

Non Parametric Measures of Statistical Dependence

José C. Príncipe

principe@cnel.ufl.edu

<http://www.cnel.ufl.edu>

Supported by NSF Grant IIS 0964197



Acknowledgements

- My Students
 - Sohan Seth (his Ph.D. thesis topic)
 - Bilal Fadlallah
- My collaborator Dr. Andreas Keil
NIH Center of Emotion and Attention U. of Florida
- Support of NSF Grant IIS 0964197

Outline

- Motivation
- Statistical Independence and Dependence
- Generalized Measure of Association
- Measure of Conditional Association
- Cognitive Interfaces
- Conclusions

The Elephant in the Room

- Most real world complex systems are defined exactly by multiscale couplings and display some form of dependence, in physiology of the human body (e.g. brain signals, respiration/heart, etc.), or in monitoring of complex engineering systems (e.g. power plant, electric grid, web,etc).
- But dependence and causation lack universally accepted definitions.
- On the other hand, independence and non-causation (or equivalently conditional independence) are precisely defined in mathematical terms.
- Define: Dependence as absence of independence.
- However, these concepts are not always reciprocal of each other, and they bear very different understanding and usage in the context of specific applications.

Statistical Definitions

- What is statistical independence?

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- What is conditional independence?

$$P(X \in A, Y \in B | Z \in C) = P(X \in A | Z \in C)$$

- What is statistical **dependence**?

$$P(X \in A, Y \in B) \neq P(X \in A)P(Y \in B)$$

- What is **conditional dependence**?

$$P(X \in A, Y \in B | Z \in C) \neq P(X \in A | Z \in C)$$

A Practical Issue

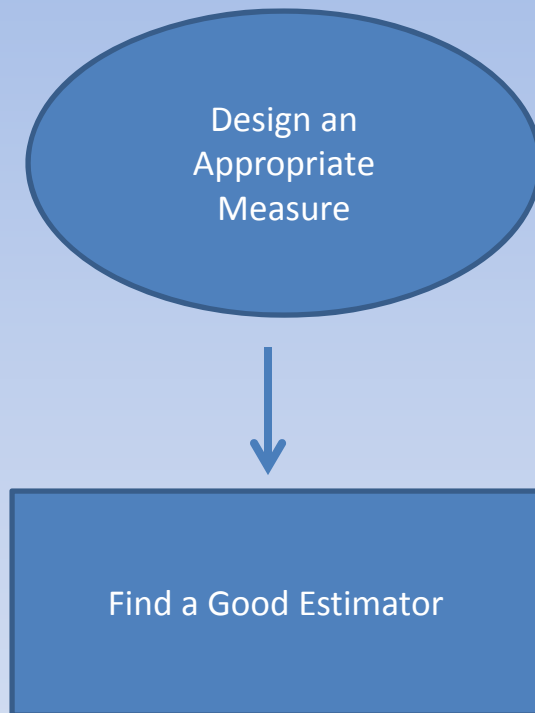
- How to translate statistical criteria on random variables to realizations (estimators)?

Obviously we would like to use the most powerful statistical measure, but unfortunately this is usually the one that has the most difficult estimator (no free lunch!).

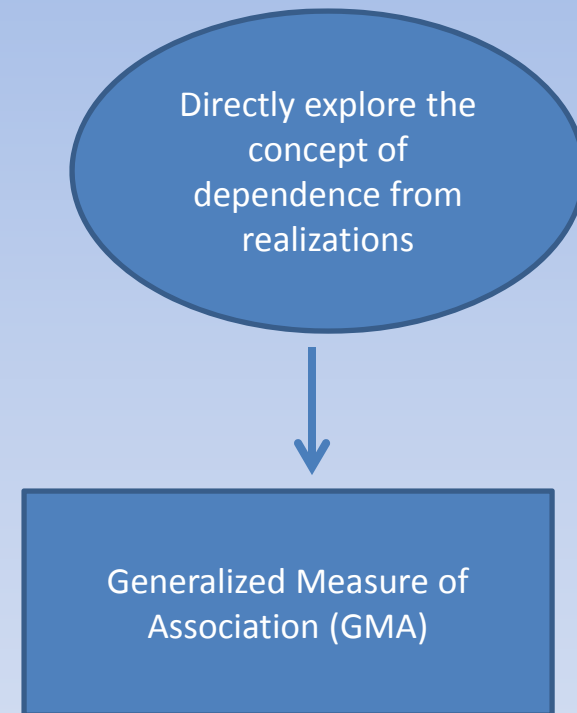
- Parametric versus non parametric estimators
- Estimators should preserve the properties of the statistical criterion
- Hopefully no free parameters, and well established statistical tests
- Some other lesser problems are the number of variables, abstractness, data types and scalability/computational complexity.

Approaches to Quantify Dependence

Traditional



Our Approach



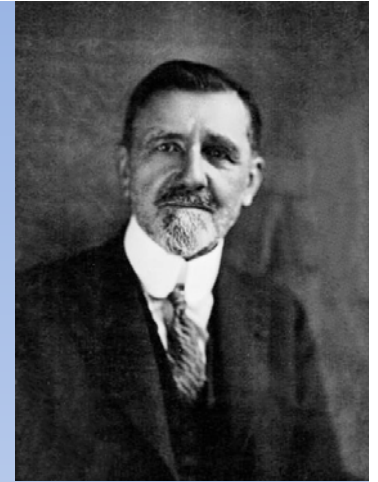
- GMA generalizes the concept of association to arbitrary metric spaces
- GMA is bounded, parameter-free, easy to compute, and *asymmetric in nature*.

State of the Art

- Independence
 - Very well studied, many measures exist. Short summary
- Conditional independence
 - Very well studied, but not many measures. Not covered
- Dependence
 - Use of correlation and mutual information is dominant. We explore new understanding, i.e. Generalized Association
- Conditional dependence
 - Not well studied. Extend Generalized Association

Concept of Independent Events

- Independence was applied during many years in a vague and intuitive sense on event spaces.
- The rule of multiplication of probabilities for independent events is an attempt to formalize the concept of independence and to build a calculus around it (but they are NOT the same).
- Borel in 1909 claimed that binary digits (or Rademacher functions) were independent. This launched the modern theory of probability, with the concept of measurable functions (the random variables) and solved this problem for (most of) us.
- Independence is in the core fabric of measure theory and hence of probability, and explains why the concept is so clear and a huge simplification that is exploited throughout statistics.



Emile Borel

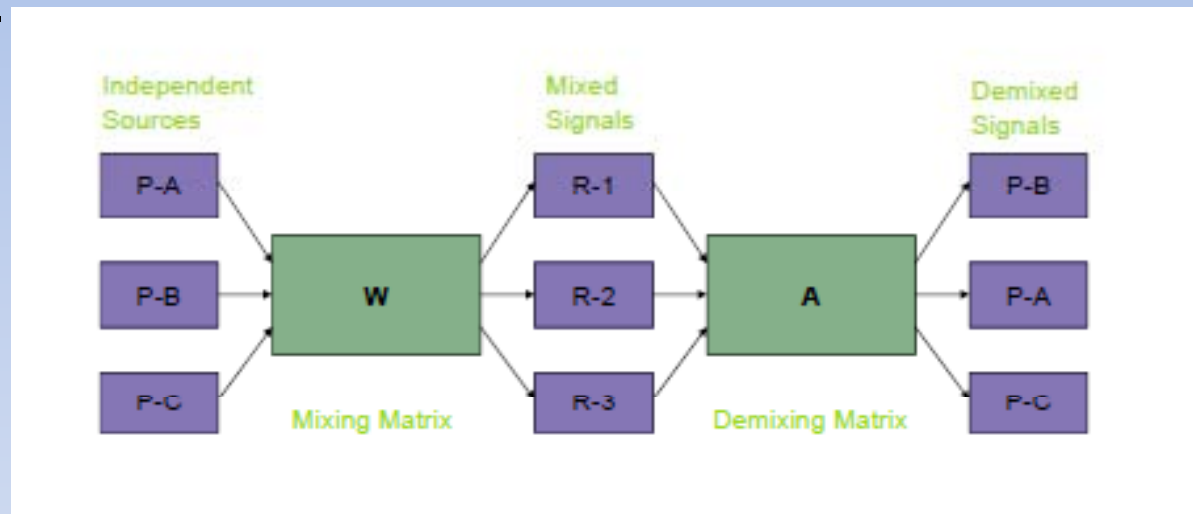
Independence in Machine Learning

- Independent Component Analysis (ICA) has been one of the most visible applications of independence in Machine Learning and Signal Processing.

Sources P and W
Unknown

Assume sources
Independent,
Mixing linear

$\dim R > \dim P$



- The early work concentrated on using surrogate costs for independence (contrast functions) such as Kurtosis, Negentropy, etc.
- We proposed Information Theoretic Learning (ITL) while Bach and Jordan proposed kernel ICA. They both exploit directly independence.

Kernel ICA

- The idea is to use the maximal correlation defined in the space of functions F as

$$\rho_F = \max_{f_1, f_2} \text{corr}(f_1(x_1), f_2(x_2)) = \max_{f_1, f_2} \frac{\text{cov}(f_1(x_1), f_2(x_2))}{\sqrt{\text{var } f_1(x_1)} \sqrt{\text{var } f_2(x_2)}}$$

and estimate it using projections in a Reproducing Kernel Hilbert Space (RKHS) as $\text{corr}(f_1(x_1), f_2(x_2)) = \text{corr}(\langle \Phi(x_1), f_1 \rangle, \langle \Phi(x_2), f_2 \rangle)$

which coincides with the first canonical correlation in the RKHS, followed by a maximization step.

- The method is very principled, complexity is still reasonable $O(N^2)$ and precision is very high for low dimensions.
- There are two free parameters: the regularization and the kernel size in the RKHS mapping

RKHS in Probability Theory

- Loeve noted the existence of a RKHS representation for time series in 1948 and Parzen presented a systematic treatment in 1959.
- Emanuel Parzen should be credited for pointing out the importance of RKHS to probability theory

“One reason for the central role of RKHS is that a basis tool in the statistical theory of stochastic processes is the theory of equivalence and singularity of normal measures, and this theory seems to be most simply expressed in terms of RKHS”. Parzen (1970)

Michel Loeve



Current work goes by the name
of statistical embeddings

Manny Parzen



ITL RKHS

- In information theoretic learning (ITL) we defined a RKHS induced by a positive definite function called the cross information potential (CIP) as $v(f_X, f_Y) = \int f_X(z) f_Y(z) dz$ which is basically the inner product of two pdfs (in L_2).

- Notice that this RKHS follows Parzen's idea and it is different from the work of Bach and Jordan. In this space the distance D between two pdfs can be estimated by

$$D^2(f_X, f_Y) = v(f_X - f_Y, f_X - f_Y)$$

- Notice that $D = 0$ iff $f_X = f_Y$ almost everywhere.
- Therefore the problem of independence can be framed in the ITL RKHS by making $f_X = f_{XY}$ and $f_Y = f_X f_Y$ (which we called quadratic mutual information (QMI)) and minimizing D .

ITL RKHS

- The straight estimator of the CIP using Parzen windows is $O(N^3)$

$$\hat{V} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \kappa_1(x_i, x_j) \kappa_2(y_i, y_j) - \frac{2}{N^3} \sum_{k=1}^N \left(\sum_{j=1}^N \kappa_1(x_k, x_j) \right) \left(\sum_{i=1}^N \kappa_2(y_k, y_i) \right) + \frac{1}{N^4} \sum_{i=1}^N \sum_{j=1}^N \kappa_1(x_i, x_j) \sum_{i=1}^N \sum_{j=1}^N \kappa_2(y_i, y_j)$$

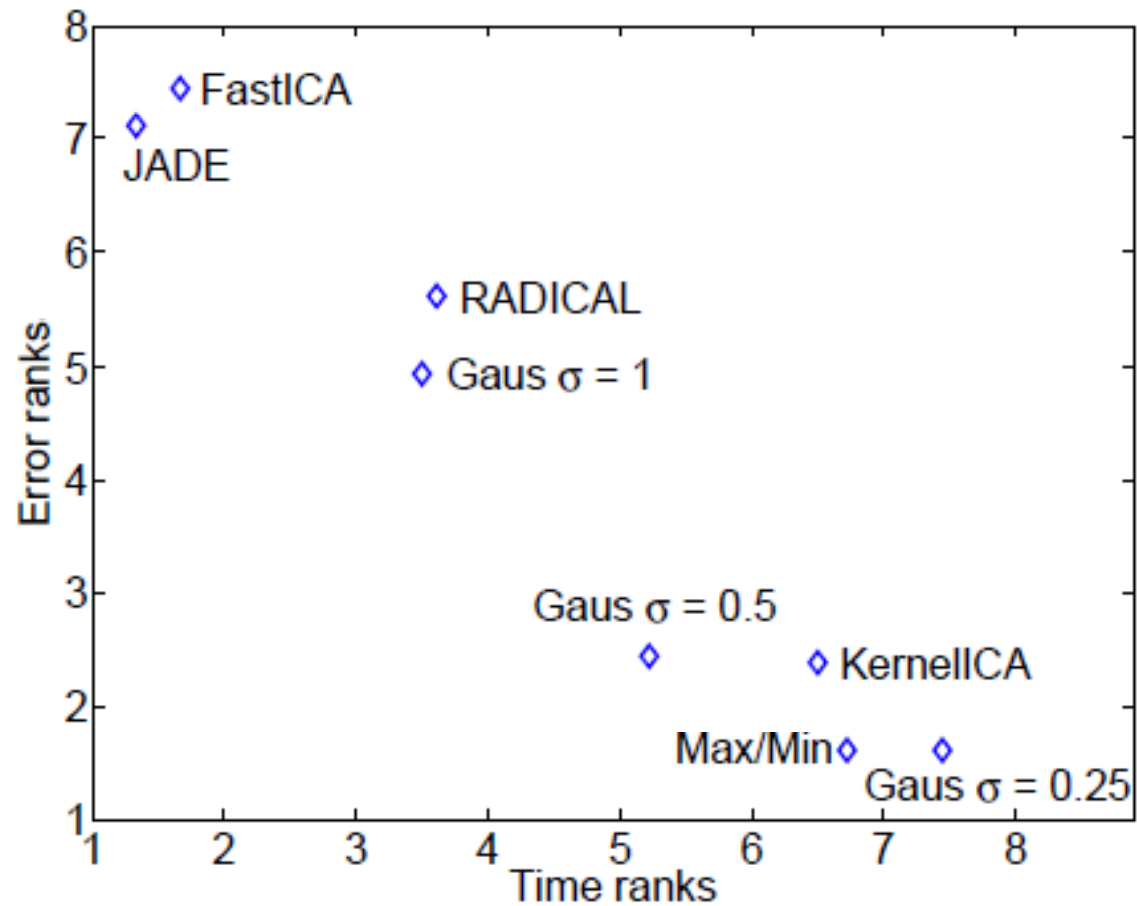
(one free parameter) but there are faster estimators using the fast Gauss transform $O(N^2)$ and the incomplete Cholesky decomposition.

- The CIP can be generalized as

$$v_g(f_X, f_Y) = \iint \kappa(x, y) f_X(x) f_Y(y) dx dy \quad \kappa(x, y) = 1 - |x - y|$$

where the kernel is a symmetric strictly positive definite function (when the kernel defaults to a delta function we get the CIP). This gives rise to both an unifying concept of all quadratic measures of independence as well as new kernels that do not have any free parameter (but data requires normalization).

ICA: A Mature Field



Concept of Dependent Events

- Dependence is coupling in the joint space of random variables.
- The term dependence appears in contexts such as “*X is highly dependent on Y*” or “*Z is more dependent on X than on Y*”.
- We infer one of three things:
 1. the random variables are not independent,
 2. the random variables are correlated,
 3. the random variables share some information.
- NONE of these three interpretations explore the dependence concept in its entirety.
 - Not independent does not quantify the degree of dependence.
 - Correlation does this, but only quantifies linear dependence over the reals
 - Mutual Information is probably appropriate for discrete r.v., but it becomes obscure for other data types and it is very difficult to estimate.

Correlation – Pearson

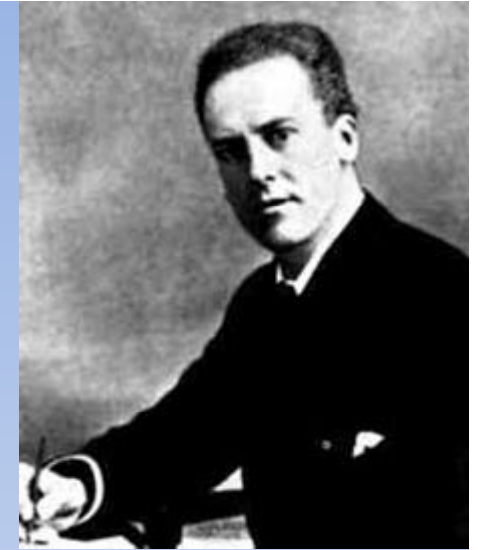
- Pearson's correlation is defined as

$$\rho_{XY} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

- It can be estimated as:

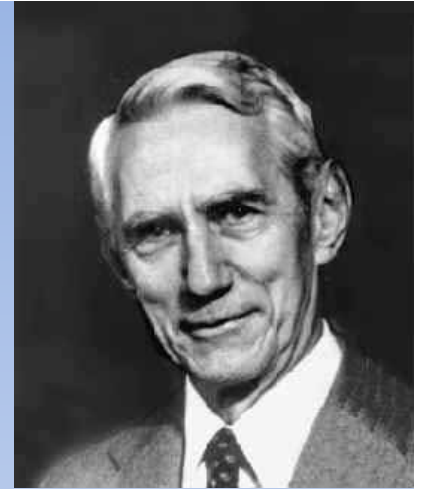
$$r_{XY} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

- No free parameters.
- Intuitive definition for both r.v. and estimators



Karl Pearson

Mutual Information



Claude Shannon

- Mutual Information

$$MI(X, Y) = \int_{\mathcal{X} \times \mathcal{Y}} \log \left(\frac{f_{XY}(x, y)}{f_X(x) f_Y(y)} \right) f_{XY}(x, y) d\mu(x, y)$$

- Conditional Mutual Information

$$CMI(X, Y) = \int_{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}} \log \left(\frac{f_{XY|Z}(x, y | z)}{f_{X|Z}(x | z) f_{Y|Z}(y | z)} \right) f_{XYZ}(x, y, z) d\mu(x, y, z)$$

- Information theoretical measures of dependence quantify the distance in probability spaces between the joint and the product of the marginals.
- In contrast to the linear correlation coefficient, MI is sensitive to dependences that do not manifest themselves in the covariance.

Mutual Information - Estimation

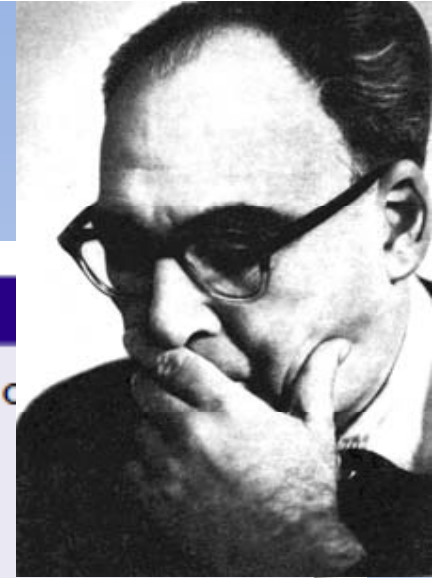
- MI is not easy to estimate for real variables! It is a fundamental problem since the Radon-Nicodým derivative is ill-posed.
- Estimators are normally designed to be consistent, but their small sample behavior is unknown so it is unclear what they are really estimating in the context of dependence or independence.
- Most approaches use Parzen estimators to evaluate MI. But unlike the measure, the estimated MI between two random variables is never invariant to one-to-one transformations.
- Alternatively, mutual information can be estimated from k -nearest neighbor statistics.

Measures of Dependence -Survey

- Many extensions to the concepts of correlation and mutual information as measures of dependence.
- Renyi's has defined dependence by a set of properties that can be applied more generally.
- Copulas are a very interesting and potentially useful concept
- Generalized correlation can be defined with kernel methods
- Measures of Association
 - Have a clear meaning in terms of both realization and r. v. and simple estimators, but they are limited to real numbers.

Measures of Dependence

Functional Definition



Alfred Rényi

Rényi's postulates

- 1 $D(X, Y)$ is defined for any X and Y , neither of them being constant with probability 1.
- 2 $D(X, Y) = D(Y, X)$.
- 3 $0 \leq D(X, Y) \leq 1$.
- 4 $D(X, Y) = 0$ if and only if X and Y are independent.
- 5 $D(X, Y) = 1$ if there is a strict dependence between X and Y , i.e. either $X = g(Y)$ or $Y = f(X)$ where f and g are Borel-measurable functions.
- 6 If the Borel-measurable functions f and g map the real axis in a one to one way onto itself, $D(f(X), g(Y)) = D(X, Y)$.
- 7 If the joint distribution of X and Y is normal, then $D(X, Y) = |\rho(X, Y)|$.

Does not deal with estimation.

Very useful framework. Maximal correlation obeys all these properties

Measures of Dependence

Sklar - Copula

- This is a very important concept answering the simple question: how to create all possible joint pdfs when the marginals $f(x)$ and $g(y)$ are known?
- Copula is exactly defined as the operator $C(\dots)$ from $[0,1] \times [0,1] \rightarrow [0,1]$ that makes $C(f(x),g(y))$ a bona fide pdf.
- So copula is intrinsically related to measures of dependence between r.v. and separates it from the marginal structure.
- It has been used in statistics to create multivariate statistical models from known marginals and provide estimates of joint evolution.
- Most work has been done with parametric copulas (which may produce dangerous bias.....)
- Not well known in engineering or machine learning.



Abe Sklar

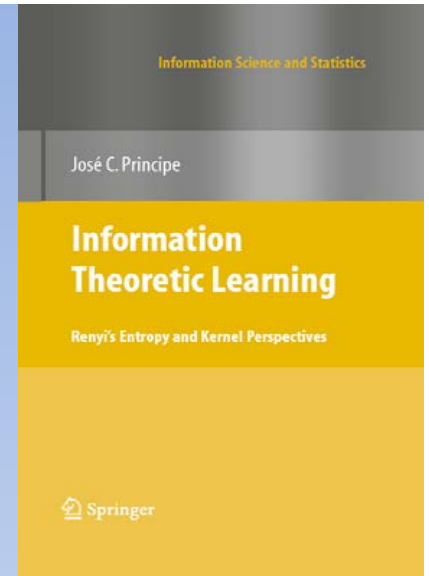
Generalized Correlation

- Correlation only quantifies similarity fully if the random variables are Gaussian distributed.
- Use the kernel framework to define a new function that measures similarity but it is not restricted to second order statistics.
- Define generalized correlation as

$$\nu(X, Y) = E_{XY}[\kappa(X, Y)] = \iint \kappa(x, y) p_{X, Y}(x, y) dx dy$$

which is a valid function for positive definite kernels in L_∞

- Still easy to understand from realizations



Generalized Correlation

- Correlation is obtained when $\kappa(x, y) = xy$
- Let us use instead a shift invariant kernel like the Gaussian

$$\kappa(x, y) = G_{\sigma}(x - y)$$

- We have defined **correntropy** as

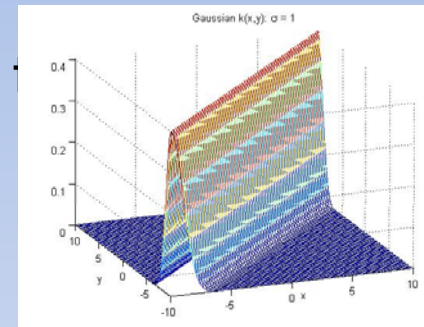
$$v_{\sigma}(X, Y) = E_{XY}[\kappa(X, Y)] = \iint G_{\sigma}(x - y) p_{X, Y}(x, y) dx dy$$

Its estimator is trivial (empirical mean), but has a free parameter

$$\hat{v}_{\sigma}(X, Y) = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(x_i - y_i)$$

Correntropy includes all even order moments of the difference variable

$$v_{\sigma}(X, Y) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E\|X - Y\|^{2n}$$



Correntropy: Dependence measure

Define **centered correntropy**

$$U(X, Y) = E_{X, Y}[\kappa(X - Y)] - E_X E_Y[\kappa(X - Y)] = \iint \kappa(x - y) \{dF_{X, Y}(x, y) - dF_X(x)dF_Y(y)\}$$

$$\hat{U}(x, y) = \frac{1}{N} \sum_{i=1}^N \kappa(x_i - y_i) - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \kappa(x_i - y_j)$$

Define **correntropy coefficient**

$$\eta(X, Y) = \frac{U(X, Y)}{\sqrt{U(X, X)U(Y, Y)}} \quad \hat{\eta}(x, y) = \frac{\hat{U}(x, y)}{\sqrt{\hat{U}(x, x)\hat{U}(y, y)}}$$

Define **parametric correntropy** with

$$V_{a, b}(X, Y) = E_{X, Y}[\kappa(aX + b - Y)] = \iint \kappa(ax + b - y) dF_{X, Y}(x, y)$$

Define **parametric centered correntropy** $a, b \in R \quad a \neq 0$

$$U_{a, b}(X, Y) = E_{X, Y}[\kappa(aX + b - Y)] - E_X E_Y[\kappa(aX + b - Y)]$$

Define **Parametric Correntropy Coefficient**

$$\eta_{a, b}(X, Y) = \eta(aX + b, Y)$$

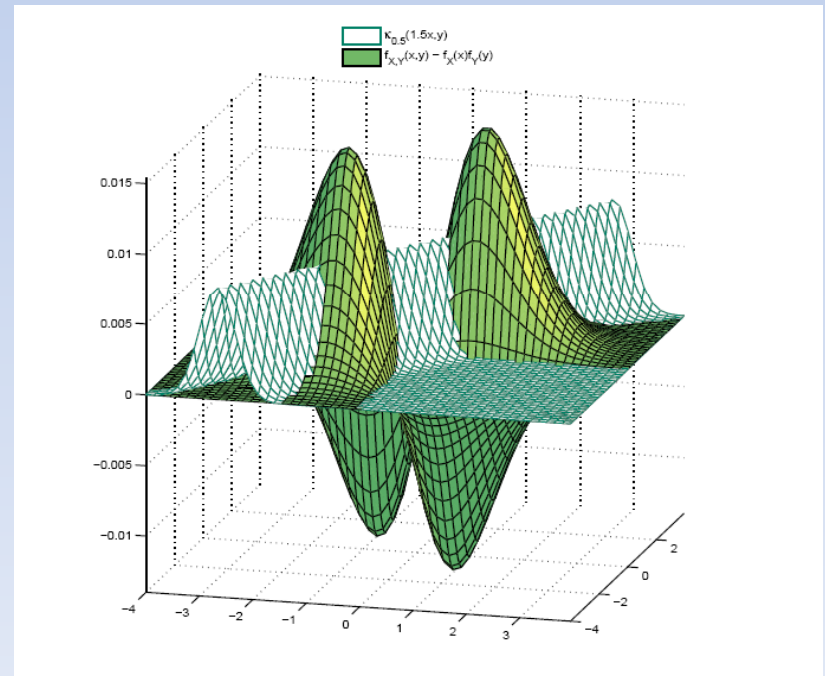
Correntropy: Dependence Measure

Theorem 1: Given two random variables X and Y : the parametric centered correntropy $U_{a,b}(X, Y) = 0$ for all $a, b \in \mathbb{R}$ if and only if X and Y are independent.

Theorem 2: Given two random variables X and Y the parametric correntropy coefficient $\eta_{a,b}(X, Y) = 1$ for certain $a = a_0$ and $b = b_0$ if and only if $Y = a_0X + b_0$.

Definition: Given two r.v. X and Y Correntropy Dependence Measure is defined as

$$\Gamma(X, Y) = \sup_{a, b \in \mathbb{R} \quad a \neq 0} |\eta_{a,b}(X, Y)|$$



Measures of Association

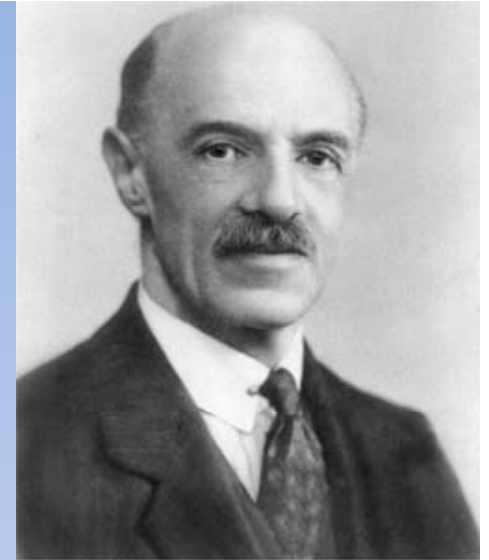
Spearman ρ

- Relaxes correlation on values to ranks
- Non-parametric rank-based measure of the degree of association between two variables.
- Defined as (in a no-ties situation):

$$\rho = 1 - \frac{6 \sum_i d_i^2}{N(N^2 - 1)}$$

where $d_i = x_i - y_i$ represents the difference between the ranks of the two observations X_i and Y_i .

- More robust to outliers, but needs to estimate ranks which is $O(N \log N)$



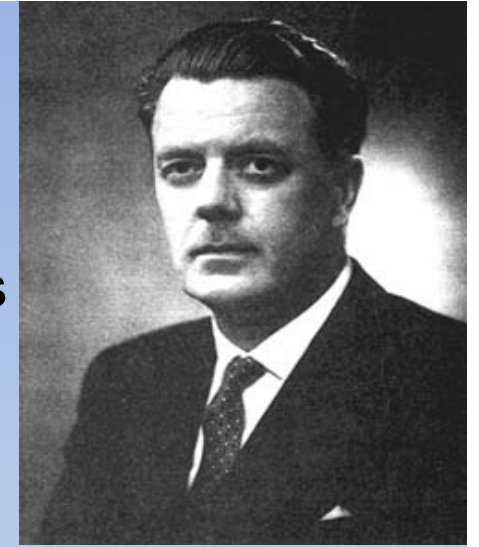
Charles Spearman

Measures of Association

Kendall τ

- Non-parametric rank-based coefficient that measures correlation between ordinal variables.

- Defined as
$$\tau = \frac{N_c - N_{nc}}{1/2N(N-1)}$$



Maurice Kendall

Where N_c refers to the number of **concordant pairs** i.e. cases verifying:

$$\{x_i > x_j \text{ and } y_i > y_j\} \text{ or } \{x_i < x_j \text{ and } y_i < y_j\}$$

And N_{nc} refers to the number of **non-concordant pairs** i.e. cases verifying:

$$\{x_i > x_j \text{ and } y_i < y_j\} \text{ or } \{x_i < x_j \text{ and } y_i > y_j\}$$

- Value increases from -1 to +1 as agreement increased between the ranking.

Generalized Association (GMA)

- The beauty of the measures of association is that they are well understood in both the random variables and their realizations. Since engineers and computer scientists work with data, having clear (and hopefully easy) estimators for realizations is an asset.
- We developed a novel rank-based measure of dependence capable of capturing nonlinear statistical structure in abstract (metric) spaces called **Generalized Measure of Association (GMA)**.
- This route still preserves statistical meaning but proofs of properties and convergence become more difficult (ongoing work).

GMA Definition

Definition

- Given two random variables (X, Y) , Y is associated with X if close realization pairs of Y , i.e. $\{y_i, y_j\}$ are associated with close realization pairs of X , i.e. $\{x_i, x_j\}$, where closeness is defined in terms of the respective metrics of the spaces where the r.v. lie.
- In other words, if two realizations $\{x_i, x_j\}$ are close in \mathbb{X} , then the corresponding realizations $\{y_i, y_j\}$ are close in \mathbb{Y} .
- This follows the spirit of Spearman and Kendall but extends the concept to any metric space because of the pairs.
- The algorithm requires only estimation of ranks over pairs which is $O(N^2 \log N)$

GMA Properties

- GMA is defined between any two r. v. that take values in two distinct metric spaces.
- The proposed measure is asymmetric.
- If $X \perp Y$ (independent) then $D_G \approx 0.5$, and if $X = Y$ then $D_G = 1$ (necessary conditions). At this point, we do not have a formal proof whether these conditions are also sufficient.
- The proposed measure is invariant to any isometric transformation of X and Y since it is solely based on pairwise distances.
- The GMA estimator is parameter free.

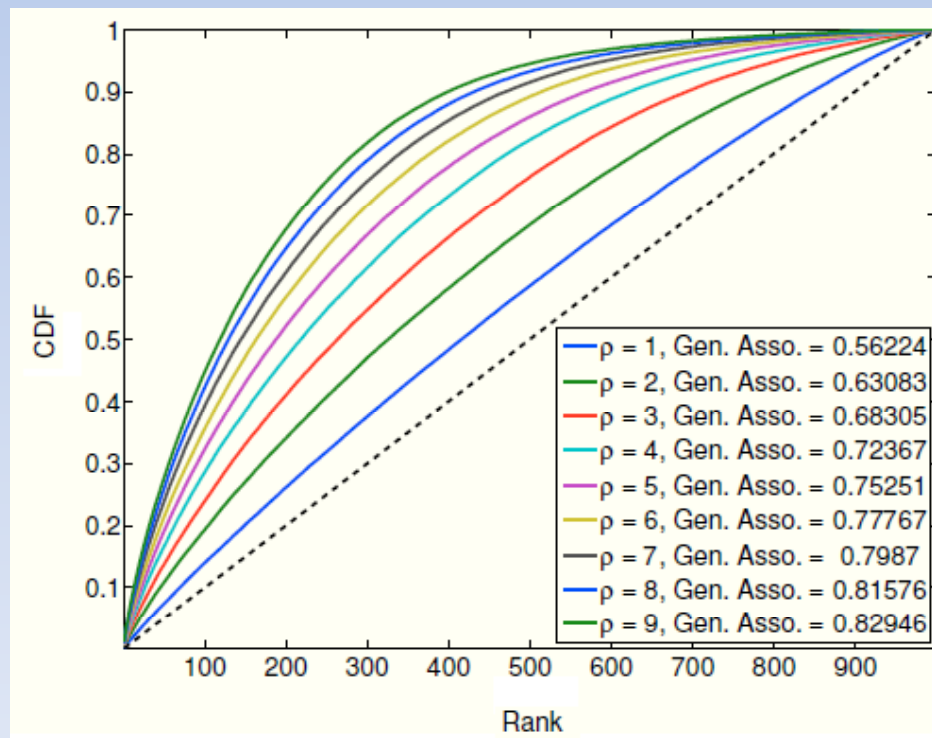
GMA Understanding

- Consider the ranks r_i as a r.v. R , then the distribution of R will quantify dependence. In fact, if the ranks are broadly distributed then dependence is small, while skewness of R will mean more dependence.
 - If ranks are uniform distributed, variables are independent, and $GMA=0.5$
 - If ranks are delta function distributed, then variables are strictly dependent and $GMA=1$.
- This assumes that ranks are never the same between pairs, which is not the case for categorical variables. We modified the procedure to work with a probabilistic rank

GMA Understanding

- The estimator for GMA can be obtained very easily: estimate the CDF of R and normalize it by n-1 (area under the curve),

$$\hat{D}_G = \frac{1}{N-1} \sum_{r=1}^{N-1} (N-r)P(R=r) \quad P(R=r) = \#\{i : r_i = r\} / N$$



Bivariate Clayton Copula

$$C(u, v) = (u^{-\rho} + v^{-\rho} - 1)^{-1/\rho}$$

GMA Algorithm- Stochastic Rank

- For two time series $\{x_t, y_t\}_{t=1}^N$, compute GMA as the area under the CDF of the ranks r_i of y_{j^*} where j^* corresponds to the index of the closest element to each time series element x_i :

1. Assign $P(R = r) = 0 \forall r \in \{1, \dots, (n-1)\}$
2. For all $i \in \{1, \dots, n\}$, repeat the following;
3. Find the set of points $\{x_{j^*} : j^* \in \mathcal{J}\}$ closest to x_i in terms of \mathfrak{d}_X i.e.

$$\mathcal{J} = \{j^* : j^* = \arg \min_{j \neq i} \mathfrak{d}_X(x_i, x_j)\}.$$

4. For all $j^* \in \mathcal{J}$, find the spread of ranks i.e. $r_{i,\max}$ and $r_{i,\min}$ of y_{j^*} in terms of \mathfrak{d}_Y such that

$$r_{i,\max} = \#\{j : \mathfrak{d}_Y(y_j, y_i) \leq \mathfrak{d}_Y(y_{j^*}, y_i)\} \text{ and}$$
$$r_{i,\min} = \#\{j : \mathfrak{d}_Y(y_j, y_i) < \mathfrak{d}_Y(y_{j^*}, y_i)\}.$$

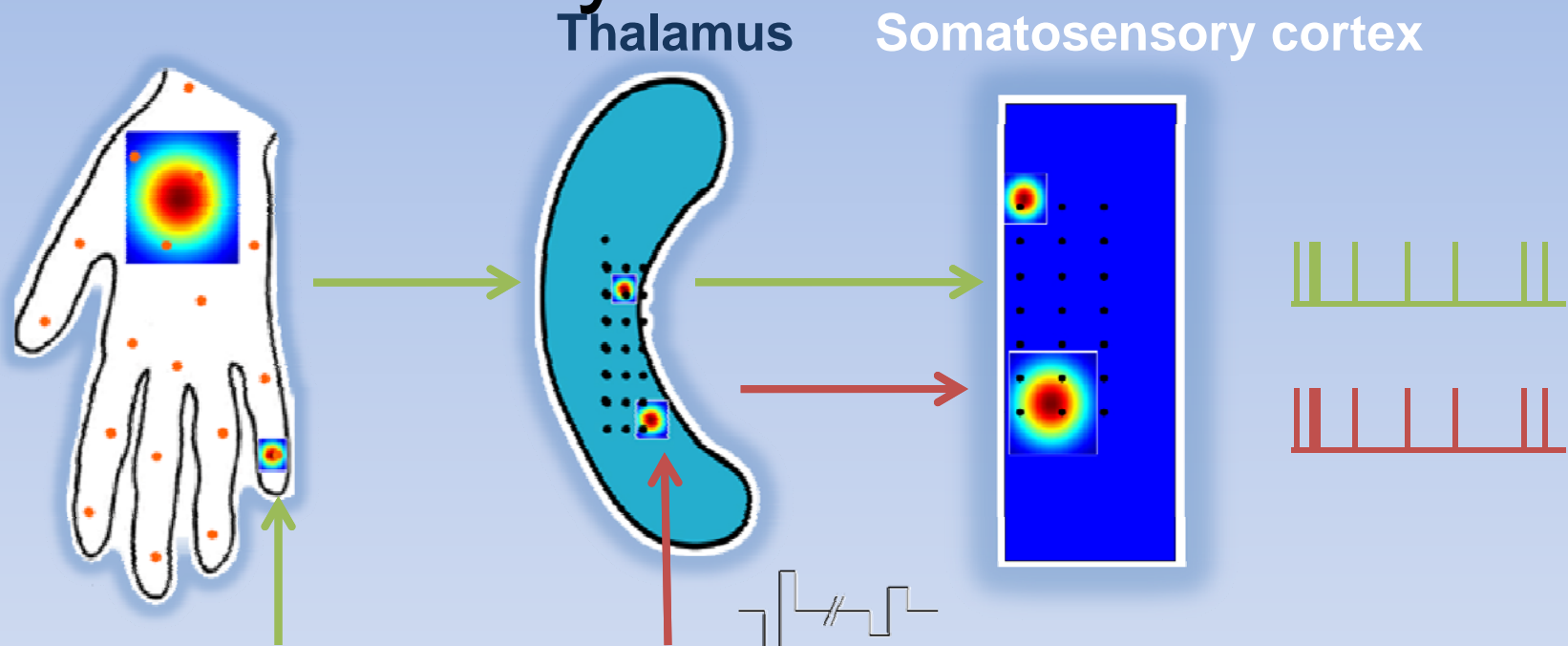
5. For all rank values $r_{i,\min} < r \leq r_{i,\max}$, assign

$$P(R = r) \leftarrow P(R = r) + 1/|\mathcal{J}| / (r_{i,\max} - r_{i,\min}) / n.$$

How to measure cause and effect in neuroscience?

- There are no well-defined estimators of mutual information in non-Euclidean spaces
- This is particularly critical in neuroscience because neurons produce spike trains that can be modeled as point processes.
- How to measure dependence between stimulus and neural response?
- For manipulations in experimental neuroscience it is very important to related the dependence between stimulus and neural response. But stimulus normally are controlled by variables in L_2 , while neurons produce spike trains....

Dependence of neural response on somatosensory stimulation

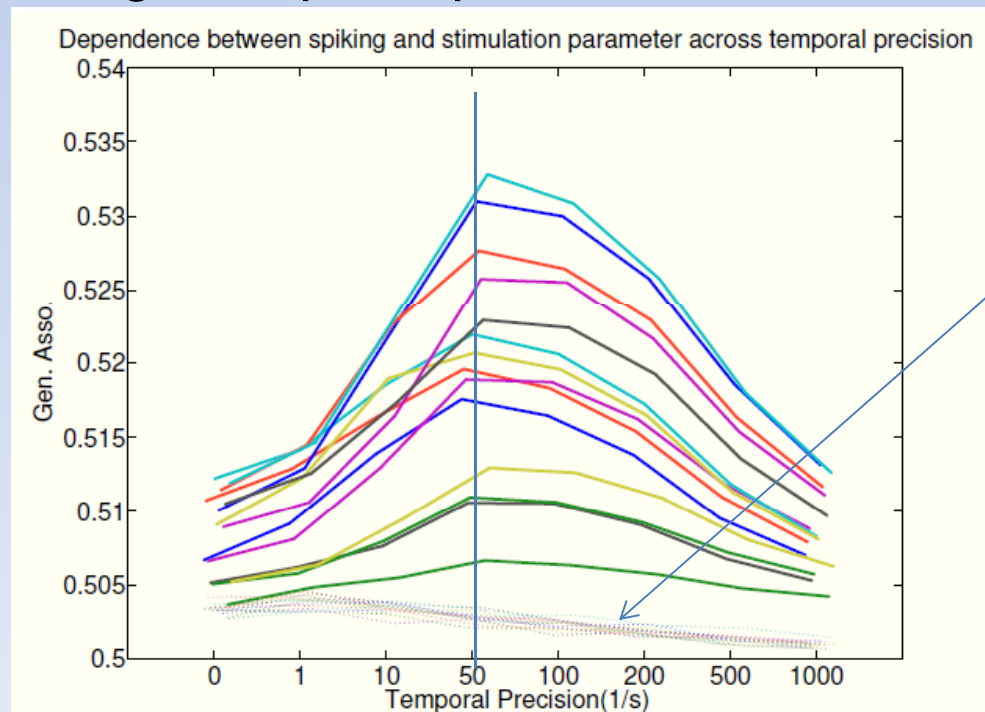


Tactile Stimulation Micro-stimulation

- We control the amplitude and duration of electrical stimulation: 19 distinct pairs of pulse duration and current amplitude were applied, with 140 responses from each pair randomly permuted throughout the recording. We analyze 480 ms of spiking data after stimulus onset on 14 cortical channels after each stimulus onset for analysis.

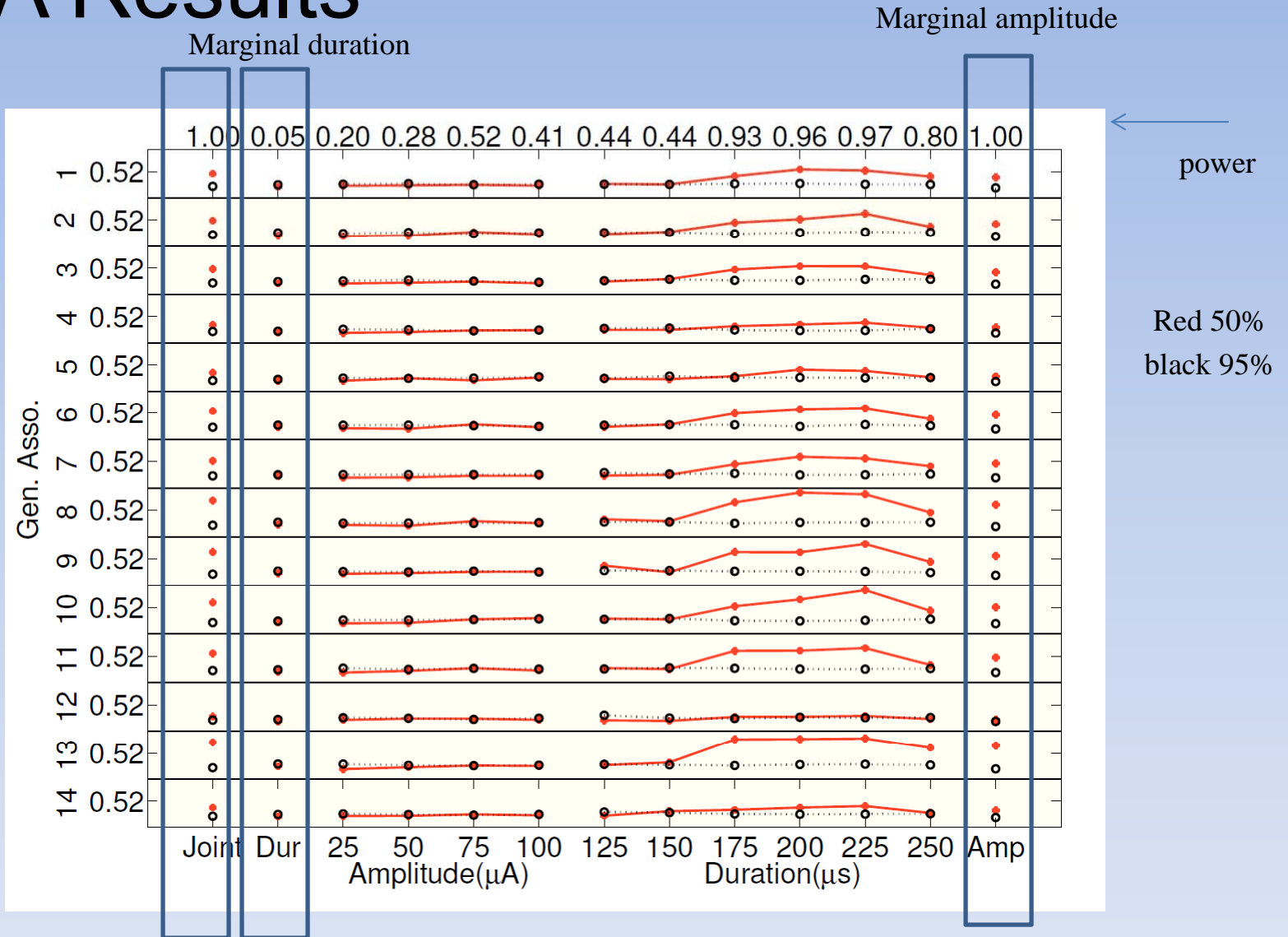
Dependence of neural response on somatosensory stimulation

- To measure the differences in amplitude and duration use L_2 norm.
- To measure the spike train dynamics we use the Victor Purpura distance, which is a function of a temporal precision parameter q . There are 100 bootstrap trials with 40 trials, out of 140, for each stimulation setting to help set q .



95 percentile of surrogates

GMA Results

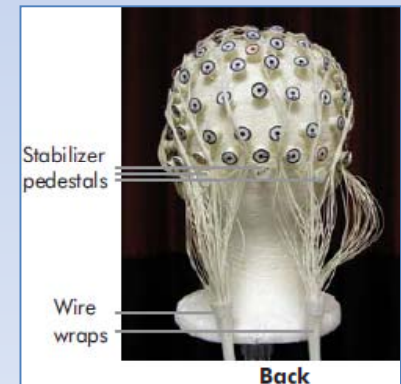
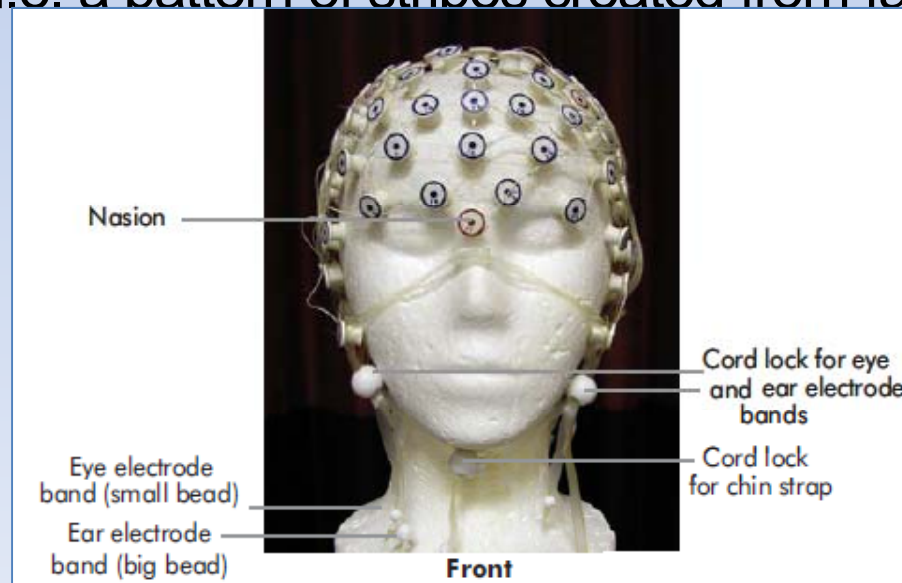
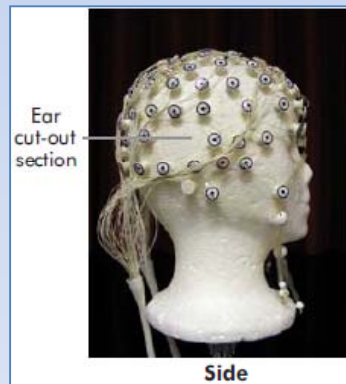


Quantifying Cognitive Brain Processes

- Most visual information processed by the primary visual cortex can be identified by category (table, pen, insect, bird)
This is however different for Faces
- We always recognize a face as someone's face. Besides, this uniqueness is usually recognized with other factors like condition and expression.
- We are specifically interested in studying **affective perception** to identify the brain circuits and cognitive processes involved in recognizing facial and non facial structures. Perhaps we can predict when the subject is looking at a face through the EEG like a cognitive BCI.

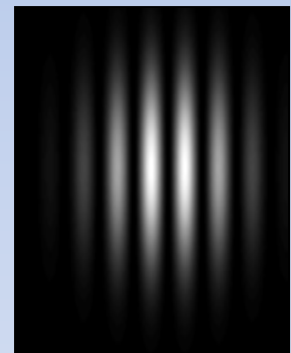
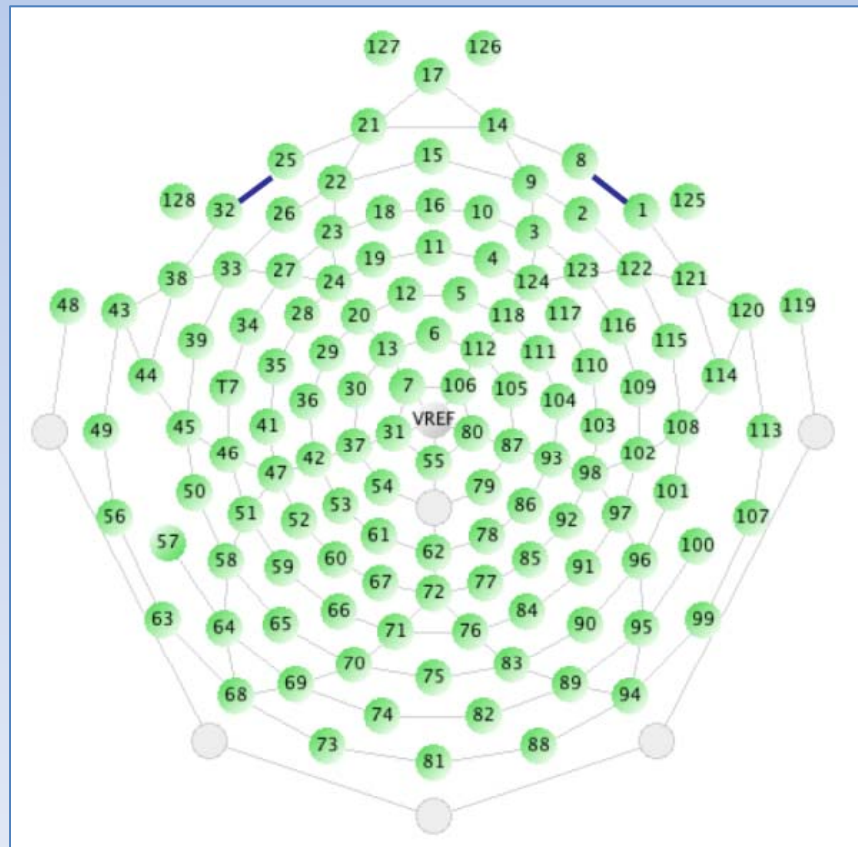
ssVEPs for Cognitive Neuroscience

- In this work, we flash the full image with affective content and study how the whole brain processes the Steady-State Visually Evoked Potentials ssVEPs information (i.e. discriminate between two types of visual stimuli).
- Stimuli consist of two types of images displayed on a 17" monitor flickering at a frequency of 17.5 ± 0.20 Hz and matched for contrast
- First condition shows a human face (*Face*) and the second shows a Gabor patch, i.e. a pattern of stripes created from face (*Gabor*)



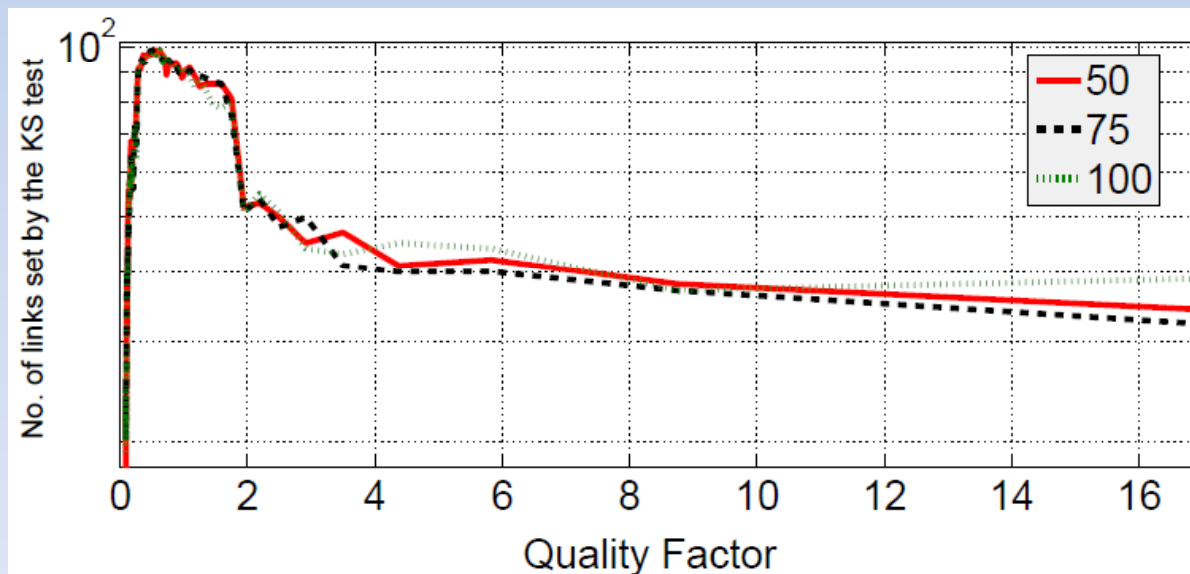
Experimental Setting

- Electrode recordings are collected from a 129 Hydro-Cell Geodesic Sensor Net (HCGSN) montage
- Sampling rate used is 1000 Hz. 15 trials are performed for each condition. Luminance is set to vary from 0 to 9.7 cd.m^{-2}



Preprocessing

- The subject was instructed not to blink and keep head movements as minimal as possible. Still, the recorded EEG signal remains *noisy* and contaminated by important *artifacts*: volume conduction (use source models), line noise (notch filters).
- A BP filter necessary to extract the 17.5 Hz component. The design of this filter is crucial not to distort the dependence. $Q=1$ is the best!!!!



Statistical tests

- We use the two sample Kolmogorov-Smirnov (KS) test to compare the vector distributions.
- The KS test is a non-parametric test to compare two sample vectors, where the test statistic is:

$$KS_{\gamma_1, \gamma_2} = \max_x |F_{\gamma_2}(x) - F_{\gamma_1}(x)|$$

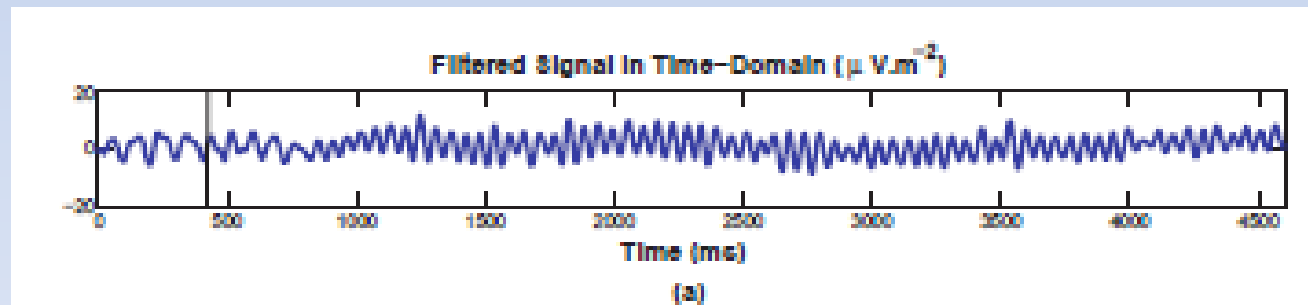
$\gamma_1(x)$ and $\gamma_2(x)$ are sample vectors

$F_{\gamma_1}(x)$ and $F_{\gamma_2}(x)$ are empirical PDFs

- We apply the KS test at an α -level of 0.1 on dependence measures vectors belonging to Face and Gabor patch (Gabor).

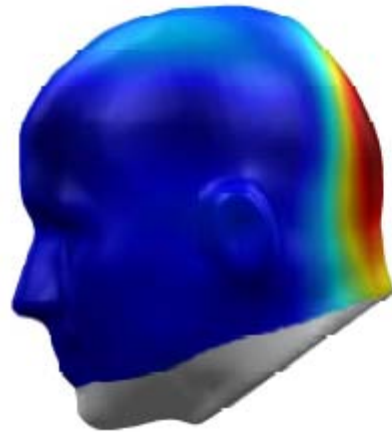
Testing

- Same settings are used to compare the results of correlation, mutual information and GMA:
 - *Pairwise dependence to channel 72*
 - *Embedding dimension set to 8 (25 samples shift for correlation)*
- Time windows of 114 *ms* or 114 *samples*. This corresponds to two cycles of F_s / F_o
- Dependence is visualized in sensor space as a function of amplitude

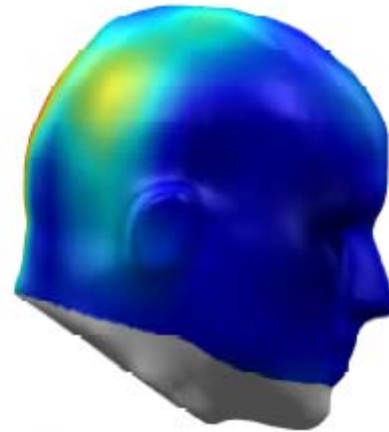


Results- Abs Corr

FACE

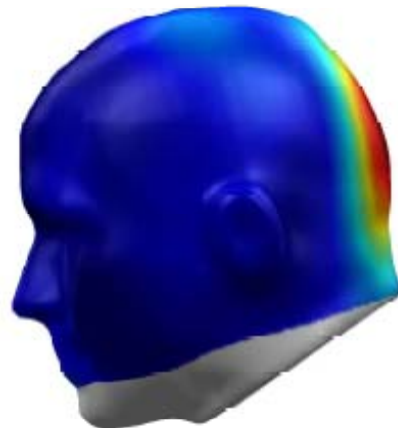


(a) Face - Left - $Corr_{abs}$

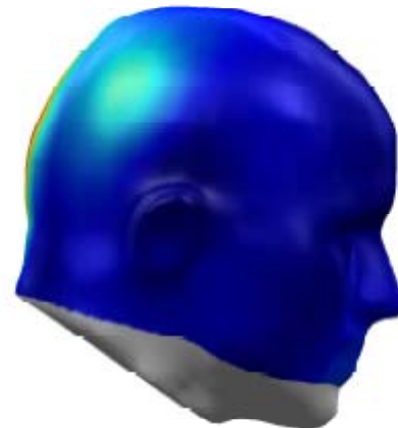


(b) Face - Right - $Corr_{abs}$

GABOR



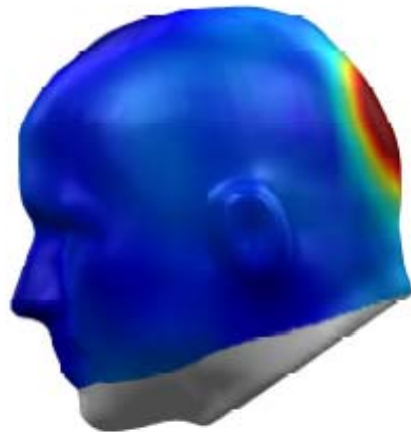
(c) Gabor - Left - $Corr_{abs}$



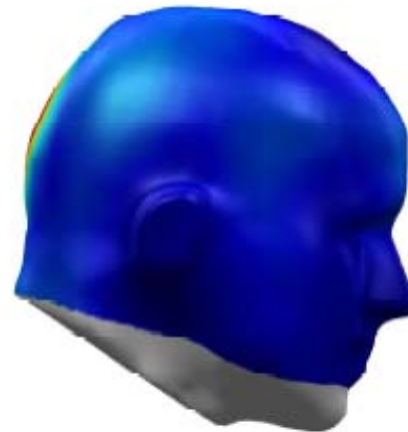
(d) Gabor - Right - $Corr_{abs}$

Results- MI

FACE

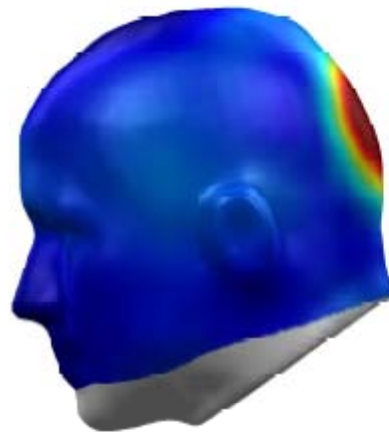


(a) Face - Left - MI

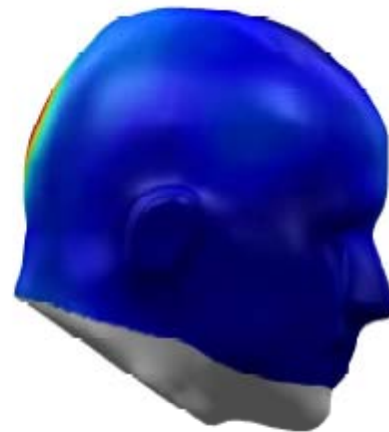


(b) Face - Right - MI

GABOR



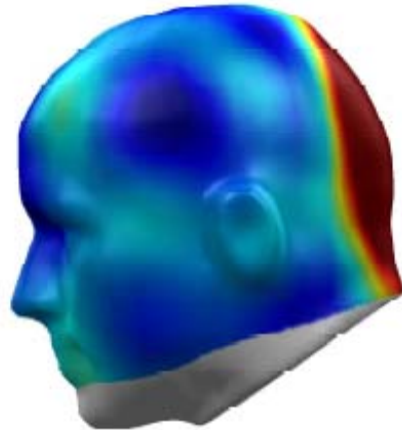
(c) Gabor - Left - MI



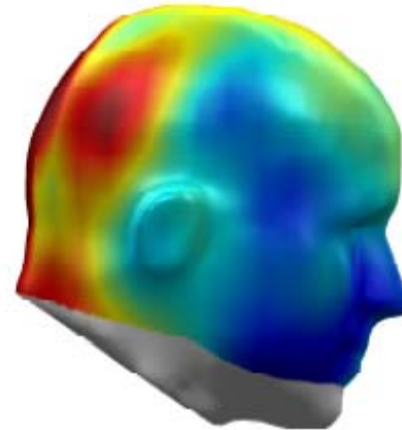
(d) Gabor - Right - MI

Results- GMA

FACE

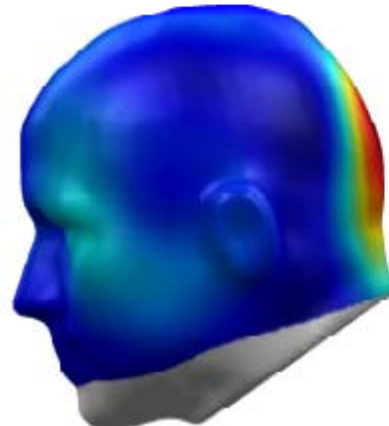


(a) Face - Left - GMA_f

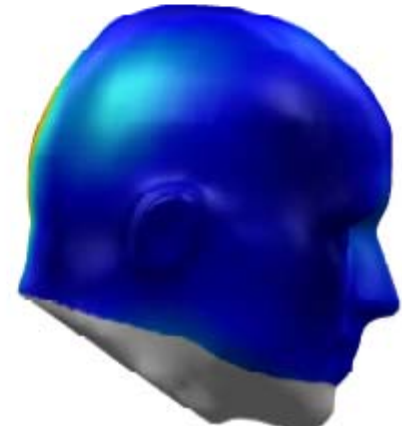


(b) Face - Right - GMA_f

GABOR

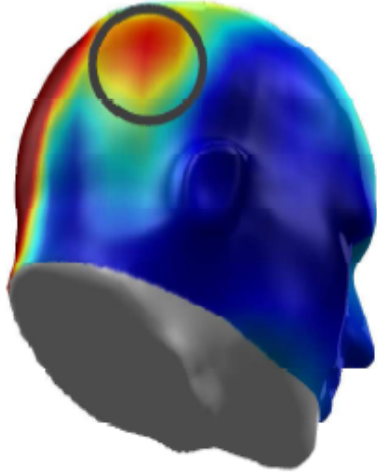


(c) Gabor - Left - GMA_f

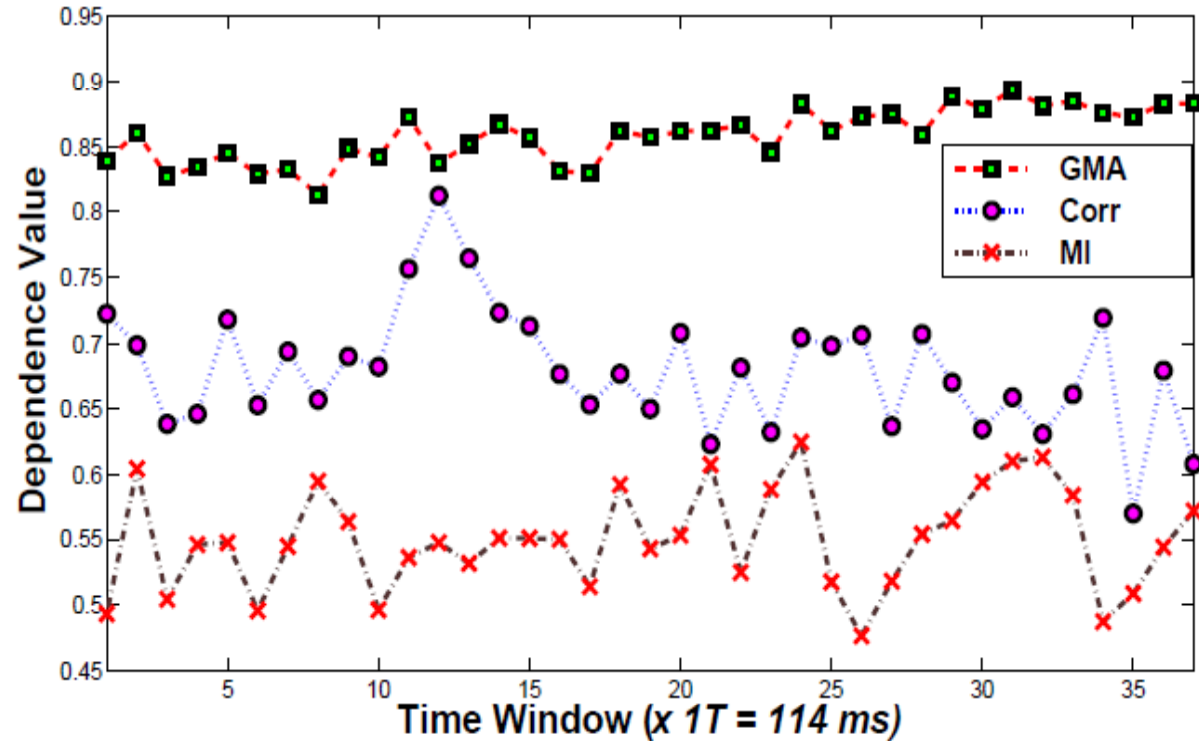


(d) Gabor - Right - GMA_f

Results- Comparison



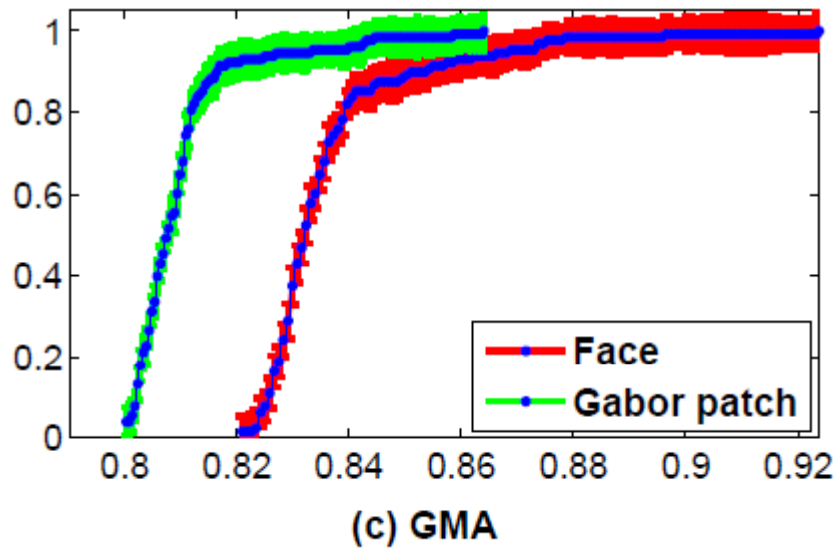
(a) Electrode subset



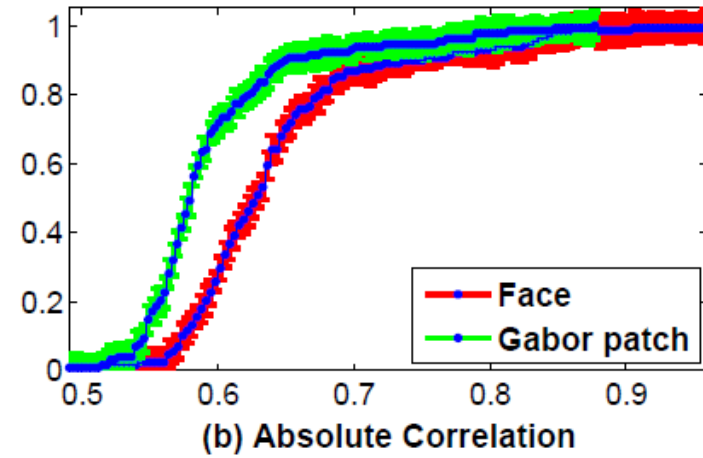
(b) Averaged dependence over time windows

Results- Empiric CDFs

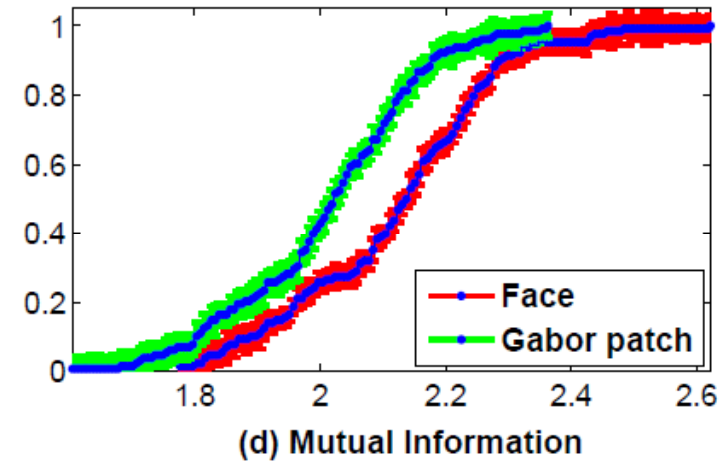
Empirical CDFs Per Condition for GMA



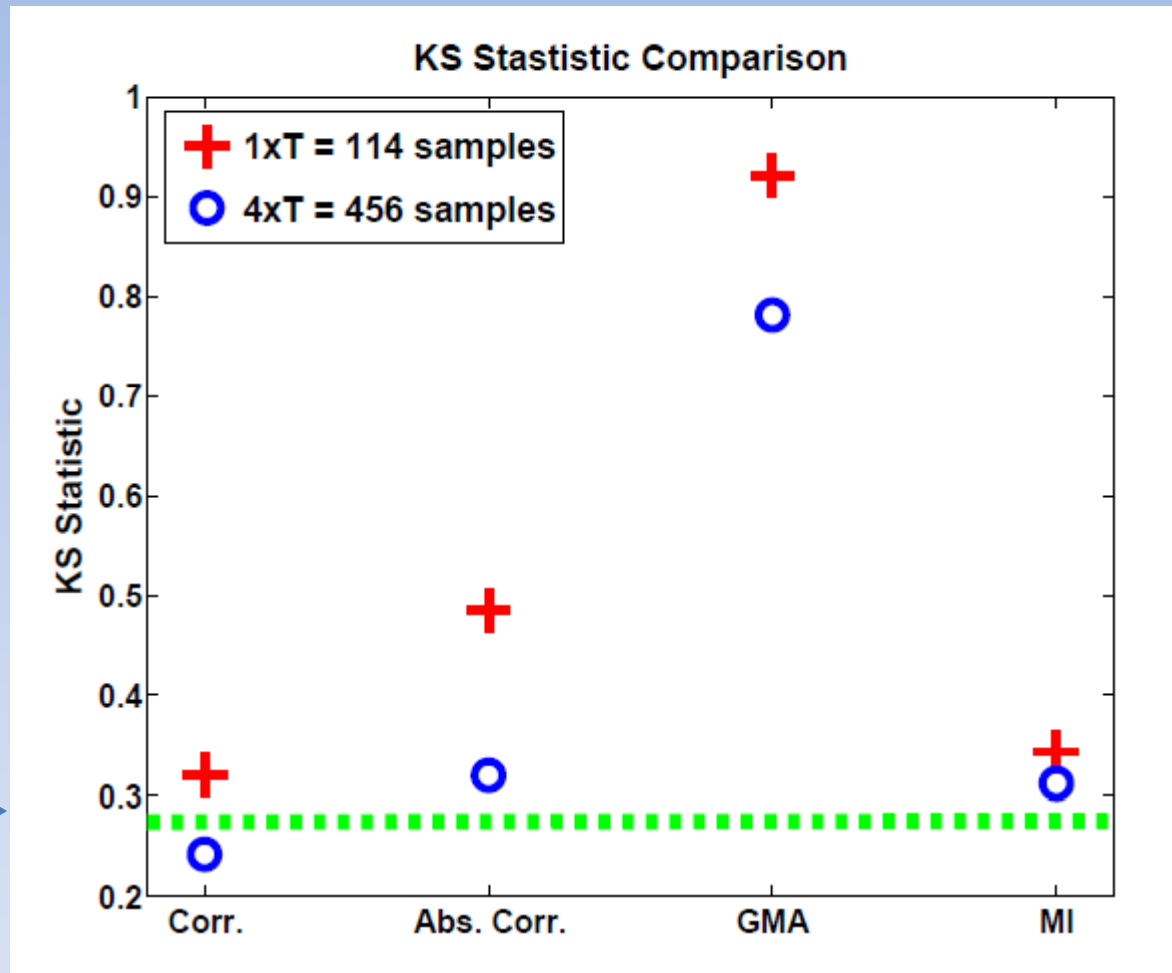
Empirical CDFs Per Condition for Abs. Correlation



Empirical CDFs Per Condition for Mutual Information



Results- K statistic



Power Spectrum

Conclusion

- This presentation addressed the issue of statistical dependence.
- We opted to emphasize the importance of understanding dependence when working with the realizations. So decided to define a new measure of association that generalizes the ones in the literature.
- The estimator for GMA has no free parameters and it is easy to predict how the measure is handling the data. It can also be generalized to conditional association, and the results are comparable with others in the literature, but are easier to compute.
- No proof that GMA is a measure of dependence is known.

Conclusion

- The measure works well in practice, and we are using it to quantify dependence amongst brain areas using the EEG. A filter with a quality factor falling in the range $[0.7-1.5]$ is suitable for maximizing the KS statistical difference between the dependence vectors distributions of Face and Gabor.
- Results show consistency in the brain regions active, over cycles of 114 ms.
- More activity is visible in the right hemisphere of the brain for the Face case, which might be explained by the activity of the right fusiform gyrus and the amygdala.
- GMA showed more discriminability for the two stimuli than Correlation and Mutual Information.