perchè restringe la ricerca della matrice di mixing alle matrici ortogonali. Ossia ora

$$Z = V \cdot X = V \cdot A \cdot S$$

La matrice (di mixing) quindi da stimare è **W** = **V A** ed è ortogonale:

$$W \cdot W^T = I$$

Il whitening riduce il numero di parametri (elementi di matrice) da stimare. Invece di avere n^2 parametri (considerando una matrice $n \times n$), il numero di parametri indipendenti in una matrice ortogonale di dimensione n è:

$$n\,(n-1)/2$$

$$Z = W \cdot S$$

Ora abbiamo una combinazione lineare delle sorgenti indipendenti. Tale somma è “piu’ gaussiana” delle componenti originarie.

Sotto l'ipotesi di nongaussianità delle fonti, grazie al teorema del limite centrale si può, infatti, affermare che le misture date dalla combinazione lineare delle fonti avranno una distribuzione di probabilità sicuramente più gaussiana di quella delle fonti stesse.

Quindi la somma $y = z^T s$ diventa “al minimo” gaussiana quando $y=s_i$ ossia z ha solo l’i-imo elemento non nullo.

W scelto in modo da massimizzare la non-gaussianità di $Y=W^TZ$

Principali indici statistici

di posizione

- **MODA**
- **MEDIANA**
- **MEDIA**

di dispersione

- **SCARTO QUADRATICO MEDIO**
- **VARIANZA**
- **RANGE**

di forma

- **ASIMMETRIA (SKEWNESS)**
- **CURTOSI (KURTOSIS)**

Stimatori di non-gaussianità:

CURTOSI $\quad kurt(y)=E\{y^4\} - 3(E\{y^2\})^2$ è nullo per variabili gaussiane

quindi si cerca il max di $|Kurt(y)|$

Entropia

Ricordiamo che è definita come (variabile aleatoria X):

$$\boxed{H[X] = -\int_{-\infty}^{+\infty} p(x)\cdot \ln\left(p(x)\right)\cdot dx}$$

e che nel caso di variabile gaussiana

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow H[X]_{gaus} = \frac{1}{2} + \frac{1}{2}\cdot \ln\left(2\pi\sigma^2\right)$$

questo è anche il valore per il quale si massimizza l'entropia ossia

$$\max\left(H[X]\right) = H[X]_{gaus}$$

Misure di non-gaussianità:

CURTOSI $kurt(y)=E\{y^4\} - 3(E\{y^2\})^2$ è nullo per variabili gaussiane

quindi si cerca il max di $|Kurt(y)|$

NEGENTROPY $J(y)=H(y_{gauss}) - H(y)$

è nulla per variabili gaussiane (quelle con la max entropia H) quind
si massimizza

MUTUAL INFORMATION $I(y_1,\dots,y_M) = \sum_i H(y_i) - H(y)$

È nulla per variabili indipendenti e non negativa → va minimizzata

FastICA

cons:

Bad: some components can be hard to interpret

Bad: run-run variability in decomposition

ICA: OPEN ISSUES

- Scelta del numero di sorgenti k
- Selezione delle sorgenti & componenti di interesse

ICA: OPEN ISSUES

- Scelta del numero di sorgenti k
- Selezione delle sorgenti & componenti di interesse

ICA: NUMBER OF COMPONENTS

- Too many ➔ over-splitting of the components
- Too few ➔ over-clumping of the components

How to choose?

- PCA
- Fixing a value
- Post-ICA clustering

ICA: OPEN ISSUES

- Scelta del numero di sorgenti k
- Selezione delle sorgenti & componenti di interesse