

ANÁLISIS DE COMPONENTES PRINCIPALES E INDEPENDIENTES APLICADOS A REDUCCIÓN DE RUIDO EN SEÑALES ELECTROCARDIOGRAFICAS.

Principal and independent component analysis applied to noise reduction in electrocardiographic signals

RESUMEN

Este artículo pretende mostrar como se puede remover ruido y artefactos de una señal electrocardiográfica utilizando métodos de análisis multivariado, como análisis de componentes principales (ICA) y análisis de componentes independientes (PCA) aplicados sobre una base de datos de electrocardiograma tomadas en humanos. El algoritmo de FastICA usando el criterio de negentropía como función de costo, muestra un mejor resultado sobre su variante llamada JADE, a su vez el algoritmo de FastICA muestra su superioridad en la eliminación de a ruido comparación a un esquema de filtrado clásico compuesto por un banco de Filtros FIR.

PALABRAS CLAVES: Artefacto, Análisis de componentes independientes, Análisis de componentes principales, Electrocardiograma, Ruido.

ABSTRACT

This article pretends to show how can be removed the noise and artifacts from electrocardiographic signal, such as Principal component analysis and Principal component analysis, they will be applied on a data base of electrocardiogram who have been taken on humans been. The FastICA algorithm using negentropy as cost function, shows a better result over its variant call JADE, in turn FastICA algorithm shows superiority in noise elimination in comparison of a classical filtering scheme composed for a FIR filter bank

KEYWORDS: Artifact, Noise, Independent component analysis, Electrocardiogram, Principal component analysis.

1. INTRODUCCIÓN

Los problemas cardiovasculares demandan la necesidad de desarrollar técnicas no invasivas como la electrocardiografía, con la cual los médicos pueden detectarlos y diagnosticar lo mas acertado a sus pacientes:

- Los ECG comúnmente son invadidos por ruido y/o artefactos los cuales afectan la señal que ha sido tomada en humanos.
- Las señales electrocardiográficas son en general grupos de datos de gran tamaño, en los cuales los métodos estadísticos como análisis de componentes independientes (ICA) han tenido buen desempeño.
- El análisis de componentes independientes puede detectar y remover una variedad de ruido y artefactos en los electrocardiogramas y realzar la calidad de la señal.

JULIANA MUÑOZ

Ingeniera Electricista.
Purchasing Department
Papeles Nacionales S.A
juliana.munoz@papelesnacionales.com

JORGE RIVERA

Ingeniero Electrónico, M. Sc
Profesor Asistente
Universidad Tecnológica de Pereira
j.rivera@utp.edu.co

EDISON DUQUE

Ingeniero Electrónico, M. Sc (c)
Profesor Asistente
Universidad Tecnológica de Pereira
eduardo@utp.edu.co

2. CONTENIDO

En esta sección se hace una breve descripción de los elementos teóricos necesarios que conforman el trabajo

2.1 Anatomía y fisiología del corazón: El corazón tiene una función vital para el organismo vivo, este se encarga de bombear la sangre a los pulmones y al resto del cuerpo [1],[2].

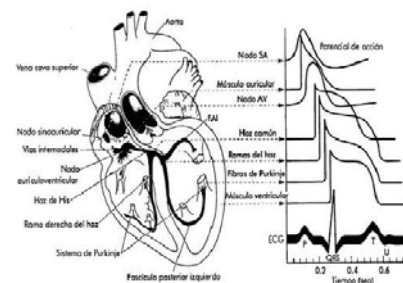


Figura 1. Fisiología del corazón y generación de la señal.

2.2 El electrocardiograma: El electrocardiograma provee una descripción cuantitativa de la actividad eléctrica del corazón esta señal variable en el tiempo es causada por la contracción y relajación de los músculos del corazón [1],[2].

La señal del ECG se obtiene por la medición de la diferencia de potencial entre varios electrodos situados en el tórax, brazos y piernas.

La señal electrocardiográfica esta conformada por: Ondas P, Q, R, S, T, U; Intervalos PR, ST, QT, QRS [2],[3],[4].

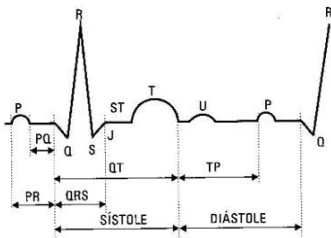


Figura 2. Típica señal electrocardiográfica con sus ondas componentes.

2.3 Perturbaciones en señales electrocardiográficas:

En este caso se utilizan dos términos ruido e interferencia.

El término ruido se utiliza cuando el origen de la interferencia es interno y debido al propio sistema de medida, entre estas tenemos ruidos por problemas de contacto con los electrodos (Produce cambios en la línea base y es de tipo aleatorio), por movimiento entre electrodo-piel (Desplazamiento transitorio de la línea base), Ruido Electromiográfico (Causan variaciones abruptas en la línea base del rango de milivoltios, estas ocurren durante la respiración).

El término interferencia se aplica a señales externas, entre estas se encuentran, interferencia debidas a la línea de alimentación (se presentan señales diferenciales no deseadas que se combinan con la frecuencia de la señal ECG), pueden ser de origen interno y externo como equipos eléctricos y/o electrónicos próximas al entorno de medida.

2.4 Reducción de perturbaciones en el ECG:

Comúnmente se han utilizado muchos tipos de filtros para lograr reducir ruido e interferencias de la señal ECG, como, los filtros digitales los cuales requieren información a priori de la estadística de la señal para ser procesada.

Suelen emplearse los filtros adaptativos o filtros lineales (pasa-bajo, pasa-alto o pasa-banda).

2. METODOS DE TRANSFORMACIÓN

2.1 Planteamiento del problema:

Usualmente, en los problemas de clasificación, las características relevantes son desconocidas a priori, y la inclusión de características irrelevantes de acuerdo

al objetivo deseado, genera un aumento en el tiempo de aprendizaje de los algoritmos. Los métodos para reducir la dimensionalidad tienen como objetivo extraer las características que son relevantes para la clasificación planteada.

2.2 Notación:

El siguiente listado muestra las notaciones usadas en este trabajo.

i, n	contador y numero de patrones
\mathbf{X}, p	el espacio de entrada, $p = \dim(\mathbf{X})$
\mathbf{Y}, q	espacio reducido, $q = \dim(\mathbf{Y})$
\mathbf{w}	vector de mezcla
$\Sigma_{\mathbf{x}}$	matriz de covarianza
$\mathbf{V}_{(i)}, \lambda_{(i)}$	autovector, autovalor

2.3 Métodos lineales: Un conjunto de datos de dimensión p puede reducirse a uno de q dimensiones, utilizando p combinaciones lineales de q variables. Las combinaciones lineales pueden ser consideradas como proyecciones lineales.

2.3.1. Análisis de componentes principales (PCA)

Sea el espacio inicial $\mathbf{x}_{1 \times p}$ un vector aleatorio con valor medio $E\{\mathbf{x}\} = \mu_{\mathbf{x}}$ y matriz de covarianza dada por

$$E\{(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{x} - \mu_{\mathbf{x}})^T\}$$

para el cual algunos de sus elementos o componentes están correlacionados, tal que,

$$\exists P(\mathbf{x}) \exists P(\mathbf{x}_i | \mathbf{x}_j) \neq 0, \forall i \neq j$$

Si se tiene que la cantidad de componentes es muy grande, la tarea de clasificación se hace difícil, por lo que en la práctica se establece el procedimiento por el cual se halla un espacio transformado \mathbf{Y} más simple, que represente la información estadística considerada importante y contenida en \mathbf{X} , dada en la forma

$$\mathbf{Y} = \mathbf{X}\mathbf{A}$$

siendo $\mathbf{A} \in p \times q$, donde $q < p$, que implica la reducción de la dimensión del espacio inicial de representación. La exigencia sobre el contenido de la mayoría de la información estadística en \mathbf{Y} , se formaliza en la siguiente función de minimización:

$$\min_{\mu_{1 \times p}, \mathbf{C}_{p \times q}} E\{(\mathbf{x} - \mu_{\mathbf{x}} - \mathbf{y}\mathbf{C})^T (\mathbf{x} - \mu_{\mathbf{x}} - \mathbf{y}\mathbf{C})\}$$

La matriz \mathbf{A} en, que halla el mínimo de la función (2.2), tiene la forma:

$$\mathbf{A} = \begin{bmatrix} \mathbf{V}_{\Sigma_{\mathbf{x}}, 1}^T \\ \vdots \\ \mathbf{V}_{\Sigma_{\mathbf{x}}, q}^T \end{bmatrix}$$

donde,

$$V_{\Sigma_x, l}, l = 1, \dots, q$$

es el vector propio de la matriz Σ_x .

Los elementos de representación de la variable Y se denominan los componentes principales de X, que tienen la forma,

$$y_l = V_{\Sigma_x, l}^T X$$

para las cuales se cumple que

$$\begin{cases} \sum y_l^2 = \lambda_{\Sigma_x, l} \\ \text{cov}\{y_j, y_i\} = 0, \forall j \neq i \end{cases}$$

Sea la varianza explicada por los q componentes se tendrá

$$\psi = \frac{\lambda_{\Sigma_x, 1} + \dots + \lambda_{\Sigma_x, q}}{\sum_{j=1}^q \lambda_{\Sigma_x, j}}$$

2.3.2 Selección del número de componentes principales

Se han sugerido distintas reglas para seleccionar el número de componentes:

1. Realizar la representación gráfica de $\lambda_{\Sigma_x, i}$ contra i , con $i = 1, \dots, q$. Se comienza seleccionando el número de los componentes hasta que los restantes tengan aproximadamente el mismo valor de $\lambda_{\Sigma_x, i}$. Se busca un punto de quiebre o codo en el gráfico, es decir, un punto a partir del cual los valores propios son aproximadamente iguales, con lo cual se determina el número de componentes que excluye a los valores asociados a las magnitudes pequeñas.

2. Seleccionar los componentes hasta una proporción determinada de varianza, por ejemplo, del 80 o el 90%. Sin embargo, es posible que un único componente recoja el 90% de la variabilidad, pero pueden existir otros componentes que sean adecuados para explicar la estructura de las variables.

3. Desechar aquellos componentes asociados a valores propios inferiores a una cota, que suele fijarse como la varianza media, $\sum_i \lambda_{\Sigma_x, i} / p$.

En particular cuando se trabaja con la matriz de correlación, el valor medio de los componentes es 1, y esta regla lleva a seleccionar los valores propios mayores que la unidad. Sin embargo, una variable que sea independiente del resto suele llevarse un componente principal y puede tener un valor propio mayor que la unidad. Sin embargo, si está no correlacionada con el resto puede ser una variable poco relevante para el análisis, y no aportar mucho a la comprensión del fenómeno global.

2.3.3. Análisis de componentes independientes (ICA)

ICA es un método estadístico de orden superior que busca proyecciones lineales, no necesariamente

ortogonales entre ellas, que sean estadísticamente independientes. La independencia estadística es una condición mucho más fuerte que la no-correlación. Mientras que la última involucra estadísticas de segundo orden, la primera depende completamente de estadística de orden superior.

Independencia estadística y no correlación

Sean $\{y_i : i = 1, \dots, q\}$ un conjunto de variables aleatorias con función de densidad de probabilidades conjunta $p(y_1, \dots, y_q)$. Suponiendo que $E\{y_i\} = 0$, se dice que las variables y son mutua y estadísticamente independientes si

$$p(y_1, \dots, y_q) = \prod_{i=1}^q p_i(y_i)$$

donde $p_i(y_i)$ corresponde a la densidad marginal de y_i .

Por otro lado, se dice que dos variables aleatorias son no correlacionadas si

$$E\{y_i y_j\} = E\{y_i\} E\{y_j\} = 0, \text{ para } i \neq j$$

La independencia estadística es en general un criterio mucho más fuerte que la no correlación.

De hecho, es fácil demostrar que si y_i y y_j son variables estadísticamente independientes,

Las variables aleatorias estadísticamente independientes son en general no correlacionadas; lo contrario no es cierto. Una excepción a esta regla ocurre cuando las variables $\{y_i, i = 1, \dots, q\}$ tienen una función de densidad de probabilidades conjunta gaussiana

Definición del modelo

El modelo ICA libre de ruido para el vector aleatorio x p -dimensional busca estimar los componentes del vector q -dimensional y y la matriz de mezcla con rango completo $A_{p \times q}$,

$y = xA$, tal que los componentes de y son lo más independientemente posibles, de acuerdo a algún criterio de independencia. El modelo ICA que incluye perturbaciones contiene componentes adicionales de ruido aditivos,

$$y = xA + \varepsilon$$

La estimación de dichos modelos es todavía tema de investigación. Existen versiones sobre-dimensionadas de ICA, donde el número q de componentes independientes es mayor que el número de variables originales p .

Sin pérdida de generalidad se puede asumir que las componentes y_i tienen media cero, $E\{y_i\} = 0$, y en consecuencia $E\{y_j\} = 0$. Para resolver A y y esta técnica asume que las señales y_i son estadísticamente independientes entre sí y, como se discutirá más adelante, que a lo más uno de los componentes x_i sea gaussiano. Es de anotar que generalmente, las funciones de probabilidad de x_i no se conocen de forma a priori (si se

1. Las notas de pie de página deberán estar en la página donde se citan. Letra Times New Roman de 8 puntos

conocieran el problema en consideración se facilitarla considerablemente).

Restricciones del modelo ICA

La validez del modelo se garantiza si se cumplen las siguientes restricciones:

1. A lo más uno de los componentes independientes y_i es gaussiano.
2. El número de características iniciales x_j , p , debe ser al menos tan grande como el número de componentes independientes y_i , q , esto es, $p \geq q$. En lo que sigue se considerará el caso $p = q$. Cuando $p > q$, el tratamiento se hace igual al caso $p = q$ haciendo una reducción de dimensiones previa utilizando técnicas como PCA.
3. La matriz A debe ser una matriz de rango completo de columna.

Limitaciones del modelo ICA

El modelo ICA presenta las siguientes limitaciones [5]:

1. Este método no da información alguna sobre la varianza de la señal original ni del signo de la misma, sin embargo, este no es un problema de trascendencia; en consecuencia las características iniciales x_j se estandarizan de modo que su varianza sea unitaria.
2. No se puede restaurar el orden de los componentes. Sin embargo, se garantiza que existe una matriz de permutación para la cual se puede recuperar el ordenamiento original de los componentes. Existen dos técnicas para introducir un orden en los componentes independientes: la primera consiste en utilizar las normas de las columnas de la matriz A , la cual da las contribuciones de los componentes independientes a las varianzas de x_j . El ordenamiento de los y_i de acuerdo con dichas normas se asemeja al dado por el método PCA. La segunda técnica consiste en utilizar la no gaussianidad de los componentes independientes. El ordenamiento de los y_i de acuerdo con dichas normas se asemeja al dado por el método de seguimiento de proyección.

3. PRUEBAS Y RESULTADOS

Las pruebas se realizaron sobre varios tipos de señales para probar la efectividad de el método ICA, se probaron con negentropía y con JADE, inicialmente se realizaron sobre tres bases de datos diferentes, a las cuales se les adiciona señales de ruido de diferentes características, amplitud, frecuencia, muestreo.

La primera prueba se realiza sobre un grupo de datos que contiene 3 tipos de señales diferentes el siguiente grupo de figuras muestra el grupo de datos original y seguido el mismo grupo con ruido y finalmente la señal recuperada.

La base de datos “toy” hace parte de las bases de datos propuestas en [12], como parte de los ejercicios para sintonizar algoritmos de BSS (*Blind Source Separation*)

La base de datos reales utilizada fue creada por el hospital general de Massachusetts (MGH), la cuál esta formada por registros electrónicos de registros hemodinámicos y electrocardiográficos de pacientes en unidades de cuidados intensivos. Dentro de los registros, se encuentran ECG, presión arterial, presión pulmonar arterial, presión central intravenosa, impedancia respiratoria, y concentraciones de CO₂. Para este trabajo en específico solo se utilizaron, los 3 registros ECG. La frecuencia de muestreo es de 360 muestras por segundo.

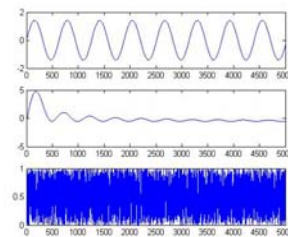


Figura 3: Toy Data Original

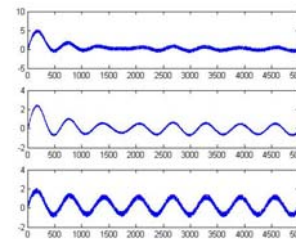


Figura 4: Toy Data Mezclada

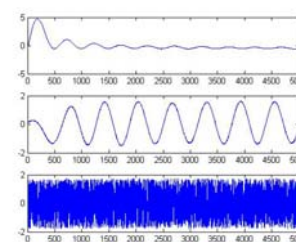


Figura 5: Toy Data Recuperada

Se puede observar en los resultados obtenidos que la señal recuperada es muy similar a la original, el ruido mezclado influye mucho sobre la forma, frecuencia y amplitud de la onda.

La segunda prueba se realizo sobre tres grupos de datos de ECG contaminadas con ruido, se evaluaron con los dos métodos, usando negentropía y el algoritmo de jade.

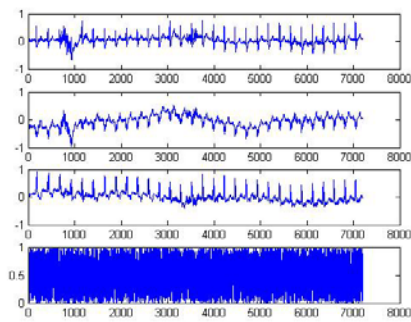


Figura 6: mgh Original

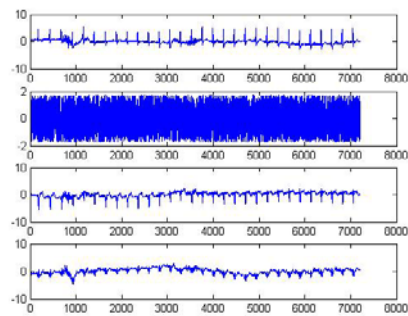


Figura 7: MGH aplicando Negentropía

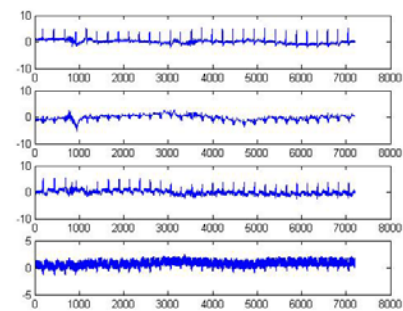


Figura 8: MGH aplicando Jade

Los resultados obtenidos muestran que se obtiene un mejor resultado con el algoritmo de Negentropía que con el algoritmo de Jade.

La tercera prueba se realizó sobre grupos de datos de ECG a las cuales se adicióno ruido aleatorio de diferente frecuencia y amplitud, se puede observar en la siguiente figura, que la señal es recuperada casi en su totalidad pero con las limitaciones que ICA tiene en el orden y en el signo de los componentes.

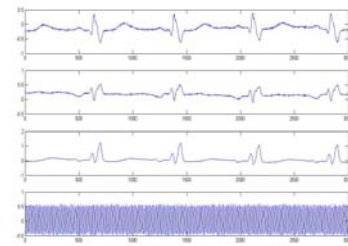


Figura 9: ECG Original

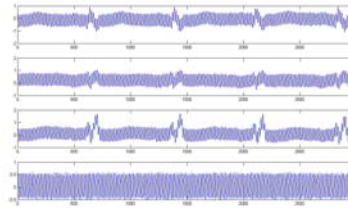


Figura 10: ECG Mezclada

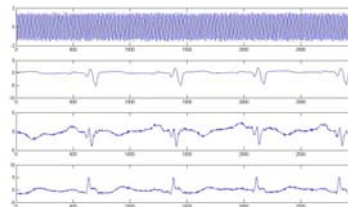


Figura 11: ECG Negentropía

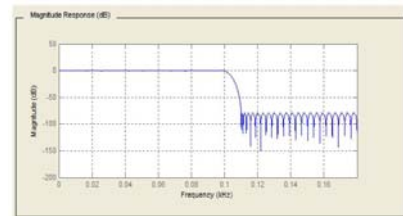


Figura 12: Diseño filtro

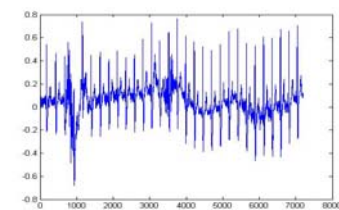


Figura 13: ECG Lead 1 original

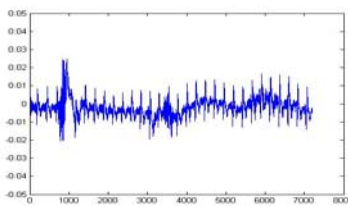


Figura 14: ECG Lead 1 filtrada

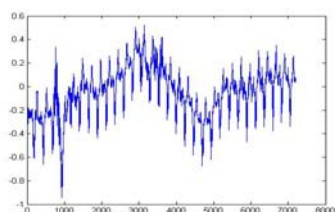


Figura 15: ECG Lead II original

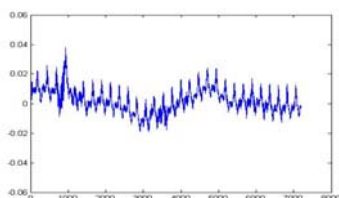


Figura 16: ECG Lead II filtrada

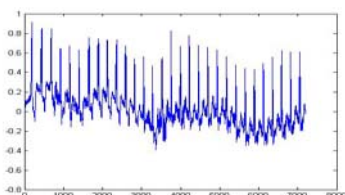


Figura 17: ECG Lead V original

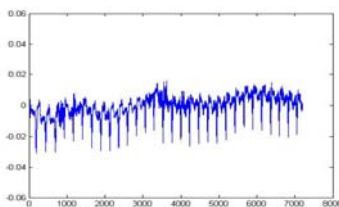


Figura 18: ECG Lead V filtrada

3. CONCLUSIONES Y RECOMENDACIONES

- Se lograron reducir las interferencias generadas por ruido en las señales Electrocardiográficas que se seleccionaron.
- El análisis de componentes y en general el campo del análisis estadístico multivariado resulta idóneo para el manejo de las perturbaciones en arreglos de señales, esto es, los supuestos estadísticos generales que se realizan sobre el ruido electrónico en un sistema de adquisición de datos se ajustan a los modelos de análisis de componentes.
- El problema de permutación de ICA, y el cambio de escala resultante, hacen que un análisis de distorsión armónica (PRD) o relación señal a ruido (SNR) se dificulte, se plantea como trabajo futuro un protocolo para aplicación de estas medidas

4. BIBLIOGRAFÍA

- [1] FAUCI, Anthony. HARRISON, Principios de Medicina Interna. Mc Graw Hill, Ed.14, 1998.
- [2] VELEZ J.; BORRERO H. Fundamentos de Medicina, Cardiología. Corporación para Investigaciones Biológicas, Medellín, Ed.6, 2004.
- [3] RIOS, A. Sistema para el Procesamiento del ECG de Doce Derivaciones en Tiempo Real. PhD thesis, Universidad Nacional de Colombia. Colombia, 1999.
- [4] ARANGO, J. Manual de Electrocardiografía. Corporación para Investigaciones Biológicas, Medellín, Colombia, 2002.
- [5] HYVARINEN, A. Survey on independent component analysis. Neural Computing Surveys, vol.2, pp.94-128, 1999.
- [6] GIRALDO, B. Contribución al Diagnostico Automático de Arritmias Cardíacas basado en el Código Minnesota. PhD thesis, Universidad Politécnica de Cataluña. España, 1996.
- [7] PARK, K.; LEE, K. Application of a wavelet adaptive filter to minimize distortion of the ST segment. Medical and Biological Engineering and Computers, no.36, 1998.
- [8] CASTELLANOS, R.; PUERTA L. Filtración Digital para la Reducción en Tiempo Real de Interferencias Durante la Adquisición de Señales de ECG. Grupo de Control y Procesamiento Digital de Señales, Manizales, 2000.
- [9] CHANG, S.; TSENG, C. Elimination of ac interference in electrocardiogram using IIR notch filter with transient suppression. IEEE Transactions on Biomedical Engineering, vol.42, Num.11, pp.1128-1132, 1995.
- [10] BURGESS, I. Developments in ECG acquisition, Preprocessing, Parameter Measurement and Recording. EMB Magazine, vol.17, pp.50-58, 1998.
- [11] ALIREZA, K.; KONRAD, A.; A Nonlinear Adaptive Method of Elimination of Power Line Interference in ECG Signals. IEEE Transactions On Biomedical Engineering, pp.540-547, 2002.
- [12] A. Cichocki, S. Amari, K. Siwek, T. Tanaka et al., ICALABToolboxes, <http://www.bsp.brain.riken.jp/ICALAB>.